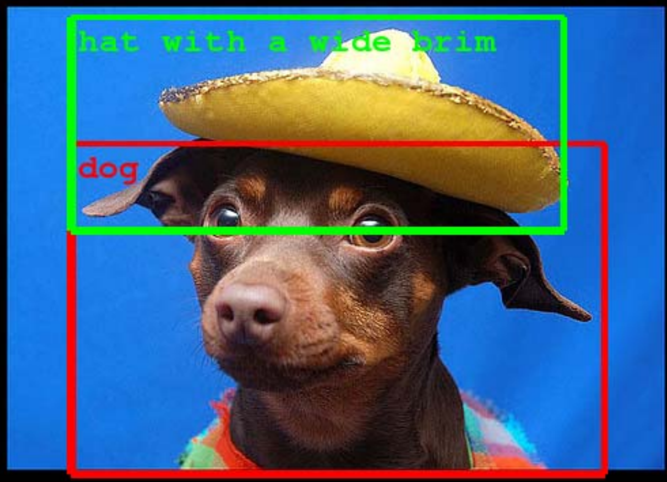
CS4495 Computer Vision

Introduction to Recognition

Aaron Bobick

School of Interactive Computing



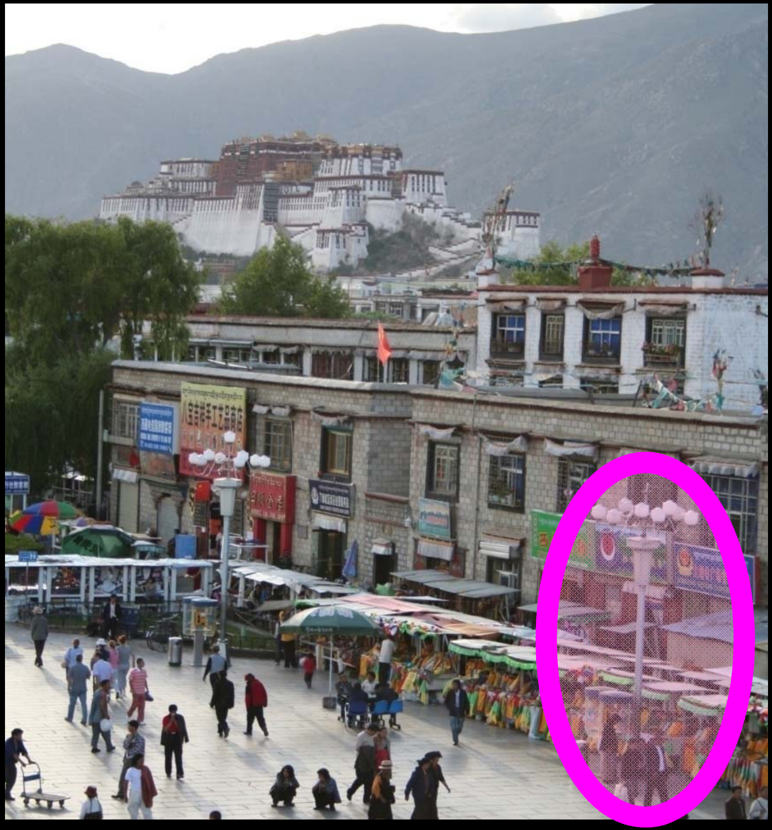
[redacted]

[redacted]

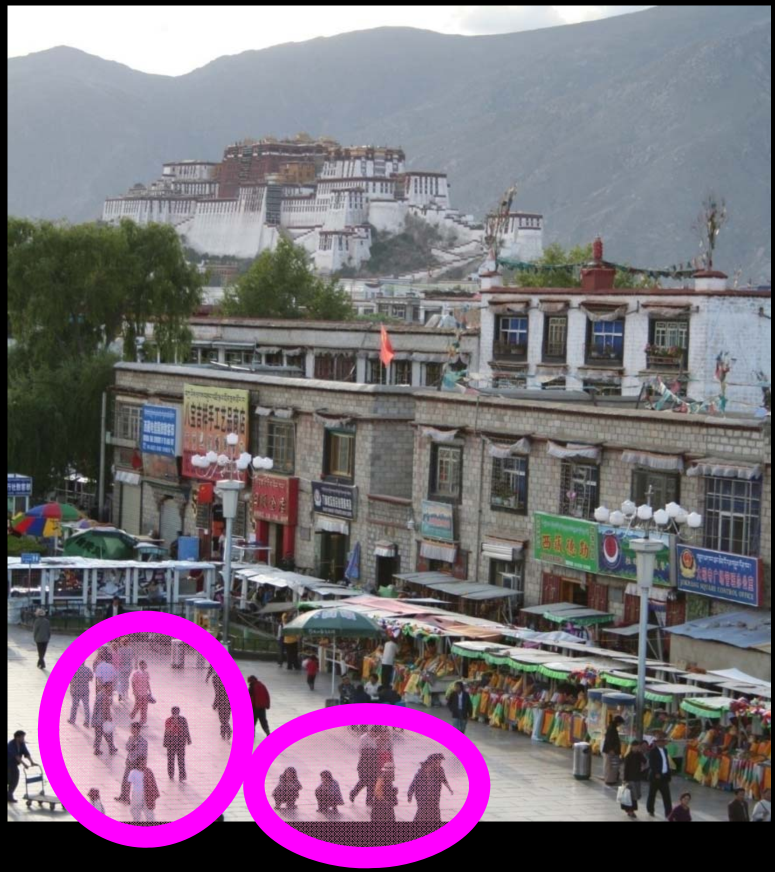
What does recognition involve?

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

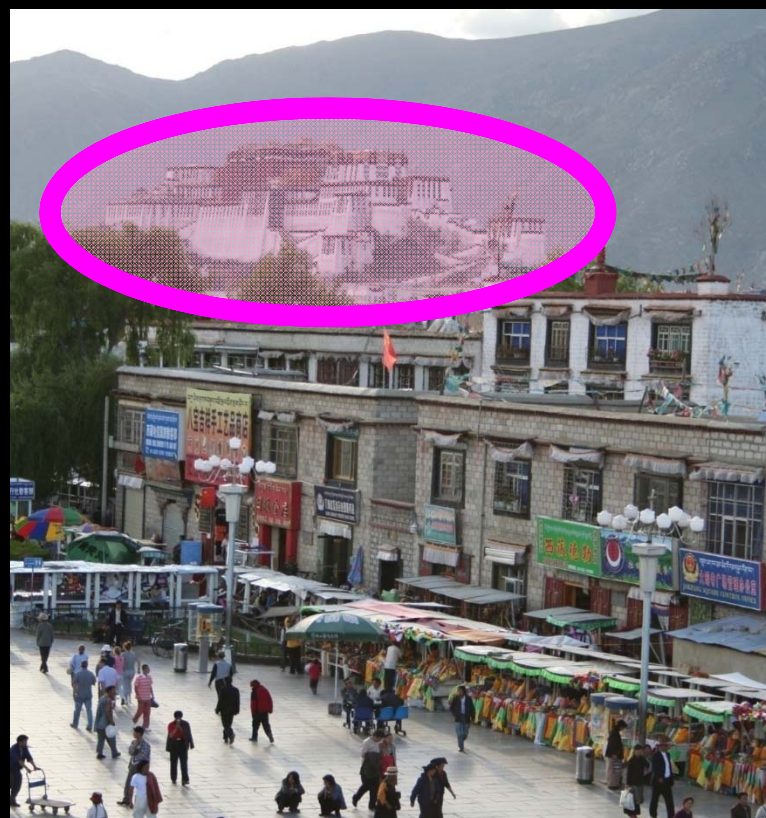
Verification: is that a lamp?



Detection: are there people?

