CS840a
Learning and Computer Vision
Prof. Olga Veksler

Lecture 6
Markov Random Fields and Belief Propagation
Today

- Markov Random Fields
- Belief Propagation
- Next time paper:
  - “Object Class Recognition by Unsupervised Scale-Invariant Learning” by R. Fergus, P. Perona, A. Zisserman
Labeling Problem

Given

1. Image with \( m \) pixels
   \[
   S = \{ l, \cdots, m \}
   \]
   \[
   \begin{array}{cccc}
   1 & 2 & \cdots & \vdots \\
   \vdots & \vdots & \ddots & \vdots \\
   \vdots & \vdots & \vdots & \ \ m \\
   \end{array}
   \]

2. Set of labels
   \[
   L = \{ l_1, \cdots, l_k \}
   \]

Find

Labeling

\[
f = \{ f_1, \cdots, f_m \}, \quad f_p \in L
\]

Pixel \( p \) is assigned \( l \)

\[
f_p = l
\]
Examples

Restoration

Stereo
A priori knowledge is frequently available

- Bayesian framework is suitable: assign prior probabilities $Pr(f)$ to labelings

$$
\begin{array}{cccc}
  f_1 & f_2 & \cdots & \\
  \cdots & \cdots & \cdots & \\
\end{array}
$$

$$
f = \{f_1, \ldots, f_m\} \in L^m
$$
**Desirable Properties of prior Pr(f)**

Pixel’s label depends on neighboring pixels
Each pixel $p$ has a neighborhood $N_p$

Property: $q \in N_p \iff p \in N_q$

Examples:
$N_p = \{t, b, l, r\}$
$N_q = \{x, z\}$
F is an MRF if:

- **Markovianity:**
  \[
  Pr\left(f_p \mid f_{s-\{p\}}\right) = Pr\left(f_p \mid f_{N_p}\right)
  \]

- **Positivity:**
  \[
  Pr(f) > 0, \quad \forall f \in L^m
  \]
Markov Random Fields

- Not obvious how to compute $Pr(f)$ from local conditional distributions $Pr(f_p | f_{N_p})$

- **Hammersley-Clifford** theorem: $f$ is an MRF if and only if it has *Gibbs* distribution

\[
Pr( f ) \propto \exp \left( - \sum_{c \in C} V_c (f) \right)
\]
A **clique** $c$ is a set of pixels 
$$c = \{p_1, \ldots, p_k | \text{p}_i \text{'s are neighbors of each other}\}$$
$C$ is the set of all cliques

- $V_c(f): L^m \rightarrow R$ is a **clique potential**

- **Clique potentials** $V_c(f)$ specify an MRF
**MRF Example**

- Only two-cliques are non-zero
- Two-clique potential \( \{p,q\} \) :

\[
V_{pq}(f_p,f_q) = u_{pq} \cdot \delta(f_p \neq f_q)
\]

\[
V_{pq}(f_p,f_q) = \begin{cases} 
0 & \text{if } f_p = f_q \\
 u_{pq} & \text{otherwise}
\end{cases}
\]

encourages piecewise-constant \( f \)

\[
Pr(f) \propto \exp \left( - \sum_{\{p,q\} \in N} u_{pq} \cdot \delta(f_p \neq f_q) \right)
\]
MAP-MRF estimation

- Given: observed data \( d \), prior \( Pr(f) \) and likelihood function \( Pr(d|f) \)

- **Maximize Posterior Probability (MAP)**
  \[
  Pr(f|d) \propto Pr(d|f) Pr(f)
  \]
  over all \( f \in L^m \)

- By Bayes’ law, \( Pr(f|d) \propto Pr(d|f) Pr(f) \)

- **MAP estimate**:
  \[
  \hat{f} = \arg \max_{f \in L^m} Pr(d|f) Pr(f)
  \]
MAP-MRF estimation

- Likelihood function:

\[ Pr(d|f) = \prod_p D_p(d_p|f_p) \]

