Markov Random Fields for Computer Vision (Part 1)

Machine Learning Summer School (MLSS 2011)

Stephen Gould
stephen.gould@anu.edu.au

Australian National University

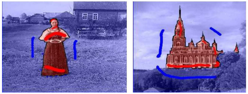
13–17 June, 2011

Pixel Labeling

Label every pixel in an image with a class label from some pre-defined set, i.e., $y_p \in \mathcal{L}$.

Pixel Labeling

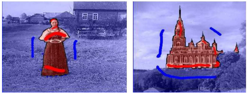
Label every pixel in an image with a class label from some pre-defined set, i.e., $y_p \in \mathcal{L}$.



Interactive figure-ground segmentation (Boykov and Jolly, 2001; Boykov and Funka-Lea, 2006)

Pixel Labeling

Label every pixel in an image with a class label from some pre-defined set, i.e., $y_p \in \mathcal{L}$.

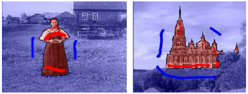


Interactive figure-ground segmentation (Boykov and Jolly, 2001; Boykov and Funka-Lea, 2006)

Surface context (Hoiem et al., 2005)

Pixel Labeling

Label every pixel in an image with a class label from some pre-defined set, i.e., $y_p \in \mathcal{L}$.



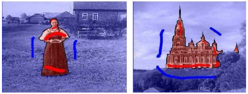
Interactive figure-ground segmentation (Boykov and Jolly, 2001; Boykov and Funka-Lea, 2006)

Surface context (Hoiem et al., 2005)

Semantic labeling (He et al., 2004; Shotton et al., 2006; Gould et al., 2009)

Pixel Labeling

Label every pixel in an image with a class label from some pre-defined set, i.e., $y_p \in \mathcal{L}$.



Interactive figure-ground segmentation (Boykov and Jolly, 2001; Boykov and Funka-Lea, 2006)

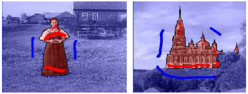
Surface context (Hoiem et al., 2005)

Semantic labeling (He et al., 2004; Shotton et al., 2006; Gould et al., 2009)

Stereo matching (Scharstein and Szeliski, 2002)

Pixel Labeling

Label every pixel in an image with a class label from some pre-defined set, i.e., $y_p \in \mathcal{L}$.



Interactive figure-ground segmentation (Boykov and Jolly, 2001; Boykov and Funka-Lea, 2006)

Surface context (Hoiem et al., 2005)

Semantic labeling (He et al., 2004; Shotton et al., 2006; Gould et al., 2009)

Stereo matching (Scharstein and Szeliski, 2002)

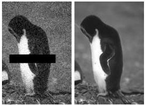


Image denoising (Felzenszwalb and Huttenlocher, 2004; Szeliski et al., 2008)

Digital Photo Montage

(Agarwala et al., 2004)

Digital Photo Montage

demonstration

Tutorial Overview

- **Part 1.** Pairwise conditional Markov random fields for the pixel labeling problem (45 minutes)
- **Part 2.** Pseudo-boolean functions and graph-cuts (1 hour)
- **Part 3.** Higher-order terms and inference as integer programming (30 minutes)

please ask lots of questions

Probability Review

Bayes Rule

$$\underbrace{P(\mathbf{y} \mid \mathbf{x})}_{\text{posterior}} = \frac{\overbrace{P(\mathbf{x} \mid \mathbf{y})}^{\text{likelihood}} \cdot \overbrace{P(\mathbf{y})}^{\text{prior}}}{P(\mathbf{x})}$$

Maximum a Posteriori (MAP) inference: $\mathbf{y}^\star = \operatorname{argmax}_{\mathbf{y}} P(\mathbf{y} \mid \mathbf{x})$.