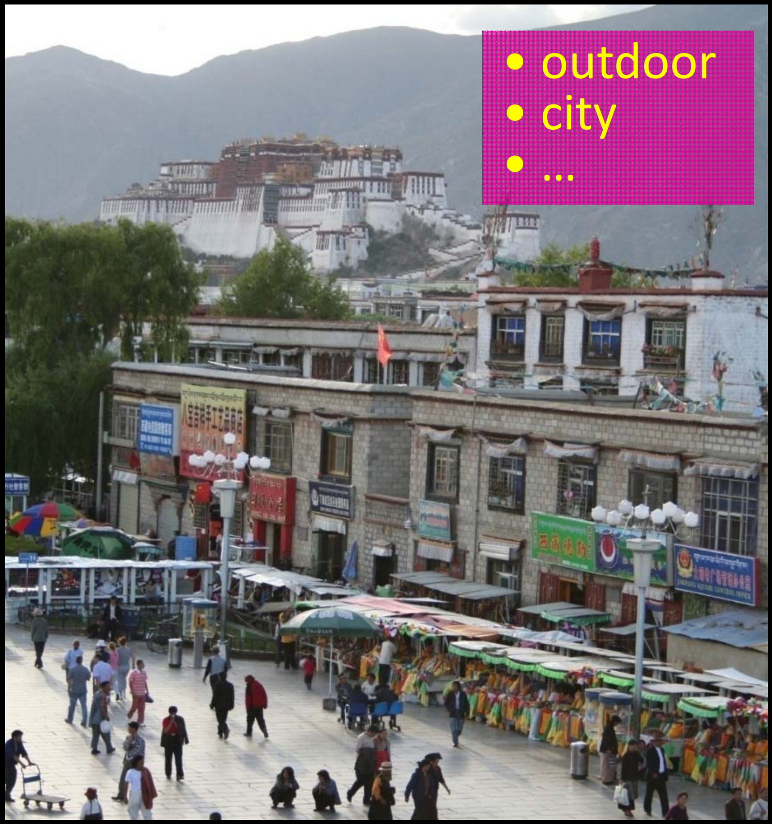


Identification: is that Potala Palace?



Object categorization

Scene and context categorization



Instance-level recognition problem

Generic categorization problem

Object Categorization

Task: *Given a (small) number of training images of a category, recognize a-priori unknown instances of that category and assign the correct category label.*

K. Grauman, B. Leibe

Object Categorization

"Fido" German shepherd dog animal living being

Which categories are the best for visual identification?

Visual Object Categories

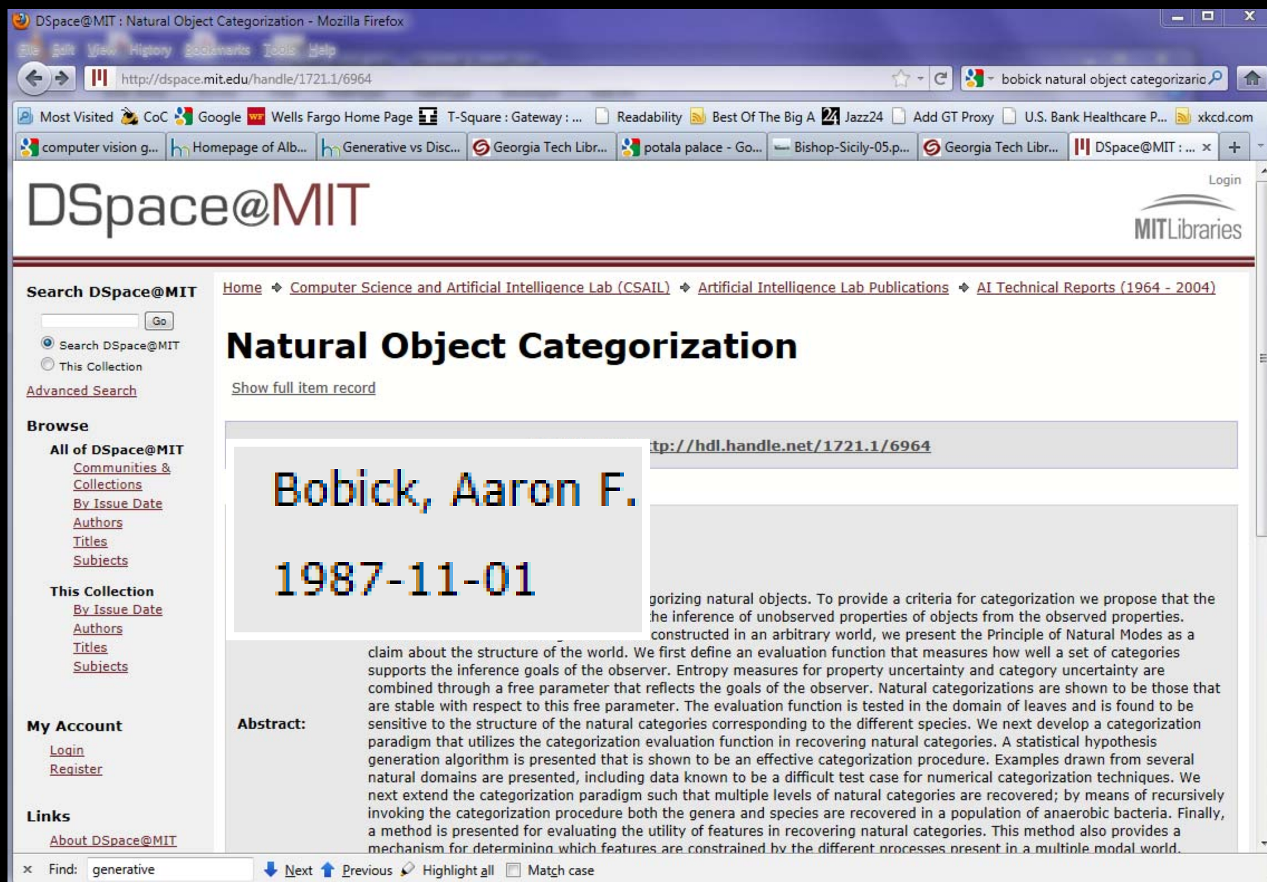
Basic Level Categories in human categorization
[Rosch 76, Lakoff 87]

- The highest level at which category members have similar perceived shape
- The highest level at which a single mental image reflects the entire category

Visual Object Categories

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

- 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



DSpace@MIT : Natural Object Categorization - Mozilla Firefox
http://dspace.mit.edu/handle/1721.1/6964
DSpace@MIT
Login
MITLibraries
Search DSpace@MIT
Go
Search DSpace@MIT
This Collection
Advanced Search
Browse
All of DSpace@MIT
Communities & Collections
By Issue Date
Authors
Titles
Subjects
This Collection
By Issue Date
Authors
Titles
Subjects
My Account
Login
Register
Links
About DSpace@MIT
Home ◆ Computer Science and Artificial Intelligence Lab (CSAIL) ◆ Artificial Intelligence Lab Publications ◆ AI Technical Reports (1964 - 2004)
Natural Object Categorization
Show full item record
tp://hdl.handle.net/1721.1/6964
Bobick, Aaron F.
1987-11-01
Abstract:
gorizing natural objects. To provide a criteria for categorization we propose that the
he inference of unobserved properties of objects from the observed properties.
constructed in an arbitrary world, we present the Principle of Natural Modes as a
claim about the structure of the world. We first define an evaluation function that measures how well a set of categories supports the inference goals of the observer. Entropy measures for property uncertainty and category uncertainty are combined through a free parameter that reflects the goals of the observer. Natural categorizations are shown to be those that are stable with respect to this free parameter. The evaluation function is tested in the domain of leaves and is found to be sensitive to the structure of the natural categories corresponding to the different species. We next develop a categorization paradigm that utilizes the categorization evaluation function in recovering natural categories. A statistical hypothesis generation algorithm is presented that is shown to be an effective categorization procedure. Examples drawn from several natural domains are presented, including data known to be a difficult test case for numerical categorization techniques. We next extend the categorization paradigm such that multiple levels of natural categories are recovered; by means of recursively invoking the categorization procedure both the genera and species are recovered in a population of anaerobic bacteria. Finally, a method is presented for evaluating the utility of features in recovering natural categories. This method also provides a mechanism for determining which features are constrained by the different processes present in a multiple modal world.
Find: generative
Next
Previous
Highlight all
Match case

