Object Recognition: History and Overview

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

~10,000 to 30,000

Biederman 1987
~10,000 to 30,000
So what does object recognition involve?
Scene categorization

- outdoor
- city
- ...

![Outdoor city scene](image-url)
Image-level annotation: are there people?

- outdoor
- city
- ...
Object detection: where are the people?
Image parsing
Modeling variability

Variability: Camera position
Illumination
Shape parameters

Within-class variations?
Within-class variations
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

\[
\text{Find transformation } T \text{ that minimizes } \sum_i \text{residual}(T(x_i), x'_i)
\]
Recall: Origins of computer vision

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)
Recall: 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

Geons (Biederman 1987)
General shape primitives?

Generalized cylinders
Ponce et al. (1989)

Zisserman et al. (1995)

Forsyth (2000)
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>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>
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
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
Sliding window approaches

- Scale / orientation range to search over
- Speed
- Context
Local features

Combining *local* appearance, spatial constraints, invariants, and classification techniques from machine learning.

Schmid & Mohr'97

Lowe'02

Mahamud & Hebert'03
Local features for recognition of object instances
Local features for recognition of object instances

• Lowe, et al. 1999, 2003
• Mahamud and Hebert, 2000
• Ferrari, Tuytelaars, and Van Gool, 2004
• Rothganger, Lazebnik, and Ponce, 2004
• Moreels and Perona, 2005
• ...
Representing categories: Parts and Structure

Parts-and-shape representation

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

Figure from [Fischler & Elschlager 73]
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?
Timeline of recognition

• 1965-late 1980s: alignment, geometric primitives
• Early 1990s: invariants, appearance-based methods
• Mid-late 1990s: sliding window approaches
• Late 1990s: feature-based methods
• Early 2000s: parts-and-shape models
• 2003 – present: bags of features
• Present trends: combination of local and global methods, modeling context, emphasis on “image parsing”
Global scene context

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

http://people.csail.mit.edu/torralba/code/spatialenvelope/
J. Hays and A. Efros, **Scene Completion using Millions of Photographs**, SIGGRAPH 2007
Scene-level context for image parsing

(a) Query Image  
(b) Retrieval Set

(c) Superpixels  
(d) Per-class Likelihoods

(e) Final Labeling

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

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.)