Probability Review

Bayes Rule

$$\underbrace{P(\mathbf{y} \mid \mathbf{x})}_{\text{posterior}} = \frac{\overbrace{P(\mathbf{x} \mid \mathbf{y})}^{\text{likelihood}} \cdot \overbrace{P(\mathbf{y})}^{\text{prior}}}{P(\mathbf{x})}$$

Maximum a Posteriori (MAP) inference: $\mathbf{y}^\star = \operatorname{argmax}_{\mathbf{y}} P(\mathbf{y} \mid \mathbf{x})$.

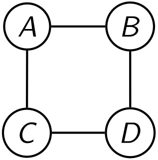
Conditional Independence

Random variables $\mathbf{y}$ and $\mathbf{x}$ are *conditionally independent* given $\mathbf{z}$ if $P(\mathbf{y}, \mathbf{x} \mid \mathbf{z}) = P(\mathbf{y} \mid \mathbf{z}) P(\mathbf{x} \mid \mathbf{z})$.

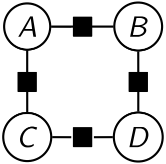
Graphical Models

We can exploit conditional independence assumptions to represent probability distributions in a way that is both *compact* and *efficient* for inference.

This tutorial is all about one particular representation, called a Markov Random Field (MRF), and the associated inference algorithms that are used in computer vision.

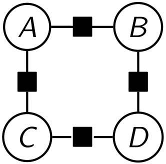


$a \perp\!\!\!\perp d \mid b, c$



$$\frac{1}{Z}\Psi(a, b)\Psi(b, d)\Psi(d, c)\Psi(c, a)$$

Graphical Models



$$\begin{aligned}\mathrm{P}(a,b,c,d) &= \frac{1}{Z}\Psi(a,b)\Psi(b,d)\Psi(d,c)\Psi(c,a) \\ &= \frac{1}{Z}\exp\left\{-\psi(a,b)-\psi(b,d)-\psi(d,c)-\psi(c,a)\right\}\end{aligned}$$

where $\psi = -\log\Psi$.

Energy Functions

Let **x** be some observations (i.e., features from the image) and let $\mathbf{y} = (y_1, \ldots, y_n)$ be a vector of random variables. Then we can write the conditional probability of **y** given **x** as

$$\mathrm{P}(\mathbf{y} \mid \mathbf{x}) = \frac{1}{Z(\mathbf{x})} \exp\left\{-E(\mathbf{y}; \mathbf{x})\right\}$$

where $Z(\mathbf{x}) = \sum_{\mathbf{y} \in \mathcal{L}^n} \exp\left\{-E(\mathbf{y}; \mathbf{x})\right\}$ is called the *partition function*.

Energy Functions

Let $\mathbf{x}$ be some observations (i.e., features from the image) and let $\mathbf{y} = (y_1, \ldots, y_n)$ be a vector of random variables. Then we can write the conditional probability of $\mathbf{y}$ given $\mathbf{x}$ as

$$\mathrm{P}(\mathbf{y} \mid \mathbf{x}) = \frac{1}{Z(\mathbf{x})} \exp\{-E(\mathbf{y}; \mathbf{x})\}$$

where $Z(\mathbf{x}) = \sum_{\mathbf{y} \in \mathcal{L}^n} \exp\{-E(\mathbf{y}; \mathbf{x})\}$ is called the *partition function*.

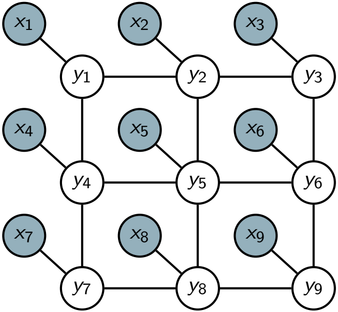
The *energy function* $E(\mathbf{y}; \mathbf{x})$ usually has some structured form:

$$E(\mathbf{y}; \mathbf{x}) = \sum_c \psi_c(\mathbf{y}_c; \mathbf{x})$$

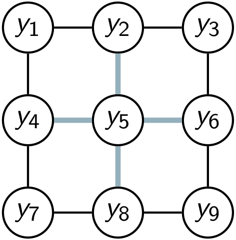
where $\psi_c(\mathbf{y}_c; \mathbf{x})$ are *clique potentials* defined over a subset of random variables $\mathbf{y}_c \subseteq \mathbf{y}$.