Technical Report 1001

Natural Object Categorization

Aaron F. Bobick

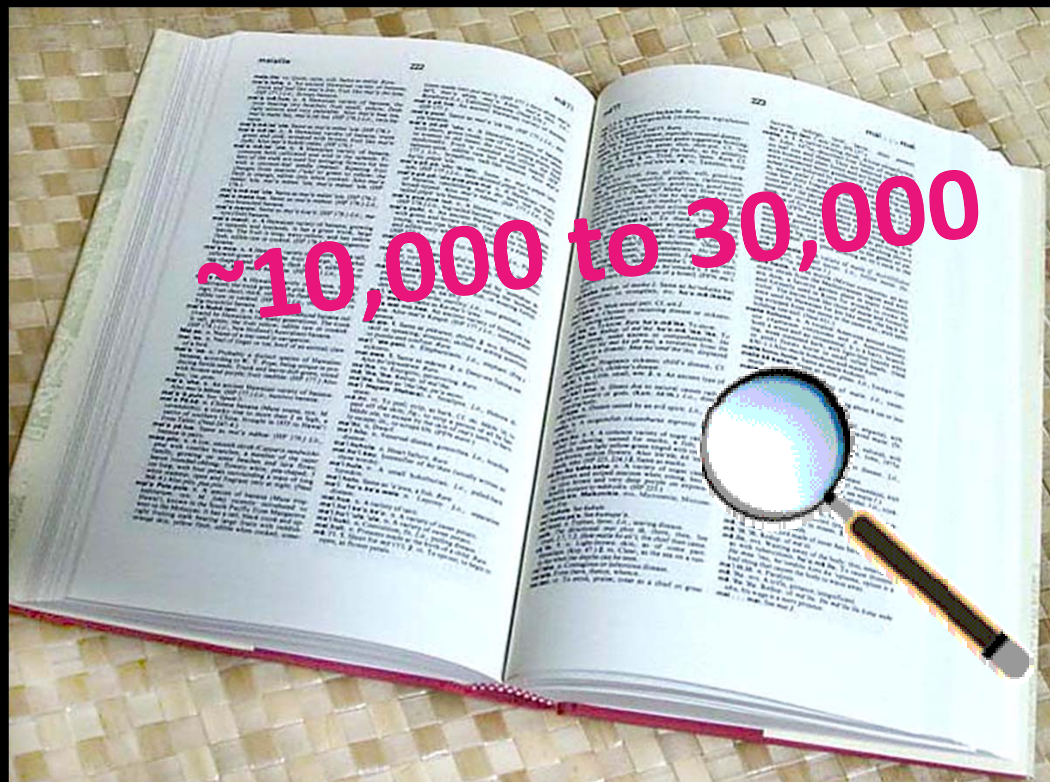
MIT Artificial Intelligence Laboratory

Object Categorization

"Fido" **German shepherd** **dog** **animal** **living being**

Which categories are the best for visual identification?

How many object categories are there?



Biederman 1987

Other Types of Categories

Functional Categories

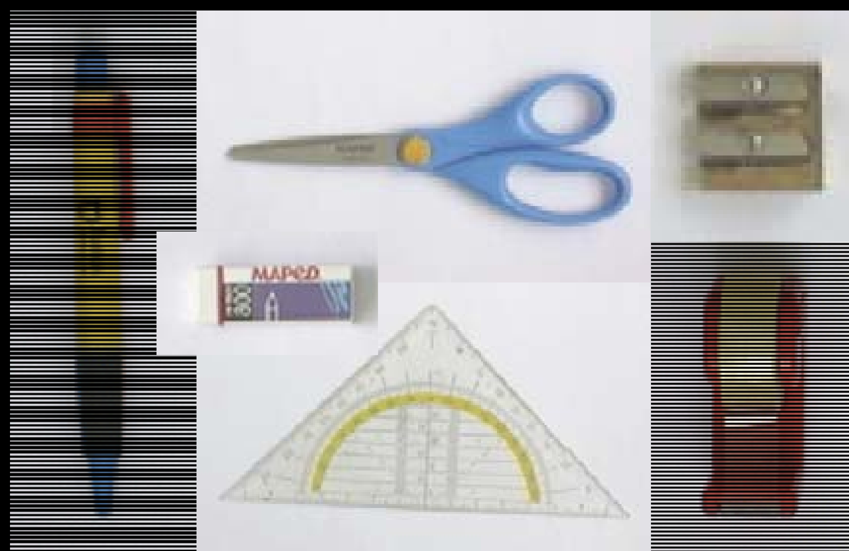
e.g. chairs = *"something you can sit on"*

K. Grauman, B. Leibe

Other Types of Categories

Ad-hoc categories

e.g. *"something you can find in an office environment"*



K. Grauman, B. Leibe

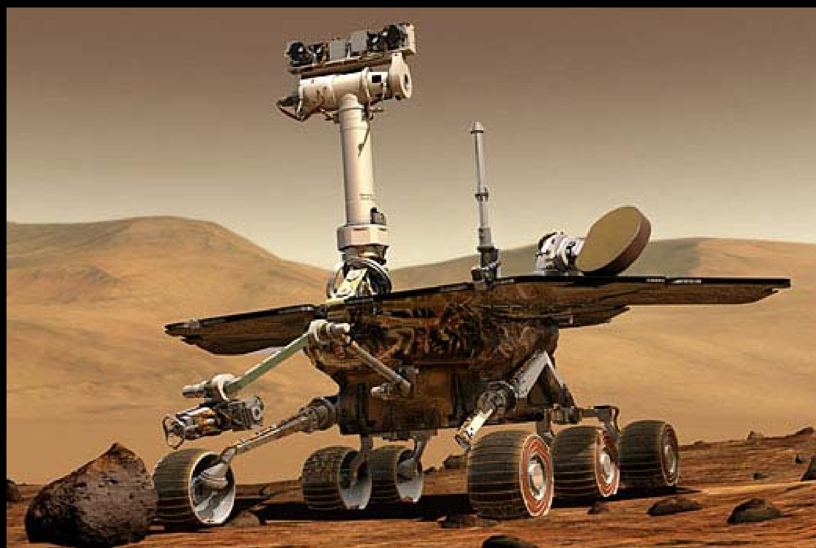
Words: Why recognition?

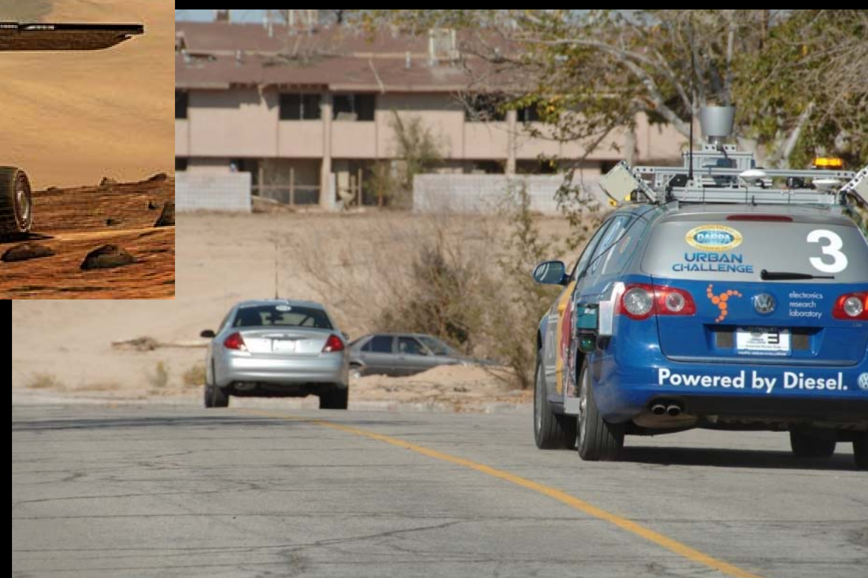
- Recognition a fundamental part of perception
  - e.g., robots, autonomous agents

- Organize and give access to visual content
  - Connect to information
  - Detect trends and themes