- Maximization problem:

\[ \arg \max_{f \in L^m} \exp \left( \sum_p \ln D_p(d_p|f_p) - \sum_{\{p,q\} \in N} u_{pq} \cdot \delta(f_p \neq f_q) \right) \]

- Equivalent to posterior energy minimization:

\[ \hat{f} = \arg \min_f \left( \sum_{\{p,q\} \in N} u_{pq} \cdot \delta(f_p \neq f_q) - \sum_p \ln D_p(d_p|f_p) \right) \]
MAP-MRF approach for computer vision was first advocated by Geman and Geman, 1984

- High optimization cost. Standard methods, like simulated annealing, work very slow because typical image size is formidable.
MAP-MRF approach for computer vision was first advocated by *Geman and Geman, 1984*

Computation is only tractable in some special cases

In general, high optimization cost. Standard methods, like simulated annealing, work very slow because typical image size is formidable
Belief Propagation (BP)

- Advocated by Pearl for tree-structured graphs in the context of Bayesian Nets
- Gives exact estimates in a tree graph
- Can be applied to general graphs, however
  - No optimality guarantee
  - May not converge
- There are 2 versions of BP
  - max-product, best to use if MAP estimate is needed
    - Map estimate is $\max_f Pr(f|d)$
  - sum-product, best to use if marginals at each node are needed
    - Marginal at node $p$ is $Pr(f_p|d)$
Max-product BP

- Iterate until convergence:
  - Each pixel $p$ (or node) sends a message vector (in parallel) to each of its neighbors $q$,
    \[
    \begin{bmatrix}
    m_{p\rightarrow q}(l_1) \\
    m_{p\rightarrow q}(l_2) \\
    \vdots \\
    m_{p\rightarrow q}(l_k)
    \end{bmatrix}
    \]
    where
    \[
    m_{p\rightarrow q}(l) = \min_{l'} \left[ D_p(l') + V_{pq}(l,l') + \sum_{r\in N_p \setminus q} m_{r\rightarrow p}(l') \right]
    \]
  - After “convergence” the final label at each pixel $p$ is computed as:
    \[
    \arg\min_l \left[ D_p(l) + \sum_{q\in N_p} m_{q\rightarrow p}(l) \right]
    \]
Sum-product BP

- Iterate until convergence:
  - Each pixel $p$ (or node) sends a message vector (in parallel) to each of its neighbors $q$,
    \[
    \begin{bmatrix}
    m_{p\rightarrow q}(l_1) & m_{p\rightarrow q}(l_2) & \ldots & m_{p\rightarrow q}(l_k)
    \end{bmatrix}
    \]
    where
    \[
    m_{p\rightarrow q}(l) = \sum_{l'} \left[ D_p(l') N_{pq}(l', l) \prod_{r \in N_p \setminus q} \sum m_{r\rightarrow p}(l') \right]
    \]
  
- After “convergence” compute the belief for each label $l$ at each pixel $p$:
    \[
    b_l(p) = D_p(l) \prod_{q \in N_p} m_{q\rightarrow p}(l)
    \]
For tree-structured graph, it can be shown that
- max-product BP gives the MAP estimate of the MRF, that is \( \max_f Pr(f \mid d) \)
- sum-product BP gives the correct posterior marginals that is \( Pr(f_p \mid d) = b(p) \)
- Start passing messages at the leaves moving up to the roof
- Compute the solution label (for max-product) or the belief vector (for sum-product) from the root down to the leaves
Several interesting speed-up techniques for max-product BP (and code) available from
Comments about BP on General Graphs

- **Pros**
  - Can be applied to any energy function and any neighborhood structure
  - Very easy to implement

- **Cons**
  - Not guaranteed to converge on (does go into infinite loops occasionally)
  - No optimality guarantees even if converges
  - Not very fast
Better Message Passing Algorithms

- Tree-reweighed message passing algorithms