Generic object recognition

Wed, April 6

Kristen Grauman

What does recognition involve?

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

Verification: is that a lamp?

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

Detection: are there people?

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

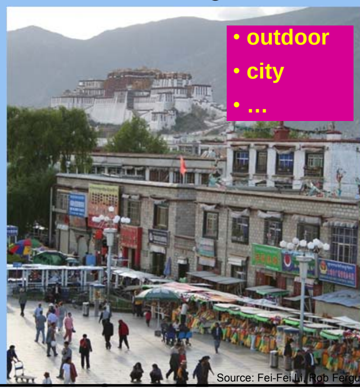
Identification: is that Potala Palace?

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

Object categorization

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

Scene and context categorization



Instance-level recognition problem

John's car

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?**

Visual Object Recognition Tutorial

K. Grauman, B. Leibe

Visual Object Categories

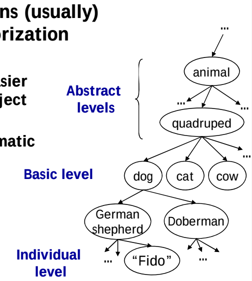
- **Basic Level Categories in human categorization**
  [Rosch 76, Lakoff 87]
  - **The highest level at which category members have similar perceived shape**
  - **The highest level at which a single mental image reflects the entire category**
  - **The level at which human subjects are usually fastest 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 Recognition Tutorial

K. Grauman, B. Leibe

Visual Object Categories

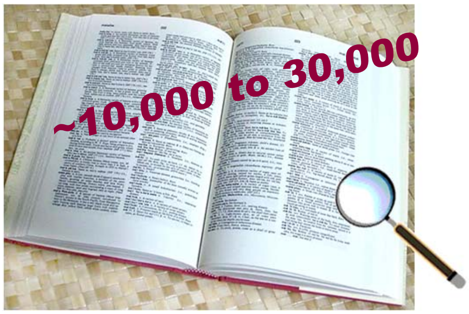
- **Basic-level categories in humans seem to be defined predominantly visually.**
- **There is evidence that humans (usually) start with basic-level categorization *before* doing identification.**
  - **⇒ Basic-level categorization is easier and faster for humans than object identification!**
  - **⇒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



Visual Object Recognition Tutorial

K. Grauman, B. Leibe

How many object categories are there?



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

Biederman 1987

Visual Object Recognition Tutorial

Other Types of Categories

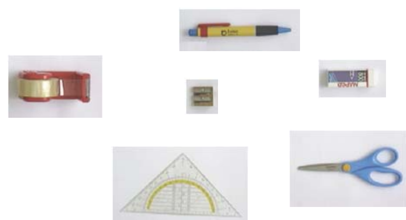
- **Functional Categories**
  - **e.g. chairs =** *"something you can sit on"*

K. Grauman, B. Leibe

Visual Object Recognition Tutorial

Other Types of Categories

- **Ad-hoc categories**
  - **e.g.** *"something you can find in an office environment"*



K. Grauman, B. Leibe

Why recognition?

- Recognition a fundamental part of perception
  - e.g., robots, autonomous agents
- Organize and give access to visual content
  - Connect to information
  - Detect trends and themes

Posing visual queries

Yeh et al., MIT

Digital Field Guides Eliminate the Guesswork

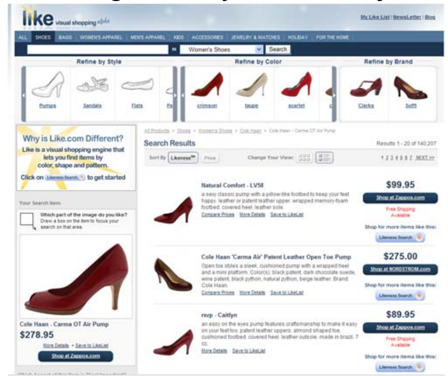
Belhumeur et al.

Kooaba, Bay & Quack et al.

Autonomous agents able to detect objects

http://www.darpa.mil/grandchallenge/gallery.asp

Finding visually similar objects



Discovering visual patterns

Objects Sivic & Zisserman

Categories Lee & Grauman

Actions Wang et al.

Kristen Grauman

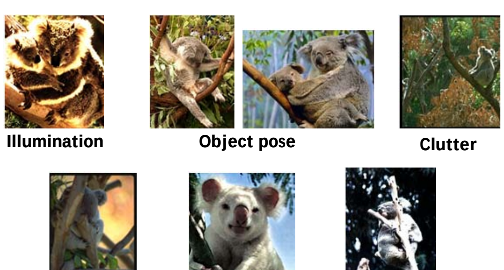
Auto-annotation

Gammeter et al.

