

Lecture 7

Introduction to Recognition

COS 429: Computer Vision

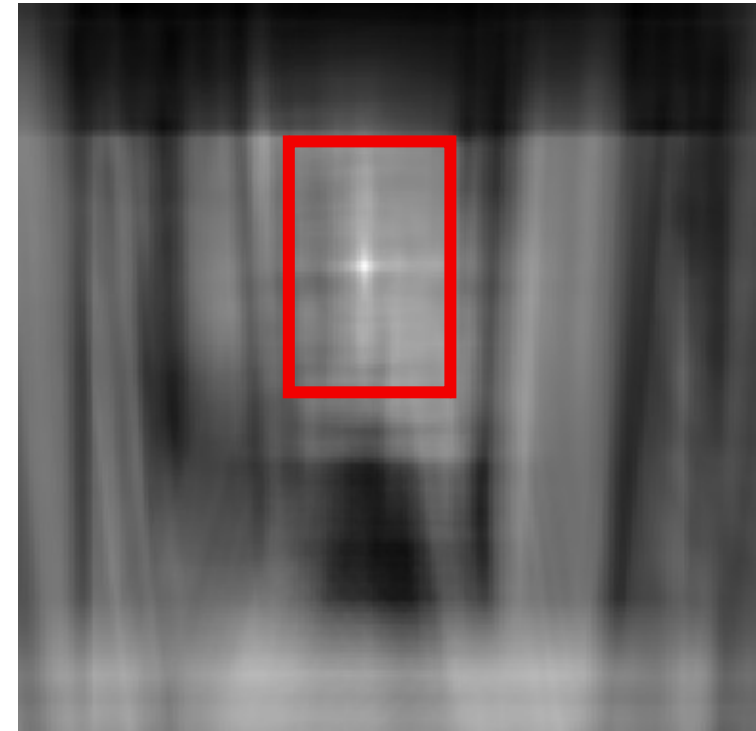


Object recognition: let's try something simple

Find the chair in this image



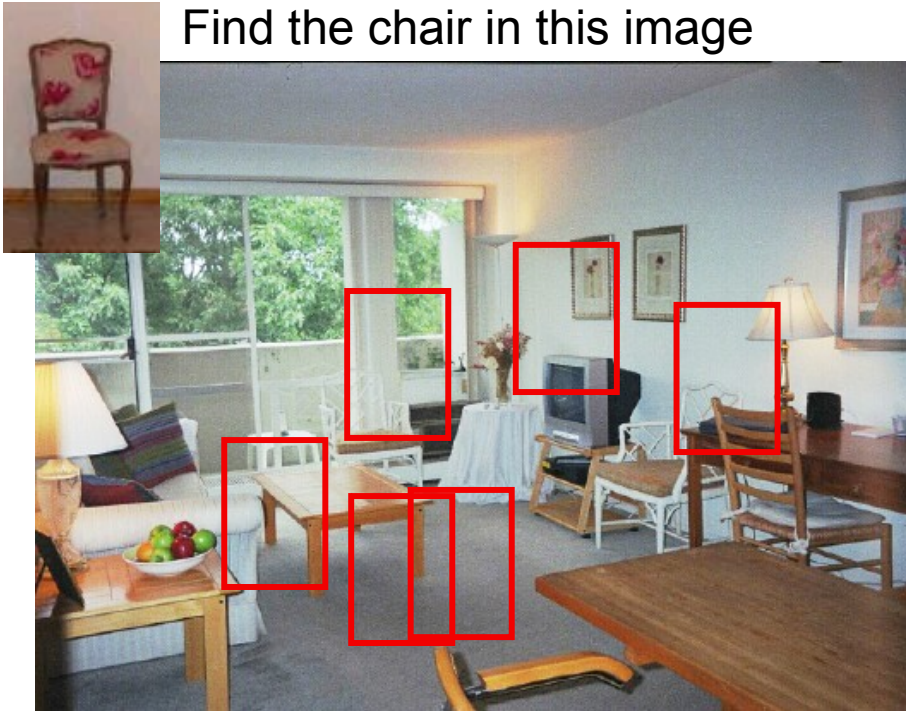
Output of normalized correlation



This is a chair

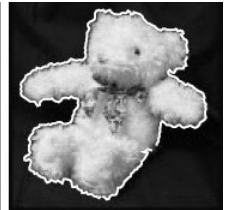


Object recognition: let's try something simple



Simple template matching
is not going to be enough

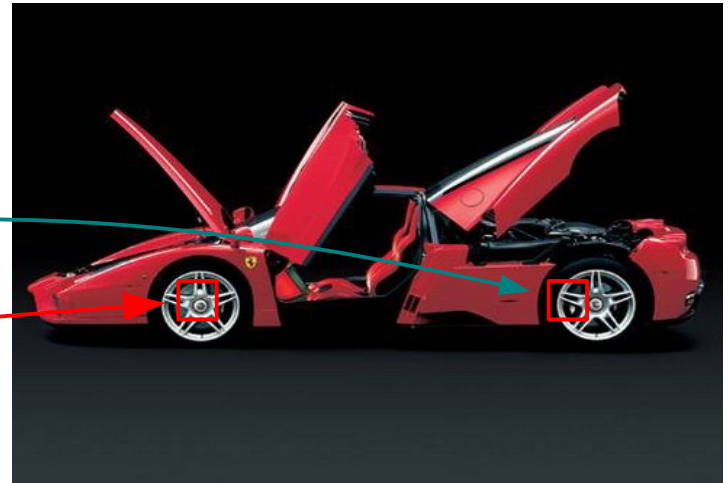
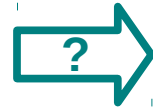
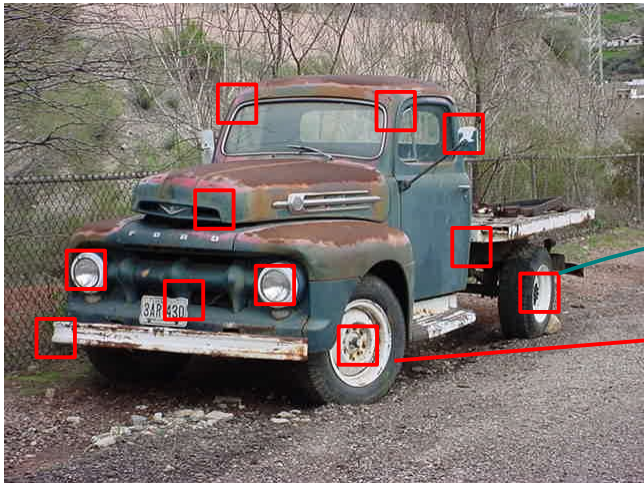
How about SIFT based alignment?



D. Lowe (1999, 2004)

SIFT Matching with RANSAC

- Good at matching same **instance** of:
 - General textured objects from similar views
 - Flat textured objects from fairly different views (using affine or homography)



- But it is not good at matching between:
 - Non-flat objects from different views
 - Distinct instances from the same category

=> Would need template for each instance from each view!

