Computer Vision III

Some of the material modified from Lampert, examples from Kohli, Ladicki, Gould, Koller, Zhu.

• Last time: Computer vision tasks as massive search problems

\[ f(x) = y^* = \arg\max_{y \in Y} g(x, y) \]
• Detection: \( f : X \rightarrow Y \) \( Y = \) all possible positions (and scales) of object

• Foreground/background segmentation: \( f : X \rightarrow Y \) \( Y = \) all possible 0/1 labelings of image \( \{0,1\}^n \)

• Labeling: \( f : X \rightarrow Y \) \( Y = \) all possible labelings of image \( \{1, \ldots, L\}^n \)

• Pose estimation: \( f : X \rightarrow Y \) \( Y = \) all possible poses \( (u, v, \theta, s) \) of image \( \{1, \ldots, P\}^K \)

\[
f(x) = y^* = \arg\max_{y \in Y} g(x, y)
\]

• Detection: \( f : X \rightarrow Y \) \( Y = \) all possible positions (and scales) of object

\[
g(x, y) = w \cdot \varphi(x, y)
\]

• Foreground/background segmentation: \( f : X \rightarrow Y \) \( Y = \) all possible 0/1 labelings of image \( \{0,1\}^n \)

• Labeling: \( f : X \rightarrow Y \) \( Y = \) all possible labelings of image \( \{1, \ldots, L\}^n \)

• Pose estimation: \( f : X \rightarrow Y \) \( Y = \) all possible poses \( (u, v, \theta, s) \) of image \( \{1, \ldots, P\}^K \)

Energy models:

\[
g(x, y) = \sum_{i=1}^{n} g_i(x, y_i) + \sum_{i,j \in E(i)} g_{i,j}(y_i, y_j, x)
\]

Probabilistic models:

\[
p(y|x) = \frac{1}{Z} \exp(g(x, y)) = \frac{1}{Z} \prod_{i} \psi_i(x, y_i) \prod_{i,j \text{ linked}} \psi_{ij}(y_i, y_j)
\]
• Exact algorithms on tree-structured graphs
  – Message passing
  – Max-product: compute $y^*$
  – Sum-product: estimate marginals $p(y_i|x)$
• Approximate on non-tree graphs: e.g., Loopy BP
• Today: Exact and approximate solutions on non-tree graph in vision

Assumptions for exact solution

Using min instead of max

$$g(x, y) = \sum_{i=1}^{n} g_i(x, y_i) + \sum_{i, j \in \mathcal{N}(i)} g_{i,j}(y_i, y_j, x)$$

• $y_i$ binary
• $g_i(.) \geq 0$
• $g_{ij}(y_i, y_j) = g_{ij}(y_j, y_i) \geq 0$ if $y_i \neq y_j$
• $g_{ij}(y_i, y_j) = g_{ij}(y_j, y_i) = 0$ if $y_i = y_j$
Exact solution

- Exact solution with mincut
- Optimal labeling of original minimization problem

Example: Binary segmentation

\[ g(x, y) = \sum_{i=1}^{n} -\log p(y_i | x_i) + \lambda \sum_{i, j \in N(i)} U(x_i, x_j) [y_i \neq y_j] \]

- \( x_i = \) color at pixel \( i \)
- \[ U(x_i, x_j) = e^{\gamma ||x_i - x_j||^2} \]
- \( \lambda \) controls the level of smoothing

[Example from Blake 2004]
More general condition

- Regularity:
  - $g_{ij}(0,0) + g_{ij}(1,1) \leq g_{ij}(0,1) + g_{ij}(1,0)$
  - Problem can be solved exactly as graph cut iff $g$ is regular
Example

- Given initial hypothesis foreground/background: $V_{fg} V_{bg}$, generate the best segmentation
- Related to algorithms for interactive segmentation

$$g_i(x, y_i) = \begin{cases} 
\infty & \text{if } y_i = 0 \ i \in V_{fg} \\
\infty & \text{if } y_i = 1 \ i \in V_{bg} \\
0 & \text{if } y_i = 1 \ i \notin V_{bg} \\
\log \frac{p_f(x_i)}{p_b(x_i)} + \alpha & \text{if } y_i = 0 \ i \notin V_{fg} 
\end{cases}$$

- As before: $g_{ij} = U(x_i, x_j)[y_i \neq y_j]$
Approximate solutions: Neighborhood search

• No exact solutions for non-binary labels
• Optimize only in local subsets of search space so that exact solution can be found in each subset

\[ \alpha \beta \text{-swap} \]

• Subset: \( Y_{\alpha \beta} (z) = \{ y \text{ s.t. } y_i = z_i \text{ if } z_i \notin \{ \alpha, \beta \}, y_i \in \{ \alpha, \beta \} \text{ otherwise} \} \)

• New binary problem:
  – Given current estimated solution \( z \)
  – Find the best assignment of labels \( \alpha \) and \( \beta \) to those sites of \( z \) that are \( \alpha \) or \( \beta \)

