Lecture 18: Context and segmentation
Announcements

• Project 3 out soon
• Final project
When is context helpful?

- Local features
- Information
- Contextual features

Slide by Antonio Torralba
Is it just for small / blurry things?

A B C

Slide by Antonio Torralba
Is it just for small / blurry things?
Is it just for small / blurry things?
Context is hard to fight!

Thanks to Paul Viola for showing me these
more “Look-alikes”
Don’t even need to see the object
Don’t even need to see the object

Chance ~ 1/30000

Slide by Antonio Torralba
The influence of an object extends beyond its physical boundaries

But object can say a lot about the scene
Object priming

Increasing contextual information

Torralba, Sinha, Oliva, VSS 2001
Object priming

Inconsistent objects

Consistent objects

Consistent

Fobj=21 c/i

Fctx=4 7 12 21

Correct object recognition rate

Increasing contextual information

Torralba, Sinha, Oliva, VSS 2001
Why context is important?

- Typical answer: to “guess” small / blurry objects based on a prior
  - most current vision systems
So, you think it’s so simple?

What is this?

slide by Takeo Kanade
Why is this a car?
...because it’s on the road!

Why is this road?
Why is this a road?
Same problem in real scenes
Why context is important?

• Typical answer: to “guess” small / blurry objects based on a prior
  • Most current vision systems

• Deeper answer: to make sense of the visual world
  • much work yet to be done!
Why context is important?

- To resolve ambiguity
  - Even high-res objects can be ambiguous
    - e.g. there are more people than faces in the world!
  - There are 30,000+ objects, but only a few dozen can happen within a given scene

- To notice “unusual” things
  - Prevents mental overload

- To infer function of unknown object
# Sources of Context

<table>
<thead>
<tr>
<th>Source of Context</th>
<th>Contextual Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local Pixel Context</td>
<td>window surround, image neighborhoods, object boundary/shape</td>
</tr>
<tr>
<td>Scene Gist Context</td>
<td>global image statistics</td>
</tr>
<tr>
<td>Geometric Context</td>
<td>3D scene layout, support surface, surface orientations, occlusions, contact points, etc.</td>
</tr>
<tr>
<td>Semantic Context</td>
<td>event/activity depicted, scene category, objects present in the scene, keywords</td>
</tr>
<tr>
<td>Photogrammetric Context</td>
<td>camera height, orientation, focal length, lens distortion, radiometric response function</td>
</tr>
<tr>
<td>Illumination Context</td>
<td>sun direction, sky color, cloud cover, shadow contrast, etc</td>
</tr>
<tr>
<td>Weather Context</td>
<td>current/recent precipitation, wind speed/direction, temperature, the season, etc.</td>
</tr>
<tr>
<td>Geographic Context</td>
<td>GPS location, terrain type, land use category, elevation, population density, etc.</td>
</tr>
<tr>
<td>Temporal Context</td>
<td>nearby frames (if video), temporally proximal images, videos of similar scenes</td>
</tr>
<tr>
<td>Cultural Context</td>
<td>photographer bias, dataset selection bias, visual clichés, etc</td>
</tr>
</tbody>
</table>

Table 1. Taxonomy of various sources of contextual information.

from Divvala et al, CVPR 2009
Local Pixel Context

Wolf & Bileschi 2006

Dalal & Triggs 2005
Scene Gist Context

Global (low-dimensional) scene statistics

Figure 8: Images at a resolution of 32x32 pixels and the segmentations provided by the participants. Figure B shows some of the recognized objects cropped. Many of those objects become unrecognizable once they are extracted from the context.
Geometric Context

Geometric Context [Hoiem et al. ’2005]
Photogrammetric Context

Hoiem et al. 2006

Lalonde et al. 2008
Illumination and Weather Context

Illumination context (Lalonde et al)
Geographic Context

Hays & Efros 2008
Temporal Context

Liu et al. 2008

Gallagher et al 2008
Cultural Context

Photographer Bias

Society Bias
Flickr Paris
“uniformly sampled” Paris
…or Notre Dame
Why Photographers are Biased?

People want their pictures to be recognizable and/or interesting
Why Photographers are Biased?

People follow photographic conventions

VS.