Conditional Markov Random Fields

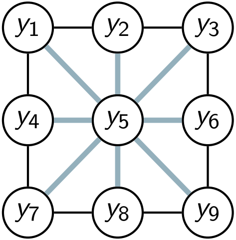
$$\begin{aligned} E(\mathbf{y}; \mathbf{x}) &= \sum_c \psi_c(\mathbf{y}_c; \mathbf{x}) \\ &= \underbrace{\sum_{i \in \mathcal{V}} \psi_i^U(y_i; \mathbf{x})}_{\text{unary}} + \underbrace{\sum_{ij \in \mathcal{E}} \psi_{ij}^P(y_i, y_j; \mathbf{x})}_{\text{pairwise}} + \underbrace{\sum_{c \in \mathcal{C}} \psi_c^H(\mathbf{y}_c; \mathbf{x})}_{\text{higher-order}}. \end{aligned}$$



Pixel Neighbourhoods



4-connected, $\mathcal{N}_4$



8-connected, $\mathcal{N}_8$

Binary MRF Example

Consider the following energy function for two binary random variables, y_1 and y_2.

0	5
1	2

0	1
1	3

	0	1
0	0	3
1	4	0

$$E(y_1, y_2) = \psi_1(y_1) + \psi_2(y_2) + \psi_{12}(y_1, y_2)$$

Binary MRF Example

Consider the following energy function for two binary random variables, y_1 and y_2.

$$
\begin{array}{c|c|}
 & \\
0 & 5 \\
1 & 2
\end{array}
\qquad
\begin{array}{c|c|}
 & \\
0 & 1 \\
1 & 3
\end{array}
\qquad
\begin{array}{c|c|c|}
 & 0 & 1 \\
0 & 0 & 3 \\
1 & 4 & 0
\end{array}
$$

$$
\begin{aligned}
E(y_1, y_2) &= \psi_1(y_1) + \psi_2(y_2) + \psi_{12}(y_1, y_2) \\
&= \underbrace{5\bar{y}_1 + 2y_1}_{\psi_1} \\
&\qquad + \underbrace{\bar{y}_2 + 3y_2}_{\psi_2} \\
&\qquad\quad + \underbrace{3\bar{y}_1 y_2 + 4y_1\bar{y}_2}_{\psi_{12}}
\end{aligned}
$$

where $\bar{y}_1 = 1 - y_1$ and $\bar{y}_2 = 1 - y_2$.

Binary MRF Example

Consider the following energy function for two binary random variables, y_1 and y_2.

0	5
1	2

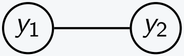
0	1
1	3

	0	1
0	0	3
1	4	0

$$
\begin{aligned}
E(y_1, y_2) &= \psi_1(y_1) + \psi_2(y_2) + \psi_{12}(y_1, y_2) \\
&= \underbrace{5\bar{y}_1 + 2y_1}_{\psi_1} \\
&\qquad + \underbrace{\bar{y}_2 + 3y_2}_{\psi_2} \\
&\qquad\quad + \underbrace{3\bar{y}_1 y_2 + 4y_1\bar{y}_2}_{\psi_{12}}
\end{aligned}
$$

where $\bar{y}_1 = 1 - y_1$ and $\bar{y}_2 = 1 - y_2$.

Graphical Model



Probability Table

y_1	y_2	E	P
0	0	6	0.244
0	1	11	0.002
1	0	7	0.090
1	1	5	0.664

Compactness of Representation

Consider a 1 mega-pixel image, e.g., 1000×1000 pixels. We want to annotate each pixel with a label from $\mathcal{L}$. Let $L = |\mathcal{L}|$.

