Labeling Problem &
Graph-Based Solution

Asem M. Ali
Labeling Problem

In labeling Problem we have a set of sites $P$ and a set of labels $L$.

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

$L$ : represents intensities, disparities, … etc.

Labeling problem is a mapping $P \rightarrow L$. We denote the labeling by $f$.

$P = f1; 2; ⋯; n g \quad L = f l1; l2; ⋯; lk g \quad f = ff1; f2; ⋯; fn g$

Set of all labeling $L^n$ is denoted by $F$.

Simple Example:

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P$ =</td>
<td>f1; 2; 3; 4g</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$L$ =</td>
<td>f50; 100; 150g</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f$ =</td>
<td>f100; 50; 50; 150g</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f$ =</td>
<td>f50; 50; 50; 150g</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f$ =</td>
<td>f100; 50; 100; 150g</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

**Image Segmentation**

- \( P = f(1; 2; \ldots; R \times C) \)
- \( L = f(0; 0; 0); \ldots; (255; 255; 255) \)

**Image Restoration**

- \( P = f(1; 2; \ldots; R \times C) \)
- \( L = f(0; 0; 0); \ldots; (255; 255; 255) \)
Labeling problem concept gives a common notation for diverse vision problem, such as:

Stereo Matching
Labeling problem concept gives a common notation for diverse vision problem, such as:

**Stereo Matching**
Labeling problem concept gives a common notation for diverse vision problem, such as:  

**Stereo Matching**

![Left image](image1.png) ![Right image](image2.png)

P = \{f_1; f_2; \ldots; R \subseteq C \}  

L = \{f_{d_{\min}} : d_{\max} \}  

**Disparity range**
Labeling problem concept gives a common notation for diverse vision problems, such as:

**Image Matching**
Shekhovtsov, et al CVPR’07

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

\[
L = f(\pm x_{\text{min}}; \pm y_{\text{min}}): (\pm x_{\text{max}}; \pm y_{\text{max}})\]

**Displacement range**

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

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

\[
L = \{1; 2; \ldots; S\}
\]
Problem Formulation

- Image has a natural structure in which pixels are arranged in 2D array.
- \( P = \{1, 2, \ldots, n\} \) set of image pixel
- Neighborhood system \( N \) in \( P \) is the set of all neighboring pairs \( \{p, q\} \) if \( p, q \in P \)
- Example up to the 5th order

1st order neighborhood system

\[
N_p = \{a, b, c, q\} \\
N_q = \{z, p, y, x\} \\
N_r = \{s, t\}
\]

3 x 5 image

The neighborhood system satisfies

a) \( p \not\in N_p \)  
   b) if \( p \in N_q \) then \( q \in N_p \)
Problem Formulation

- $L = f_1; 2; \ldots; K g$ set of labels. $K$ # labels/classes (e.g., $K = 2$)
- $I = f I_1; I_2; \ldots; I_n g$ observed image.
- $f = f f_1; f_2; \ldots; f_n g$ labeled image. $f : P \rightarrow L$
- $F$ set of all labelings $L^n$ (e.g., in this case $2^{36}$ different labeling sets)
- $F = f F_1; F_2; \ldots; F_n g$ set of random variables defined on $P$, and $f$ is a configuration of the field $F$
Problem Formulation

- F is a Markov Random Field (MRF) w.r.t N if its probability mass function
  \( P(F = f) \) abbreviated by \( P(f) \) satisfies

1. \( P(F = f) > 0 \) for all \( f \in F \), Positivity
2. \( P(F_p = f_p | F_{P_i f_{pg}} = f_{P_i f_{pg}}) = P(F_p = f_p | F_{N_p} = f_{N_p}) \), Markov Property
3. \( P(F_p = f_p | F_{N_p} = f_{N_p}) \) is the same for all sites \( p \), Homogeneity

- Markov property establishes the local model

- GRF describes the properties of an image in terms of the joint distribution of labels for all pixels, and provides a global model for the image.
Problem Formulation

- **GRF** provides a global model for an image by specifying the (joint) probability distribution:

\[
P(f) = \frac{1}{Z} \exp\left(\sum_{c \in C} V_c(f)\right);
\]

where
- $Z$ is the normalizing constant called the partition function,
- $V_c$ is the potential function, clique function, summation over cliques “Gibbs energy” $C$ set of all cliques.