• Alternatively: Binary “label”: swap/no-swap
\[ y^{t+1} = \arg\min_{y \in \mathcal{Y}_{\alpha\beta}(y^t)} \sum_{\forall y_i^t \in \{\alpha, \beta\}} g_i(x, y_i^t) + \sum_{\forall y_i^t \in \{\alpha, \beta\}} g_i(x, y_i) \]

\[ + \sum_{\forall y_i^t, y_j^t \in \{\alpha, \beta\}} g_{ij}(y_i^t, y_j^t, x) + \sum_{\forall y_i^t \in \{\alpha, \beta\}, y_j^t \in \{\alpha, \beta\}} g_{ij}(y_i^t, y_j^t, x) \]

\[ + \sum_{\forall y_i^t \in \{\alpha, \beta\}, y_j^t \in \{\alpha, \beta\}} g_{ij}(y_i^t, y_j, x) + \sum_{\forall y_i^t \in \{\alpha, \beta\}, y_j^t \in \{\alpha, \beta\}} g_{ij}(y_i, y_j, x) \]
$y^{t+1} = \text{argmin}_{y \in \mathcal{Y}} \sum_{y_i \in (\alpha, \beta)} g_i(x, y_i) + \sum_{y_i \in (\alpha, \beta)} g_i(x, y_i) + \sum_{y_i, y_j \in (\alpha, \beta)} g_{ij}(y_i, y_j, x)$

Constant terms: Ignore
Unary terms: $\sum_{z_k \in (\alpha, \beta)} h_k(x, z_k)$
Binary terms: $\sum_{z_k \in (\alpha, \beta), z_l \in (\alpha, \beta)} h_{kl}(z_k, z_l, x)$

- Iterate over all $\alpha \beta$
- Optimal at each step
  $g_{ij}(\alpha, \alpha) + g_{ij}(\beta, \beta) \leq g_{ij}(\alpha, \beta) + g_{ij}(\beta, \alpha)$
- No guarantee in general for global solution
- In practice: Close to optimal in many problems

Example from [Kolmogorov et al.]
Generalization: Higher-order potential

- Restricting to pairwise interactions greatly reduces the effectiveness of the model
- We’d like to represent consistency of labels over larger patches or regions of the image

Replace:

$$g(x, y) = \sum_{i=1}^{n} g_i(x, y_i) + \sum_{i,j \in N(i)} g_{i,j}(y_i, y_j, x)$$

By

$$g(x, y) = \sum_{i=1}^{n} g_i(x, y_i) + \sum_{c \in C} g_c(y_c, x_c)$$

Example from Christoph Lampert
\[ g(x, y) = \sum_{i=1}^{n} g_i(x, y_i) + \sum_{c \in C} g_c(y_c, x_c) \]

Pairwise Potts
\[ g_{ij}(y_i, y_j, x) > 0 \text{ if } y_i \neq y_j \]
\[ g_{ij}(y_i, y_j, x) = 0 \text{ if } y_i = y_j \]

\[ p^n \text{ Potts} \]
\[ g_c(y_c, x_c) = \gamma_o \text{ if } y_i \neq y_j \text{ for some } i, j \in c \]
\[ g_c(y_c, x_c) = \gamma_k \text{ if } y_i = k \text{ for all } i \in c \]
\[ \gamma_o > \gamma_k \]

\[ \alpha \beta \text{ expansion still polynomial with } p^n \text{ Potts model} \]

[ Kohli et al., CVPR 2007]

**Example**

\[
\begin{cases}
0 & \text{if } y_i = y_j = k \text{ for all } i, j \in c \\
|c|^{\theta_h}(\theta_p^h + \theta_p^h G(c)) & \text{if } y_i \neq y_j \text{ for some } i, j \in c
\end{cases}
\]

Region quality

\[ G(c) = \exp \left( -\theta_B \frac{\sum_{i \in c} (f(i) - \mu)^2}{|c|} \right) \]

Variance of feature over region

[ Kohli, Ladicky, Torr, IJCV 2009]
Example

\[
G(c) = \exp \left( -\theta_h \frac{\| \sum_{i \in c} (f(i) - \mu)^2 \|}{|c|} \right)
\]

Region quality Variance of feature over region

- Darker = higher quality

Robust model

- Enforcing that all the pixels in $c$ have the same label is too strict
- $g_c$ will switch from 0 to $\gamma_{\text{max}}$ as soon as one label disagrees
- Allow some labels to disagree, with some penalty

if $y_i \neq y_j$ for some $i, j \in c$

\[
g_c = \begin{cases} 
N_i(x_c) \frac{1}{Q} \gamma_{\text{max}} & \text{if } N_i(x_c) \leq Q \\
\gamma_{\text{max}} & \text{otherwise},
\end{cases}
\]

$N_i(x_c) = \text{number of pixels in } c \text{ who disagree with the majority label}$