- There are L^{10^6} possible ways to label such an image.
- A naive encoding—i.e., one big table—would require $L^{10^6} - 1$ parameters.
- A pairwise MRF over $\mathcal{N}_4$ requires $10^6 L$ parameters for the unary terms and $2 \times 1000 \times (1000 - 1)L^2$ parameters for the pairwise terms, i.e., $O(10^6 L^2)$. Even less are required if we share parameters.

Inference and Energy Minimization

We are usually interested in finding the most probable labeling,

$$\mathbf{y}^\star = \operatorname*{argmax}_{\mathbf{y}} \mathrm{P}\left(\mathbf{y} \mid \mathbf{x}\right) = \operatorname*{argmin}_{\mathbf{y}} E\left(\mathbf{y}; \mathbf{x}\right).$$

This is known as *maximum a posteriori* (MAP) inference or *energy minimization*.

Inference and Energy Minimization

We are usually interested in finding the most probable labeling,

$$\mathbf{y}^\star = \operatorname*{argmax}_{\mathbf{y}} \mathrm{P}\left(\mathbf{y} \mid \mathbf{x}\right) = \operatorname*{argmin}_{\mathbf{y}} E\left(\mathbf{y}; \mathbf{x}\right).$$

This is known as *maximum a posteriori* (MAP) inference or *energy minimization*.

A number of techniques can be used to find $\mathbf{y}^\star$, including:

- message-passing (dynamic programming)
- integer programming (part 3)
- graph-cuts (part 2)

However, in general, inference is NP-hard.

Characterizing Markov Random Fields

Markov random fields can be categorized via a number of different dimensions:

- **Label space:** binary vs. multi-label; homogeneous vs. heterogeneous.
- **Order:** unary vs. pairwise vs. higher-order.
- **Structure:** chain vs. tree vs. grid vs. general graph; neighbourhood size.
- **Potentials:** submodular, convex, compressible.

These all affect tractability of inference.

Markov Random Fields for Pixel Labeling

$$\mathrm{P}\left(\mathbf{y} \mid \mathbf{x}\right) \propto \mathrm{P}\left(\mathbf{x} \mid \mathbf{y}\right) \mathrm{P}\left(\mathbf{y}\right) = \exp\left\{-E\left(\mathbf{y}; \mathbf{x}\right)\right\}$$

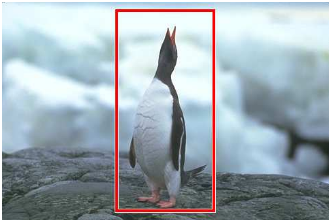
$$\overbrace{E\left(\mathbf{y}; \mathbf{x}\right)}^{\text{energy}} = \underbrace{\sum_{i \in \mathcal{V}} \psi_i^U(y_i; \mathbf{x})}_{\text{unary}} + \underbrace{\lambda \sum_{ij \in \mathcal{N}_8} \psi_{ij}^P(y_i, y_j; \mathbf{x})}_{\text{pairwise}}$$

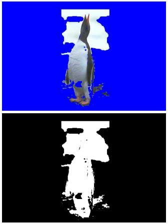
$$\psi_i^U(y_i; \mathbf{x}) = \overbrace{-\sum_{\ell \in \mathcal{L}} [\![y_i = \ell]\!] \log \mathrm{P}\left(x_i \mid \ell\right)}^{\text{likelihood}}$$

$$\psi_{ij}^P(y_i, y_j; \mathbf{x}) = \underbrace{[\![y_i \neq y_j]\!]}_{\text{Potts prior}}$$

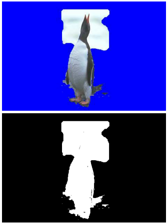
Here the prior acts to "smooth" predictions (independent of $\mathbf{x}$).