- A clique is a set of pixels in which all pairs of pixels are mutual neighbors.
Problem Formulation

- In the pairwise interaction models, Gibbs energy is defined in terms of clique of size 2.
- The image is represented by a MRF with joint distribution:

\[
P(f) = Z^{-1} \exp\left( \sum_{f_p, f_q \in N} V(f_p, f_q) \right)
\]

(A)

**Maximum-A-Posteriori (MAP) Estimation**

- The input image \( I \) and the labeled image \( f \) are described by a joint Markov-Gibbs random field (MGRF).
- MGRF model is fitted within the Bayesian framework of MAP estimation to estimate

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

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

\[
P(I | f) = \prod_{p \in P} \frac{P(I_p | f_p)}{P(f_p)}
\]

(B)
Problem Formulation

Maximum-A-Posteriori (MAP) Estimation

From (A) & (B) the MAP estimator

\[ f^* = \arg\max_{f \in \mathcal{F}} \prod_{p \in \mathcal{P}} \log(P(I_p | f_p)) \cdot \prod_{f_p, q \in \mathcal{N}} V(f_p; f_q) \]

Equivalent to minimize the energy

\[ E(f) = \prod_{f_p, q \in \mathcal{N}} V(f_p; f_q) \cdot \prod_{p \in \mathcal{P}} \log(P(I_p | f_p)) \]

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 Solution

- **Modern energy minimization methods such as:**
  - Graph cuts (Zabih PAMI’01)
  - Loopy Belief Propagation (LBP) (Felzenszwalb CVPR’04)
  - Tree-ReWeighted message passing (TRW) (Wainwright Info Theory’05)
  - Quadratic Pseudo-Boolean Optimization (QPBOP) (Kolmogorov CVPR’07)

- **Classical methods such as:**
  - Iterated Conditional Modes (ICM) (Besag’74)
  - Simulated Annealing (Geman & Geman’84)
Problem Solution

“Fast Approximate Energy Minimization via Graph Cuts” Boykov, Olga, and Zabih, PAMI’01

GC

SA


Stereo

Segmentation
Problem Solution

“Comparison of Energy Minimization Algorithms for Highly Connected Graphs”, Kolmogorov and Rother, ECCV’06

Graph Cut (a)  TRW (b)  BP (c)

“Optimizing Binary MRFs via Extended Roof Duality”  Rother, Kolmogorov, Lempitsky, Szummer, CVPR’07

<table>
<thead>
<tr>
<th>Applications</th>
<th>Sim. An.</th>
<th>ICM</th>
<th>GC</th>
<th>BP</th>
<th>BP+I</th>
<th>P+BP+I</th>
<th>QPBO</th>
<th>QPBOP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagram recognition (4.8con)</td>
<td>0 (0.28s)</td>
<td>999 (0s)</td>
<td>119 (0s)</td>
<td>25 (0s)</td>
<td>0 (0s)</td>
<td>0 (0s)</td>
<td>56.3% (0s)</td>
<td>0% (0s)GM</td>
</tr>
<tr>
<td>New View Synthesis (8con)</td>
<td>- (-s)</td>
<td>999 (0.2s)</td>
<td>2 (0.3s)</td>
<td>18 (0.6s)</td>
<td>0 (2.3s)</td>
<td>0 (1.2s)</td>
<td>3.9% (0.7s)</td>
<td>0% (1.4s)GM</td>
</tr>
<tr>
<td>Super-resolution (8con)</td>
<td>7 (52s)</td>
<td>68 (0.02s)</td>
<td>999 (0s)</td>
<td>0.03 (0.01s)</td>
<td>0.001 (0.06s)</td>
<td>0 (0.03s)</td>
<td>0.5% (0.016s)</td>
<td>0% (0.047s)GM</td>
</tr>
<tr>
<td>Image Segm. 9BC + 1 Fgd Pixel (4con)</td>
<td>983 (50s)</td>
<td>999 (0.07s)</td>
<td>0 (28s)</td>
<td>28 (0.2s)</td>
<td>0 (31s)</td>
<td>0 (10.5s)</td>
<td>99.9% (0.08s)</td>
<td>0% (10.5s)GM</td>
</tr>
<tr>
<td>Image Segm. 9BC; 4RC (4con)</td>
<td>900 (50s)</td>
<td>999 (0.04s)</td>
<td>0 (14s)</td>
<td>24 (0.2s)</td>
<td>0 (3s)</td>
<td>0 (1.48s)</td>
<td>1% (1.46s)</td>
<td>0% (1.48s)GM</td>
</tr>
<tr>
<td>Texture restoration (15con)</td>
<td>15 (165s)</td>
<td>636 (0.26)</td>
<td>999 (0.05s)</td>
<td>19 (0.18s)</td>
<td>0.01 (2.4s)</td>
<td>0 (14s)</td>
<td>16.5% (1.4s)</td>
<td>0% (14s)GM</td>
</tr>
</tbody>
</table>
Graph Cuts
Graph Cuts

Graph Cut Basic Definition & Notation

The weighted graph $G = (V; E)$

- $V$ is the set of vertices in graph correspond to pixels, or other features.
- $f_t; s_g$ (sink & source) are two distinguished vertices called terminals.
- $E$ a subset of pairs $(p; q)$ of elements from $V$ “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
A cut $C \subseteq E$ is a set of edges such that terminals are separated in the induced graph $G(C) = \langle V \setminus C \cup \{s, t\}, E \setminus C \rangle$.

