



Computer vision

Computer Vision

Lecture 5: Object Recognition I

Dr. Dina Khattab

dina.khattab@cis.asu.edu.eg

Scientific Computing Department

Instructor:	Dr. Dina Khattab
Email:	<u>dina.khattab@cis.asu.edu.eg</u>
Office:	Main Building – 4 th floor – Room 302
Office Hours:	Monday 12:00 - 2:00 PM Thursday 11:00 AM to 12:00 PM

Agenda

- Object Recognition
 - Generative Models (Naïve Bayes)
 - App: Skin recognition
 - Discriminative Models (Linear classifiers)
 - Bag of Words (BOW)

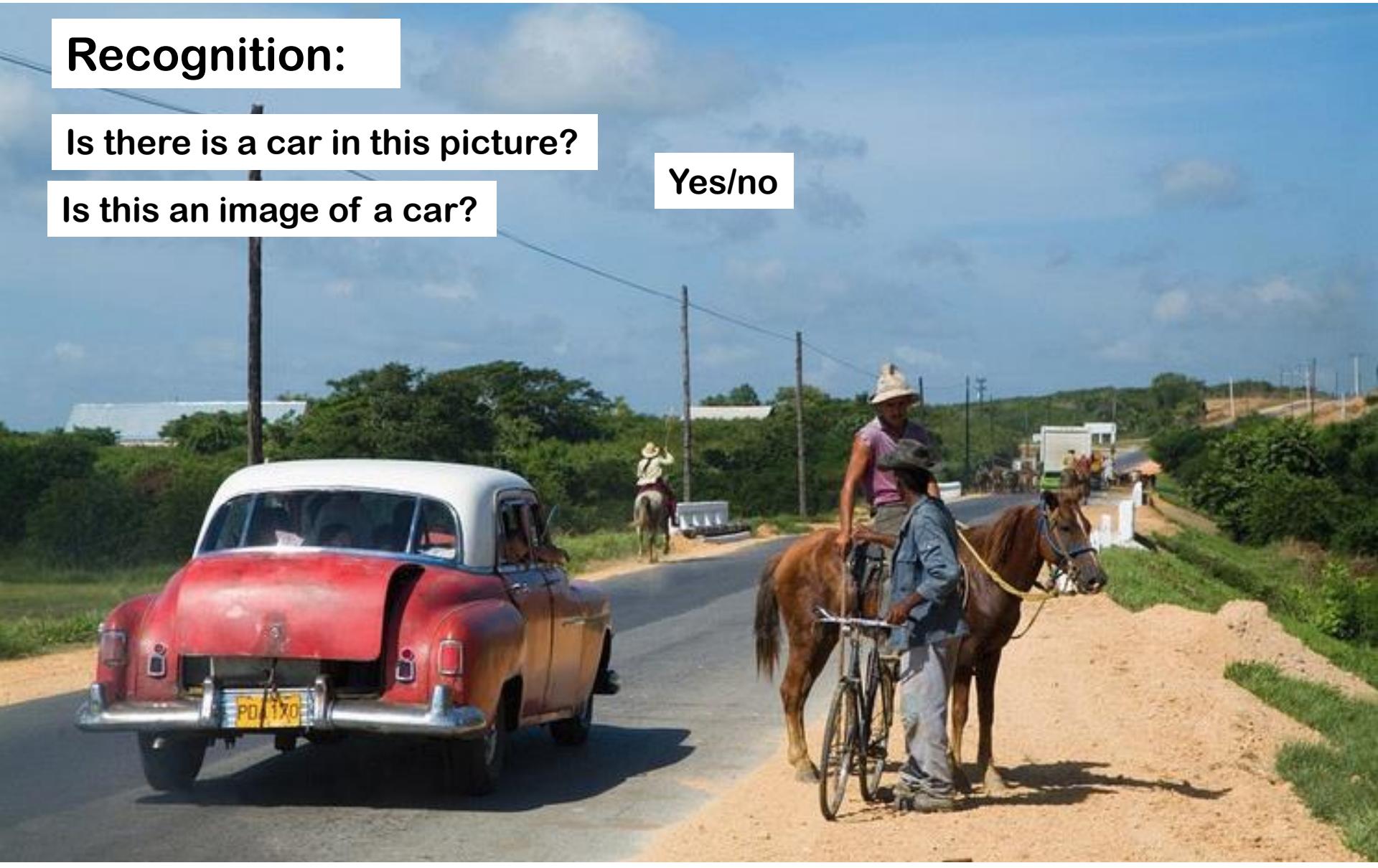
Visual Recognition

Recognition:

Is there is a car in this picture?

Yes/no

Is this an image of a car?



Visual Recognition

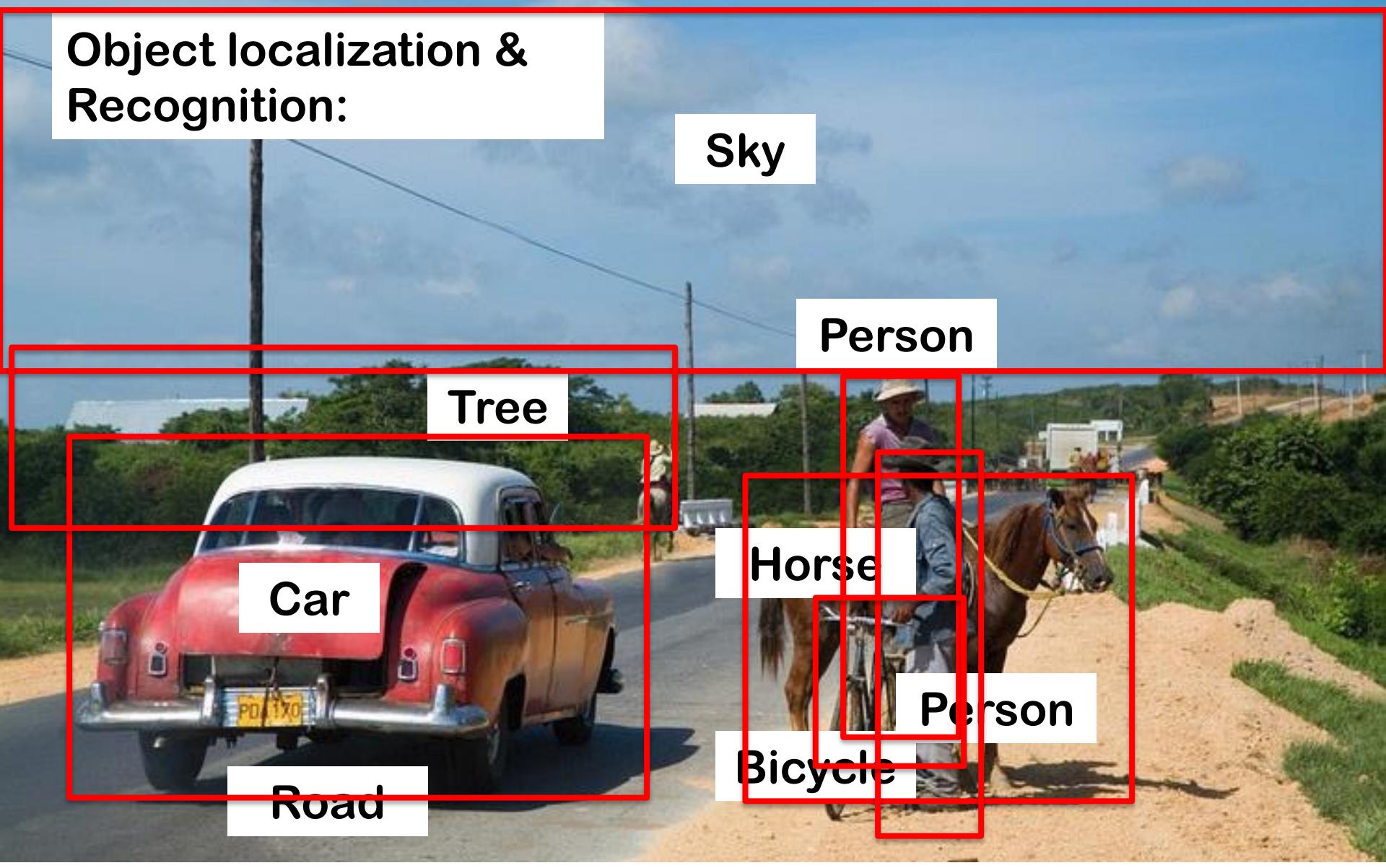
Detection

Where is the car?



