Object Recognition: Overview and History

Slides adapted from Fei-Fei Li, Rob Fergus, Antonio Torralba, and Jean Ponce
How many visual object categories are there?

Biederman 1987

~10,000 to 30,000
~10,000 to 30,000
Specific recognition tasks
Scene categorization

- outdoor/indoor
- city/forest/factory/etc.
Image annotation/tagging

- street
- people
- building
- mountain
- ...

[Image of a street scene with annotations]
Object detection

- find pedestrians
Image parsing

sky

mountain

building

tree

banner

street lamp

market

people
Image understanding?
Recognition is all about modeling variability

Variability: Camera position
Illumination
Shape parameters

Within-class variations?
Within-class variations
History of ideas in recognition

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

Shape: assumed known

Roberts (1965); Lowe (1987); Faugeras & Hebert (1986); Grimson & Lozano-Perez (1986); Huttenlocher & Ullman (1987)
Recall: Alignment

- 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 as an alignment problem: Block world

Fig. 1. A system for recognizing 3-d polyhedral scenes. a) L.G. Roberts. b) A blocks world scene. c) Detected edges using a 2x2 gradient operator. d) A 3-d polyhedral description of the scene, formed automatically from the single image. e) The 3-d scene displayed with a viewpoint different from the original image to demonstrate its accuracy and completeness. (b) - e) are taken from [64] with permission MIT Press.)

J. Mundy, Object Recognition in the Geometric Era: a Retrospective, 2006
Alignment: Huttenlocher & Ullman (1987)
Variability

Invariance to: Camera position
Illumination
Internal parameters

Duda & Hart (1972); Weiss (1987); Mundy et al. (1992-94); Rothwell et al. (1992); Burns et al. (1993)
Example: invariant to similarity transformations computed from four points

Projective invariants (Rothwell et al., 1992):

General 3D objects do not admit monocular viewpoint invariants (Burns et al., 1993)
Representing and recognizing object categories is harder...

ACRONYM (Brooks and Binford, 1981)
Binford (1971), Nevatia & Binford (1972), Marr & Nishihara (1978)
Recognition by components

Biederman (1987)

Primitives (geons)  Objects

Cube
- Straight Edge
- Straight Axis Constant

Wedge
- Straight Edge
- Expanded

Pyramid
- Straight Edge
- Expanded

Cylinder
- Curved Edge
- Straight Axis Constant

Barrel
- Curved Edge
- Straight Axis Expanded & Constant

Arch
- Straight Edge
- Curved Axis Constant

Cone
- Curved Edge
- Expanded

Expanded Cylinder
- Curved Edge
- Expanded

Handle
- Curved Edge
- Straight Axis Constant

Expanded Handle
- Curved Edge
- Expanded

General shape primitives?

Generalized cylinders
Ponce et al. (1989)

Zisserman et al. (1995)

Forsyth (2000)
History of ideas in recognition

• 1960s – early 1990s: the geometric era
• 1990s: appearance-based models
Empirical models of image variability

**Appearance-based techniques**

Turk & Pentland (1991); Murase & Nayar (1995); etc.
Eigenfaces (Turk & Pentland, 1991)

<table>
<thead>
<tr>
<th>Experimental Condition</th>
<th>Correct/Unknown Recognition Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forced classification</td>
<td>96/0, 85/0, 64/0</td>
</tr>
<tr>
<td>Forced 100% accuracy</td>
<td>100/19, 100/39, 100/60</td>
</tr>
<tr>
<td>Forced 20% unknown rate</td>
<td>100/20, 94/20, 74/20</td>
</tr>
</tbody>
</table>
Color Histograms

Appearance manifolds

H. Murase and S. Nayar, Visual learning and recognition of 3-d objects from appearance, IJCV 1995
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
- Argawal 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

Large-scale image search
Combining local features, indexing, and spatial constraints

Image credit: K. Grauman and B. Leibe
Large-scale image search
Combining local features, indexing, and spatial constraints
Large-scale image search
Combining local features, indexing, and spatial constraints

Google Goggles in Action
Click the icons below to see the different ways Google Goggles can be used.

Available on phones that run Android 1.6+ (i.e. Donut or Eclair)
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]
Constellation models

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
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

Object → Bag of ‘words’
Objects as texture

• All of these are treated as being the same

• No distinction between foreground and background: scene 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
- Mid-2000s: bags of features
- Present trends: combination of local and global methods, data-driven methods, context
Global scene descriptors

- The “gist” of a scene: Oliva & Torralba (2001)

[Diagram showing edge orientation and frequency with examples of scenes]

http://people.csail.mit.edu/torralba/code/spatialenvelope/
Data-driven methods

J. Hays and A. Efros, Scene Completion using Millions of Photographs, SIGGRAPH 2007
Data-driven methods

J. Tighe and S. Lazebnik, ECCV 2010
Geometric context

(a) Input image
(b) $P(\text{person}) = \text{uniform}$
(c) Surface estimate
(d) $P(\text{person} \mid \text{geometry})$
(e) $P(\text{viewpoint} \mid \text{objects})$
(f) $P(\text{person} \mid \text{viewpoint})$
(g) $P(\text{person} \mid \text{viewpoint, geometry})$

Discriminatively trained part-based models

What “works” today

• Reading license plates, zip codes, checks
What “works” today

- Reading license plates, zip codes, checks
- Fingerprint recognition
What “works” today

- Reading license plates, zip codes, checks
- Fingerprint recognition
- Face detection
What “works” today

• Reading license plates, zip codes, checks
• Fingerprint recognition
• Face detection
• Recognition of flat textured objects (CD covers, book covers, etc.)