[Kohli, Ladicky, Torr, IJCV 2009]
Robust model

- Enforcing that all the pixels in $c$ have the same label is to strict
- $g_c$ will switch from 0 to $\gamma_{\text{max}}$ as soon as one label disagrees
- Allow some labels to disagree, with some penalty

If $y_i \neq y_j$ for some $i, j \in c$

$$g_c = \begin{cases} \frac{N_i(x_c)}{Q} \gamma_{\text{max}} & \text{if } N_i(x_c) \leq Q, \\ \gamma_{\text{max}} & \text{otherwise,} \end{cases}$$

---

Kohli, Ladicky, Torr, IJCV 2009

---

Unary terms $g_1(\cdot)$ only
Superpixels

Superpixels
Pairwise terms only $g_{ij}(\cdot)$
Large clique model $\mathcal{P}^n$

(g)

Robust model

[Kohli, Ladicky, Torr, IJCV 2009]
Approximate solutions: Relaxation

- Solve the problem in $\mathcal{Z} \supseteq \mathcal{Y}$ to get bound on the original problem
- (Common) example: LP formulation
• $\mu_i$ binary indicator variable with $L$ possible values:
  $\mu_i(z) = 1$ if $y_i = z$, 0 else
• $\mu_{ij}$ binary indicator variable with $L \times L$ possible values:
  $\mu_{ij}(z_i, z_j) = 1$ if $y_i = z_i$, $y_j = z_j$, 0 else

\[
\begin{align*}
g(x, y) &= \sum_{i=1}^{n} g_i(x, y_i) + \sum_{i,j \in N(i)} g_{i,j}(y_i, y_j, x) \\
&= \sum_i \sum_z g_i(x, z)\mu_i(z) + \sum_{i,j \in N(i)} \sum_{z,z'} g_{i,j}(z, z', x)\mu_{ij}(z, z') \\
&= \sum_i \sum_z \theta_i(z)\mu_i(z) + \sum_{i,j \in N(i)} \sum_{z,z'} \theta_{ij}(z, z')\mu_{ij}(z, z')
\end{align*}
\]

Need additional constraints on the larger set of variables:

$\sum_z \mu_i(z) = 1$, $\sum_{z'} \mu_{ij}(z, z') = \mu_i(z)$
$\sum_z \mu_{ij}(z, z') = \mu_j(z')$
\[
\max_{\mu} \sum_{i,z} \theta_i(z) \mu_i(z) + \sum_{i,z,z'} \theta_{ij}(z,z') \mu_{ij}(z,z')
\]

s.t. \( \sum_z \mu_i(z) = 1 \quad \sum_{z,z'} \mu_{ij}(z,z') = \mu_i(z) \quad \sum_z \mu_{ij}(z,z') = \mu_j(z') \)

NP-hard for \( \mu \) binary

Efficient solutions for the relaxed version
\( \mu_{ij} \in [0,1] \quad \mu_i \in [0,1] \)

Optimality??

- Many efficient solvers for relaxed LP problem
- Tight solution if the graph has no cycles
- If regular, fractional solutions cannot be optimal
- Useful approach in practice, depending on problem

Example task

[Komodakis and Paragios, ECCV 2008]

[Graphs showing relative energy vs. coupling strength for different coupling strengths ρ: (a) ρ = 1%, (b) ρ = 25%, (c) ρ = 50%]
Example

- Inference over graph of regions

\[
\sum_c g_c(y_c) + \sum_{c,c'} g_{cc'}(y_c, y_{c'})
\]

\[
\sum_{k,c} g_c(k) \mu_c(k) + \sum_{i,i',c,c'} g_{cc'}(k, k') \mu_{cc'}(k, k')
\]

\[
\sum_k \mu_c(k) = 1, \quad \sum_{z_{i,i',c,c'}} \mu_{cc'}(k, k') = \mu_c(k)
\]

\[
\sum_k \mu_{cc'}(k, k') = \mu_c(k')
\]

\[
g_{cc'}(y_c, y_{c'})
\]

Problem

- A fixed set of regions is not going to work
- Need to use many candidate segmentation
Example

- Use a large set of (overlapping) segmentation of the image
- Special label 0 to indicate that a region does not have a label
- Need to add coverage constraint: a pixel belongs to exactly one labeled region

For all pixels $p$: $\sum_{c \in C} \sum_{i=1}^{L} \mu_c(i) = 1$

Note: all the previous sums are from 0 to $L$. This one is from 1 to $L$.

• Solved with relaxed LP
• Graph has cycles $\rightarrow$ Need modification (adding constraints on triplets of regions)
Approximate solutions: Sampling

- Rejection
- Importance
- MCMC
- Gibbs
- Sequential (particles)

- Exact solutions
- Approximate solutions:
  - Neighborhood search
  - Relaxation
  - Sampling
- Using larger supports and inference over sets of regions