

Lecture 7

Introduction to Recognition

(and image blending from last time)

COS 429: Computer Vision

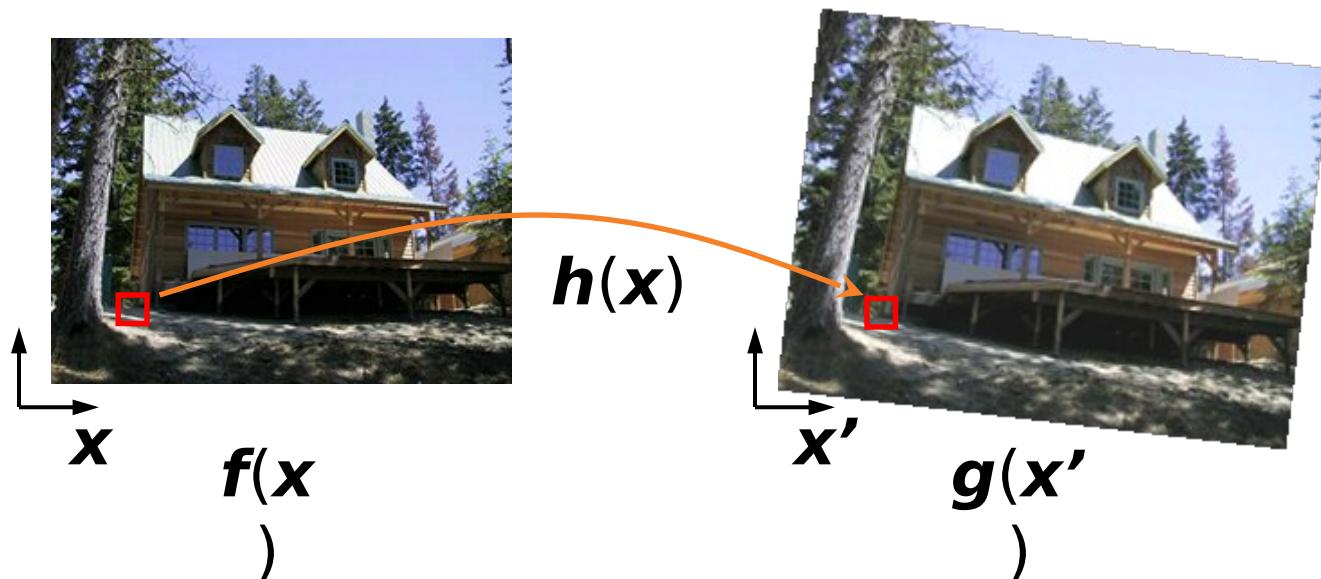


Feature-Based Alignment

- Find keypoints; compute SIFT descriptors
- Generate candidate keypoint matches
- Use RANSAC to select a subset of matches
- Fit to find best image transformation
- Warp images according to transformation
- Blend images in overlapping regions

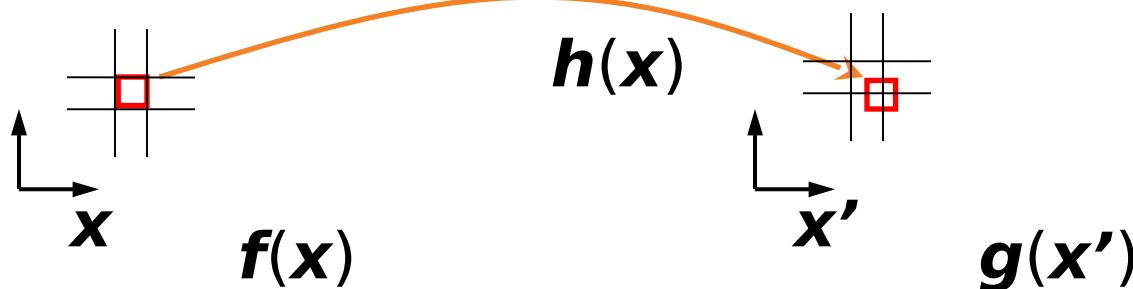
Forward Warping

- Send each pixel $f(x)$ to its corresponding location $x' = h(x)$ in $g(x')$
 - What if pixel lands “between” two pixels?