Challenges 1: view point variation



Michelangelo 1475-1564

Slides: course object recognition
ICCV 2005

Challenges 2: illumination



Challenges 3: occlusion



Magritte, 1957

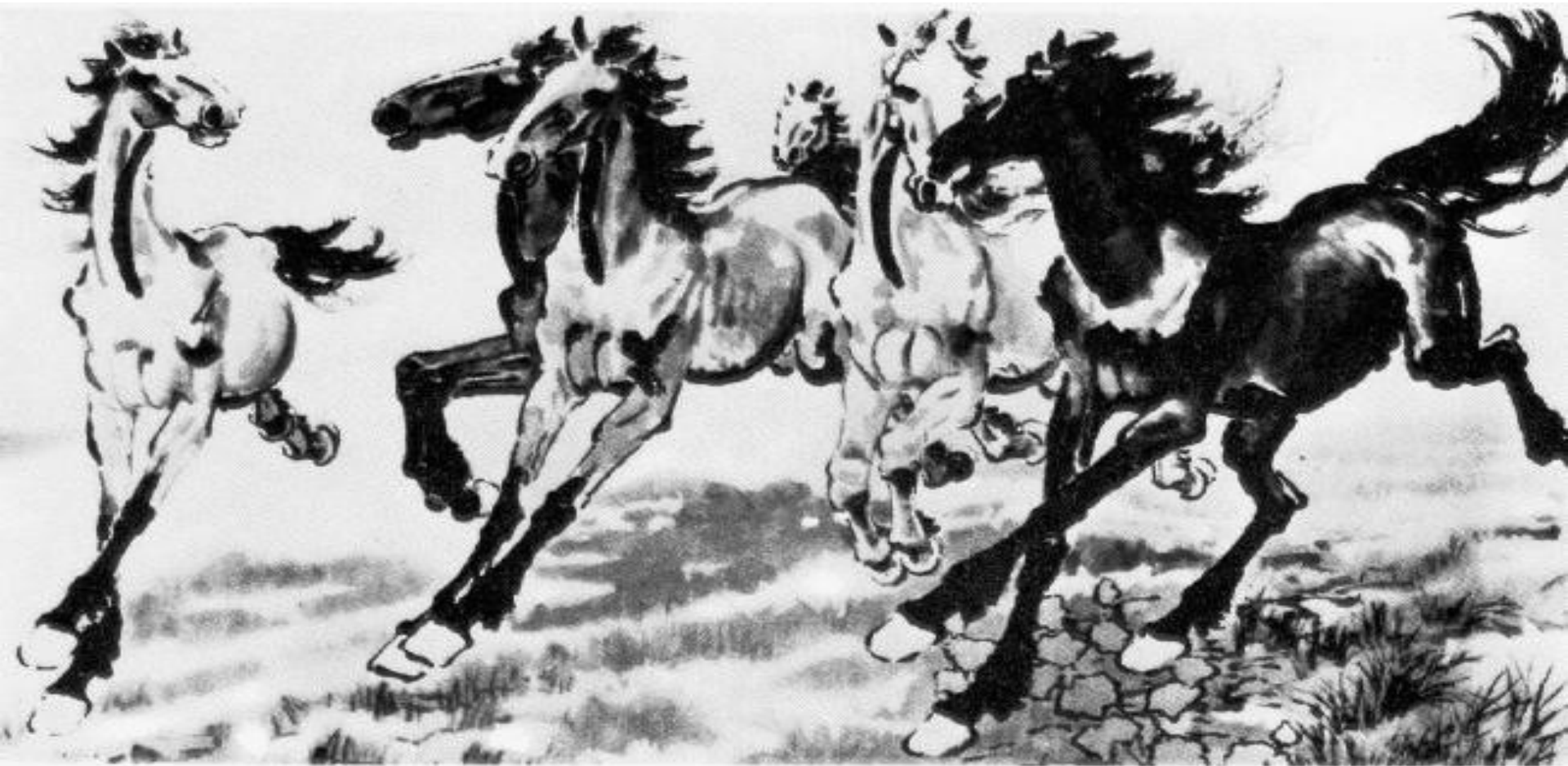
Slides: course object recognition
ICCV 2005

Challenges 4: scale



Slides: course object recognition
ICCV 2005

Challenges 5: deformation



Challenges 6: background clutter

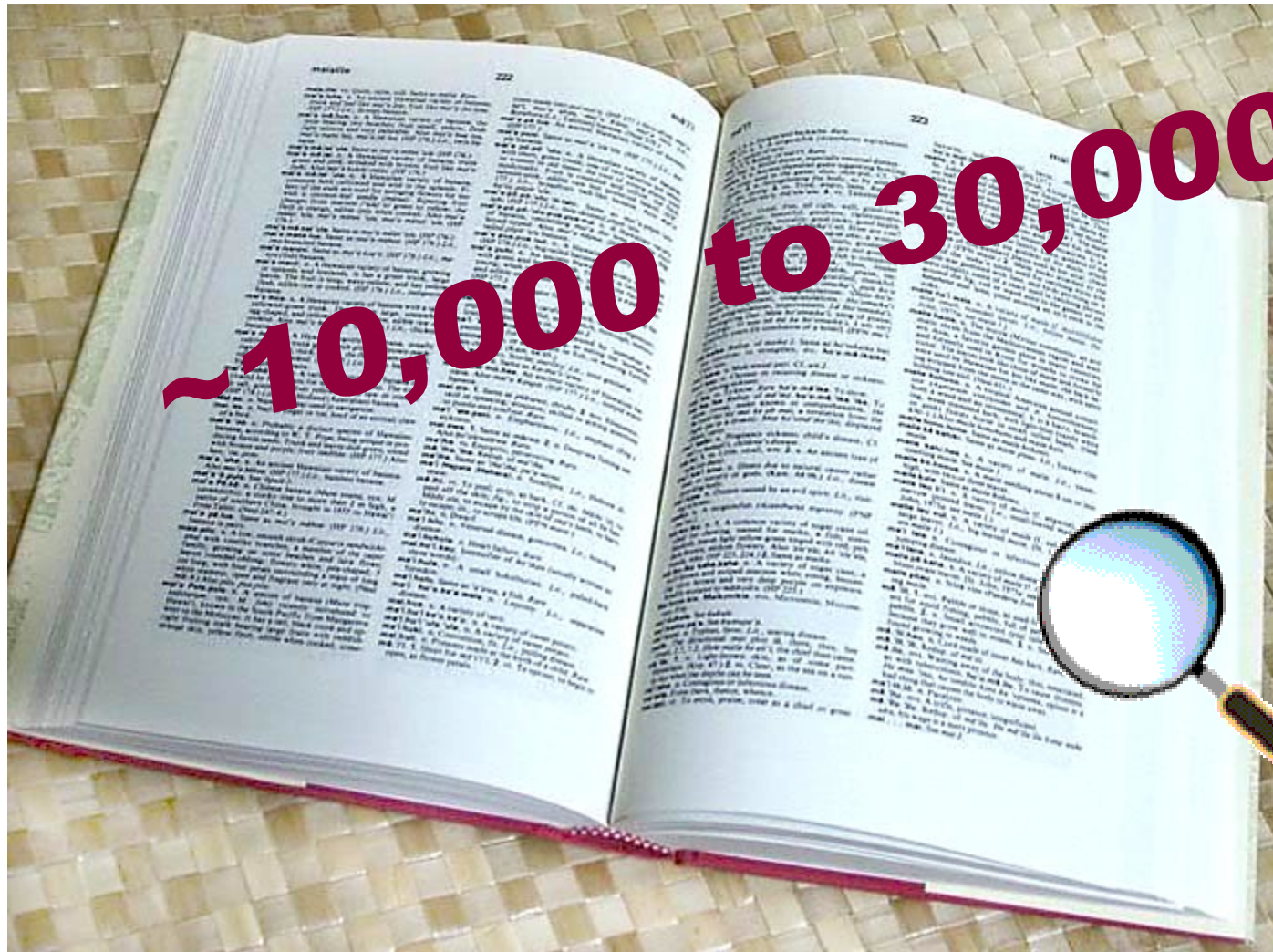


Brady, M. J., & Kersten, D. (2003). Bootstrapped learning of novel objects. *J Vis*, 3(6), 413-422

Challenges 7: intra-class variation



How many visual object categories are there?

