Consistency Potentials for Scene Understanding: from Pairwise to Higher-order

Stephen Gould, ANU

Graphical Models for Scene Understanding: Challenges and Perspectives, ICCV 2013
2 December 2013
Multi-class Pixel Labeling

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

- **FG/BG segmentation**
  - [Boykov and Jolly, 2001; Rother et al., 2004]

- **Geometric context**
  - [Hoiem et al., 2005]

- **Semantic segmentation**
  - [He et al., 2004; Shotton et al., 2006; Gould et al., 2009]

- **Stereo reconstruction**
  - [Scharstein and Szeliski, 2005]

- **Digital photo montage**
  - [Agarwala et al., 2004]

- **Denoising and Inpainting**
Pixelwise Pixel Labeling

\[ P(y \mid x) = \prod_i P(y_i \mid x_i) \]
Pixelwise Pixel Labeling

\[ P(y \mid x) = \prod_i P(y_i \mid x_i) \]
Introducing (Data Dependent) Priors

- Options for improving accuracy:
  - (i) use more features, more data, more complex models
  - (ii) use priors to guide the labeling towards a more plausible solution

- Most common priors enforce smoothness (e.g., pairwise)
- Data dependent priors can take into account image features
Conditional Markov Random Fields

\[ P(y \mid x) = \frac{1}{Z(x)} \exp \{-E(y; x)\} \]

energy function

\[ E(y; x) = \sum_c \psi_c(y_c) \]
\[ = \sum_i \psi^U_i (y_i; x_i) + \sum_{ij} \psi^P_{ij} (y_i, y_j) + \sum_c \psi^H_c (y_c) \]
Binary CRFs and Pseudo-Boolean Fcns

- A pseudo-Boolean function is a mapping

\[ f : \{0, 1\}^n \rightarrow \mathbb{R} \]

- Can be written (uniquely) as a \textit{multi-linear polynomial} or (non-uniquely) in \textit{posiform}

- A binary pairwise MRF is just a quadratic (bilinear) pseudo-Boolean function (QPBF)

- \textit{Submodular} QPBFs can be minimized by graph cuts
  - identified by negative coefficients on pairwise terms

[ Boros and Hammer, 2001 ]
Graph-Cuts

- construct a graph where every st-cut corresponds to a joint assignment to the variables
- the cost of the cut should equal the energy of the assignment
- the minimum-cut then corresponds to the energy minimizing assignment

\[
E(u, v) = 2u + 5\bar{u} + 3v + \bar{v} + 3\bar{u}v + 4u\bar{v}
\]
Graph-Cuts

- construct a graph where every *st-cut* corresponds to a joint assignment to the variables
- the *cost* of the cut should equal the energy of the assignment
- the *minimum-cut* then corresponds to the energy minimizing assignment

\[ E(u, v) = 6\bar{u} + 5v + 7uv\bar{v} \]
Energy Minimization via Graph-Cuts

• Start with a pixel labeling problem
• Formulate as multi-label CRF inference
• (move-making: $\alpha$-expansion, $\alpha\beta$-swap, ICM)
  – Convert to a sequence of binary pairwise CRF inference problems
  – Write CRF as a quadratic pseudo-Boolean function
  – Solve by finding the minimum cut (maximum flow)

• (relaxation)
• (approximation)

[Boykov et al., 2001]
Contrast Sensitive Pairwise Smoothness

\[
\psi_{ij}^P(y_i, y_j) = \begin{cases} 
0 & \text{if } y_i = y_j \\
\frac{\lambda}{d_{ij}} \exp \left\{ - \frac{\|x_i - x_j\|^2}{2\beta} \right\} & \text{if } y_i \neq y_j
\end{cases}
\]
Pairwise Smoothness Results

Image

Independent (unary only)

Pairwise CRF
Why Not Use Superpixels?

• **Ideal:** Suppose an oracle told us which pixels belong together. Then all we would need to do is predict the class labels.

• **Problem:** no over-segmentation algorithm is perfect. Even if they were, our label predictions may be wrong.

