Lecture 9
Recognition

Dr Christopher Brunner
AMME4710 – Computer Vision and Image Processing
Semester 2, 2013

By the end of this lecture you will be able to:

• Appreciate the importance of image recognition and its applications
• Understand the main challenges and approaches for image recognition
• Describe algorithms for face detection and face recognition
Resources

• Readings
  – Szeliski, Computer Vision, Ch. 14
  http://szeliski.org/Book/

• Additional material (optional)
  – Viola and Jones, Robust Real-time Object Detection, IJCV 2001

• Slides credits
  – S. Lazebnik, K. Grauman, T. Darrell, S. Seitz, L. Fei-Fei, R. Fergus, A. Torralba
What do we mean by “object recognition”?
Detection: are there people?
Identification: is that Potala Palace?
Verification: is that a lamp?
Object categorization

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

- outdoor
- city
- ...

[Image of outdoor city scene]
Applications: Computational photography

- Face detection
- Auto focus/exposure
- Smile/Blink/Wink detection
- Scene detection
Applications: Assisted driving

Pedestrian and car detection

- Collision warning systems with adaptive cruise control,
- Lane departure warning systems,
- Parking assist systems
Applications: image search
Posing visual queries

Yeh et al., MIT

Digital Field Guides Eliminate the Guesswork

Belhumeur et al.

Google goggles

Kooaba, Bay & Quack et al.
Autonomous agents able to detect objects

Discovering visual patterns

Objects

Sivic & Zisserman

Categories

Lee & Grauman

Actions

Wang et al.

Kristen Grauman
Auto-annotation (tagging)

Figure 9. Results of automatic object-level annotation with bounding boxes. Groundtruth annotation is shown with dashed lines, correct detection with solid green lines, false detections with solid red lines. Auto-annotation with related Wikipedia articles is also shown. All results are also labeled with their GPS position and estimated tags (not shown here).

Gammeter et al.

T. Berg et al.

President George W. Bush makes a statement in the Rose Garden while Secretary of Defense Donald Rumsfeld looks on, July 23, 2005: Rumsfeld said the United States would release graphic photographs of the dead sons of Saddam Hussein to prove they were killed by American troops. Photo by Larry Downing/Reuters

British director Sam Mendes and his partner actress Kate Winslet arrive at the London premiere of 'The Road to Perdition', September 18, 2002. The film stars Tom Hanks as a Chicago hit man who has a separate family life and co-stars Paul Newman and Jude Law. REUTERS/Dan Chung

Incumbent California Gov. Gray Davis (news - web site) leads Republican challenger Bill Simon by 10 percentage points — although 17 percent of voters are still undecided, according to a poll released October 22, 2002 by the Public Policy Institute of California. Davis is shown speaking to reporters after his debate with Simon in Los Angeles, on Oct. 7. (Jim Raynor/Reuters)
Outline

• Introduction
• Machine Learning (Classifiers)
• Face Detection
INTRODUCTION
Instance-level recognition problem
Generic categorization problem
Object Categorization

• Task Description
  – “Given a small number of training images of a category, recognize a-priori unknown instances of that category and assign the correct category label.”

• Which categories are feasible visually?

“Fido”  German shepherd  dog  animal  living being

K. Grauman, B. Leibe
Visual Object Categories

Basic Level Categories in human categorization
[Rosch 76, Lakoff 87]

• The highest level at which category members have similar perceived shape.
• The highest level at which a single mental image reflects the entire category.
• The level at which human subjects are usually fastest at identifying category members.
• The first level named and understood by children.
• The highest level at which a person uses similar motor actions for interaction with category members.
Visual Object Categories

- Basic-level categories in humans seem to be defined predominantly visually.
- There is evidence that humans (usually) start with basic-level categorization before doing identification.
  - Basic-level categorization is easier and faster for humans than object identification!
  - How does this transfer to automatic classification algorithms?
Object recognition
Is it really so hard?

Find the chair in this image

Output of normalized correlation

This is a chair

Slide: A. Torralba
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

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.

Slide: A. Torralba
How many object categories are there?

~10,000 to 30,000

Source: Fei-Fei Li, Rob Fergus, Antonio Torralba.

Biederman 1987
~10,000 to 30,000
Other Types of Categories

• Functional Categories
  - e.g. chairs = “something you can sit on”
Other Types of Categories

• Ad-hoc categories
  - e.g. “something you can find in an office environment”
Why do we care about categories?

Perception of function: We can perceive the 3D shape, texture, material properties, without knowing about objects.

But, the concept of category also encapsulates information about what can we do with those objects.