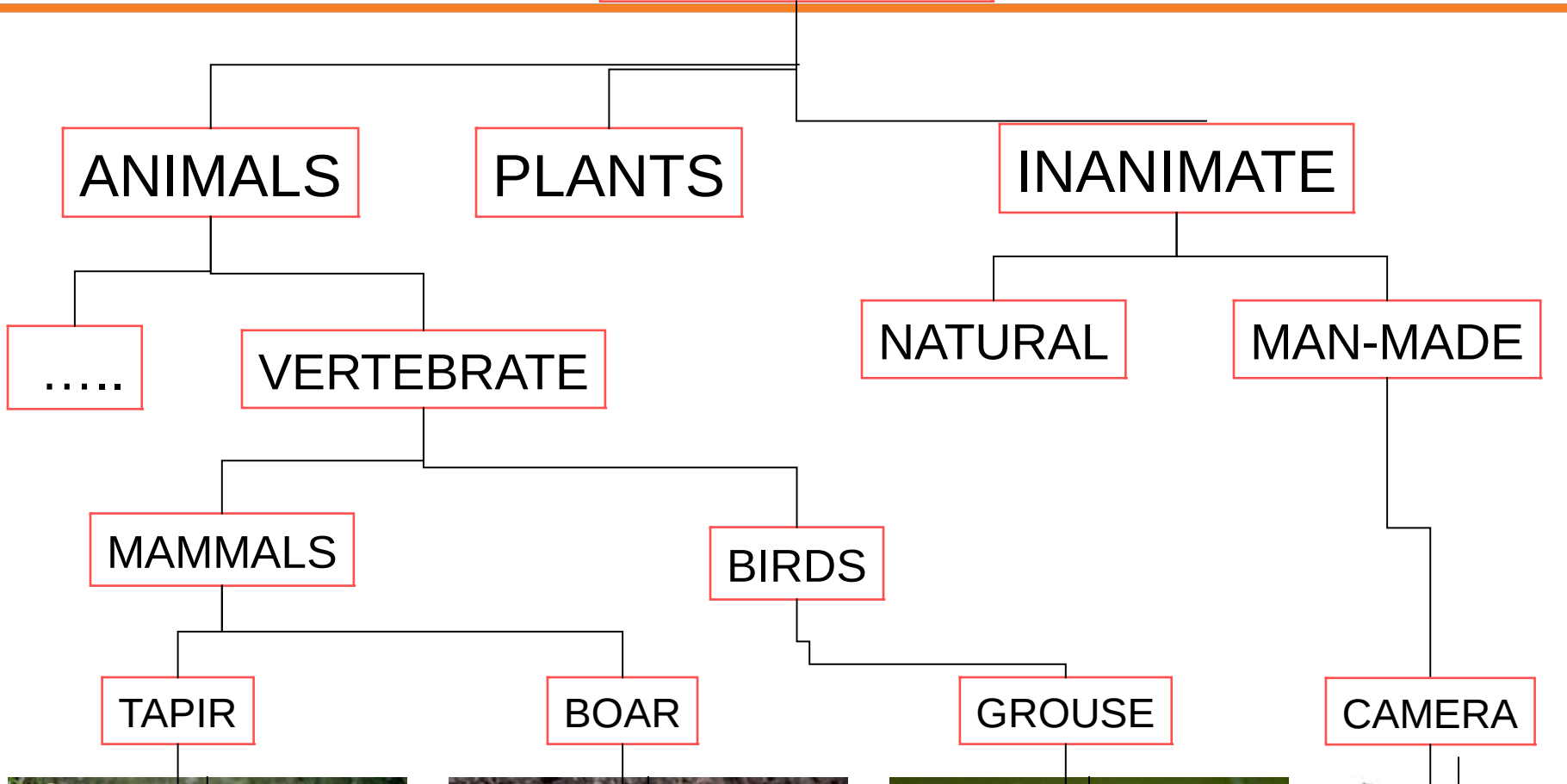


Biederman 1987



~10,000 to 30,000

OBJECTS



What do we want to recognize in an image?



Slide from: Svetlana Lazebnik

Scene categorization or classification

- outdoor
- city/forest/factory/etc.



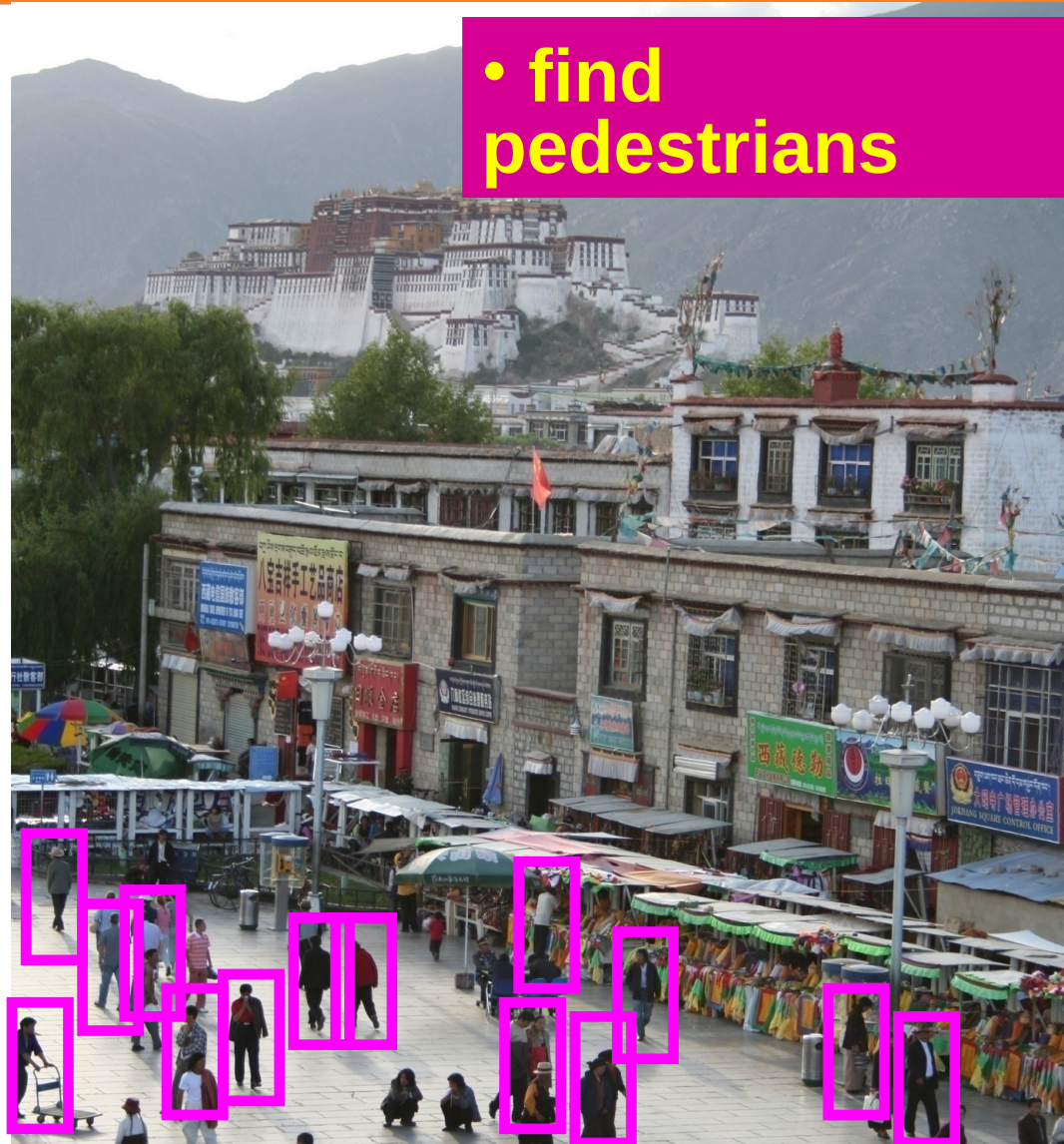
Image annotation / tagging / attributes



- street
- people
- building
- mountain
- tourism
- cloudy
- brick
- ...

Slide from: Svetlana Lazebnik

Object detection



Slide from: Svetlana Lazebnik

Image parsing / semantic segmentation



Slide from: Svetlana Lazebnik

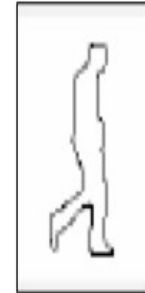
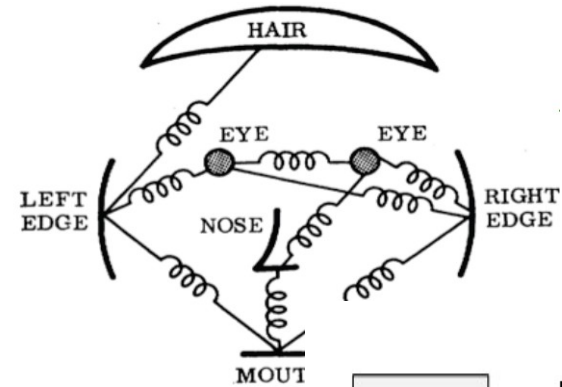
Scene understanding?