- Because it is a very human way of thinking about things…

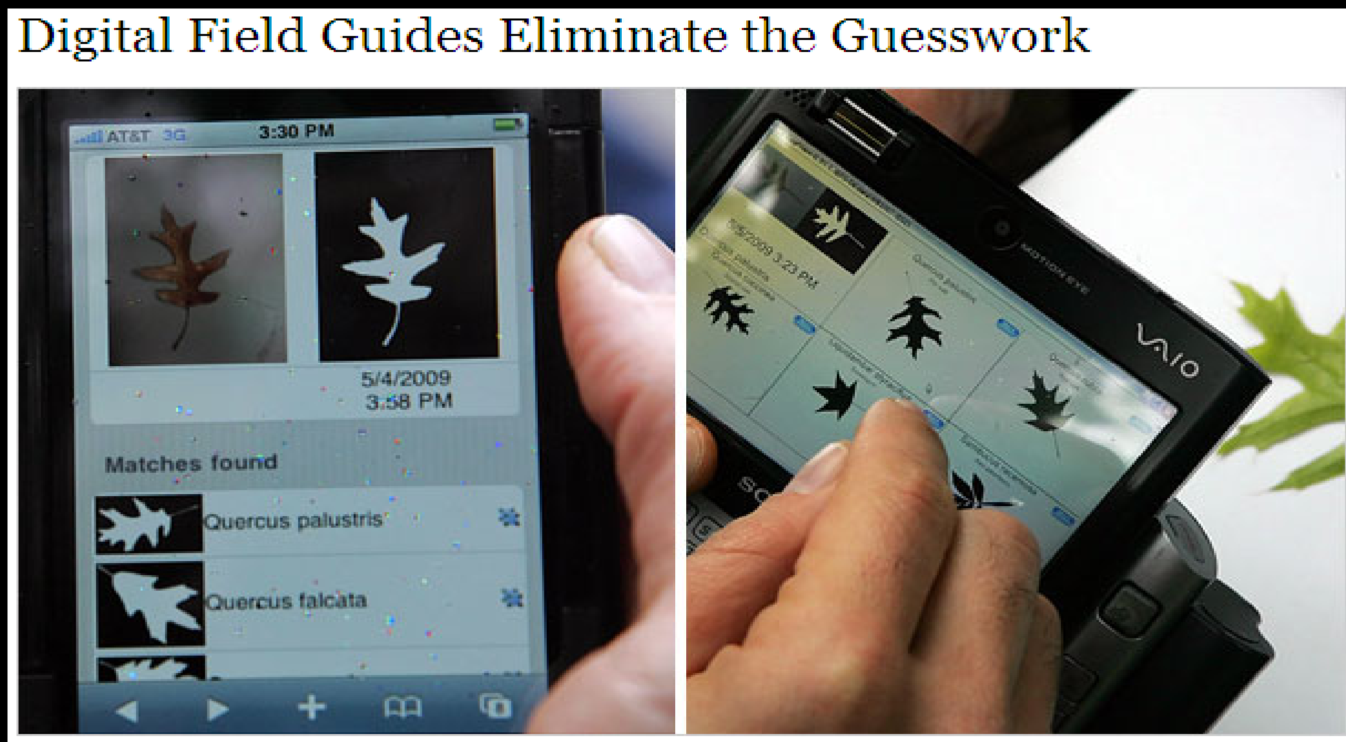
Autonomous agents able to detect objects





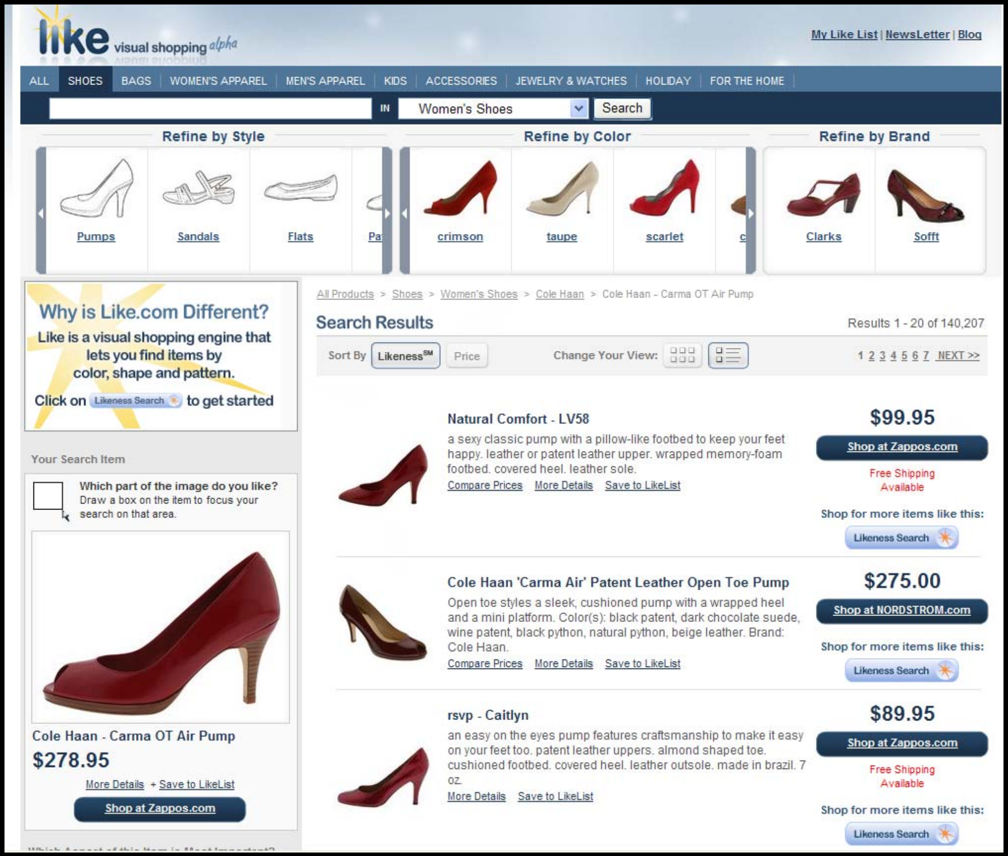
Labeling people

Posing visual queries



Belhumeur et al.

Finding visually similar objects



So why is this hard?