Simon & Seitz 2008
“100 Special Moments” by Jason Salavon

Little Leaguer

Kids with Santa

The Graduate

Newlyweds
Empirical Evaluation of Context
Divvala, Hoiem, Hays, Efros, Hebert,
CVPR 2009
Change in Accuracies

<table>
<thead>
<tr>
<th>Type</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>97%</td>
</tr>
<tr>
<td>Large</td>
<td>15%</td>
</tr>
<tr>
<td>Occluded</td>
<td>79%</td>
</tr>
<tr>
<td>Non-Occluded</td>
<td>-5%</td>
</tr>
<tr>
<td>Difficult</td>
<td>57%</td>
</tr>
<tr>
<td>Total</td>
<td>47%</td>
</tr>
</tbody>
</table>

- Percentage change in average precisions before and after including context information
- Context helps in case of impoverished appearance
Detecting difficult objects

Start recognizing the scene

Torralba, Murphy, Freeman. NIPS 2004.
Detect first simple objects (reliable detectors) that provide strong contextual constraints to the target (screen -> keyboard -> mouse)

Torralba, Murphy, Freeman. NIPS 2004.
Detecting difficult objects (object parts)

Detect first simple objects (reliable detectors) that provide strong contextual constraints to the target (screen -> keyboard -> mouse)

Torralba, Murphy, Freeman. NIPS 2004.
...on the other hand
Who needs context anyway?
We can recognize objects even out of context

Slide by Antonio Torralba
From images to objects

What defines an object?

- Subjective problem, but has been well-studied
- Gestalt Laws seek to formalize this
  - proximity, similarity, continuation, closure, common fate
  - see notes by Steve Joordens, U. Toronto
Extracting objects

How could we do this automatically (or at least semi-automatically)?
The Gestalt school

- Grouping is key to visual perception
- Elements in a collection can have properties that result from relationships
  - “The whole is greater than the sum of its parts”

subjective contours

occlusion

familiar configuration

http://en.wikipedia.org/wiki/Gestalt_psychology
The ultimate Gestalt?
Gestalt factors

- These factors make intuitive sense, but are very difficult to translate into algorithms
Semi-automatic binary segmentation
Figure 2: Image demonstrating how the live-wire segment adapts and snaps to an object boundary as the free point moves (via cursor movement). The path of the free point is shown in white. Live-wire segments from previous free point positions \((t_0, t_1, \text{ and } t_2)\) are shown in green.
Intelligent Scissors [Mortensen 95]

• Approach answers a basic question
  – Q: how to find a path from seed to mouse that follows object boundary as closely as possible?

Figure 2: Image demonstrating how the live-wire segment adapts and snaps to an object boundary as the free point moves (via cursor movement). The path of the free point is shown in white. Live-wire segments from previous free point positions ($t_0$, $t_1$, and $t_2$) are shown in green.
GrabCut

Grabcut  [Rother et al., SIGGRAPH 2004]
Is user-input required?

Our visual system is proof that automatic methods are possible
  • classical image segmentation methods are automatic

Argument for user-directed methods?
  • only user knows desired scale/object of interest
Automatic graph cut [Shi & Malik]

**Fully-connected graph**
- node for every pixel
- link between every pair of pixels, \( p, q \)
- cost \( c_{pq} \) for each link
  - \( c_{pq} \) measures *similarity*
    - similarity is *inversely proportional* to difference in color and position
Segmentation by Graph Cuts

Break Graph into Segments

- Delete links that cross between segments
- Easiest to break links that have low cost (similarity)
  - similar pixels should be in the same segments
  - dissimilar pixels should be in different segments
Cuts in a graph

Link Cut
- set of links whose removal makes a graph disconnected
- cost of a cut:
  \[
  \text{cut}(A, B) = \sum_{p \in A, q \in B} c_{p,q}
  \]

Find minimum cut
- gives you a segmentation
But min cut is not always the best cut...
Cuts in a graph

Normalized Cut

- a cut penalizes large segments
- fix by normalizing for size of segments

\[ N\text{cut}(A, B) = \frac{\text{cut}(A, B)}{\text{volume}(A)} + \frac{\text{cut}(A, B)}{\text{volume}(B)} \]

- \( \text{volume}(A) = \text{sum of costs of all edges that touch } A \)
Interpretation as a Dynamical System

Treat the links as springs and shake the system
  • elasticity proportional to cost
  • vibration “modes” correspond to segments
    – can compute these by solving an eigenvector problem
Interpretation as a Dynamical System

Treat the links as springs and shake the system
- elasticity proportional to cost
- vibration “modes” correspond to segments
  - can compute these by solving an eigenvector problem
  - [http://www.cis.upenn.edu/~jshi/papers/pami_ncut.pdf](http://www.cis.upenn.edu/~jshi/papers/pami_ncut.pdf)
Color Image Segmentation