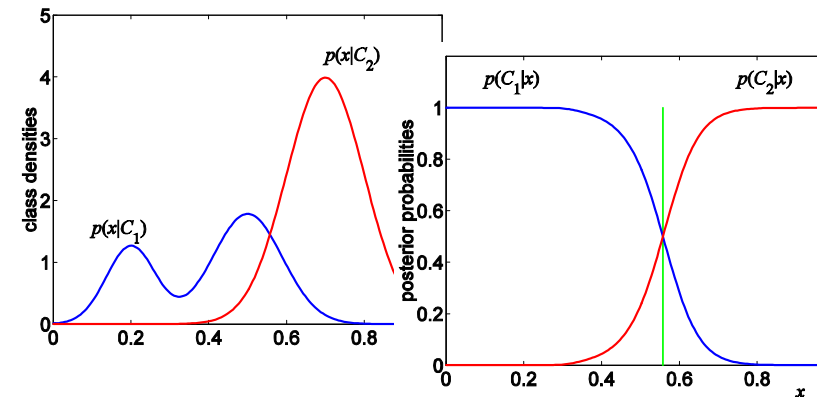
Slide from: Svetlana Lazebnik

Typical Components

- **Hypothesis** generation
 - Sliding window, Segmentation, feature point detection, random, search
- **Encoding** of (local) image data
 - Colors, Edges, Corners, Histogram of Oriented Gradients, Wavelets, Convolution Filters
- **Relationship** of different parts to each other
 - Blur or histogram, Tree/Star, Pairwise/Covariance
- **Learning** from labeled examples
 - Selecting representative examples (templates), Clustering, Building a cascade
 - Classifiers: Bayes, Logistic regression, SVM, AdaBoost, ...
 - Generative vs. Discriminative
- **Verification** - removing redundant, overlapping, incompatible examples
 - Non-Max Suppression, context priors, geometry



Exemplar Summary



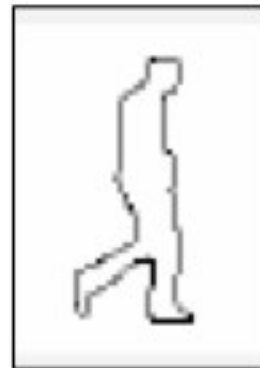
Example 1: Chamfer matching (Pedestrian Detection)



Input Image



Edge Detection



Template



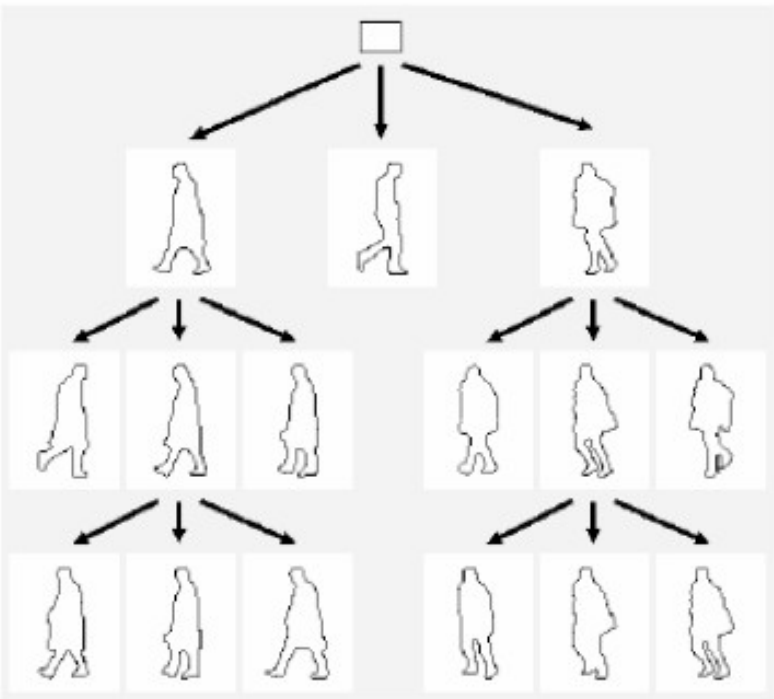
Best Match

$$D_{chamfer}(T, I) \equiv \frac{1}{|T|} \sum_{t \in T} d_I(t)$$

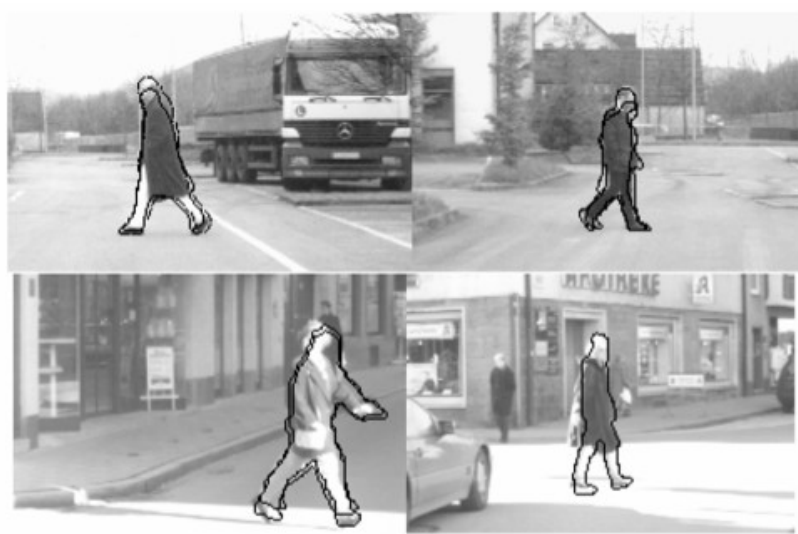


Distance Transform

Example 1: Chamfer matching (Pedestrian Detection)



Hierarchy of templates



Example 2: Viola/Jones (Face Detection)

Robust Realtime Face Detection, IJCV 2004, Viola and Jones

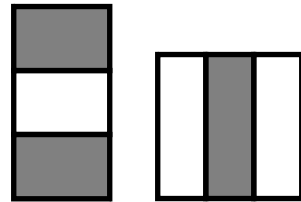
Features: “Haar-like Rectangle filters”

- Differences between sums of pixels in adjacent rectangles

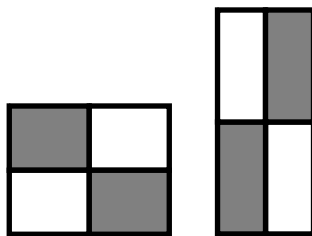
-1 +1



2-rectangle features



3-rectangle features



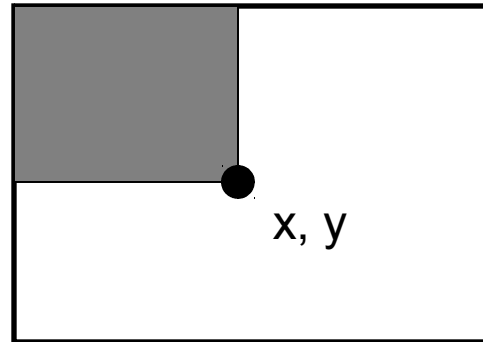
4-rectangles features



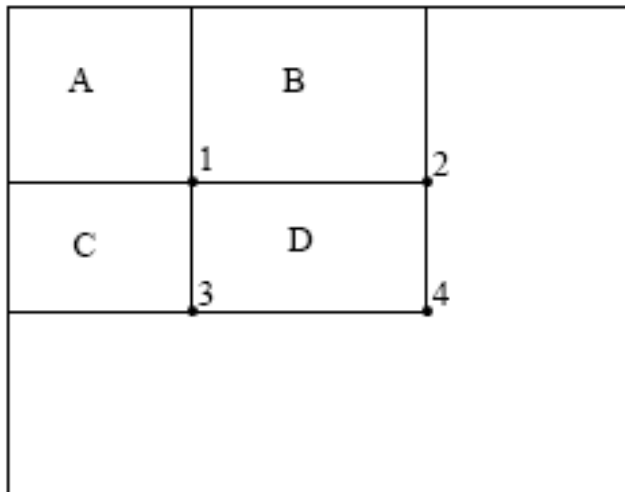
$60,000 \times 100 = 6,000,000$
Unique Features

Example 2: Viola/Jones - Integral Images

- $ii = \text{cumsum}(\text{cumsum}(im, 1), 2)$



$ii(x,y)$ = Sum of the values in the grey region



How to compute $B-A$?

How to compute $A+D-B-C$?