Challenges: Robustness

Illumination

Object pose

Clutter

Kristen Grauman

Challenges: Robustness

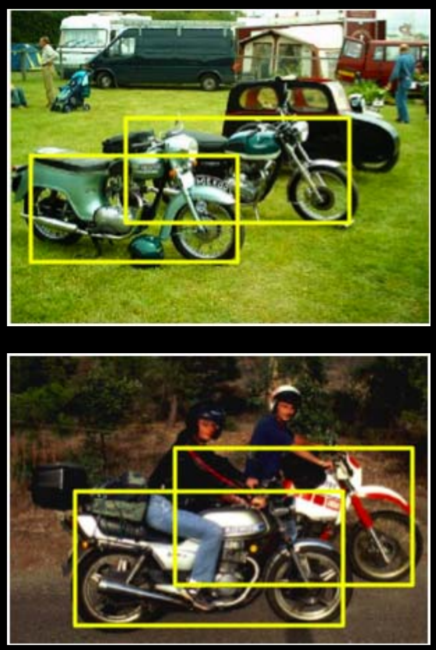
Occlusions

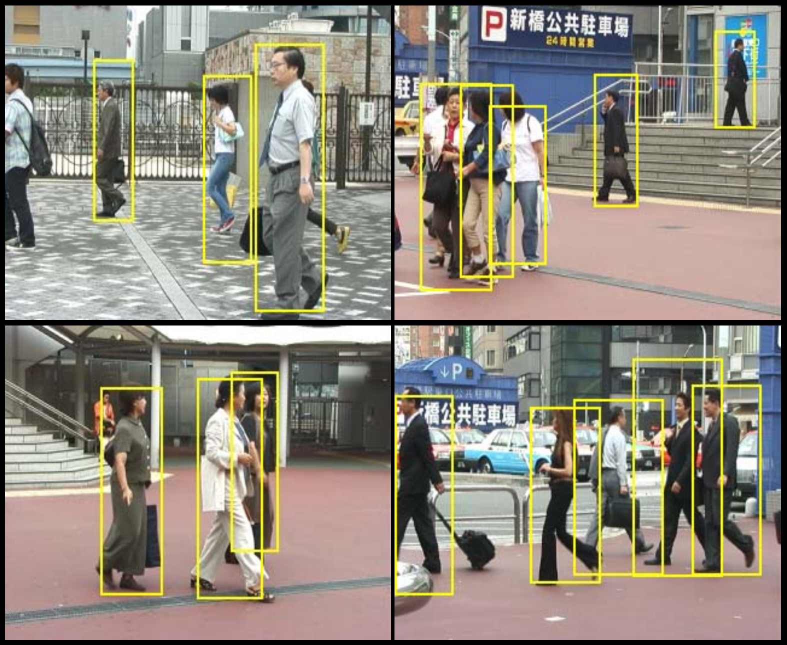
Intra-class appearance

Viewpoint

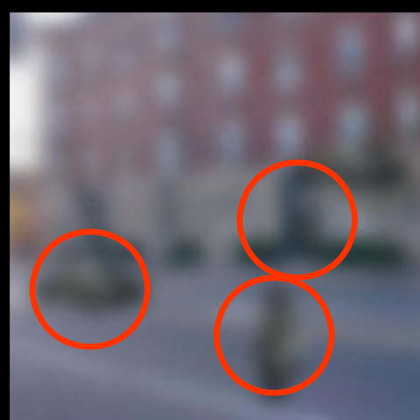
Challenges: Robustness

Realistic scenes are crowded, cluttered, have overlapping objects.



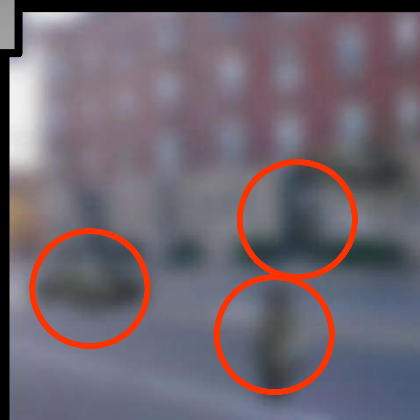


Challenges: Importance of context



Fei-Fei,
Fergus & Torralba

Challenges: Importance of context



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)

Kristen Grauman

Challenges: complexity

- Billions of images indexed by Google Image Search
- In 2011, 6 billion photos uploaded *per month*
- Approx one billion million camera phones sold in 2013
- About half of the cerebral cortex in primates is devoted to processing visual information [Felleman and van Essen 1991]

Kristen Grauman

So what works?

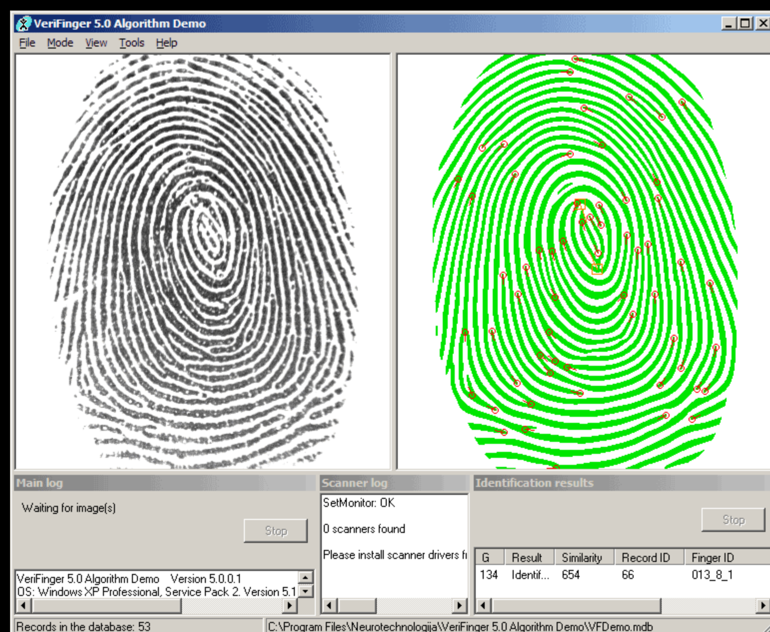
What worked most reliably “yesterday”

- Reading license plates (real easy), zip codes, checks

Lana Lazebnik

What worked most reliably "yesterday"

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



Lana Lazebnik

What worked most reliably "yesterday"

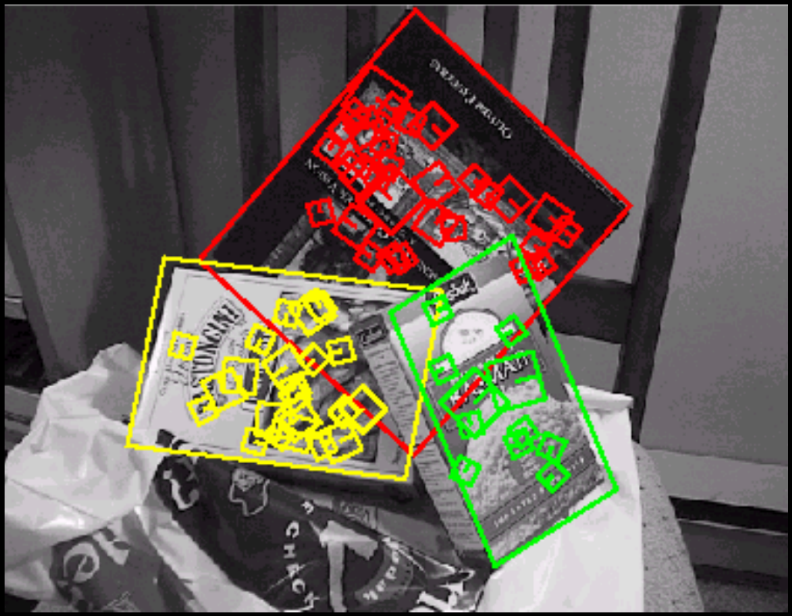
- Reading license plates, zip codes, checks
- Fingerprint recognition
- Face detection

  (Today recognition)

What worked most reliably "yesterday"

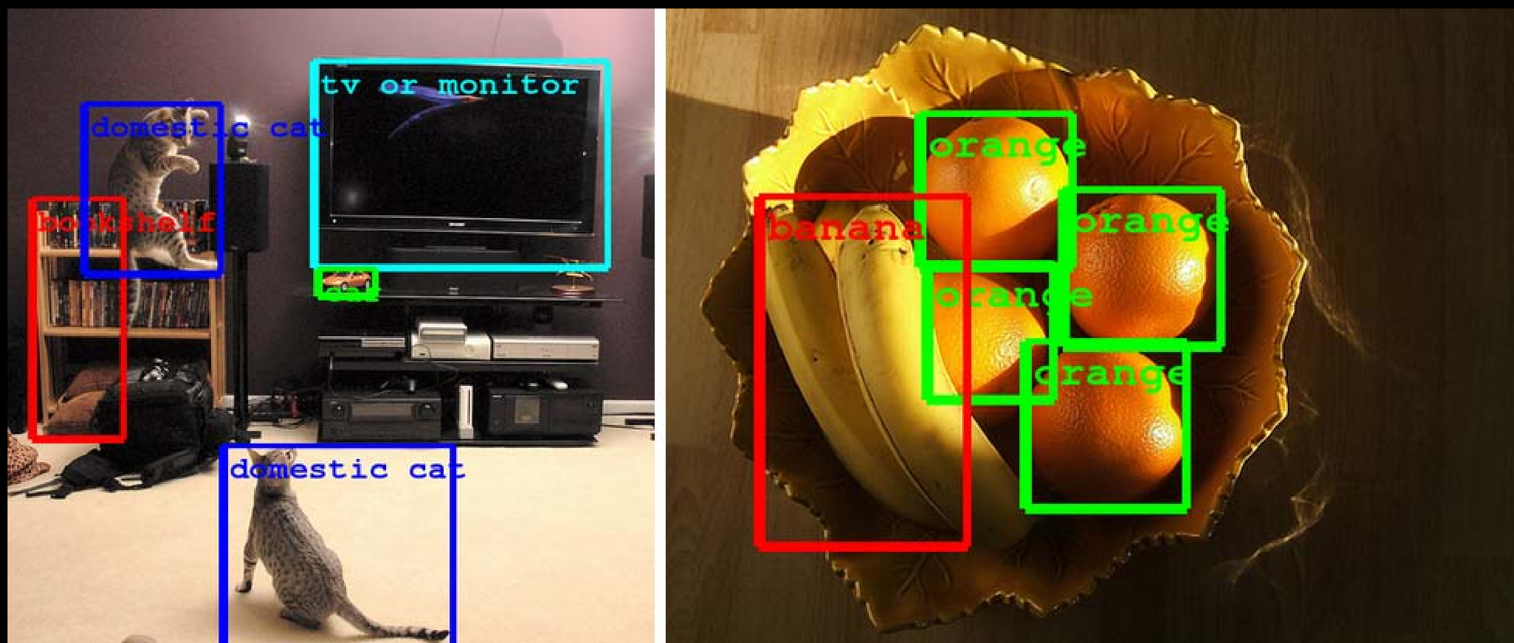
- Reading license plates, zip codes, checks
- Fingerprint recognition
- Face detection

  (Today recognition)
- Recognition of flat textured objects
  (CD covers, book covers, etc.)

