Generic object recognition
Wed, April 6
Kristen Grauman

What does recognition involve?

Verification: is that a lamp?

Detection: are there people?

Identification: is that Potala Palace?

Object categorization

mountain

building

street lamp

people

vendor

tree

banner
Scene and context categorization

• outdoor
• city
• ...

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

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 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?
How many object categories are there?

How many object categories are there?

~10,000 to 30,000

~10,000 to 30,000

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

Biederman 1987

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 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

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

Finding visually similar objects

Discovering visual patterns

Auto-annotation

Challenges: robustness

Challenges: robustness

Realistic scenes are crowded, cluttered, have overlapping objects.
Challenges: importance of context

- slide credit: Fei-Fei, Fergus & Torralba

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]

Challenges: learning with minimal supervision

- What works most reliably today
  - Reading license plates, zip codes, checks
  - Fingerprint recognition

- What works most reliably today
  - Reading license plates, zip codes, checks
  - Fingerprint recognition
What works most reliably today

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

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

Source: 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

Source: Kristen Grauman

Consider the two-class (binary) decision problem

– \( L(4 \rightarrow 9) \): Loss of classifying a 4 as a 9
– \( L(9 \rightarrow 4) \): Loss of classifying a 9 as a 4

Risk of a classifier \( s \) is expected loss:

\[
R(s) = \Pr(4 \rightarrow 9 \mid s) L(4 \rightarrow 9) + \Pr(9 \rightarrow 4 \mid s) L(9 \rightarrow 4)
\]

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

Source: Kristen Grauman
Supervised classification

Feature value $x$

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:

$$P(\text{class} = 9 \mid x)L(9 \rightarrow 4) + P(\text{class} = 4 \mid x)L(4 \rightarrow 4)$$

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

$$P(\text{class} = 4 \mid x)L(4 \rightarrow 9)$$

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

$$P(4 \mid x)L(4 \rightarrow 9) > P(9 \mid x)L(9 \rightarrow 4)$$

How to evaluate these probabilities?

---

Example: learning skin colors

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

- $P(x|\text{skin})$
- $P(x|\text{not skin})$

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?
Bayes rule

\[ P(skin \mid x) = \frac{P(x \mid skin)P(skin)}{P(x)} \]

\[ P(skin \mid x) \propto P(x \mid skin)P(skin) \]

*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.

Classify pixels based on these probabilities

- if \( p(skin|x) > \theta \), classify as skin
- if \( p(skin|x) < \theta \), classify as not skin

---

**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