• **Solution:** use superpixels as soft constraints.
Generalized Potts Model

- **Pairwise Potts Potential:**
  - Penalize if two pixels disagree

- **Higher-order Potts Potential:**
  - Penalize if **any** two pixels in a clique disagree
  - Penalty paid **once**

[Kohli et al., 2007]
Higher-order Smoothness Potentials

**Generalized Potts**
[Kohli et al., 2007]

**Robust Potts Model**
[Kohli et al., 2008; Ladicky et al., 2009]

**Arbitrary Concave**
[Kohli and Kumar, 2010; Gould, 2011]
Binary Lower Linear Envelope MRFs

\[ \psi^H(y) = \min_k \left\{ a_k \sum_i y_i + b_k \right\} \]
Inference (Binary Case)

\[
\psi^H(y) = \min \left\{ \eta \sum_i y_i, M \right\}
\]

\[
\min_y \psi^H(y) = \min_{y,z} Mz + (1 - z)\eta \sum_i y_i
\]

\[
= \min_{y,z} Mz + \sum_i \eta y_i \bar{z}
\]
Inference (Binary Case)

\[ \psi^H(y) = \min_k \left\{ a_k \sum_i y_i + b_k \right\} \]

\[
\min_y \psi^H(y) = \min_{y,z} a_1 \sum_i y_i + b_1 + \sum_k z_k \left( (a_k - a_{k-1}) \sum_i y_i + (b_k - b_{k-1}) \right)
\]
Inference (Full CRF---Binary Case)

\[
E(y; x) = \sum_i \psi_i^U(y_i; x_i) + \sum_{ij} \psi_{ij}^P(y_i, y_j) + \sum_c \psi_c^H(y_c)
\]

sum of submodular potentials is submodular
Learning (Binary Case)

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{2} \| \theta \|^2 + \frac{C}{T} \sum_t \xi_t \\
\text{subject to} & \quad \theta^T \delta \phi_t(y) \geq \Delta_t(y) - \xi_t \\
& \quad \xi_t \geq 0 \\
& \quad D^2 \theta \geq 0
\end{align*}
\]

\[
(\phi(y))_m = \begin{cases} 
1 & \text{if } \sum_i y_i = m \\
0 & \text{otherwise}
\end{cases}
\]
Learning (Binary Case)

\[ a_k = \theta_k - \theta_{k-1} \]

\[ b_k = \theta_k - a_k k \]

\[ (\phi(y))_m = \begin{cases} 1 & \text{if } \sum_i y_i = m \\ 0 & \text{otherwise} \end{cases} \]
Learning Variants (Binary Case)

• Sampled lower linear envelope (2\textsuperscript{nd} order)
  \[ a_k = \theta_k - \theta_{k-1} \]
  \[ b_k = \theta_k - a_k k \]

• Slope (1\textsuperscript{st} order)
  \[ b_1 = 0 \]
  \[ a_k = \theta_k \quad (\theta_k \leq \theta_{k-1}) \]

• Curvature (0\textsuperscript{th} order)
  \[ b_1 = 0, \ a_1 = \theta_1 \]
  \[ a_k = \theta_k + a_{k-1} \quad (\theta_k \leq 0) \]
Weighted Smoothness Potentials

\[ \psi^H(y; x) = \lambda(x) \min_k \left\{ a_k \sum_i \frac{w_i(x)}{W} y_i + b_k \right\} \]

- \( \psi^H \) is the weighted smoothness potential function.
- \( \lambda(x) \) is the penalty term.
- \( a_k \) and \( b_k \) are coefficients.
- \( w_i(x) \) and \( W \) are weights.
- \( y_i \) are the observations.
Aside: Relationship to RBM

\[
E(y, z) = - \sum_i a_i y_i - \sum_j b_j z_j - \sum_{ij} w_{ij} y_i z_j
\]

**Restricted Boltzmann Machine**

\[
P(y) \propto \sum_z \exp \{-E(y, z)\}
\]

- \(a_i, b_j\) arbitrary
- \(w_{ij}\) arbitrary

**Lower Linear Envelope**

\[
P(y) \propto \max_z \exp \{-E(y, z)\}
\]