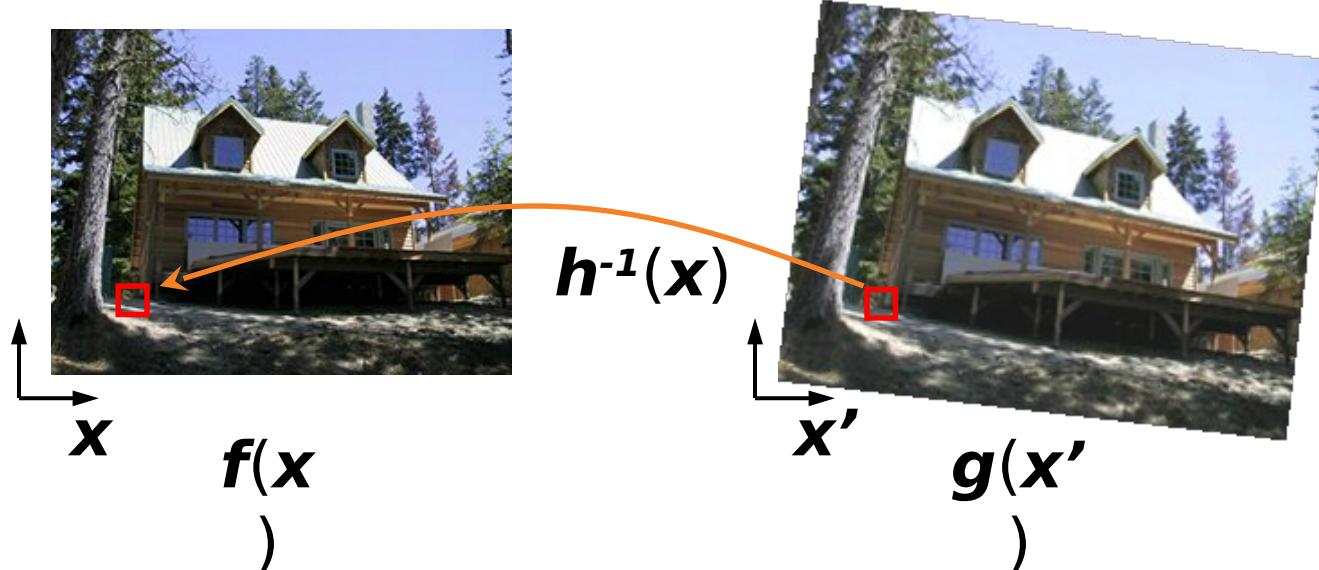
Forward Warping

- Send each pixel $f(x)$ to its corresponding location $x' = h(x)$ in $g(x')$
 - What if pixel lands “between” two pixels?
 - Answer: add “contribution” to several pixels, normalize later (*splatting*)



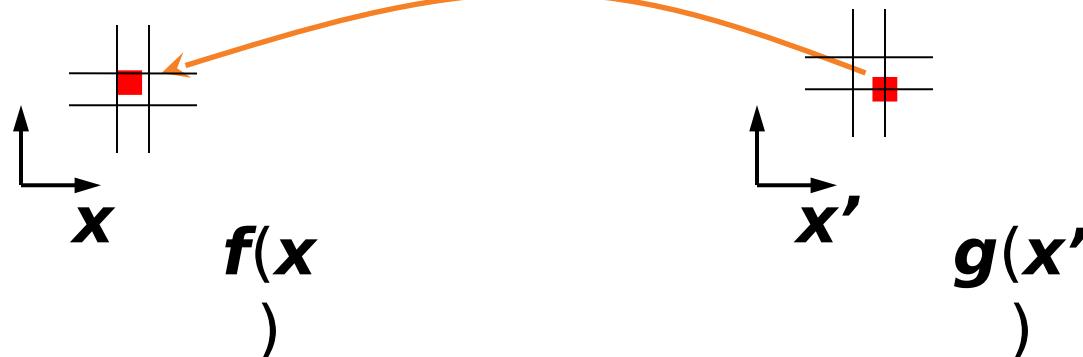
Inverse Warping

- Get each pixel $\mathbf{g}(\mathbf{x}')$ from its corresponding location $\mathbf{x}' = \mathbf{h}(\mathbf{x})$ in $\mathbf{f}(\mathbf{x})$
 - What if pixel comes from “between” two pixels?



Inverse Warping

- Get each pixel $\mathbf{g}(\mathbf{x}')$ from its corresponding location $\mathbf{x}' = \mathbf{h}(\mathbf{x})$ in $\mathbf{f}(\mathbf{x})$
 - What if pixel comes from “between” two pixels?
 - Answer: *resample color value from interpolated (prefiltered) source image*



Interpolation

- Possible interpolation filters:
 - nearest neighbor
 - bilinear
 - bicubic (interpolating)
 - sinc / FIR
- See COS 426 for details on how to avoid “jaggies”



Last Time: Feature-Based Alignment

- Find keypoints; compute SIFT descriptors
- Generate candidate keypoint matches
- Use RANSAC to select a subset of matches
- Fit to find best image transformation
- Warp images according to transformation
- Blend images in overlapping regions