No proper subset of $C$ separate the terminals in $G(C)$.

Cost of the cut $C$, denoted $\|C\|$, 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.
Min-Cut = Min Capacity = Max-Flow

Cut Cost = 7

Cut Cost = 20

Cut Cost = 30
Graph Cuts

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 as a minimization technique

\[ E(f) = \bigoplus_{f \in \mathbb{P}} V(f_p; f_q) \bigoplus_{p \in \mathbb{P}} \log(P(I_p | f_p)) \]

- 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 \(i \log P(I_p | f_p)\)
- Compute s-t MinCut
Graph Cuts

Graph cut as a minimization technique (Example)

- One row image
  
  ![One row image](image)

- Observed image
  
  ![Observed image](image)

- Threshold image
  
  ![Threshold image](image)

- Piecewise Constant Prior “Potts’ model”
  
  \[
  V(f_p; f_q) = \begin{cases} 
  50 & \text{if } f_p \not= f_q; \\
  0 & \text{if } f_p = f_q
  \end{cases}
  \]

- Data penalty term
  
  \[
  P(I_p \mid f_p) / \exp(i \cdot j \mid p \cdot i \cdot f_p j)
  \]

- N-link weight

- T-link weight

Graph Cuts

Graph cut as a minimization technique (Example)

\[ E(f) = \sum_{p,q \in N} 50 \pm (f_p \not= f_q) + \sum_{p \in P} j |_{p,i} f_p j: \]

Best labeling

![Best labeling grid]

\[ E = 50 + 50 + 90 + 10 + 40 + 20 + 120 + 30 + 10 + 30 + 20 + 50 \]

\[ = 520 \]

Threshold labeling

![Threshold labeling grid]

\[ E = 50 + 50 + 50 + 50 + 50 + 60 + 10 + 40 + 20 + 30 + 30 + 10 + 30 + 20 + 50 \]

\[ = 550 \]
Graph Cuts

Graph cut as a minimization technique (Example)

Max-Flow = 600
Graph Cuts

Graph cut as a minimization technique (Example)

Max-Flow = 880
**Multi way Graph-cut algorithm**

**\( \alpha \)-expansion algorithm (Boykov et al.’01):**

Minimizes an energy function with non binary variables by repeatedly minimizing an energy function with binary variables using Max-flow/min-cut method.

Which function can be minimized by \( \alpha \)-expansion Algorithm?

\( \alpha \)-expansion algorithm can be applied to pair-wise interactions that are *metric* on the space of labels.

\[
V_{pq}(l_1;l) = 0 \quad V_{pq}(l_1;l_2) > 0 \quad \text{if} \quad l_1 \notin l_2 \\
V_{pq}(l_1;l_2) = V_{pq}(l_2;l_1) \\
V_{pq}(l_1;l_2) + V_{pq}(l_2;l_3) \leq V_{pq}(l_1;l_3)
\]
**Algorithm**

1. Start with any arbitrary labeling $f^\mu$
2. Set $success = 0$
3. For each label $\alpha \in \mathcal{L}$ (random order)
   
   - (a) find $\tilde{f} = \arg\min f \in f$ among $f$ within one $\alpha$-expansion of $f^\mu$
   - (b) if $E(f^\mu) > E(\tilde{f})$ set $f^\mu = \tilde{f}$ & $success = 1$

4. If $success = 1$ goto 2
5. Return $f^\mu$

---

1- Every pixel either keeps its old label or switch to $\alpha$
2- There is no $\alpha$-expansion move, for any label $\alpha$, with lower energy.
Some Results

Segmentation

Restoration

Stereo Matching

Displacement
Thank You
Maximum-A-Posteriori (MAP) Estimation

- Iterative research of MAP estimate stochastic (e.g., simulated annealing) or deterministic (e.g., iterated conditional modes)

Simulated Annealing (Geman & Geman'84)
- 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.

Iterated Conditional Modes (ICM) (Besag’74)
- Pixels are processed sequentially, and for each pixel the algorithm selects the label that maximize $P(I_p | f_p)P(f_p | f_{N_p})$
- Faster than simulated Annealing
- Very sensitive to the initial labeling
- Local energy optimization technique