Slide from: Derek Hoiem

Example 2: Feature selection with Adaboost

1. Create a large pool of features

2. Select features that are discriminative and work well together:

- “Weak learner” = feature + threshold + parity
- Choose weak learner that minimizes error on the weighted training set
- Reweight

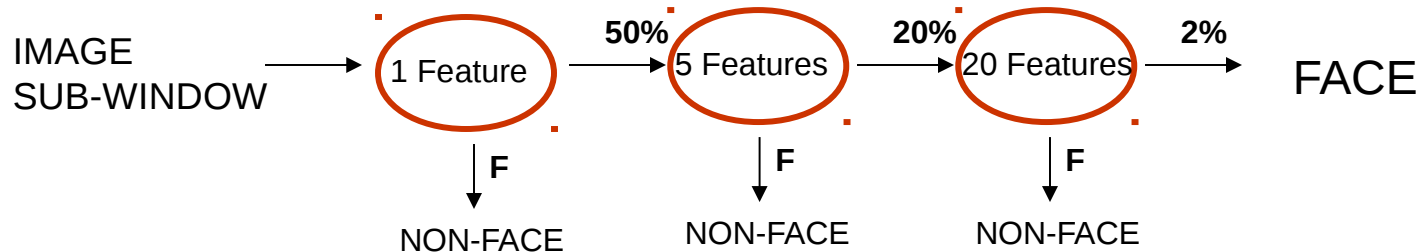
Trained Classifier
(for each stage of cascade)

$$y_t(x) = \begin{cases} +1 & \text{if } h_t(x) > \theta_t \\ -1 & \text{otherwise} \end{cases}$$

$$Y(x) = \sum \alpha_t y_t(x)$$

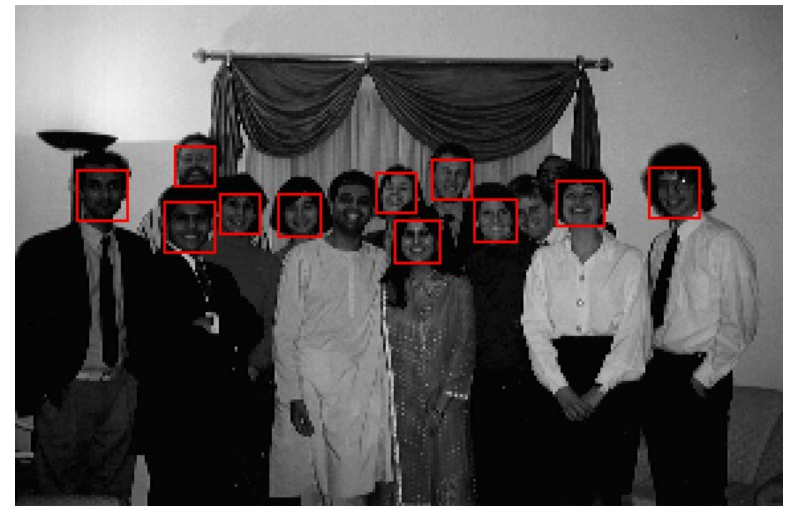
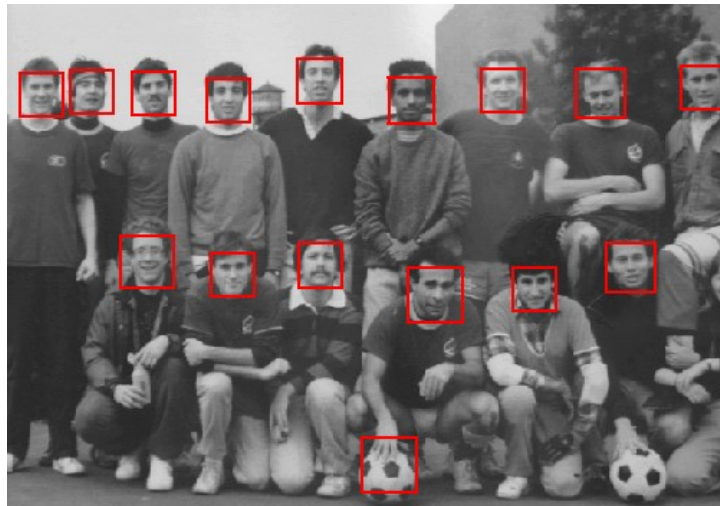
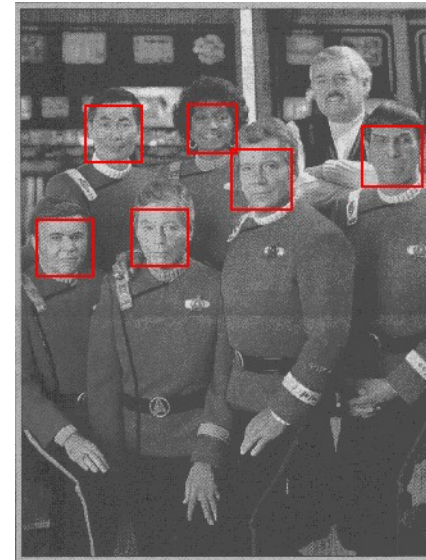
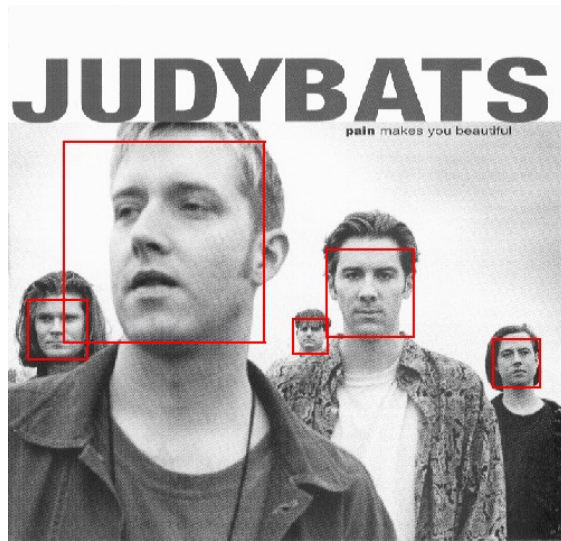
$$\text{Decision} = \begin{cases} \text{face,} & \text{if } Y(x) > 0 \\ \text{non-face,} & \text{otherwise} \end{cases}$$

Example 2: Viola/Jones Cascaded Classifier



- first classifier: 100% detection, 50% false positives.
- second classifier: 100% detection, 40% false positives
(20% cumulative)
 - using data from previous stage.
- third classifier: 100% detection, 10% false positive rate
(2% cumulative)
- Put cheaper classifiers up front

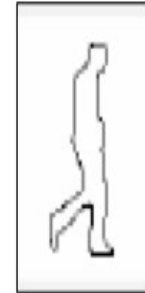
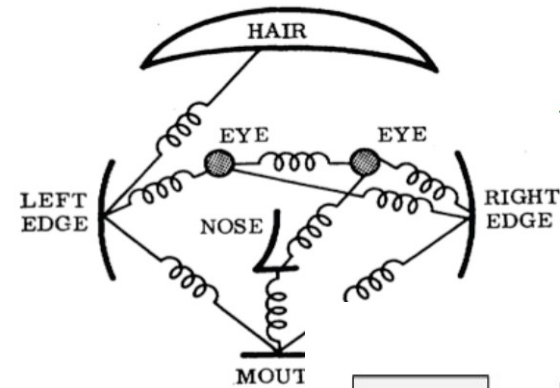
Example 2: Viola/Jones results



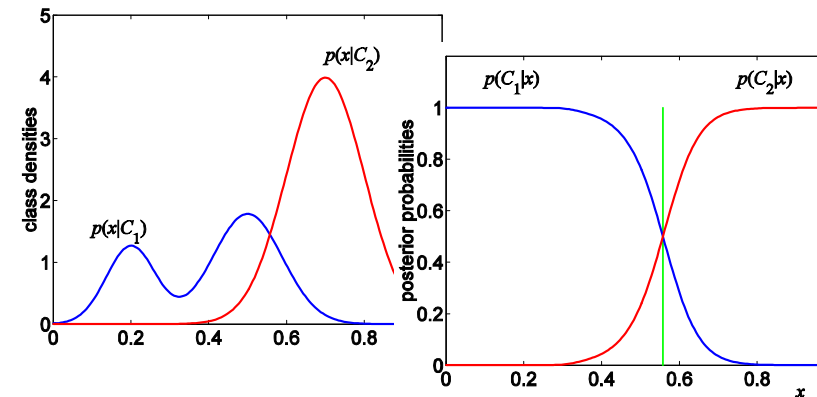
Run-time: 15fps (384x288 pixel image on a 700 Mhz Pentium III)

Typical Components

- **Hypothesis** generation
 - Whole image, Sliding window, Segmentation, Feature point detection, Search...
- **Encoding** of (local) image data
 - Colors, Edges, Corners, Histogram of Oriented Gradients, Wavelets, Convolution Filters...
- **Relationship** of different parts to each other
 - Histogram, Tree/Star, Pairwise/Covariance...
- **Learning** from labeled examples
 - Selecting representative examples (templates), Clustering, Building a cascade
 - Classifiers: Bayes, Logistic regression, SVM, AdaBoost, ...
 - Generative vs. Discriminative
- **Verification** - removing redundant, overlapping, incompatible examples
 - Non-Max Suppression, context priors, geometry



Exemplar Summary



(No Geometry) Example: Color Histograms



Swain and Ballard, [Color Indexing](#), IJCV 1991.

