Introduction to Recognition
Announcements

• Project 4 (Stereo) has been released (demo)
  – Due next Friday, April 28, by 11:59pm
  – To be done in pairs
Where we go from here

• What we know: Geometry
  – What is the shape of the world? How does that shape appear in images?

• What’s next: Recognition
  – What are we looking at?
What do we mean by “object recognition”?

Next 15 slides adapted from Li, Fergus, & Torralba’s excellent short course on category and object recognition.
Verification: is that a lamp?
Detection: are there people?
Identification: is that Potala Palace?
Object categorization

- mountain
- tree
- building
- banner
- street lamp
- vendor
- people
Scene and context categorization

- outdoor
- city
- ...

[Image: A busy outdoor scene in a city with buildings and people walking.]
Activity / Event Recognition

what are these people doing?
Object recognition
Is it really so hard?

Find the chair in this image

This is a chair

Output of normalized correlation
Object recognition
Is it really so hard?

Find the chair in this image

Pretty much garbage
Simple template matching is not going to make it
Object recognition
Is it really so hard?

A “popular method is that of template matching, by point to point correlation of a model pattern with the image pattern. These techniques are inadequate for three-dimensional scene analysis for many reasons, such as occlusion, changes in viewing angle, and articulation of parts.” Nivatia & Binford, 1977.
Why not use SIFT matching for everything?

- Works well for object *instances*

- Not great for generic object *categories*
And it can get a lot harder

How do human do recognition?

• We don’t completely know yet
• But we have some experimental observations.
Observation 1

- We can recognize familiar faces even in low-resolution images
Observation 2:

- High frequency information is not enough
What is the single most important facial features for recognition?
Observation 4:

• Image Warping is OK
The list goes on

Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know About

• http://web.mit.edu/bcs/sinha/papers/19results_sinha_etal.pdf
Why is this hard?

Variability: Camera position
Illumination
Shape parameters

Set of Images
How many object categories are there?

~10,000 to 30,000
Challenge: variable viewpoint

Michelangelo 1475-1564
Challenge: variable illumination
Challenge: scale
Challenge: deformation
Challenge: Occlusion

Magritte, 1957
Challenge: background clutter

Kilmeny Niland. 1995
Challenge: intra-class variations
History of ideas in recognition

• 1960s – early 1990s: the geometric era
Variability: Camera position, Illumination

Alignment

Shape: assumed known

Roberts (1965); Lowe (1987); Faugeras & Hebert (1986); Grimson & Lozano-Perez (1986); Huttenlocher & Ullman (1987)
Instance Recognition

- Alignment: fitting a model to a transformation between pairs of features (matches) in two images

\[
\sum_i \text{residual } (T(x_i), x'_i)
\]
Recognition by components

Biederman (1987)

Primitives (geons)

- Cube
  - Straight Edge
  - Straight Axis Constant
- Wedge
  - Straight Edge
  - Straight Axis Expanded
- Pyramid
  - Straight Edge
  - Straight Axis Expanded
- Cylinder
  - Curved Edge
  - Straight Axis Expanded
- Barrel
  - Curved Edge
  - Straight Axis Exp & Cont
- Arch
  - Straight Edge
  - Curved Axis Constant
- Cone
  - Curved Edge
  - Straight Axis Expanded
- Expanded Cylinder
  - Curved Edge
  - Straight Axis Expanded
- Handle
  - Curved Edge
  - Curved Axis Constant
- Expanded Handle
  - Curved Edge
  - Curved Axis Expanded

Objects

- Watering can
- Telephone
- Lamp
- Sailboat

History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
Eigenfaces (Turk & Pentland, 1991)

<table>
<thead>
<tr>
<th>Experimental Condition</th>
<th>Correct/Unknown Recognition Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lighting</td>
<td>Orientation</td>
</tr>
<tr>
<td>Forced classification</td>
<td>96/0</td>
</tr>
<tr>
<td>Forced 100% accuracy</td>
<td>100/19</td>
</tr>
<tr>
<td>Forced 20% unknown rate</td>
<td>100/20</td>
</tr>
</tbody>
</table>

Svetlana Lazebnik
Limitations of global appearance models

- Requires global registration of patterns
- Not robust to clutter, occlusion, geometric transformations
History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- 1990s – present: sliding window approaches
Sliding window approaches
Sliding window approaches

- Turk and Pentland, 1991
- Belhumeur, Hespanha, & Kriegman, 1997
- Schneiderman & Kanade, 2004
- Viola and Jones, 2000
- Schneiderman & Kanade, 2004
- Agrawal and Roth, 2002
- Poggio et al. 1993
History of ideas in recognition

• 1960s – early 1990s: the geometric era
• 1990s: appearance-based models
• Mid-1990s: sliding window approaches
• Late 1990s: local features
Local features for object instance recognition

History of ideas in recognition

• 1960s – early 1990s: the geometric era
• 1990s: appearance-based models
• Mid-1990s: sliding window approaches
• Late 1990s: local features
• Early 2000s: parts-and-shape models
Parts-and-shape models

• Model:
  – Object as a set of parts
  – Relative locations between parts
  – Appearance of part

Figure from [Fischler & Elschlager 73]
Pictorial structure model

Fischler and Elschlager(73), Felzenszwalb and Huttenlocher(00)

\[
\Pr(P_{\text{tor}}, P_{\text{arm}}, \ldots | \text{Im}) \propto \prod_{i,j} \Pr(P_i | P_j) \prod_i \Pr(\text{Im}(P_i))
\]

part geometry

part appearance
Discriminatively trained part-based models

History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- Mid-1990s: sliding window approaches
- Late 1990s: local features
- Early 2000s: parts-and-shape models
- Mid-2000s: bags of features
Bag-of-features models

Svetlana Lazebnik
Bag-of-features models

Object → Bag of ‘words’
History of ideas in recognition

• 1960s – early 1990s: the geometric era
• 1990s: appearance-based models
• Mid-1990s: sliding window approaches
• Late 1990s: local features
• Early 2000s: parts-and-shape models
• Mid-2000s: bags of features
• Present trends: data-driven methods, deep learning