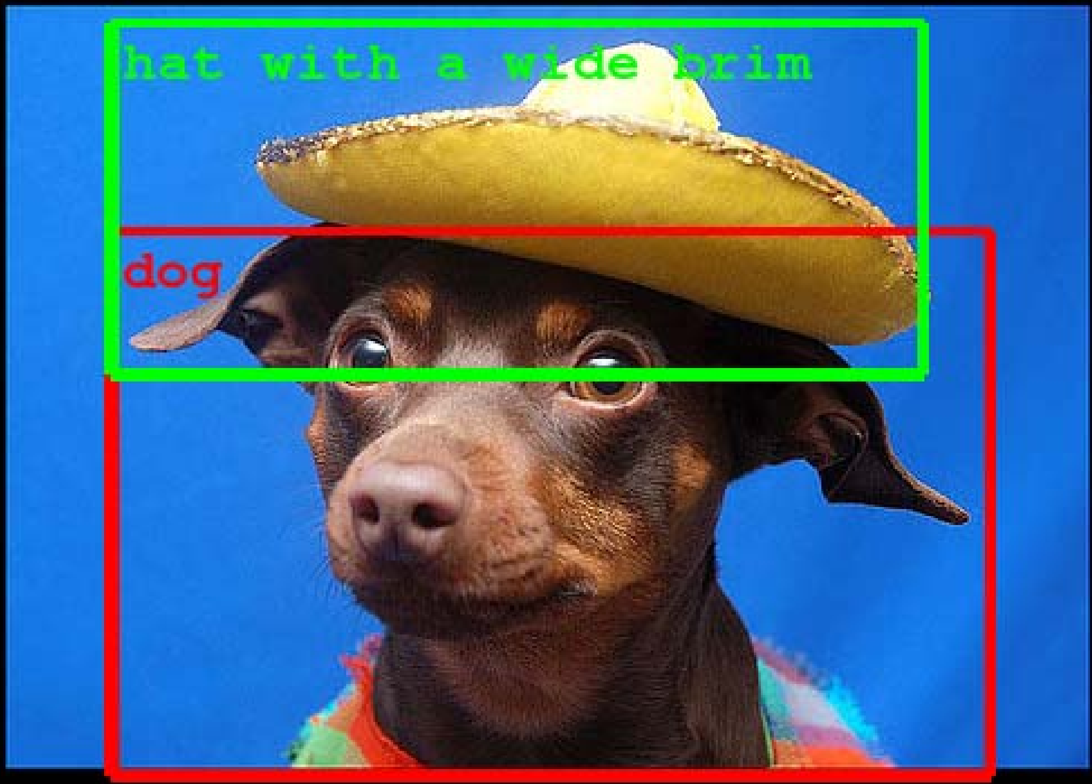


Lana Lazebnik

Just in: GoogleNet 2014



Just in: GoogleNet – no context needed?



[redacted]

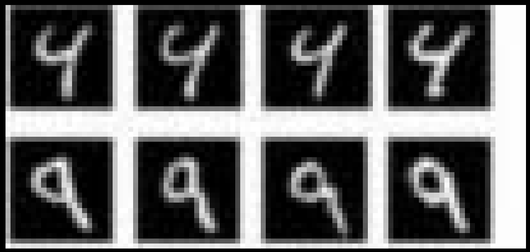
Supervised classification

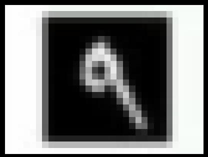
Given a collection of labeled examples, come up with a function that will predict the labels of new examples.

Training examples

"four"

"nine"





?

Novel input

Kristen Grauman

Supervised classification

How good is the function we come up with to do the classification? (What does "good" mean?)

Depends on:

- What mistakes does it make
- Cost associated with the mistakes

Kristen Grauman

Supervised classification

Since we know the desired labels of training data, we want to *minimize the expected misclassification*

Supervised classification

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***)

Supervised classification: Generative

Given labeled training examples, predict labels for new examples

- Notation: $(4 \rightarrow 9)$ - object is a '4' but you call it a '9'
- We'll assume the cost of $(X \rightarrow X)$ is zero.

Kristen Grauman

Supervised classification: Generative

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

Kristen Grauman

Supervised classification: Generative

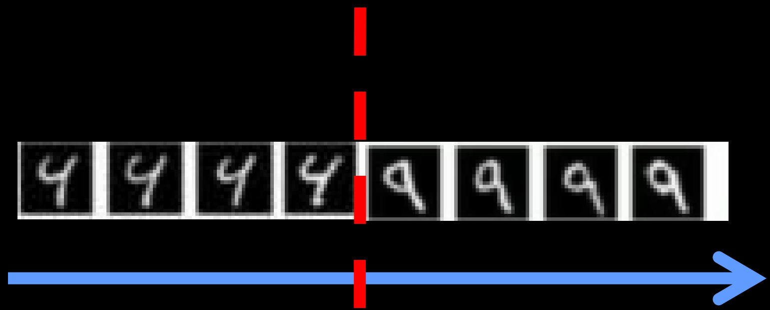
Risk of a classifier strategy ***S*** is expected loss:

$$R(\mathrm{S}) = \Pr\left(4 \to 9 \mid \text{using S}\right) L\left(4 \to 9\right) + \Pr\left(9 \to 4 \mid \text{using S}\right) L\left(9 \to 4\right)$$

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

Kristen Grauman

Supervised classification: minimal risk



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

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

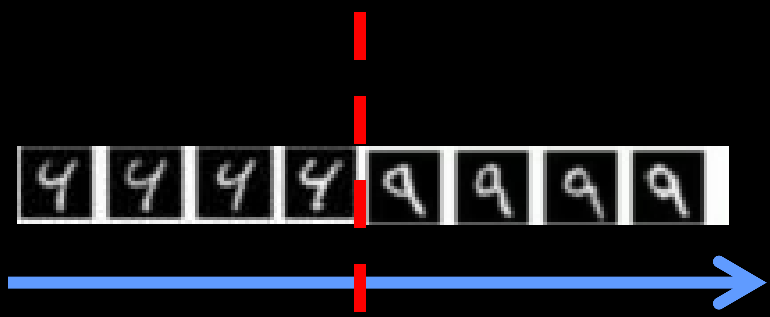
$$= P(\text{class is } 9|\mathbf{x})\ L(9 \to 4) + P(\text{class is } 4\ |\mathbf{x}) L(4 \to 4)$$