(No Geometry) Example: Bad of Words

Object



**Bag of
'words'**



Slide from: Svetlana Lazebnik

Objects as texture

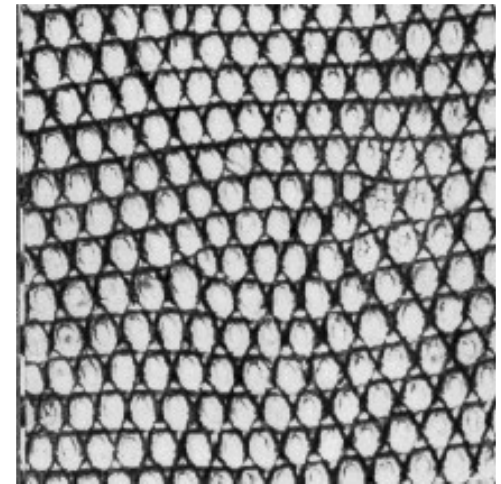
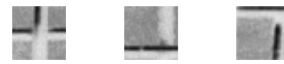
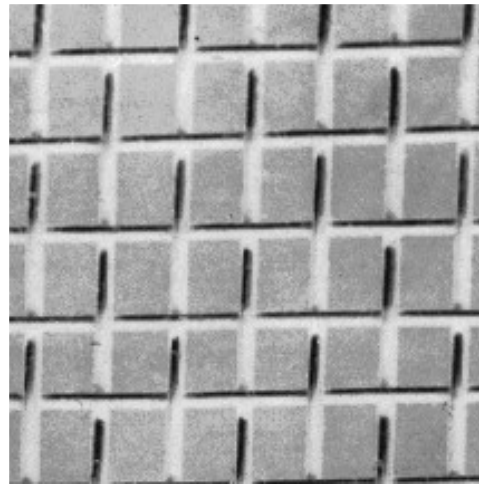
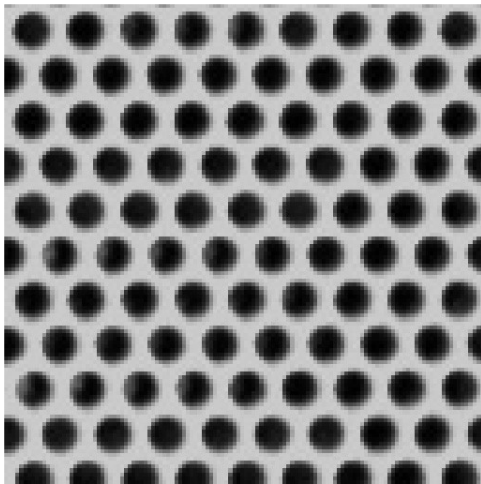
- All of these are treated as being the same



- No distinction between foreground and background: scene recognition?

Origin 1: Texture recognition

- Texture is characterized by the repetition of basic elements or *textons*
- For stochastic textures, it is the identity of the textons, not their spatial arrangement, that



Julesz, 1981; Cula & Dana, 2001; Leung & Malik 2001; Mori, Belongie & Malik, 2001; Schmid 2001; Varma & Zisserman, 2002, 2003; Lazebnik, Schmid & Ponce, 2003

Origin 2: Bag-of-words models

2007-01-23: State of the Union Address

George W. Bush (2001-)

abandon accountable **affordable** afghanistan africa aided ally anbar armed army **baghdad** bless **challenges** chamber chaos
choices civilians coalition commanders **commitment** confident confront congressman constitution corps debates deduction

deficit deliver
expand extr

1962-10-22: Soviet Missiles in Cuba

John F. Kennedy (1961-63)

insurgents ira
palestinian pay

abandon achieving adversaries aggression agricultural appropriate armaments **arms** assessments atlantic ballistic berlin
buildup burdens cargo college commitment communist constitution consumers cooperation crisis **cuba** dangers

september sh
violence vio

declined **defensive** economic situation

elimination emerge
halt hazards **hem**

modernization neglig

recession rejection r

surveillance tax te

1941-12-08: Request for a Declaration of War

Franklin D. Roosevelt (1933-45)

abandoning acknowledge aggression aggressors airplanes armaments **armed army** assault assembly authorizations bombing
britain british cheerfully claiming constitution curtail december defeats defending delays democratic dictators disclose

economic empire endanger **facts** false forgotten fortunes france **freedom** fulfilled fullness fundamental gangsters
german germany god guam harbor hawaii **hemisphere** hint hitler hostilities immune improving indies innumerable

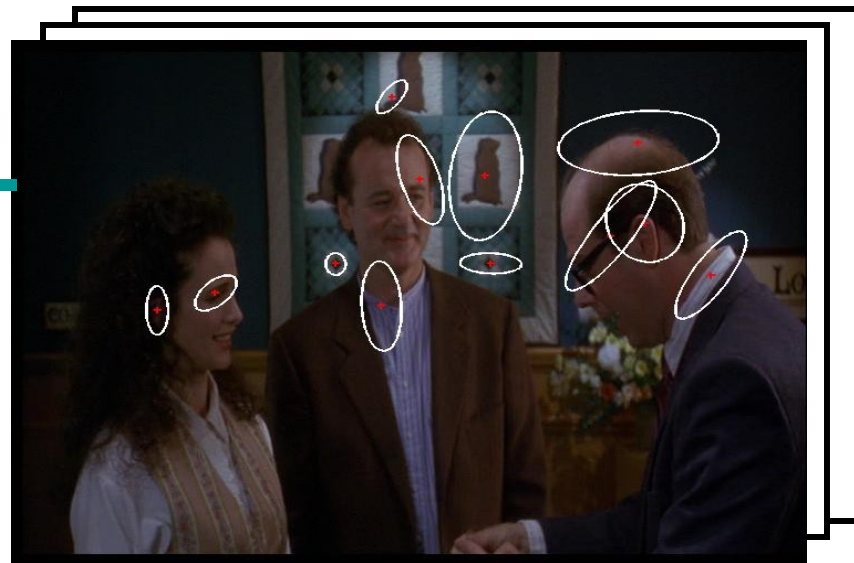
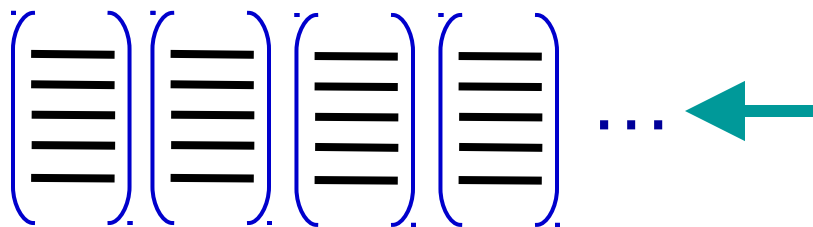
invasion **islands** isolate **japanese** labor metals midst midway **navy** nazis obligation offensive

officially **pacific** partisanship patriotism pearl peril perpetrated perpetual philippine preservation privilege reject
repaired **resisting** retain revealing rumors seas soldiers speaks speedy stamina **strength** sunday sunk supremacy tanks taxes