Prior Strength

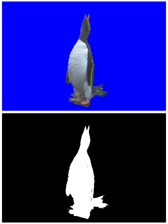




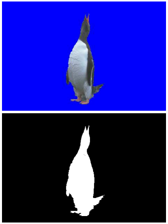
$\lambda = 1$



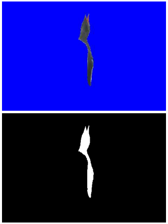
$\lambda = 4$



$\lambda = 16$



$\lambda = 128$

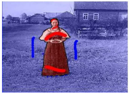
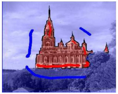


$\lambda = 1024$

Interactive Segmentation Model

- **Label space:** foreground or background

$$\mathcal{L} = \{0, 1\}$$

- **Unary term:** Gaussian mixture models for foreground and background

$$\psi_i^U(y_i; \mathbf{x}) = \sum_k \tfrac{1}{2}|\Sigma_k| + \tfrac{1}{2}\left(x_i - \mu_k\right)^T \Sigma_k^{-1}\left(x_i - \mu_k\right) - \log \lambda_k$$

- **Pairwise term:** contrast-dependent smoothness prior

$$\psi_{ij}^P(y_i, y_j; \mathbf{x}) = \begin{cases} \lambda_0 + \lambda_1 \exp\left(-\frac{\|x_i - x_j\|^2}{2\beta}\right), & \text{if } y_i \neq y_j \\ 0, & \text{otherwise} \end{cases}$$

Geometric/Semantic Labeling Model

- **Label space:** pre-defined label set, e.g.,

$$\mathcal{L} = \{\mathrm{sky}, \mathrm{tree}, \mathrm{grass}, \ldots\}$$

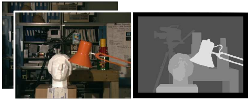
- **Unary term:** Boosted decision-tree classifiers over "texton-layout" features [Shotton et al., 2006]

$$\psi_i^U(y_i = \ell; \mathbf{x}) = \theta_\ell \log \mathrm{P}\left(\phi_i(\mathbf{x}) \mid \ell\right)$$

- **Pairwise term:** contrast-dependent smoothness prior

$$\psi_{ij}^P(y_i, y_j; \mathbf{x}) = \begin{cases} \lambda_0 + \lambda_1 \exp\left(-\frac{\|x_i - x_j\|^2}{2\beta}\right), & \text{if } y_i \neq y_j \\ 0, & \text{otherwise} \end{cases}$$

Stereo Matching Model



- **Label space:** pixel disparity

$$\mathcal{L} = \{0, 1, \ldots, 127\}$$

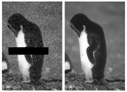
- **Unary term:** sum of absolute differences (SAD) or normalized cross-correlation (NCC)

$$\psi_i^U(y_i; \mathbf{x}) = \sum_{(u,v)\in W} \left|\mathbf{x}_{\mathrm{left}}(u, v) - \mathbf{x}_{\mathrm{right}}(u - y_i, v)\right|$$

- **Pairwise term:** "discontinuity preserving" prior

$$\psi_{ij}^P(y_i, y_j) = \max\left\{|y_i - y_j|, d_{\max}\right\}$$

Image Denoising Model



- **Label space:** pixel intensity or colour

$$\mathcal{L} = \{0, 1, \ldots, 255\}$$

- **Unary term:** square distance

$$\psi_i^U(y_i; \mathbf{x}) = \|y_i - x_i\|^2$$

- **Pairwise term:** truncated L_2 distance

$$\psi_{ij}^P(y_i, y_j) = \max\left\{\|y_i - y_j\|^2, d_{\max}^2\right\}$$

Digital Photo Montage Model

- **Label space:** image index

$$\mathcal{L} = \{1, 2, \ldots, K\}$$

- **Unary term:** none!
- **Pairwise term:** seem penalty

$$\psi^P_{ij}(y_i, y_j; \mathbf{x}) = \|\mathbf{x}_{y_i}(i) - \mathbf{x}_{y_j}(i)\| + \|\mathbf{x}_{y_i}(j) - \mathbf{x}_{y_j}(j)\|$$

(or edge-normalized variant)

end of part 1