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

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

Kristen Grauman

Supervised classification: minimal risk



Feature value x

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

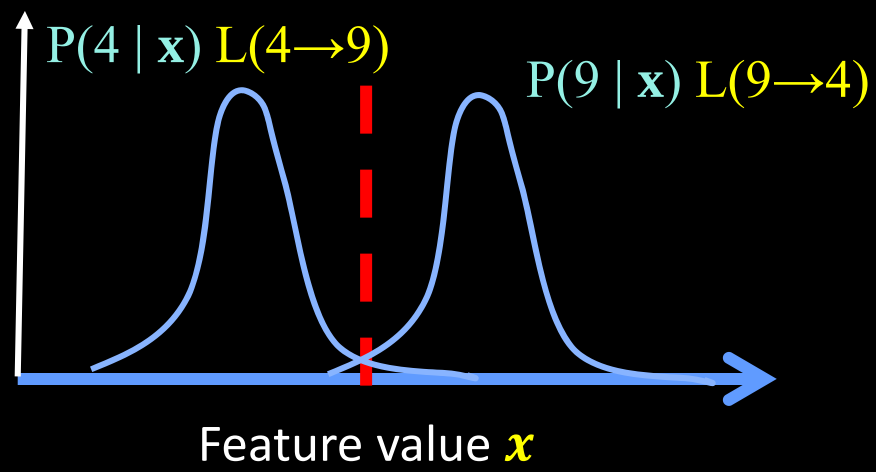
So, best decision boundary is at point **x** where:

$$P(\text{class is } 9|\mathbf{x})\ L(9 \rightarrow 4) = \mathrm{P}(\text{class is } 4|\ \mathbf{x}) L(4 \rightarrow 9)$$

To classify a new point, choose class with lowest expected loss; i.e., choose "four" if: $P(4\,|\,\mathbf{x})L(4 \rightarrow 9) > P(9\,|\,\mathbf{x})L(9 \rightarrow 4)$

Kristen Grauman

Supervised classification: minimal risk



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

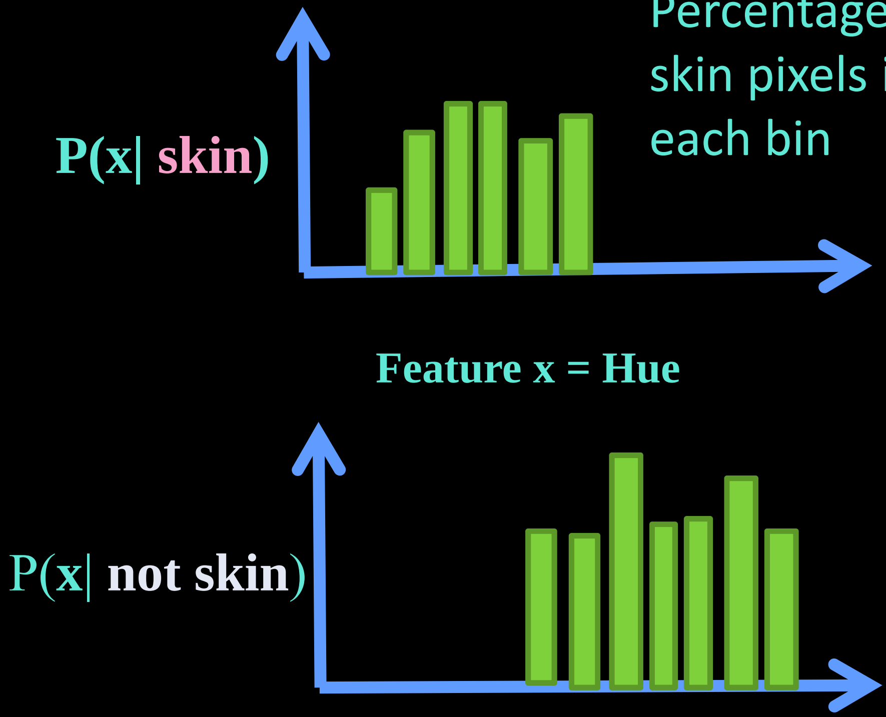
So, best decision boundary is at point **x** where:

$$\boxed{P(\text{class is } 9|\mathbf{x})}\, L(9 \rightarrow 4) = \boxed{P(\text{class is } 4|\ \mathbf{x})}\, L(4 \rightarrow 9)$$

How to evaluate these probabilities?

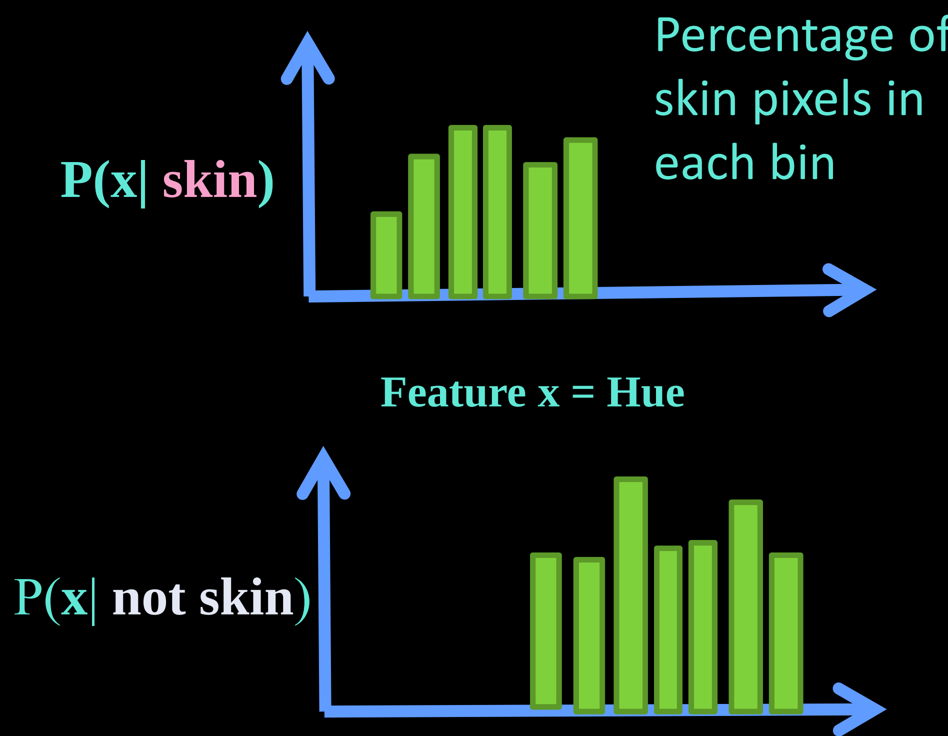
Kristen Grauman

Example: learning skin colors



Kristen Grauman

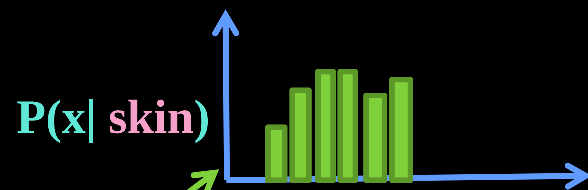
Example: learning skin colors



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

Kristen Grauman

Bayes rule



$$\underbrace{P(\text{skin} \mid x)}_{\text{posterior}} = \frac{\underbrace{P(x \mid \text{skin})}_{\text{likelihood}} \, \underbrace{P(\text{skin})}_{\text{prior}}}{P(x)}$$

$$P(\text{skin} \mid x) \propto P(x \mid \text{skin}) \, \boxed{P(\text{skin})}$$

Where does the prior come from?

Bayes rule in (ab)use

Likelihood ratio test (assuming cost of errors is the same):

If $P(skin|\boldsymbol{x})$ > $P(\sim skin|\boldsymbol{x})$ classify $\mathbf{x}$ as skin

… so ….