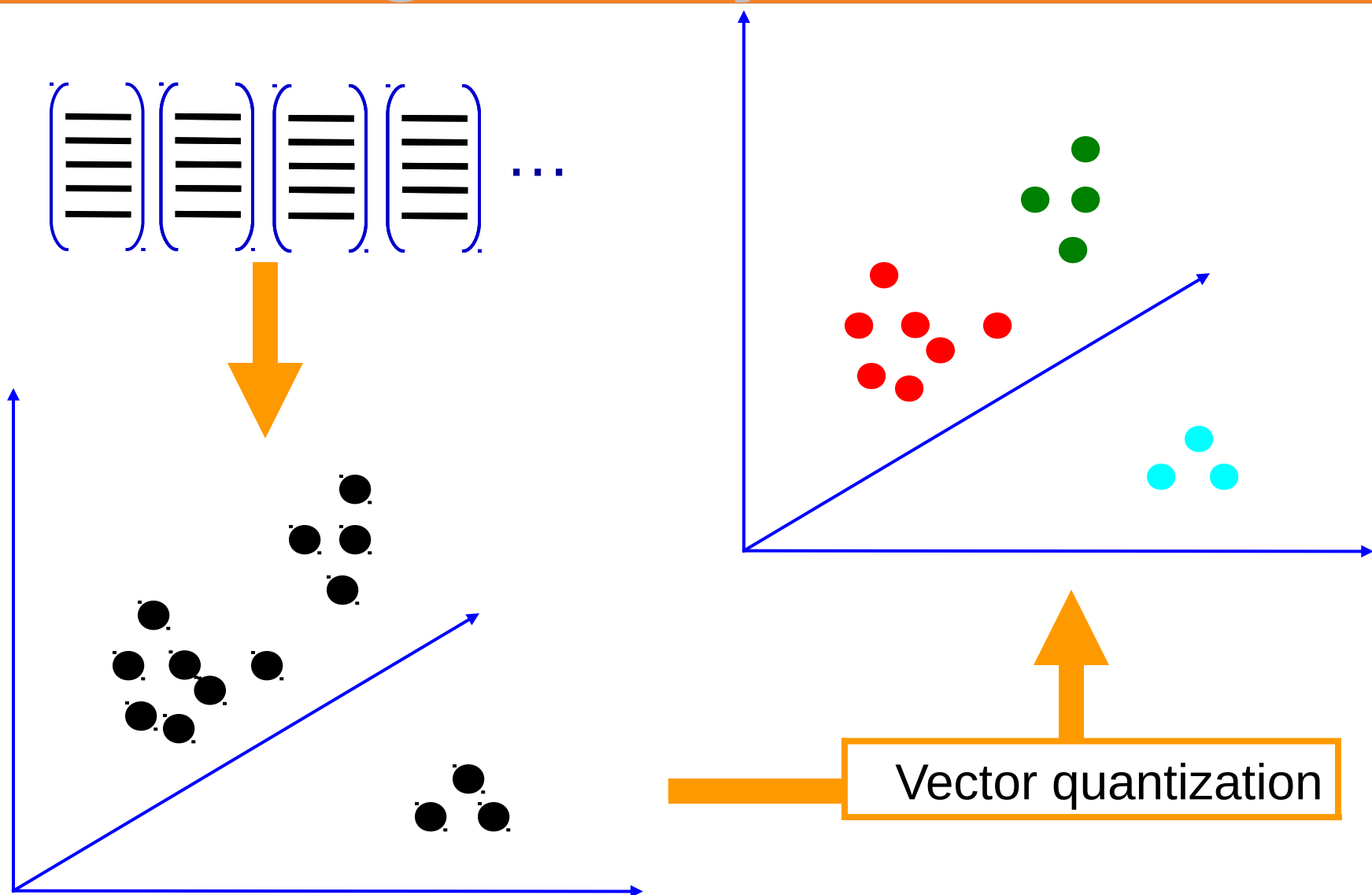
treachery true tyranny undertaken victory **war** wartime washington

- Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)

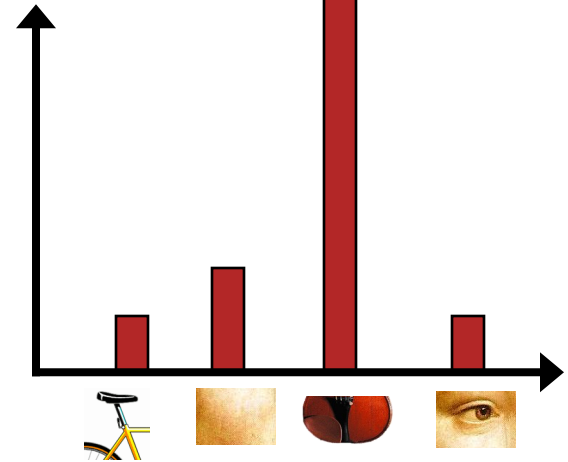
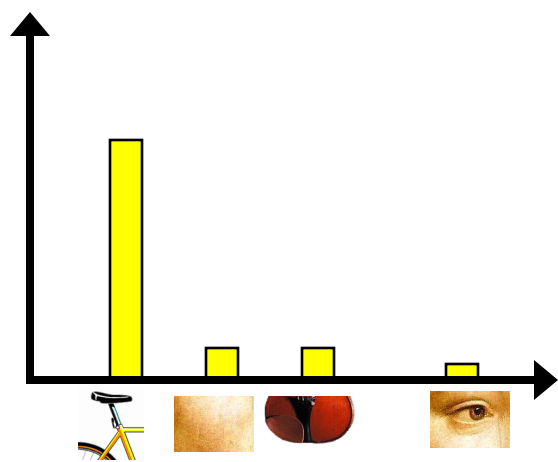
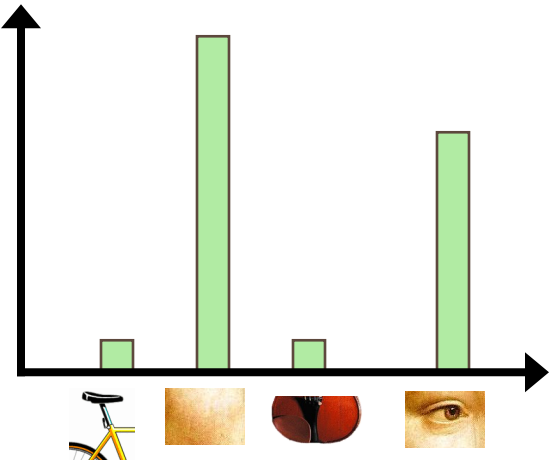
Interest Point Features



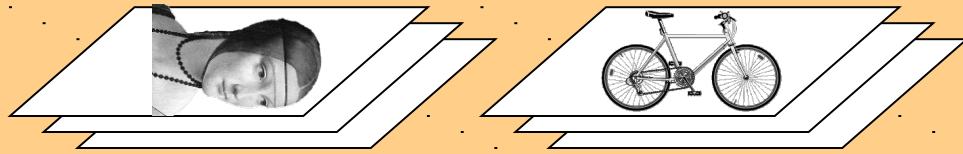
Clustering (usually k-means)



Slide credit: Josef Sivic



learning



feature detection
& representation

codewords dictionary

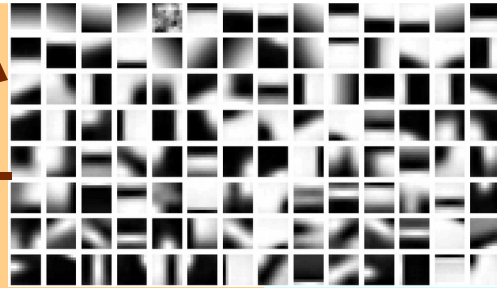


image representation



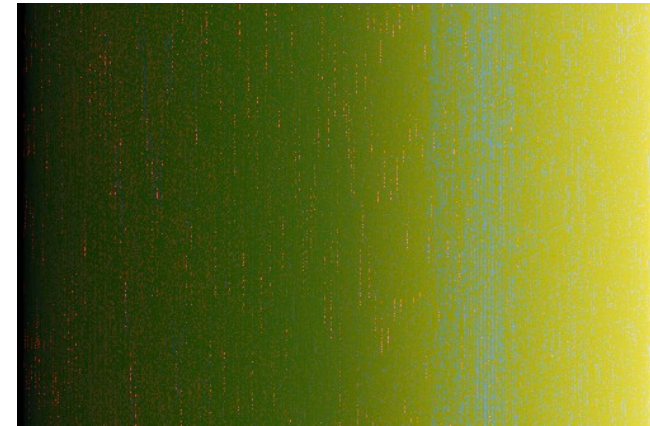
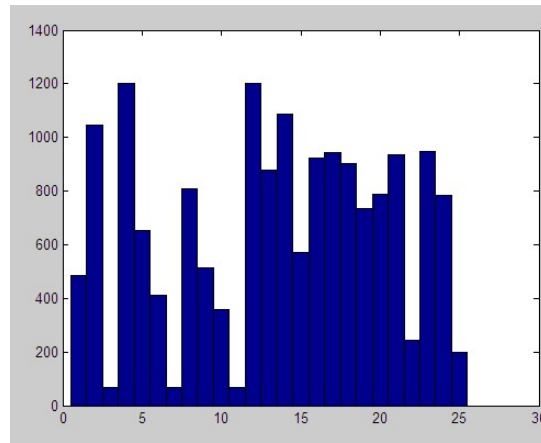
**category models
(and/or) classifiers**

