Scene Understanding

Introduction & Overview
Outline

• Motivation
• The problems
• Scene perception
• Approaches
Objects in context
Objects in context

<table>
<thead>
<tr>
<th>Context Conditions</th>
<th>Block Size (in Pixels)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>125 X 125</td>
</tr>
<tr>
<td></td>
<td>75 X 75</td>
</tr>
<tr>
<td></td>
<td>50 X 50</td>
</tr>
<tr>
<td></td>
<td>25 X 25</td>
</tr>
<tr>
<td></td>
<td>10 X 10</td>
</tr>
<tr>
<td></td>
<td>1 X 1</td>
</tr>
</tbody>
</table>

No Context Conditions
Συνεργός
Scene classification

- Natural / Urban
- Sunny / Cloudy
- Open / Closed
- Object content
- Location
- ...

[Images of a park and an urban street]
Spatial boundary
3D scene understanding (from a single image)
3D scene understanding (from a single image)
3D scene understanding (from a single image)
Image memorability

• We see hundreds of images a day
• We even remember some of them (but forget many others)

• Subjective? Not completely
• “not an inexplicable phenomenon”
Image memorability
Image memorability

- Understanding and Predicting Image Memorability at a Large Scale (ICCV 2015)
Scene perception

• 1969: Can remember and describe images seen for 0.1 seconds

• 1981: Three levels of representation
  1. Prominent object
  2. Multiple-component
  3. Global scene-emergent features

• Object-centered
  vs. Space-centered
Time frame of scene perception

- Global-to-local analysis
- 20-30 ms: Natural/Urban, Open/Closed
- 100ms: Basic category (beach, forest, etc.)
- 200ms+: Object identities

PT = 107 ms

Inside a house, like a living room, with chairs and sofas and tables, no ppl. (Subject HS)

PT = 500 ms

A room full of musical instruments. A piano in the foreground, a harp behind that, a guitar hanging on the wall (to the right). It looked like there was also a window behind the harp, and perhaps a bookcase on the left. (Subject RW)
Temporal context

- Classification of ambiguous images depend on previously perceived scenes
GIST descriptor

- Meaningful information from a glimpse
- “naturalness, openness, roughness, expansion, ruggedness”
GIST descriptor

- Segment image by 4x4 grid, compute oriented histograms
3D Scene from single images

• Geometric classes
• Support, Vertical, Sky
• Vertical subclasses: Left, center, right
  Porous, solid
3D Scene from single images

- Divide image into “superpixels”
- Group superpixels into larger regions
- Compute features for each region
- Compute confidence levels for different geometric classes
Photo pop-up

• Find connected components for vertical regions
• Fit line segments to base of each component
• Compute depth at each point
References