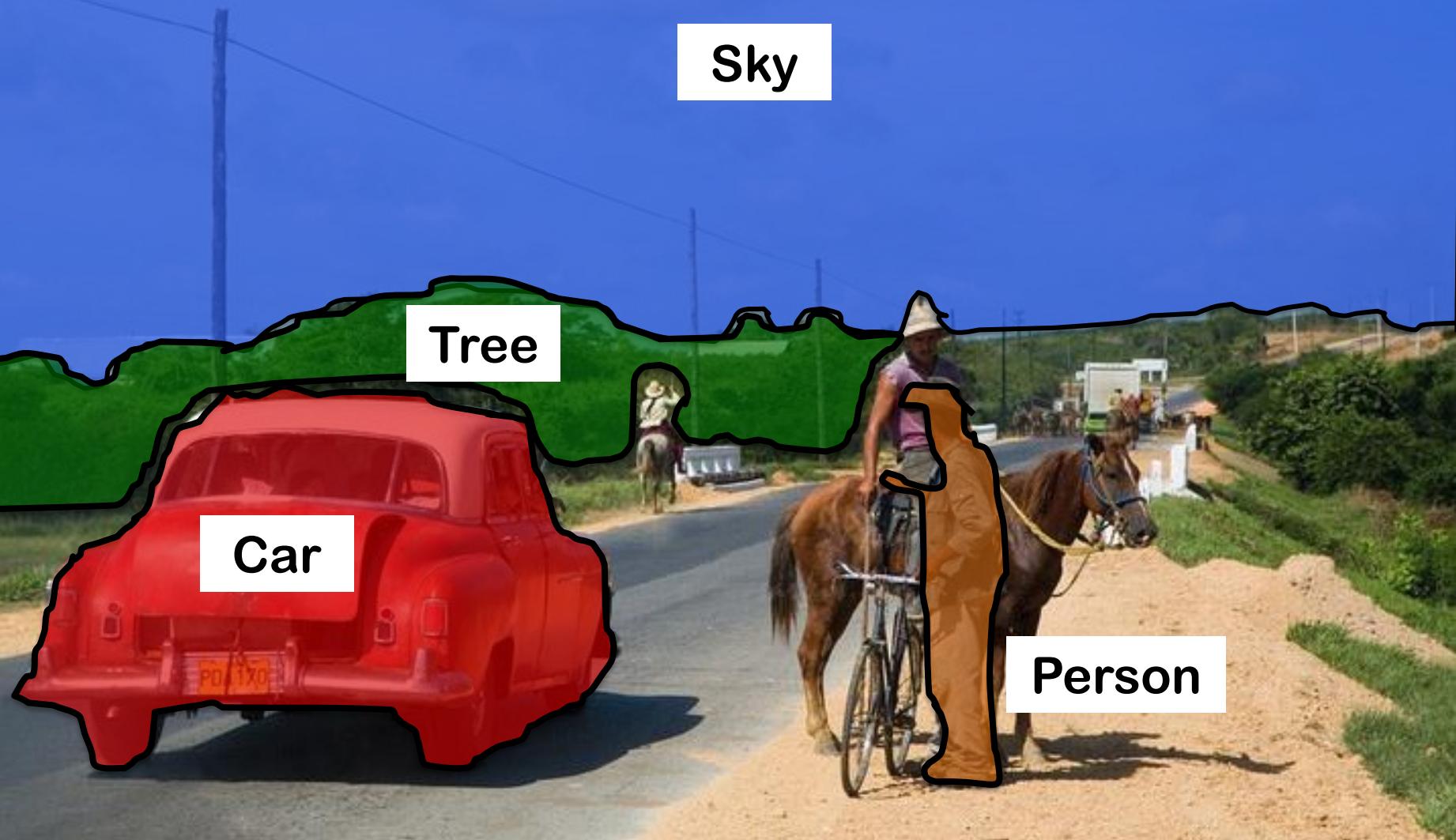
Visual Recognition

Object localization &
Recognition:



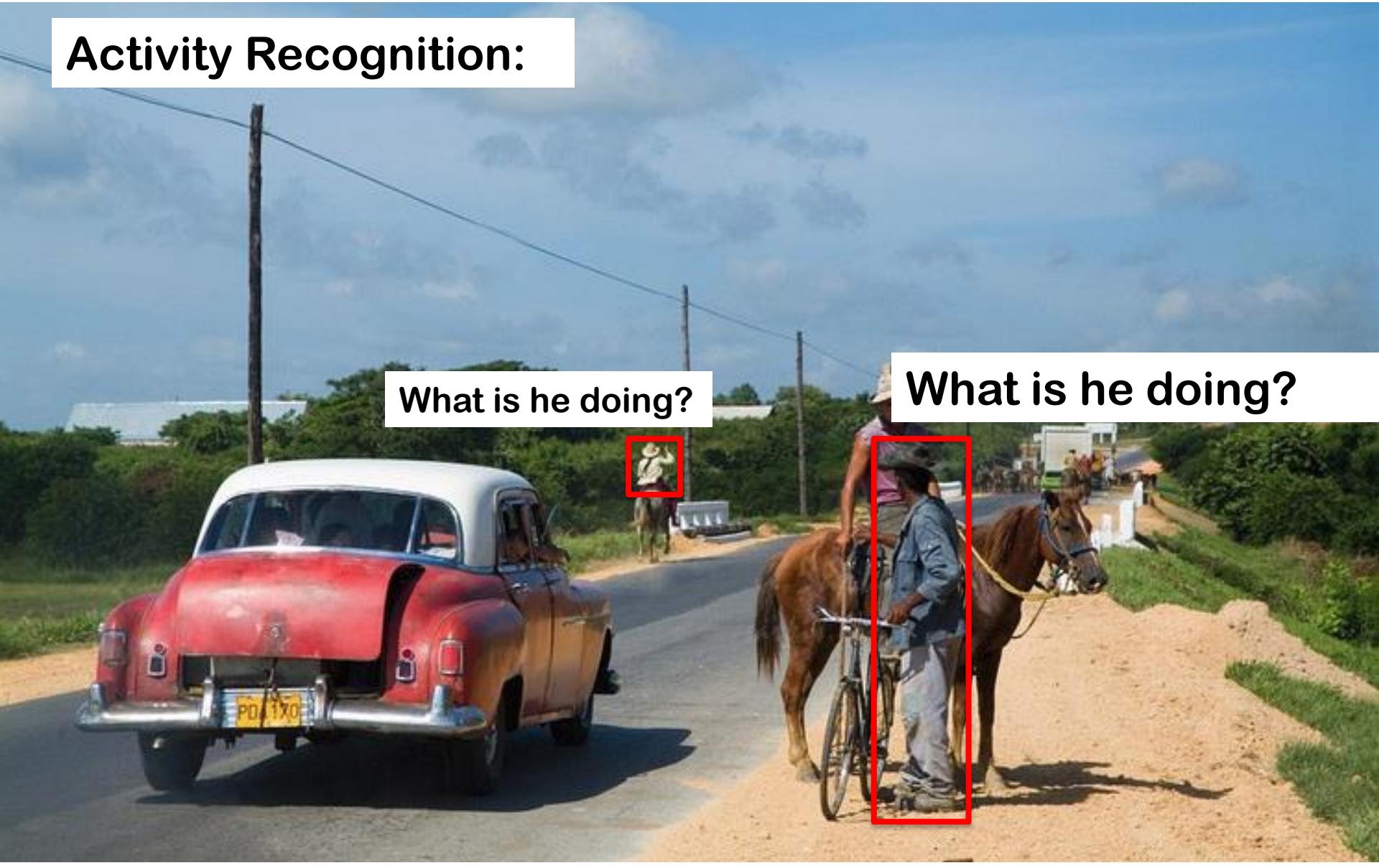
Visual Recognition

Segmentation



Visual Recognition

Activity Recognition:



Challenges: Robustness

- **Viewpoint variation:** There are many potential ways to view an object, and the change in viewpoint can lead an object to look very different.