recognition



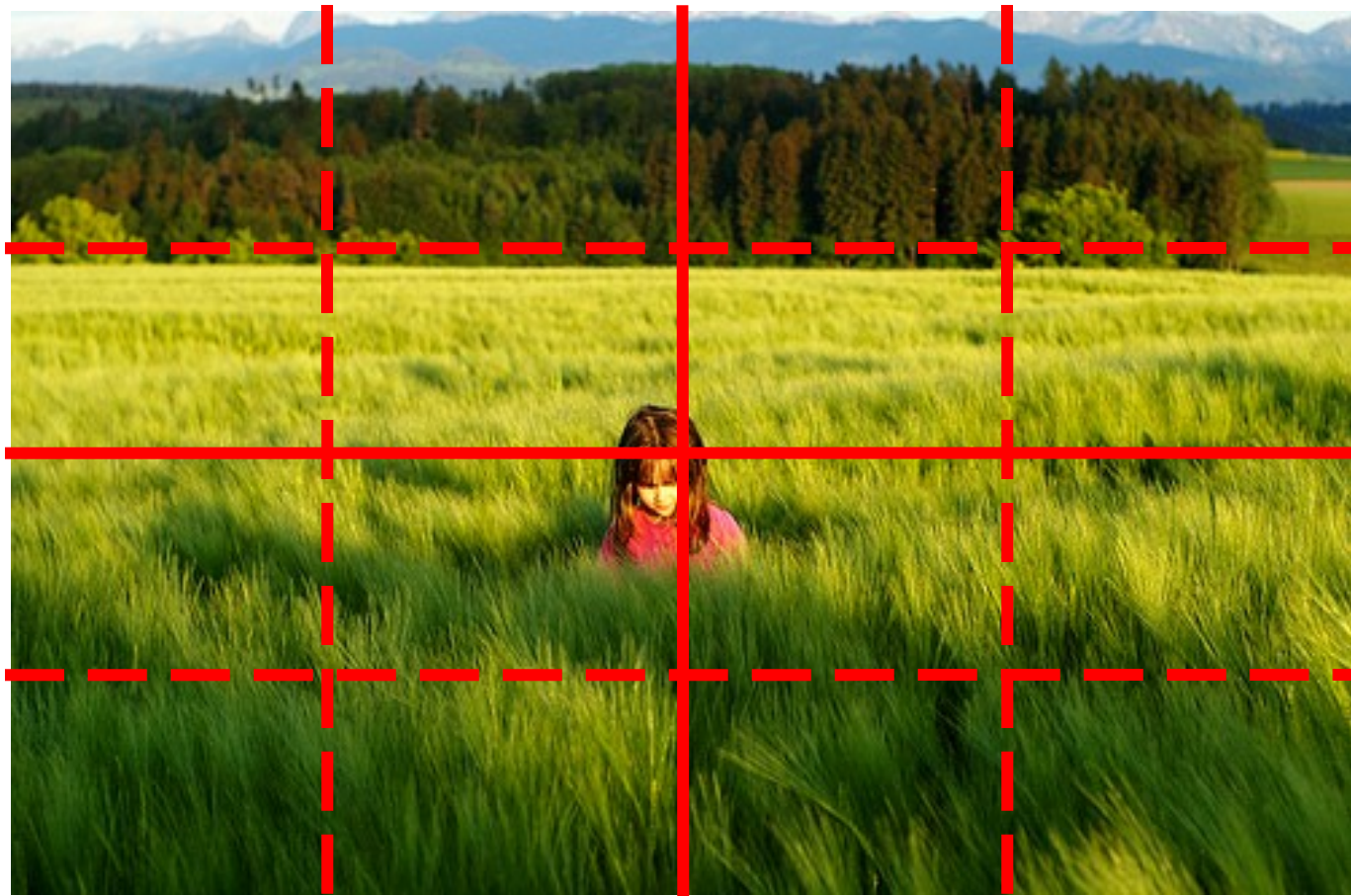
**category
decision**

The (obvious) problem with ignoring Geometry



All of these images have the same color histogram

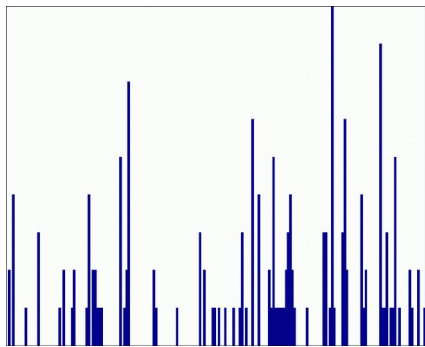
Adding Geometry back: Spatial pyramid



Compute histogram in each spatial bin

Spatial pyramid representation

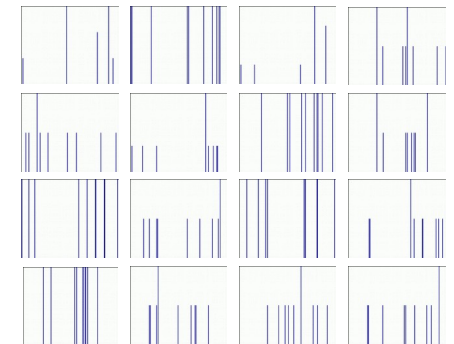
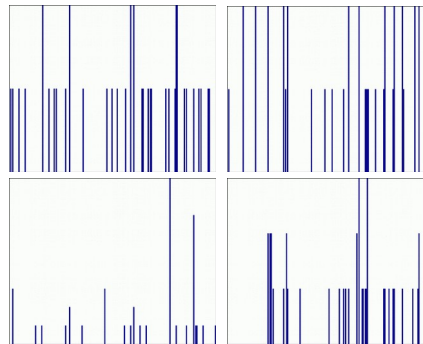
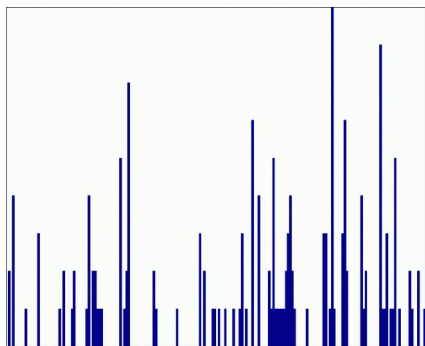
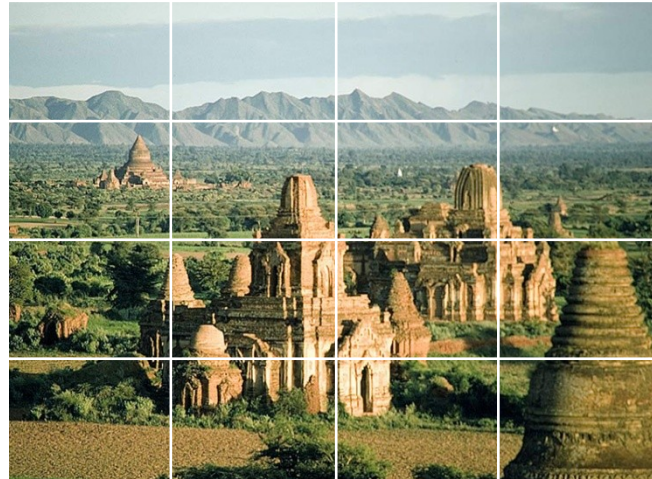
- Extension of a bag of features
- Locally orderless representation at several levels of resolution



Lazebnik, Schmid & Ponce (CVPR 2006)

Spatial pyramid representation

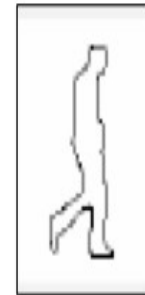
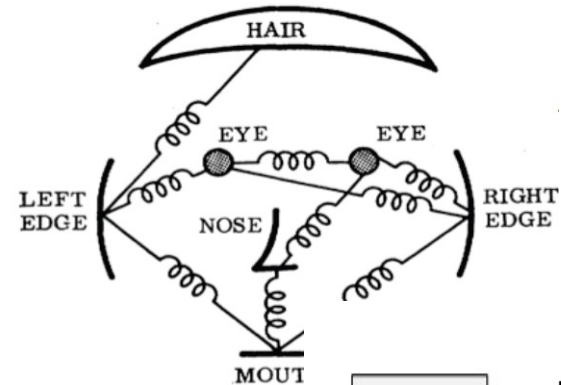
- Extension of a bag of features
- Locally orderless representation at several levels of resolution



Lazebnik, Schmid & Ponce (CVPR 2006)

More Next Time...

- **Hypothesis** generation
 - Sliding window, Segmentation, feature point detection, random, search
- **Encoding** of (local) image data
 - Colors, Edges, Corners, Histogram of Oriented Gradients, Wavelets, Convolution Filters
- **Relationship** of different parts to each other
 - Blur or histogram, Tree/Star, Pairwise/Covariance
- **Learning** from labeled examples
 - Selecting representative examples (templates), Clustering, Building a cascade
 - Classifiers: Bayes, Logistic regression, SVM, AdaBoost, ...
 - Generative vs. Discriminative
- **Verification** - removing redundant, overlapping, incompatible examples
 - Non-Max Suppression, context priors, geometry



Exemplar Summary

