Segmentation: MRFs and Graph Cuts

Computer Vision
CS 143, Brown

James Hays

Many slides from Kristin Grauman and Derek Hoiem
Today’s class

• Segmentation and Grouping
• Inspiration from human perception
  – Gestalt properties
• MRFs
• Segmentation with Graph Cuts
Grouping in vision

- Goals:
  - Gather features that belong together
  - Obtain an intermediate representation that compactly describes key image or video parts
Examples of grouping in vision

Determine image regions

Group video frames into shots

Object-level grouping

Figure-ground

Slide: Kristin Grauman
Grouping in vision

• Goals:
  – Gather features that belong together
  – Obtain an intermediate representation that compactly describes key image (video) parts

• Top down vs. bottom up segmentation
  – Top down: pixels belong together because they are from the same object
  – Bottom up: pixels belong together because they look similar

• Hard to measure success
  – What is interesting depends on the app.
What things should be grouped?
What cues indicate groups?
Gestalt psychology or Gestaltism

• German: *Gestalt* - "form" or "whole"
• Berlin School, early 20th century
  – Kurt Koffka, Max Wertheimer, and Wolfgang Köhler
• Gestalt: whole or group
  – Whole is greater than sum of its parts
  – Relationships among parts can yield new properties/features
• Psychologists identified series of factors that predispose set of elements to be grouped (by human visual system)
Gestaltism

The Muller-Lyer illusion
We perceive the interpretation, not the senses
Principles of perceptual organization

- Not grouped
- Proximity
- Similarity
- Similarity
- Common Fate
- Common Region

From Steve Lehar: The Constructive Aspect of Visual Perception
Principles of perceptual organization

Parallelism

Symmetry

Continuity

Closure
Similarity
Symmetry

Common fate

Image credit: Arthus-Bertrand (via F. Durand)
Proximity
Grouping by invisible completion

From Steve Lehar: The Constructive Aspect of Visual Perception
Gestalt cues

- Good intuition and basic principles for grouping
- Basis for many ideas in segmentation and occlusion reasoning
- Some (e.g., symmetry) are difficult to implement in practice
These intensities define the three groups. We could label every pixel in the image according to which of these primary intensities it is. i.e., segment the image based on the intensity feature. What if the image isn’t quite so simple?
• Now how to determine the three main intensities that define our groups?
• We need to \textit{cluster}. 
Clustering

• With this objective, it is a “chicken and egg” problem:
  – If we knew the **cluster centers**, we could allocate points to groups by assigning each to its closest center.
  – If we knew the **group memberships**, we could get the centers by computing the mean per per group.
Smoothing out cluster assignments

- Assigning a cluster label per pixel may yield outliers:

  ![Original Image](image1)
  ![Labeled Image](image2)

  labeled by cluster center’s intensity

- How to ensure they are spatially smooth?
Solution

Encode dependencies between pixels

Normalizing constant

\[ P(y; \theta, data) = \frac{1}{Z} \prod_{i=1..N} f_1(y_i; \theta, data) \prod_{i,j \in \text{edges}} f_2(y_i, y_j; \theta, data) \]

Labels to be predicted  Individual predictions  Pairwise predictions
Writing Likelihood as an “Energy”

\[
P(y; \theta, data) = \frac{1}{Z} \prod_{i=1..N} p_1(y_i; \theta, data) \prod_{i,j \in \text{edges}} p_2(y_i, y_j; \theta, data)
\]

\[
\text{Energy}(y; \theta, data) = \sum_{i} \psi_1(y_i; \theta, data) + \sum_{i,j \in \text{edges}} \psi_2(y_i, y_j; \theta, data)
\]

“Cost” of assignment \(y_i\)

“Cost” of pairwise assignment \(y_i, y_j\)
Markov Random Fields

Node $y_i$: pixel label

Edge: constrained pairs

Cost to assign a label to each pixel

Cost to assign a pair of labels to connected pixels

$$\text{Energy}(y; \theta, \text{data}) = \sum_i \psi_1(y_i; \theta, \text{data}) + \sum_{i,j \in \text{edges}} \psi_2(y_i, y_j; \theta, \text{data})$$
Markov Random Fields

• Example: “label smoothing” grid

\[
\begin{pmatrix}
0 & 0 & K \\
0 & K & 0 \\
K & 0 & 0
\end{pmatrix}
\]

Unary potential

0: -logP(y_i = 0 ; data)
1: -logP(y_i = 1 ; data)

Pairwise Potential

\[
\psi_1(y_i; \theta, \text{data}) + \sum_{i,j \in \text{edges}} \psi_2(y_i, y_j; \theta, \text{data})
\]
Solving MRFs with graph cuts

\[ \text{Energy}(y; \theta, \text{data}) = \sum_i \psi_1(y_i; \theta, \text{data}) + \sum_{i, j \in \text{edges}} \psi_2(y_i, y_j; \theta, \text{data}) \]
Solving MRFs with graph cuts

\[
\text{Energy}(\mathbf{y}; \theta, \text{data}) = \sum_i \psi_1(y_i; \theta, \text{data}) + \sum_{i,j \in \text{edges}} \psi_2(y_i, y_j; \theta, \text{data})
\]
GrabCut segmentation

User provides rough indication of foreground region.