Michelangelo 1475-1564

Challenges: Robustness

- **Illumination:** Different levels of light, particularly low light or a different light direction, will cause shadows to shift and the details of an object to become obscured.

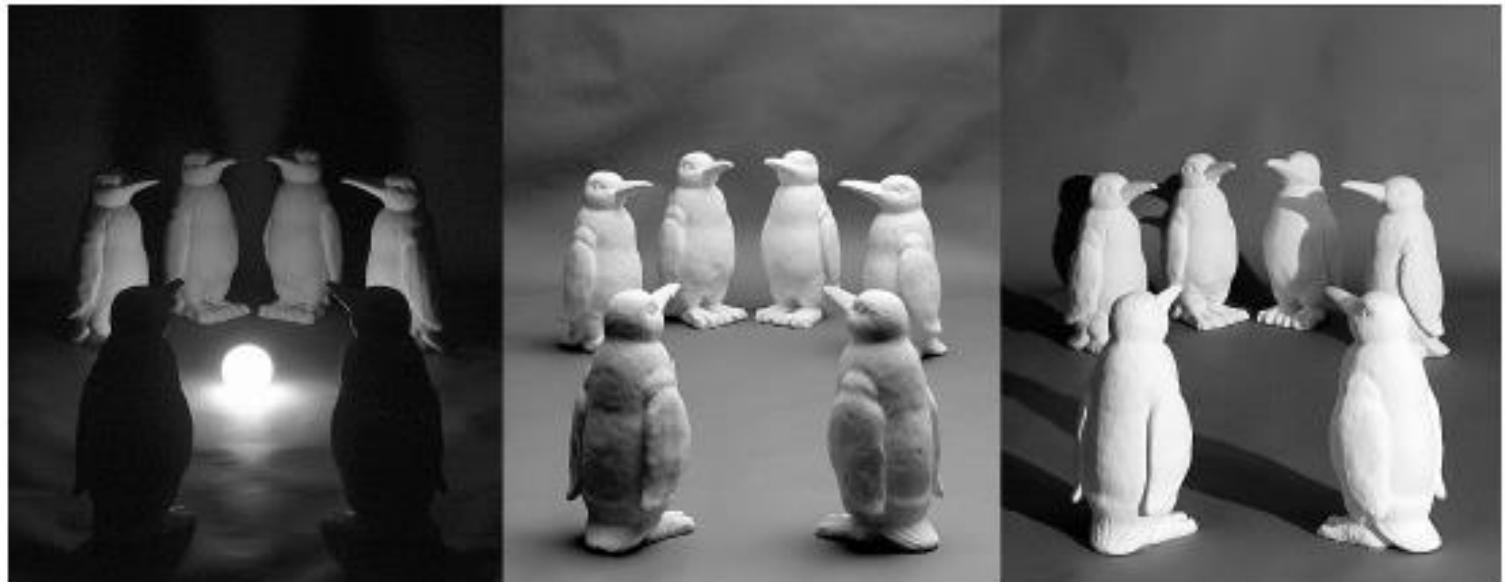


Image credit: J. Koenderink

Challenges: Robustness

- **Scale** Objects belonging to one category can come in a variety of sizes. If a classification is only trained on a particular size of object, then it will fail to recognize the same object in a different size.



Challenges: Robustness

- **Object pose:** Objects can change form and look very different, while still being considered the same object. For example, a person can be photographed in a number of poses, but is still considered a person if they're bending over or if their arms are crossed.



Challenges: Robustness

- **Occlusion** Objects may be occluded, which could hide aspects of their characteristic geometry.



Challenges: Robustness

- **Background clutter** Similarities between the texture, color, and shapes in the background and the foreground can make it difficult to detect an object .



Challenges: Robustness

- **Intra-class variation** There can be significant shape variations even within one class of objects. For example, everything from a barstool to a lounge chair can be considered a chair.



Supervised Classification

- Two General Strategies:

1. Generative:

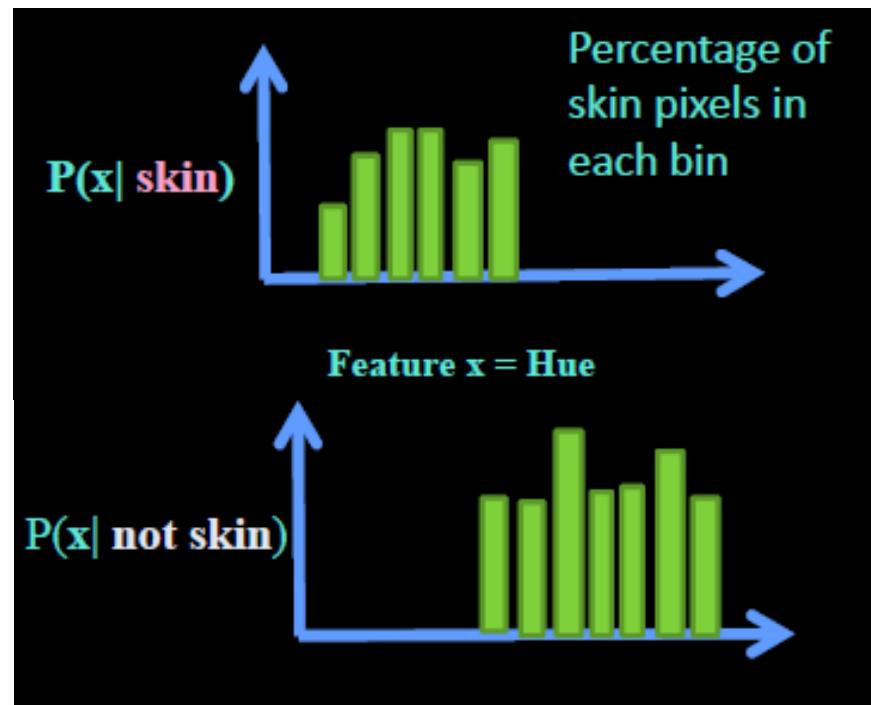
- Use the training data to build representative probability model for each class. (Model the distribution that generates the data).

2. Discriminative:

- Construct a good decision boundary between different classes.