T. Berg et al.

Kristen Grauman

Challenges: robustness



Illumination

Object pose

Clutter

Occlusions

Intra-class appearance

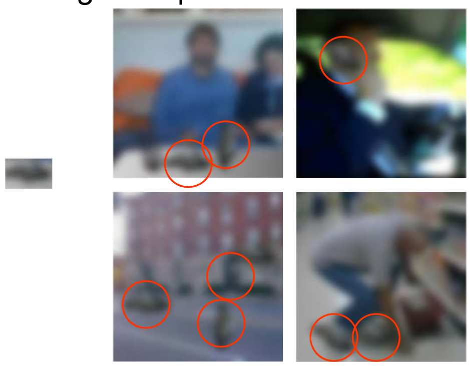
Viewpoint

Kristen Grauman

Challenges: robustness

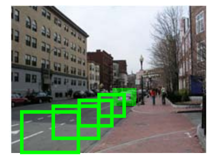
Realistic scenes are crowded, cluttered, have overlapping objects.

Challenges: importance of context



slide credit: Fei-Fei, Fergus & Torralba

Challenges: importance of context

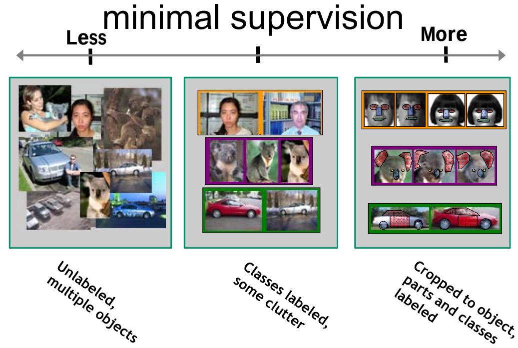


Challenges: complexity

- Thousands to millions of pixels in an image
- 3,000-30,000 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]

Kristen Grauman

Challenges: learning with minimal supervision



Kristen Grauman

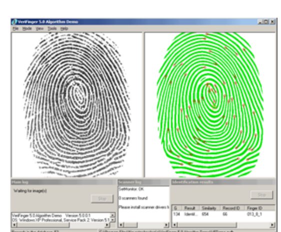
What works most reliably today

- Reading license plates, zip codes, checks

Source: Lana Lazebnik

What works most reliably today

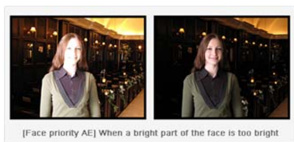
- Reading license plates, zip codes, checks
- Fingerprint recognition



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.)

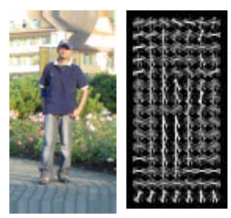
Source: Lana Lazebnik

Generic category recognition: basic framework

- Build/train object model
  - Choose a representation
  - Learn or fit parameters of model / classifier
- Generate candidates in new image
- Score the candidates

Kristen Grauman

Generic category recognition: representation choice



Window-based

Part-based

Kristen Grauman

Supervised classification

- Given a collection of *labeled* examples, come up with a function that will predict the labels of new examples.

“four”
“nine”

Training examples

?
Novel input

- How good is some function we come up with to do the classification?
- Depends on
  - Mistakes made
  - Cost associated with the mistakes

Kristen Grauman

Supervised classification

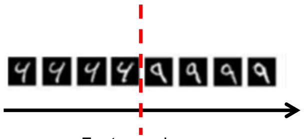
- Given a collection of *labeled* examples, come up with a function that will predict the labels of new examples.
- Consider the two-class (binary) decision problem
  - L(4→9): Loss of classifying a 4 as a 9
  - L(9→4): Loss of classifying a 9 as a 4
- **Risk** of a classifier *s* is expected loss:

$$R(s) = \Pr(4 \to 9 \mid \text{using } s) L(4 \to 9) + \Pr(9 \to 4 \mid \text{using } s) L(9 \to 4)$$

- We want to choose a classifier so as to minimize this total risk

Kristen Grauman

Supervised classification



Optimal classifier will minimize total risk.

At decision boundary, either choice of label yields same expected loss.

If we choose class "four" at boundary, expected loss is:

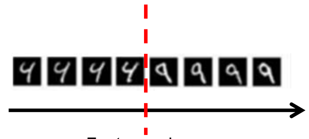
$$= P(\text{class is } 9 \mid \mathbf{x})\, L(9 \to 4) + P(\text{class is } 4 \mid \mathbf{x}) L(4 \to 4)$$

If we choose class "nine" at boundary, expected loss is:

$$= P(\text{class is } 4 \mid \mathbf{x})\, L(4 \to 9)$$

Kristen Grauman

Supervised classification



Optimal classifier will minimize total risk.

At decision boundary, either choice of label yields same expected loss.