When we recognize an object we can make predictions about its behavior in the future, beyond of what is immediately perceived.
Recognition is all about modeling variability

Variability: Camera position
            Illumination
            Shape parameters

Within-class variations?
Challenges: intra-class variation
Challenges: viewpoint variation

Michelangelo 1475-1564
Challenges: illumination variation
Challenges: occlusion

Magritte, 1957
Challenges: scale
Challenges: deformation

Xu, Beihong 1943
Challenges: background clutter

Klimt, 1913
Challenges: robustness

- Illumination
- Object pose
- Clutter
- Occlusions
- Intra-class appearance
- Viewpoint

Kristen Grauman
Challenges: robustness

Realistic scenes are crowded, cluttered, have overlapping objects.
Challenges: importance of context
Challenges: learning with minimal supervision

Unlabeled, multiple objects

Classes labeled, some clutter

Cropped to object, parts and classes

Kristen Grauman
Challenges: complexity

• Thousands to millions of pixels in an image
• 10,000-30,000 human recognizable object categories
• 30+ degrees of freedom in the pose of articulated objects (humans)
• Billions of images indexed by Google Image Search
• 18 billion+ prints produced from digital camera images in 2004
• 295.5 million camera phones sold in 2005
• About half of the cerebral cortex in primates is devoted to processing visual information [Felleman and van Essen 1991]
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
What works most reliably today

• Reading license plates, zip codes, checks

Source: Lana Lazebnik
What works most reliably today

- Reading license plates, zip codes, checks
- Fingerprint recognition

Source: Lana Lazebnik
What works most reliably today

- Reading license plates, zip codes, checks
- Fingerprint recognition
- Face detection

Source: Lana Lazebnik
What works most reliably today

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

Source: Lana Lazebnik
MACHINE LEARNING
Ontologies - categories and relationships

• continuous maintenance
• no guarantee of coverage
• difficult categories
The machine learning framework

- Apply a prediction function to a feature representation of the image to get the desired output:

\[
\begin{align*}
    f(\text{apple}) &= \text{“apple”} \\
    f(\text{tomato}) &= \text{“tomato”} \\
    f(\text{cow}) &= \text{“cow”}
\end{align*}
\]
The machine learning framework

\[ y = f(x) \]

- **Training:** given a *training set* of labeled examples \( \{(x_1,y_1), \ldots, (x_N,y_N)\} \), estimate the prediction function \( f \) by minimizing the prediction error on the training set
- **Testing:** apply \( f \) to a never before seen *test example* \( x \) and output the predicted value \( y = f(x) \)
Steps

Training

Training Images

Training Images

Image Features

Training Labels

Training

Learned model

Testing

Test Image

Image Features

Prediction

Learned model

Slide credit: D. Hoiem
Recognition task and supervision

• Images in the training set must be annotated with the “correct answer” that the model is expected to produce

Contains a motorbike
Classification

• Assign input vector to one of two or more classes

• Any decision rule divides input space into decision regions separated by decision boundaries
Features

- Raw pixels
- Histograms
- Descriptors e.g. SIFT
- …
Formulation

• Formulation: binary classification

$\begin{align*}
\text{Features } x &= x_1 \quad x_2 \quad x_3 \quad \cdots \quad x_N \\
&\quad x_{N+1} \quad x_{N+2} \quad \cdots \quad x_{N+M} \\
\text{Labels } y &= -1 \quad +1 \quad -1 \quad -1 \\
&\quad \ ? \quad \ ? \quad \ ?
\end{align*}$

Training data: each image patch is labeled as containing the object or background

Test data

• Classification function

$\hat{y} = F(x)$  Where $F(x)$ belongs to some family of functions

• Minimize misclassification error

(Not that simple: we need some guarantees that there will be generalization)
Object categorization: the probabilistic approach

\[ p(\text{zebra} \mid \text{image}) \]

vs.

\[ p(\text{no zebra} \mid \text{image}) \]

• Bayes rule:

\[
\frac{p(\text{zebra} \mid \text{image})}{p(\text{no zebra} \mid \text{image})} = \frac{p(\text{image} \mid \text{zebra})}{p(\text{image} \mid \text{no zebra})} \cdot \frac{p(\text{zebra})}{p(\text{no zebra})}
\]

posterior ratio

likelihood ratio

prior ratio

Fei-Fei, Fergus, Torralba, CVPR 2007 SC
Discriminative methods

Object detection and recognition is formulated as a classification problem. The image is partitioned into a set of overlapping windows ... and a decision is taken at each window about if it contains a target object or not.
Discriminative vs. generative

• Generative model

• Discriminative model

• Classification function
Discriminative methods

Nearest neighbor
Shakhnarovich, Viola, Darrell 2003
Berg, Berg, Malik 2005

10^6 examples

Neural networks
LeCun, Bottou, Bengio, Haffner 1998
Rowley, Baluja, Kanade 1998

Support Vector Machines and Kernels
Guyon, Vapnik
Heisele, Serre, Poggio, 2001

Conditional Random Fields
McCallum, Freitag, Pereira 2000
Kumar, Hebert 2003
Spectrum of Supervision

Unsupervised  "Weakly" supervised  Fully supervised

Definition depends on task
Generalization

• How well does a learned model generalize from the data it was trained on to a new test set?
Generalization

• Components of generalization error
  – **Bias**: how much the average model over all training sets differ from the true model?
    • Error due to inaccurate assumptions/simplifications made by the model
  – **Variance**: how much models estimated from different training sets differ from each other

• **Underfitting**: model is too “simple” to represent all the relevant class characteristics
  – High bias and low variance
  – High training error and high test error

• **Overfitting**: model is too “complex” and fits irrelevant characteristics (noise) in the data
  – Low bias and high variance
  – Low training error and high test error