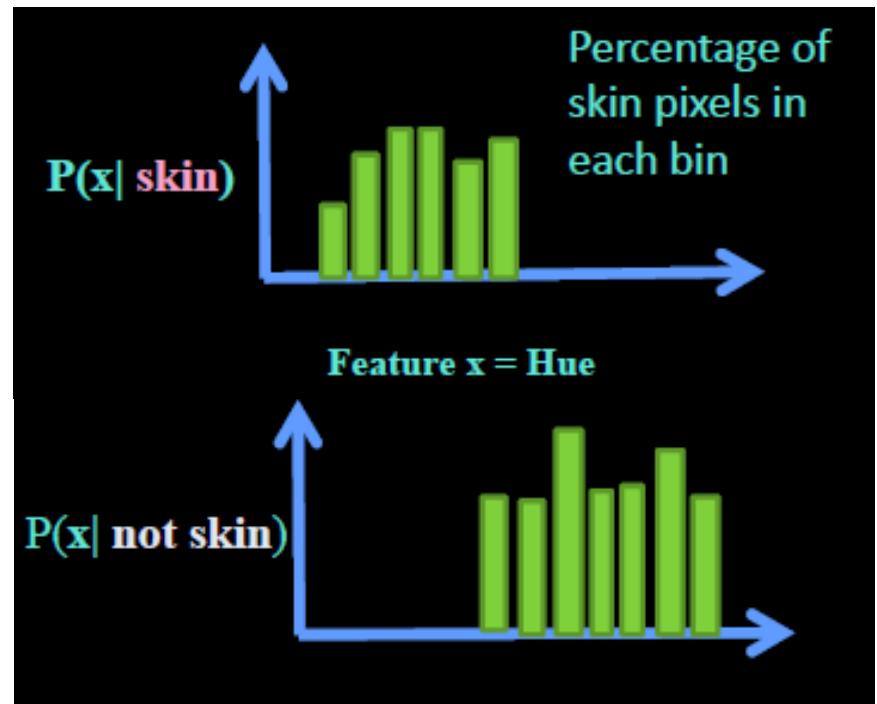
Supervised Classification: Generative (Naïve Bayes)

- Ex. Learning Skin colors

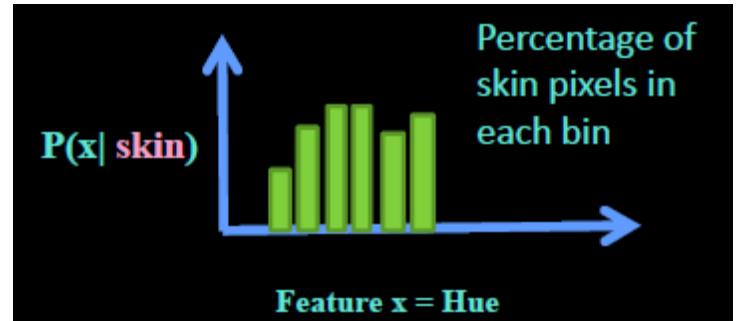


Ex. Learning Skin colors

- $P(\text{skin} \mid x) ?!!$



Bayes Rule



Posterior

Likelihood

Prior

- $P(\text{skin}|x) = \frac{p(x|\text{skin})P(\text{skin})}{p(x)}$
- $P(\text{skin}|x) \propto p(x|\text{skin})P(\text{skin})$

Bayes Rule

- If $P(\text{skin} | x) > P(\sim \text{skin} | x)$

OR $P(x | \text{skin})p(\text{skin}) > P(x | \sim \text{skin})p(\sim \text{skin})$

then **classify x as skin**

We can get an evaluation If $P(\text{skin} | x) > \theta$

More general generative models

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

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

Generative Models

Advantages

- Pure probabilistic models.
- Parametric modeling of likelihood permits using small number of examples.
- New classes don't disturb previous models.

Disadvantages

- Where to get the priors?
- Modeling non-c points?
- It doesn't help if you have lots of data.

Some challenges for generative models

- But for the modern world there are some liabilities:
- Many signals are *high-dimensional* and *representing the complete density of class is data-hard.*
 - In some sense, we don't care about modeling the classes, we *only care about making the right decisions.*
 - *Model the hard cases-the ones near the boundaries.*
- We don't typically know which features of instances actually *models* the different classes.

So..

- We want to focus on *discriminating* between the class types.
- We want the machine to somehow *learn* the features that matter.
- This gets us to *discriminative classification*

Discriminative Classification

Find a division (surface) in feature space that separates the classes

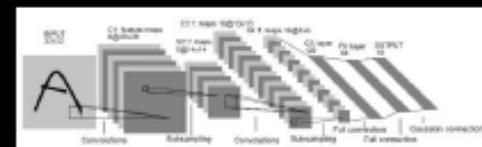
Nearest neighbor



10^6 examples

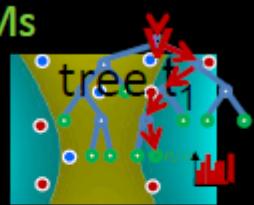
Shakhnarovich, Viola, Darrell 2003
Berg, Berg, Malik 2005 ...

Neural networks



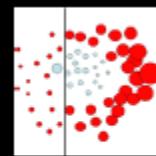
LeCun, Bottou, Bengio, Haffner 1998
Rowley, Baluja, Kanade 1998 ...

SVMs



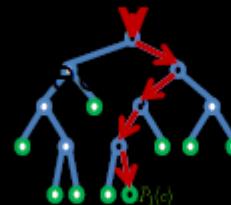
Guyon, Vapnik, Heisele,
Serre, Poggio, 2001, ...

Boosting



Viola, Jones 2001,
Torralba et al. 2004,
Opelt et al. 2006, ...

Random Forests



Breiman, 1984
Shotton, et al CVPR 2008

Image Classification using SVM

BAG OF VISUAL WORDS (BOW)