So, best decision boundary is at point **x** where

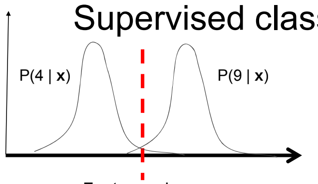
$$P(\text{class is } 9 \mid \mathbf{x})\, L(9 \to 4) = \text{P}(\text{class is } 4 \mid \mathbf{x}) L(4 \to 9)$$

To classify a new point, choose class with lowest expected loss; i.e., choose "four" if

$$P(4 \mid \mathbf{x}) L(4 \to 9) > P(9 \mid \mathbf{x}) L(9 \to 4)$$

Kristen Grauman

Supervised classification



Optimal classifier will minimize total risk.

At decision boundary, either choice of label yields same expected loss.

So, best decision boundary is at point **x** where

$$P(\text{class is } 9 \mid \mathbf{x})\, L(9 \to 4) = \text{P}(\text{class is } 4 \mid \mathbf{x}) L(4 \to 9)$$

To classify a new point, choose class with lowest expected loss; i.e., choose "four" if

$$\boxed{P(4 \mid \mathbf{x})} L(4 \to 9) > \boxed{P(9 \mid \mathbf{x})} L(9 \to 4)$$

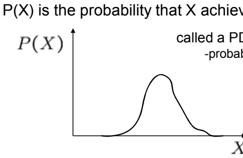
How to evaluate these probabilities?

Kristen Grauman

Probability

Basic probability

- X is a random variable
- P(X) is the probability that X achieves a certain value



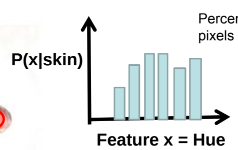
called a PDF
-probability distribution/density function

- $0 \le P(X) \le 1$
- $\int_{-\infty}^{\infty} P(X)dX = 1$ (continuous X) or $\sum P(X) = 1$ (discrete X)
- Conditional probability: P(X | Y)
  - probability of X given that we already know Y

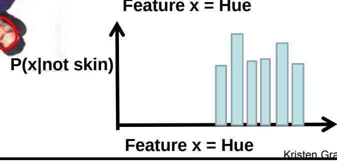
Source: Steve Seitz

Example: learning skin colors

- We can represent a class-conditional density using a histogram (a "non-parametric" distribution)



Percentage of skin pixels in each bin



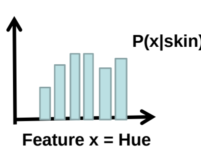
Kristen Grauman

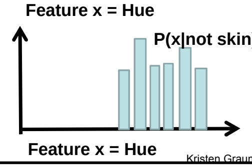
Example: learning skin colors

- We can represent a class-conditional density using a histogram (a "non-parametric" distribution)

Now we get a new image, and want to label each pixel as skin or non-skin.

What's the probability we care about to do skin detection?





Kristen Grauman

Bayes rule

$$\overbrace{P(skin \mid x)}^{\text{posterior}} = \frac{\overbrace{P(x \mid skin)}^{\text{likelihood}}\overbrace{P(skin)}^{\text{prior}}}{P(x)}$$

$$P(skin \mid x) \; \alpha \; P(x \mid skin)P(skin)$$

Where does the prior come from?

Why use a prior?

Example: classifying skin pixels

Now for every pixel in a new image, we can estimate probability that it is generated by skin.

Brighter pixels → higher probability of being skin

Classify pixels based on these probabilities

- if $p(\text{skin}|\boldsymbol{x}) > \theta$, classify as skin
- if $p(\text{skin}|\boldsymbol{x}) < \theta$, classify as not skin

Kristen Grauman

Example: classifying skin pixels

Figure 6: A video image and its flesh probability image

Figure 7: Orientation of the flesh probability distribution marked on the source video image

Gary Bradski, 1998

Kristen Grauman

Example: classifying skin pixels

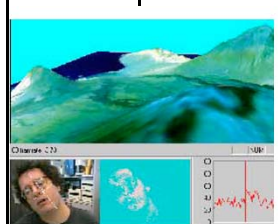


Figure 13: CAMSHIFT-based face tracker used to over a 3D graphic's model of Hawaii

Figure 12: CAMSHIFT-based face tracker used to play Quake 2 hands free by inserting control variables into the mouse queue

Using skin color-based face detection and pose estimation as a video-based interface

Gary Bradski, 1998

Kristen Grauman

Supervised classification

- Want to minimize the expected misclassification
- Two general strategies
  - Use the training data to build representative probability model; separately model class-conditional densities and priors (*generative*)
  - Directly construct a good decision boundary, model the posterior (*discriminative*)

Coming up

Pset 4 is posted, due in 2 weeks

Next week:

Face detection

Categorization with local features and part-based models