Goal: Automatically provide a pixel-level segmentation.
Grab cuts and graph cuts

<table>
<thead>
<tr>
<th>User Input</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magic Wand (198?)</td>
<td>Regions</td>
</tr>
<tr>
<td>Intelligent Scissors Mortensen and Barrett (1995)</td>
<td>Boundary</td>
</tr>
<tr>
<td>GrabCut</td>
<td>Regions &amp; Boundary</td>
</tr>
</tbody>
</table>

Source: Rother
Colour Model

Gaussian Mixture Model (typically 5-8 components)

Source: Rother
Graph cuts

Boykov and Jolly (2001)

**Cut:** separating source and sink; **Energy:** collection of edges

**Min Cut:** Global minimal energy in polynomial time

Source: Rother
Colour Model

Gaussian Mixture Model (typically 5-8 components)

Source: Rother
GrabCut segmentation

1. Define graph
   - usually 4-connected or 8-connected

2. Define unary potentials
   - Color histogram or mixture of Gaussians for background and foreground
     \[
     unary\_potential(x) = -\log \left( \frac{P(c(x); \theta_{\text{foreground}})}{P(c(x); \theta_{\text{background}})} \right)
     \]

3. Define pairwise potentials
   \[
   edge\_potential(x, y) = k_1 + k_2 \exp \left\{ -\frac{\|c(x) - c(y)\|^2}{2\sigma^2} \right\}
   \]

4. Apply graph cuts

5. Return to 2, using current labels to compute foreground, background models
What is easy or hard about these cases for graphcut-based segmentation?
Easier examples

GrabCut – Interactive Foreground Extraction
More difficult Examples

Camouflage & Low Contrast

Initial Rectangle

Initial Result

Fine structure

Harder Case

GrabCut – Interactive Foreground Extraction
Lazy Snapping (Li et al. SG 2004)
Using graph cuts for recognition

TextonBoost (Shotton et al. 2009 IJCV)
Using graph cuts for recognition

Unary Potentials + edge potentials → Alpha Expansion

TextonBoost (Shotton et al. 2009 IJCV)
Limitations of graph cuts

- Associative: edge potentials penalize different labels

  Must satisfy

  \[ E^{i,j}(0,0) + E^{i,j}(1,1) \leq E^{i,j}(0,1) + E^{i,j}(1,0) \]

- If not associative, can sometimes clip potentials

- Approximate for multilabel
  - Alpha-expansion or alpha-beta swaps
Graph cuts: Pros and Cons

• Pros
  – Very fast inference
  – Can incorporate data likelihoods and priors
  – Applies to a wide range of problems (stereo, image labeling, recognition)

• Cons
  – Not always applicable (associative only)
  – Need unary terms (not used for generic segmentation)

• Use whenever applicable
More about MRFs/CRFs

• Other common uses
  – Graph structure on regions
  – Encoding relations between multiple scene elements

• Inference methods
  – Loopy BP or BP-TRW: approximate, slower, but works for more general graphs
Further reading and resources

• Graph cuts
  – Classic paper: What Energy Functions can be Minimized via Graph Cuts? (Kolmogorov and Zabih, ECCV '02/PAMI '04)

• Belief propagation

• Normalized cuts and image segmentation (Shi and Malik)

• N-cut implementation
  [http://www.seas.upenn.edu/~timothee/software/ncut/ncut.html](http://www.seas.upenn.edu/~timothee/software/ncut/ncut.html)
Next Class

• Gestalt grouping

• More segmentation methods
Recap of Grouping and Fitting
Edge and line detection

- Canny edge detector = smooth $\rightarrow$ derivative $\rightarrow$ thin $\rightarrow$ threshold $\rightarrow$ link

- Generalized Hough transform = points vote for shape parameters

- Straight line detector = canny + gradient orientations $\rightarrow$ orientation binning $\rightarrow$ linking $\rightarrow$ check for straightness
Robust fitting and registration

Key algorithms

- RANSAC, Hough Transform
Clustering

Key algorithm

- K-means
EM and Mixture of Gaussians

Tutorials:

http://www.cs.duke.edu/courses/spring04/cps196.1/.../EM/tomasiEM.pdf
http://www-clmc.usc.edu/~adsouza/notes/mix_gauss.pdf
Segmentation

- Mean-shift segmentation
  - Flexible clustering method, good segmentation

- Watershed segmentation
  - Hierarchical segmentation from soft boundaries

- Normalized cuts
  - Produces regular regions
  - Slow but good for oversegmentation

- MRFs with Graph Cut
  - Incorporates foreground/background/object model and prefers to cut at image boundaries
  - Good for interactive segmentation or recognition
Next section: Recognition

• How to recognize
  – Specific object instances
  – Faces
  – Scenes
  – Object categories