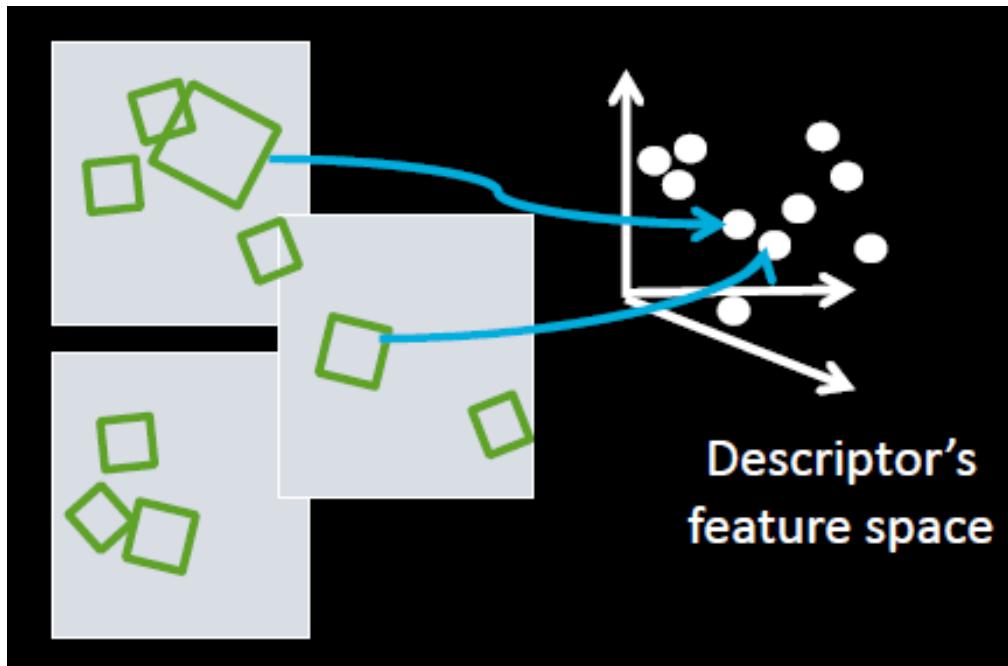
Image Classification using SVM

- Image → Feature extraction → classifier



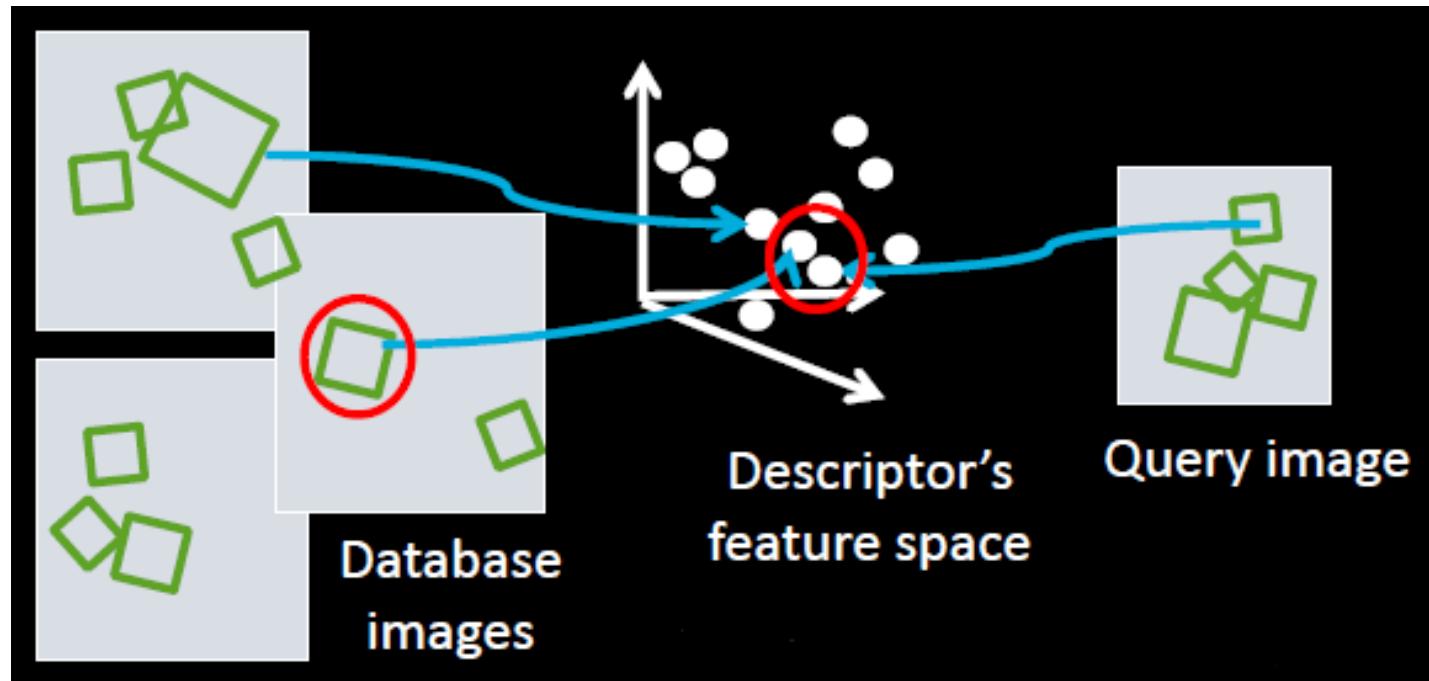
Indexing Local Features

Each patch / region has a descriptor, which is a point in some high-dimensional feature space (e.g., SIFT)



Indexing Local Features

When we see close points in feature space, we have similar descriptors, which indicates similar local content.



Easily can have millions of features to search!

Indexing Local Features

- With potentially **thousands** of features per image, and **hundreds to millions** of images to search, how to efficiently find those that are relevant to a new image?

Indexing local features: inverted file index

- For text documents, an efficient way to find all *pages* on which a *word* occurs is to use an index...
- We want to find all *images* in which a *feature* occurs.
- To use this idea, we'll need to map our features to “visual words”

Index	
“Along I-75,” From Detroit to Florida; Inside back cover	Butterfly Center; McGuire; 134
“Drive I-95,” From Boston to Florida; Inside back cover	CAA (see AAA)
1929 Spanish Trail Postcard;	CCC; They; 111,113,115,135,142
101-102,104	Citizen; 147
811 Traffic Information; 83	Caloosahatchee River; 152
ATA (Banner 18) - 156 Assess; 86	Name; 160
AAA (and CAA); 83	Canaveral National Seashore; 173
AAA National Office; 86	Cannon Creek Airport; 130
Affiliations;	Clancy Road; 106,109
Colored 25 mile Maps; cover	Gape Conover; 174
Exit Services; 98	Castillo San Marcos; 166
Thesaurus; 85	Cave Diving; 131
Africa; 177	Cayo Costa, Name; 160
Agricultural Inspection Service; 128	Celebration; 93
An-Tan-Thi-Ki Museum; 180	Charlotte County; 149
Air Conditioning, Fast; 112	Charlotte Harbor; 199
Alabama; 124	Chattahoochee; 116
Alaska; 132	Charlottesville, Name; 115
County; 131	Circus Museum, Ringling; 147
Alafia River; 143	Citrus; 88,97,130,136,145,150
Alaska, Name; 126	CityPlace, Ft. Palm Beach; 180
Alfred B. Maclay Gardens; 108	City Maps;
Alligator Alley; 154-155	Ft Lauderdale-Dowrys; 194-196
Alligator Hamm, St Augustine; 160	Jacksonville; 163
Alligator Hole Metacarta; 167	Kissimmee Express; 162-163
Alligator, Buddy; 155	Miami Expressways; 194-195
Alligators; 106-105,108,147,156	Orlando Expressways; 192-193
Anastasia Island; 170	Pensacola; 20
Anchovies; 105-108,148	Tellurian; 191
Apalachicola River; 112	Tampa, St. Petersburg; 63
Appleton Map of Art; 156	St. Augustine; 191
Archae; 102	Civil War; 100,108,127,138,141
Antique Nights; 94	Cleanwater Marine Aquarium; 187
Art Museum, Ringling; 147	Collier County; 154
Aruba Beach Cafe; 183	Collier, Barron; 152
Auxilla River Project; 106	Colonial Spanish Quarters; 169
BeeCook-Web WMA; 151	Columbia County; 101,128
Bethleem Mtn Marina; 184	Cognac Building Material; 165
Baker County; 99	Cookhouse Swamp, Name; 151
Burke's Mallmorn; 182	Cowboys; 85
Bunge Canal; 189	Croo Tap It; 144
Bee Line Epsy; 80	Cracker, Florida; 88,95,132
Belle Outlet Mall; 89	Crestview Epsys; 11,85,98,143
Bernard Castro; 156	Cuban Bread; 104
Big 'T'; 163	Dade Battlefield; 149
Big Cypress; 155,158	Dade, Mt. Francis; 133-140,161
Big Foot Monster; 105	Dania Beach Hurricane; 188
	Driving Lines; 65
	Duval County; 153
	Eau Gallie; 175
	Edison, Thomas; 152
	Edlin APB; 118-119
	Eight Roads; 176
	Elendris; 144-148
	Emerson Point Wedge; 129
	Emergency Callbox; 69
	Epiphytes; 142,148,157,159
	Essential Bay; 119
	Bridge (1-10); 119
	County; 120
	Estero; 153
	Everglades; 90,95,109-140,154-160
	Draining of; 186,187
	Wetland; 180
	Wonder Garden; 154
	Falling Waters SP; 155
	Fantasy of Flight; 95
	Fever Dykes SP; 171
	Fines, Forest; 106
	Fines, Prescribed; 148
	Fisherman's Village; 151
	Flagler County; 171
	Hagler, Harry; 97,168,167,171
	Florida Aquarium; 169
	Florida,
	12,000 years ago; 187
	Covers SP; 114
	Map of all Expressways; 2-3
	Map of Natural History; 184
	National Cemetery; 141
	Plan of Alton; 177
	Platform; 187
	Sheaff's Boys Camp; 126
	Sports Hall of Fame; 180
	Sun 'n Fun Museum; 97
	Supreme Court; 107
	Florida's Turnpike (FTP); 178,188
	25 mile Strip Maps; 60
	Administrative; 189
	Coin System; 169
	Exit Services; 189
	HEFT; 70,181,186
	History; 188
	Names; 129
	Service Plazas; 190
	Spur SP91; 70

Bags of Features for Object Recognition

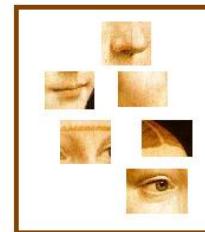


face, flowers, building

- "Bag of Words" is a way to simplify object representation as a collection of their subparts for purposes such as classification.