If $P(\boldsymbol{x}|skin)P(skin) > P(\boldsymbol{x}|\sim skin)(P(\sim skin)$
classify $\boldsymbol{x}$ as skin (Bayes rule)

(if the costs are different just re-weight)

Bayes rule in (ab)use

… but I don't really know prior $P(skin)$…

… but I can assume it some constant Ω …

… so with some training data I can ***estimate*** Ω …

…. and with the same training data I can ***measure*** the ***likelihood densities*** of ***both*** $P(\boldsymbol{x}|skin)$ and $P(\boldsymbol{x}|{\sim}skin)$ …

So…. I can more or less come up with a rule…

Steve Seitz

Example: classifying skin pixels

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

If $p(skin|x) > \theta$ classify as skin; otherwise not

Brighter pixels are higher probability of being skin

Kristen Grauman

Example: classifying skin pixels

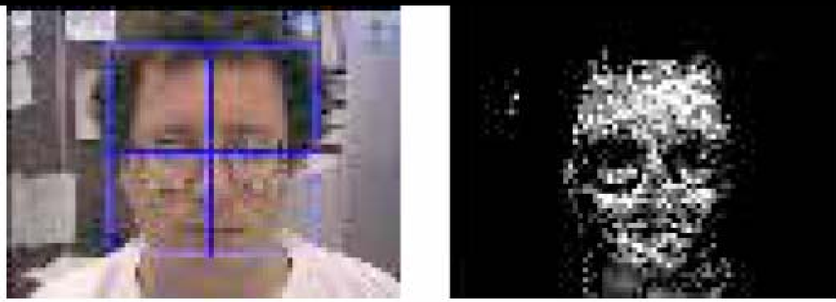


Figure 6: A video image and its flesh probability image

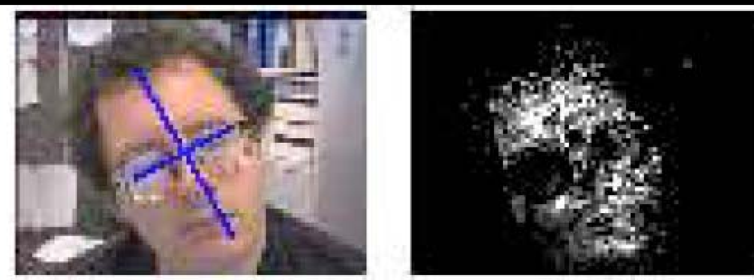


Figure 7: Orientation of the flesh probability distribution marked on the source video image

Gary Bradski, 1998

Example: classifying skin pixels

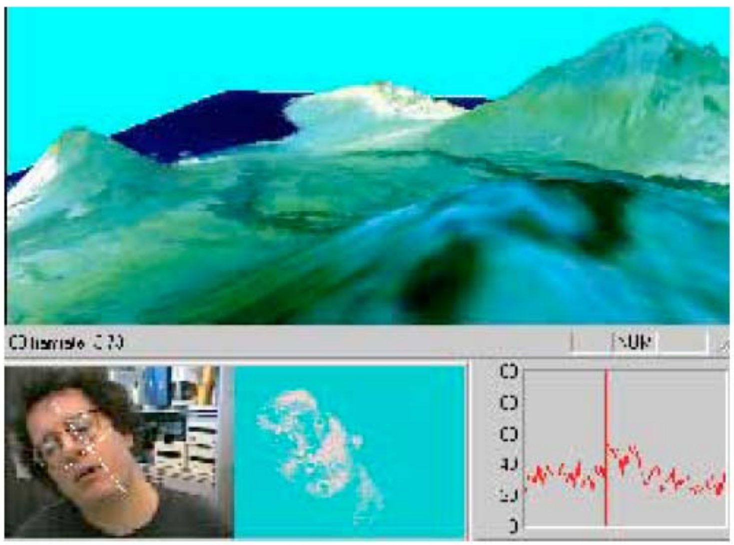


Figure 13: CAMSHIFT-based face tracker used to "fly" over a 3D graphic's model of Hawaii

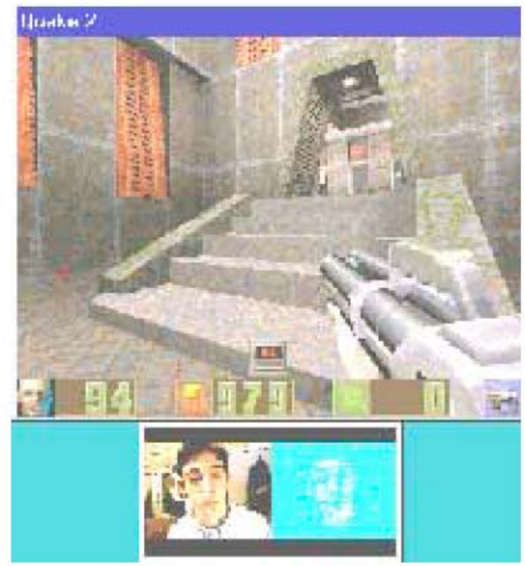


Figure 12: CAMSHIFT-based face tracker used to play Quake 2 hands free by inserting control variables into the mouse queue

Gary Bradski, 1998

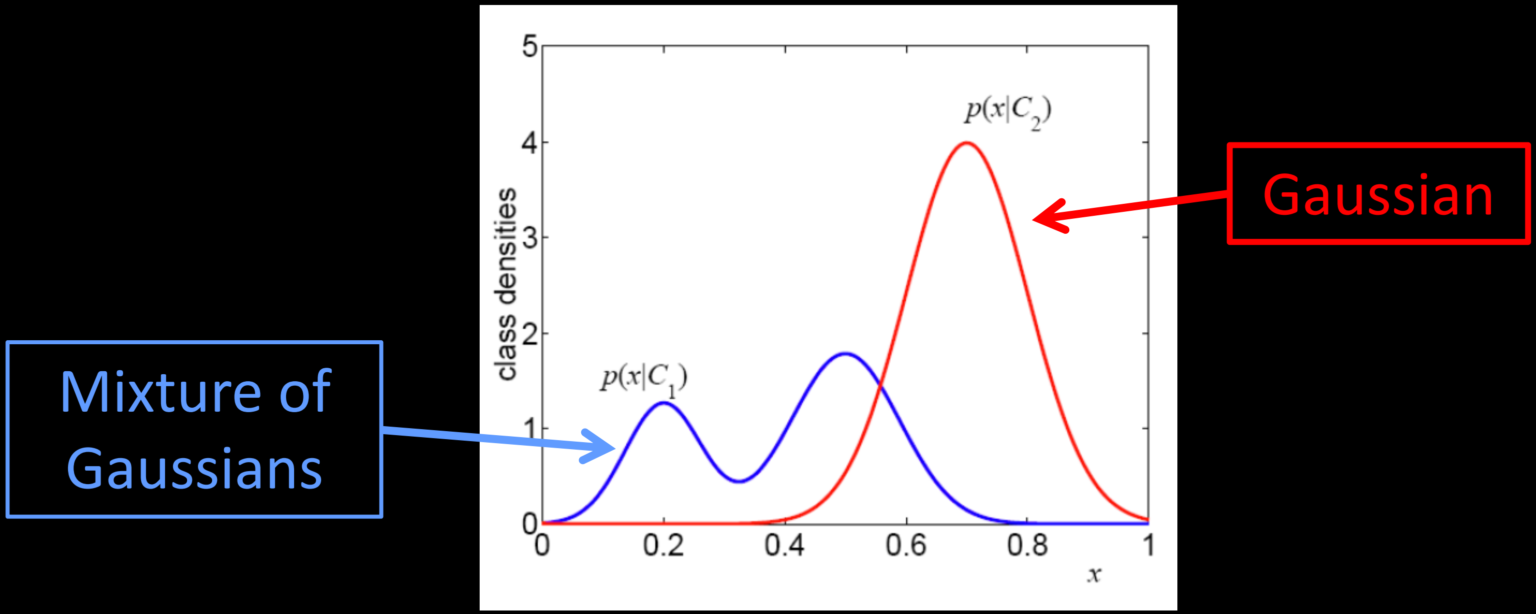
More general generative models

For a given measurement $\mathbf{x}$ and set of classes c_i choose c^* by:

$$c^* = \arg\max_c p(c \mid \mathbf{x}) = \arg\max_c p(c)\,p(\mathbf{x} \mid c)$$

Continuous generative models

- If $\mathbf{x}$ is continuous, need *likelihood* density model of $p(\mathbf{x}|c)$
- Typically parametric – Gaussian or mixture of Gaussians



Continuous generative models

- Why not just some histogram or some KNN (Parzen window) method?
  - You might…
  - But you would need lots and lots of data everywhere you might get a point
  - The whole point of modeling with a parameterized model is not to need lots of data.

Summary of generative models:

+ Firm probabilistic grounding

+ Allows inclusion of prior knowledge

+ Parametric modeling of likelihood permits using small number of examples

+ New classes do not perturb previous models

+ Others:

Can take advantage of unlabelled data

Can be used to generate samples

Summary of generative models:

- And just where did you get those priors?
- Why are you modeling those obviously non-C points?
- The example hard cases aren't special
- If you have lots of data, doesn't help

Next time…

- A really cool way of building a generative model for face recognition (not detection)
- And then *discriminative* models…