- \(a_i, b_j\) arbitrary
- \(w_{ij}\) positive
Defining the Higher-order Cliques

oversegment to produce contiguous superpixels

, ... , importantly, higher-order cliques can overlap

cluster (e.g., k-means) to produce non-contiguous regions
Extending to Multiple Labels

- Aggregation by summation

\[
\psi^H(y) = \sum_{\ell \in \mathcal{L}} \min_k \left\{ a_k \sum_i [[y_i = \ell]] + b_k \right\}
\]

- Aggregation by minimization

\[
\psi^H(y) = \min_k \left\{ a_k \sum_i [[y_i = \ell_k]] + b_k \right\}
\]

- Move-making (approximate) inference

[Park and Gould, ECCV 2012]
Dual Decomposition Inference

\[ E(y) = \sum \psi^H(y) \]

\[ E_{\text{slave}}(y) = \psi^H(y) + \sum_i \lambda_i(y_i) \]

[Komodakis et al., PAMI 2010]
Dual Decomposition Inference (Details)

\[
\min_y \min_k \left\{ a_k \sum_i \left[ y_i = \ell_k \right] + b_k \right\} + \sum_i \lambda_i(y_i)
\]

\[
= \min_y \min_k \left\{ \sum_i \left( a_k + \lambda_i(\ell_k) \right) \left[ y_i = \ell_k \right] + \sum_i \left( \min_{\ell \neq \ell_k} \lambda_i(\ell) \right) \left[ y_i \neq \ell_k \right] + b_k \right\}
\]

- cost of setting \(i\)-th variable to label for \(k\)-th linear function
- cost of not setting \(i\)-th variable to label for \(k\)-th linear function
Semantic Segmentation Results

increasing pairwise prior

increasing higher-order prior
Semantic Segmentation Results

increasing pairwise prior

increasing higher-order prior
Semantic Segmentation Results

Increasing pairwise prior

Increasing higher-order prior
Semantic Segmentation Results

Increasing pairwise prior

Increasing higher-order prior
Semantic Segmentation Results

increasing pairwise prior

increasing higher-order prior
“GrabCut” Results

increasing pairwise prior

increasing higher-order prior
“GrabCut” Results

Increasing pairwise prior

Increasing higher-order prior
“GrabCut” Results

increasing pairwise prior

increasing higher-order prior
“GrabCut” Results

increasing pairwise prior

increasing higher-order prior
Higher-order Matching Potentials

\[ \psi(y) = \min \left\{ \eta \sum_{i,j \in M} [y_i \neq y_j] w_{ij}, M \right\} \]

[Source: Gould, CVPR 2012]
Inference with Matching Potentials

\[ \psi(y) = \min_z \eta \sum_{i,j \in M} w_{ij} z y_i (1 - y_j) + w_{ij} z (1 - y_i) y_j + M (1 - z) \]

- **Problem**: non-submodular terms (in move making steps when labels already agree before the move)

- **Solution**: approximate with (tight) upper-bound by setting \( z = 1 \)

[Gould, CVPR 2012]
Cross-Image Consistency Potentials

\[ E(y_1, y_2; x_1, x_2) = E(y_1; x_1) + E(y_2; x_2) + \sum_c \psi^{\text{MATCH}}(P_c, Q_c) \]
Cross-Image Results

unary

pairwise

match

+ more
Summary and Challenges for (Higher-order) Consistency Potentials

• Priors/constraints provide a mechanism for scene understanding that simply adding more features cannot

• Many other (higher-order) consistency potentials, e.g.,
  – Cardinality [Tarlow et al., 2010], label co-occurrence [Ladicky et al., 2010], label cost [Delong et al., 2010], densely connected [Krahenbuhl and Koltun, 2011], connectivity [Vincete et al., 2008]

• Biggest challenge is in learning the parameters of these
  – Currently, piecewise learning and cross-validation works best

• Opportunities: higher-order (supermodular) loss functions [Tarlow and Zemel, 2011; Pletscher and Kohli, 2012]