- Take a bunch of images, extract features, and build up a “dictionary” or “visual vocabulary” – a list of common features
- **Bag of features: outline**

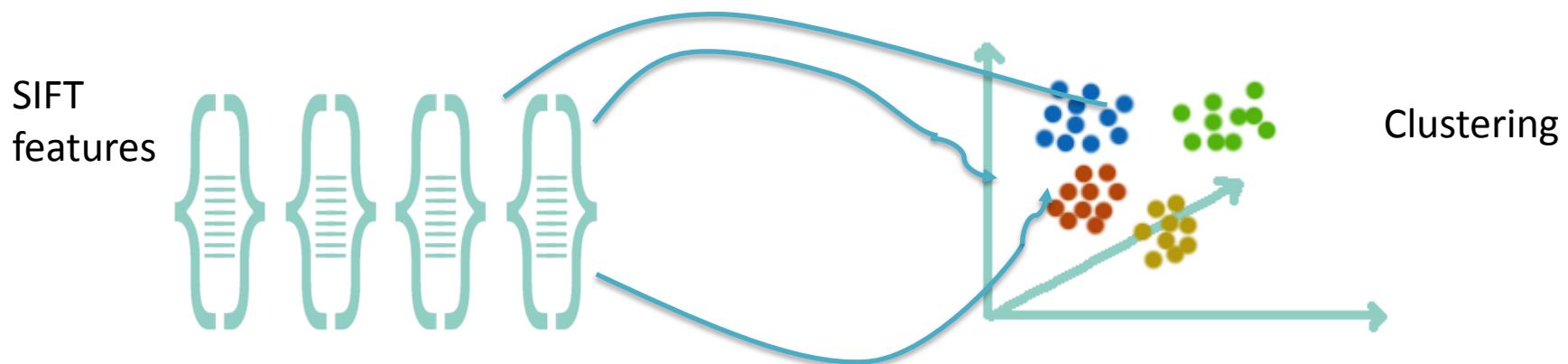
1. Extract features



2. Learn “visual vocabulary”

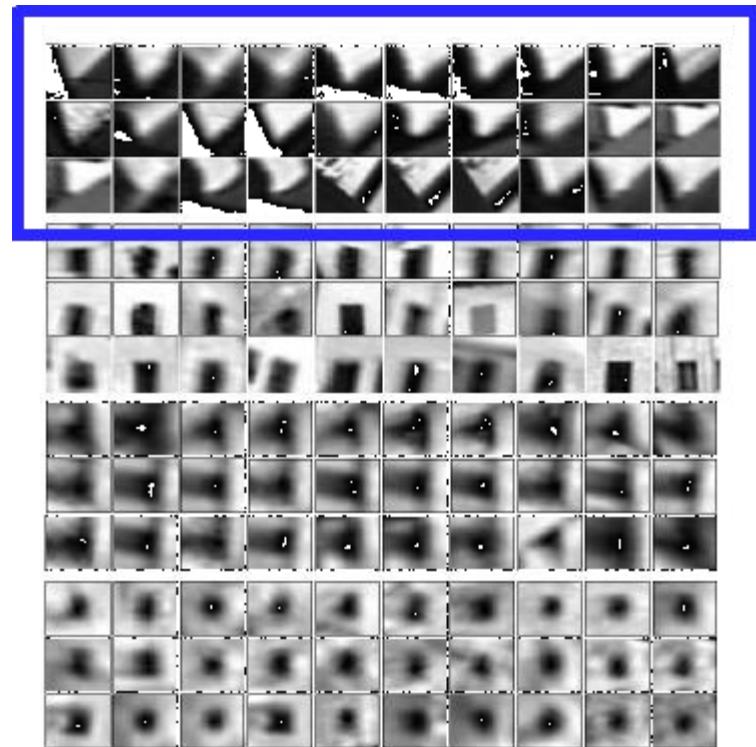
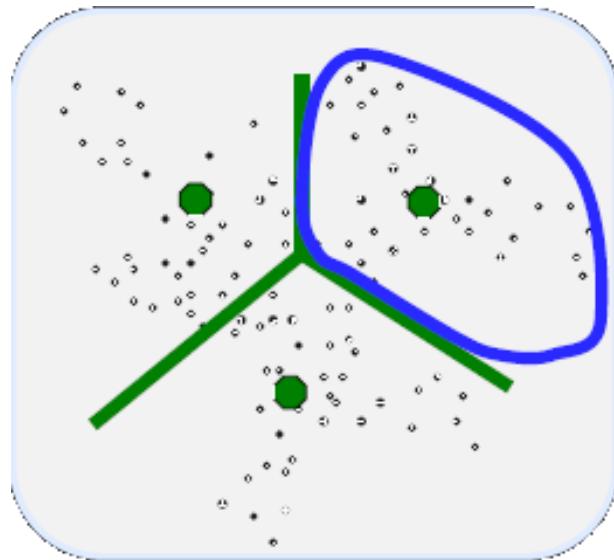


3. Quantize features using visual vocabulary



Visual Words

Example: each group of patches belongs to the same visual word



visual words are the centers of the clusters , which is the average of all patches in this cluster

From Clustering to Vector Quantization

- Clustering is a common method for learning a visual vocabulary or codebook
 - Unsupervised learning process
 - Each cluster center produced by k-means becomes a codevector
 - Provided the training set is sufficiently representative, the codebook will be “universal”
- The codebook is used for quantizing features
 - A *vector quantizer* takes a feature vector and maps it to the index of the nearest codevector in a codebook
 - Codebook = visual vocabulary
 - Codevector = visual word

Object

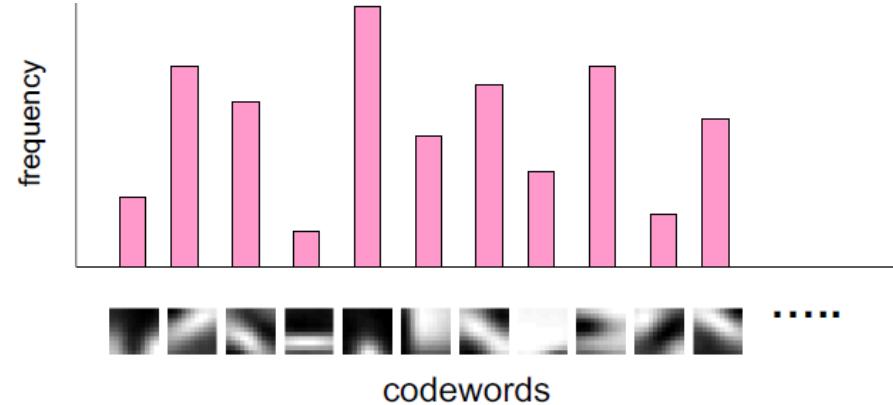
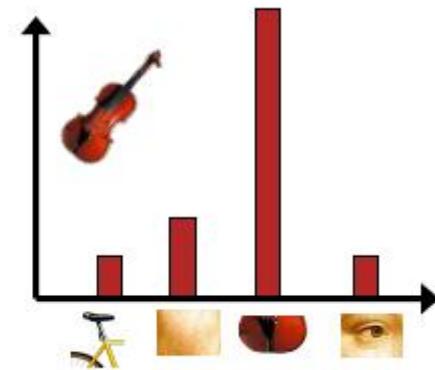
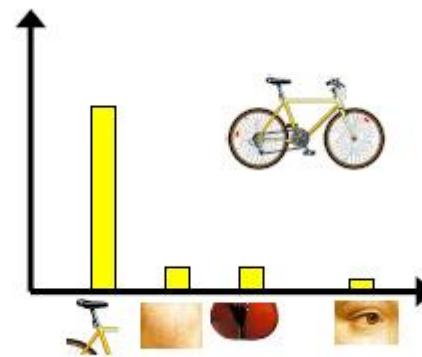
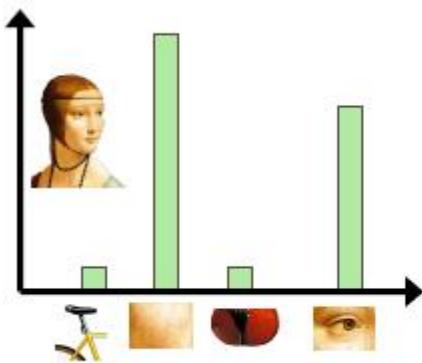
Bag of ‘words’



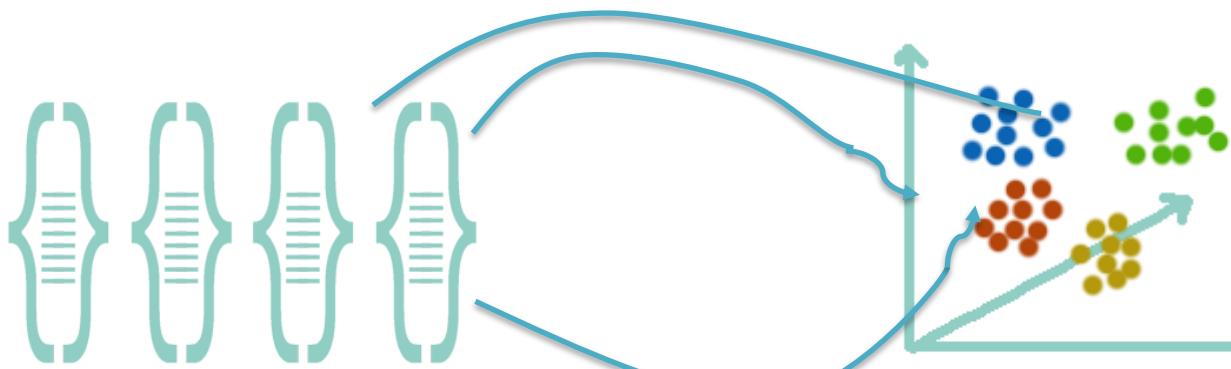
- How many of these code words show up in a particular picture?
- What is the distribution of these visual words in each image?
- We need to build a histogram for each image, so when we get a new image, we can find the images that has the same or similar histogram distribution of visual words.

Bag of features: outline

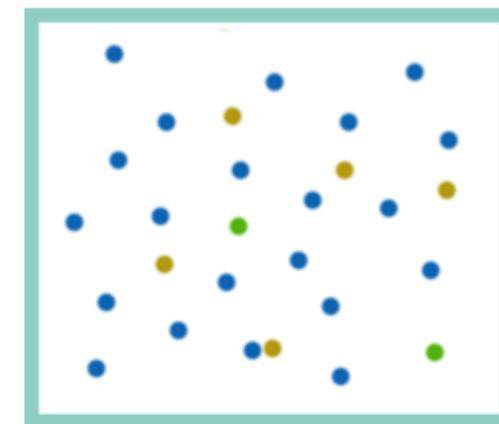
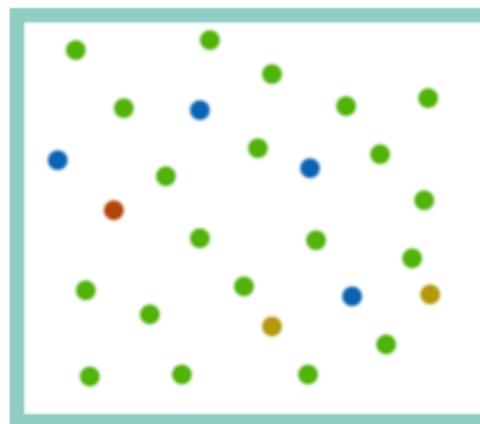
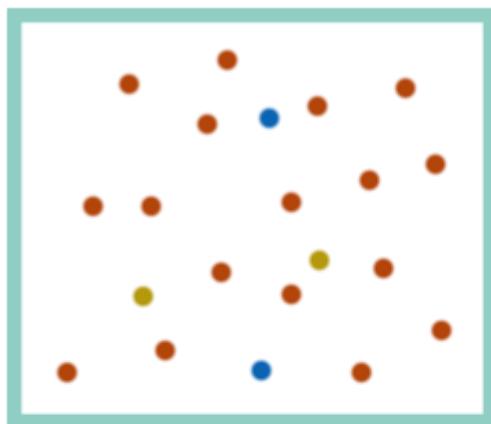
4. Represent images by frequencies of “visual words”
(Summarize entire image based on its distribution/histogram of word occurrences)



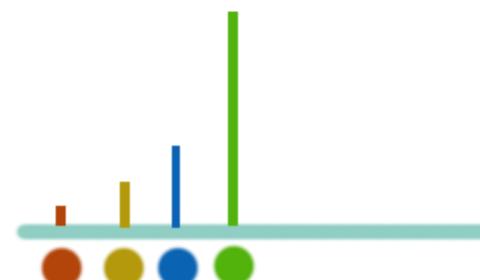
SIFT
features



Clustering



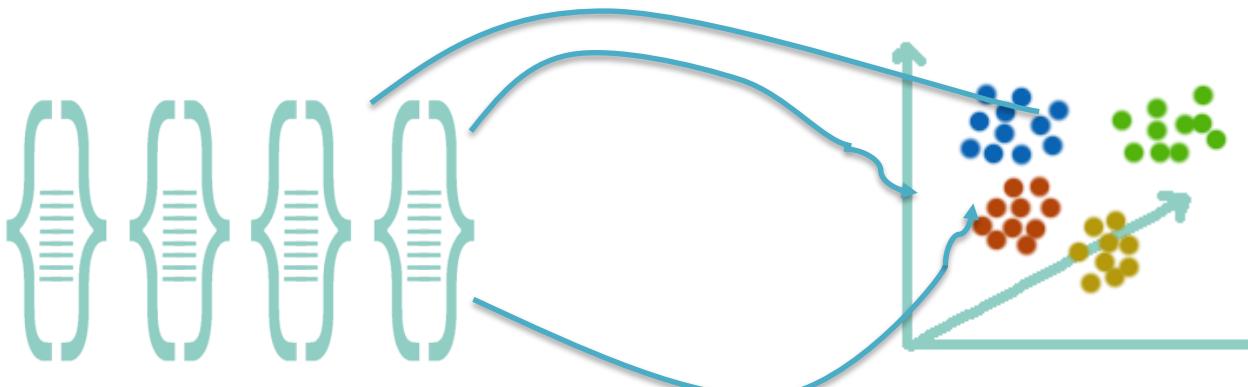
...



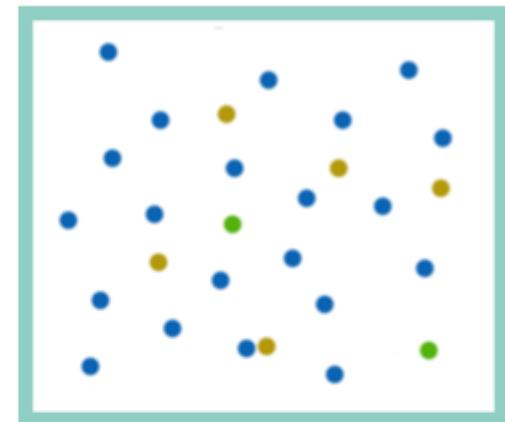
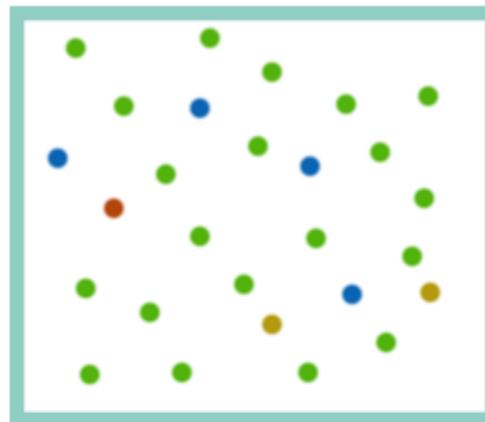
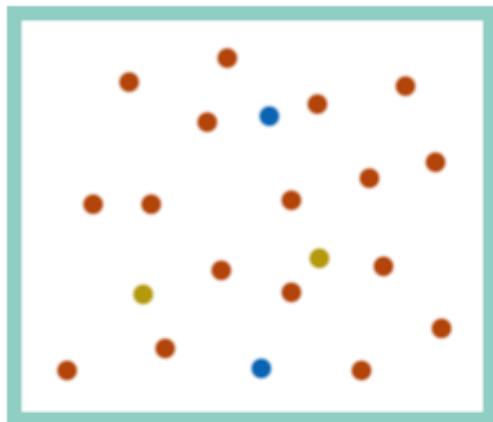
...

5. convert each histogram into a normalized feature vector

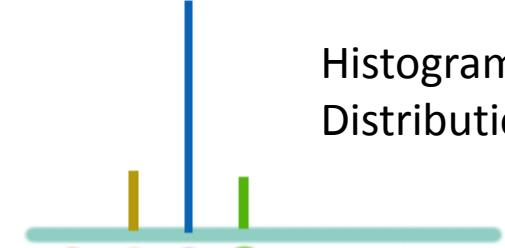
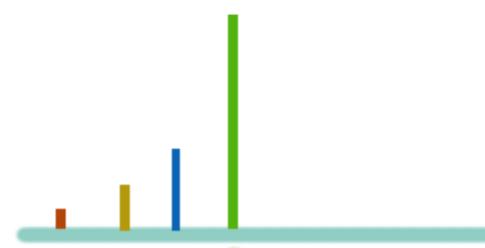
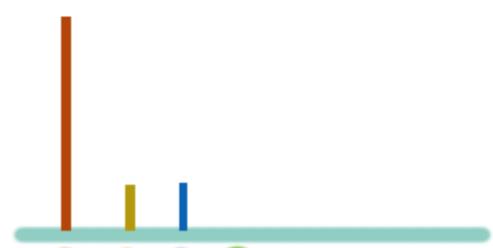
SIFT
features



Clustering



...



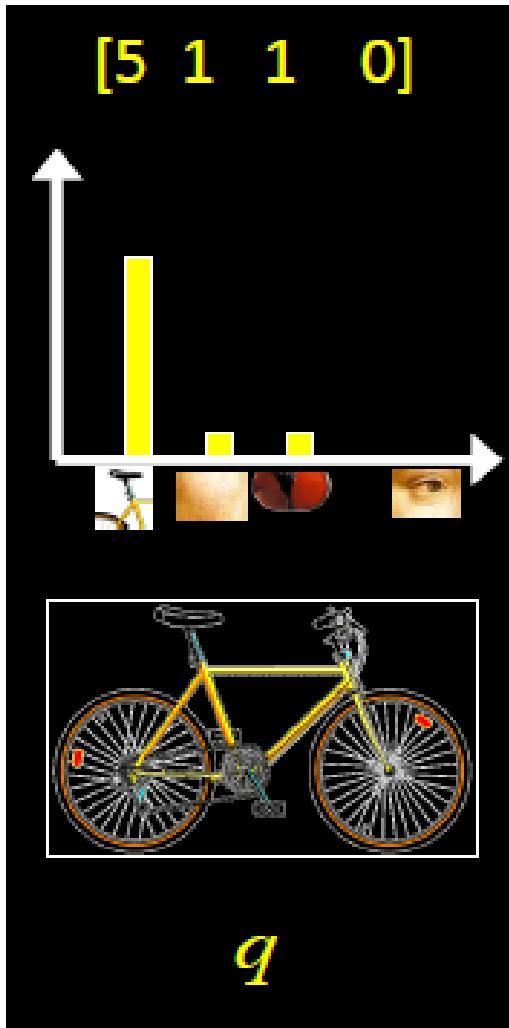
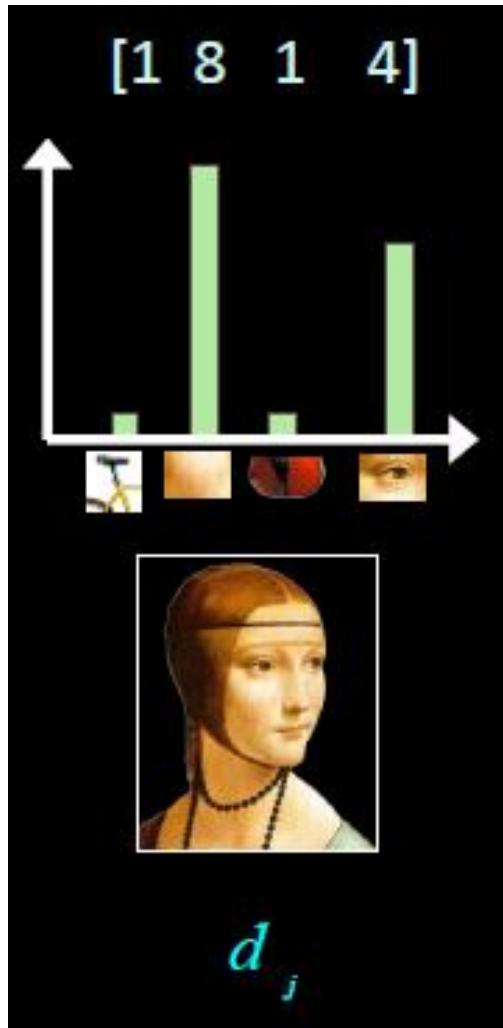
...



Feature
Vectors

...

Image Classification using BOW using k-NN or SVM



$$sim(d_j, q) = \frac{\langle d_j, q \rangle}{\|d_j\| \|q\|}$$

$$= \frac{\sum_{i=1}^V d_j(i) * q(i)}{\sqrt{\sum_{i=1}^V d_j(i)^2} * \sqrt{\sum_{i=1}^V q(i)^2}}$$

for vocabulary of V words

Object Classification with Bag of Words

- Performance on Caltech 101 dataset with linear SVM on bag-of-word vectors:



<i>True classes →</i>	<i>faces (frontal)</i>	<i>airplanes (side)</i>	<i>cars (rear)</i>	<i>cars (side)</i>	<i>motorbikes (side)</i>
<i>faces(frontal)</i>	94	0.4	0.7	0	1.4
<i>airplanes (side)</i>	1.5	96.3	0.2	0.1	2.7
<i>cars (rear)</i>	1.9	0.5	97.7	0	0.9
<i>cars(side)</i>	1.7	1.9	0.5	99.6	2.3
<i>motorbikes (side)</i>	0.9	0.9	0.9	0.3	92.7

[Csurka et al., '04]

Example bag-of-words matches



Example bag-of-words matches



Advantages

- We can create our visual vocabulary from a different dataset than the dataset that we are interested in classifying/clustering, so if first dataset is representative of the second, this algorithm will be successful (visual vocabulary can be universal).

Issues:

- How to choose vocabulary size (hyper-parameter)?
 - Too small: visual words not representative of all patches
 - Too large: quantization artifacts, overfitting

Credit for

CS 4495 Computer Vision (Spring 2015)

A. Bob - College of Computing, Georgia Tech.

*CSE 455 Computer Vision (Winter 2017) by Linda Shapiro -
University of Washington.*

*CS131 “Computer Vision: Foundations and Applications” by
University of Stanford (Fall 2019)*