Blending

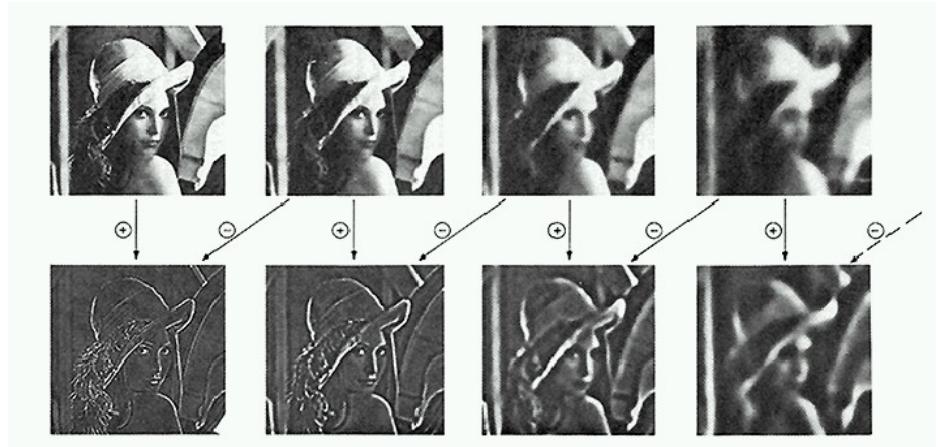
- Blend over too small a region: **seams**
- Blend over too large a region: **ghosting**

Multiresolution Blending

- Different blending regions for different levels in a pyramid [Burt & Adelson]
 - Blend low frequencies over large regions (minimize seams due to brightness variations)
 - Blend high frequencies over small regions (minimize ghosting)

Pyramid Creation

- “Gaussian” Pyramid
- “Laplacian” Pyramid
 - Created from Gaussian pyramid by subtraction
$$L_i = G_i - \text{expand}(G_{i+1})$$



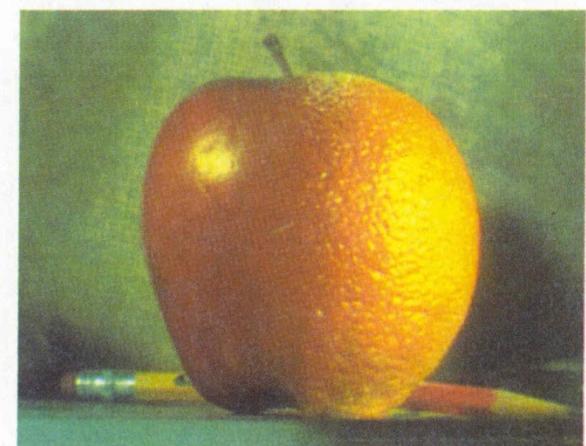
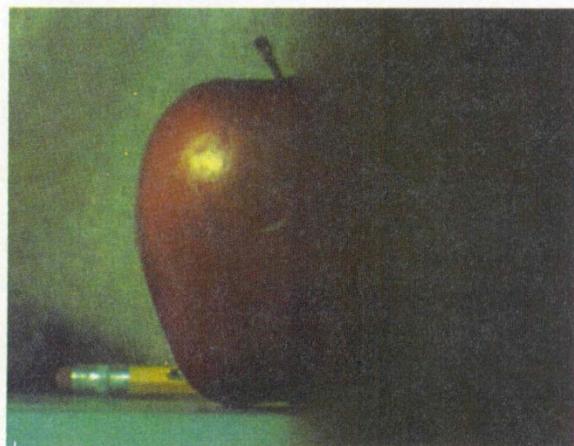
Octaves in the Spatial Domain

Lowpass Images



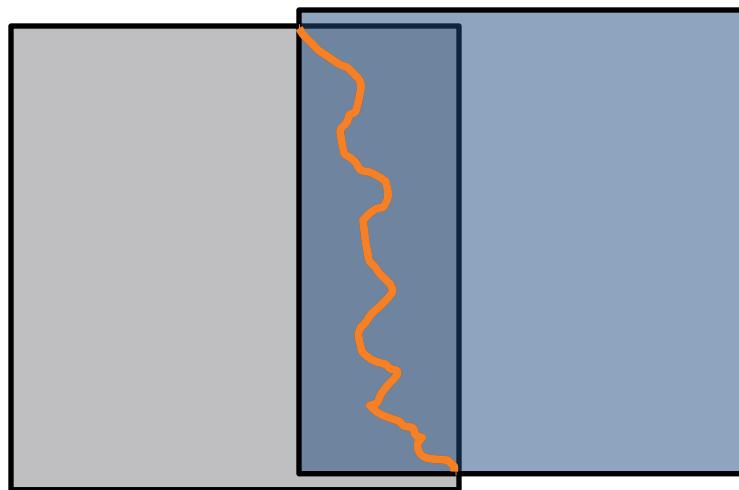
Bandpass Images

Pyramid Blending



Minimum-Cost Cuts

- Instead of blending high frequencies along a straight line, blend along line of minimum differences in image intensities



Minimum-Cost Cuts



Moving object, simple blending  blur

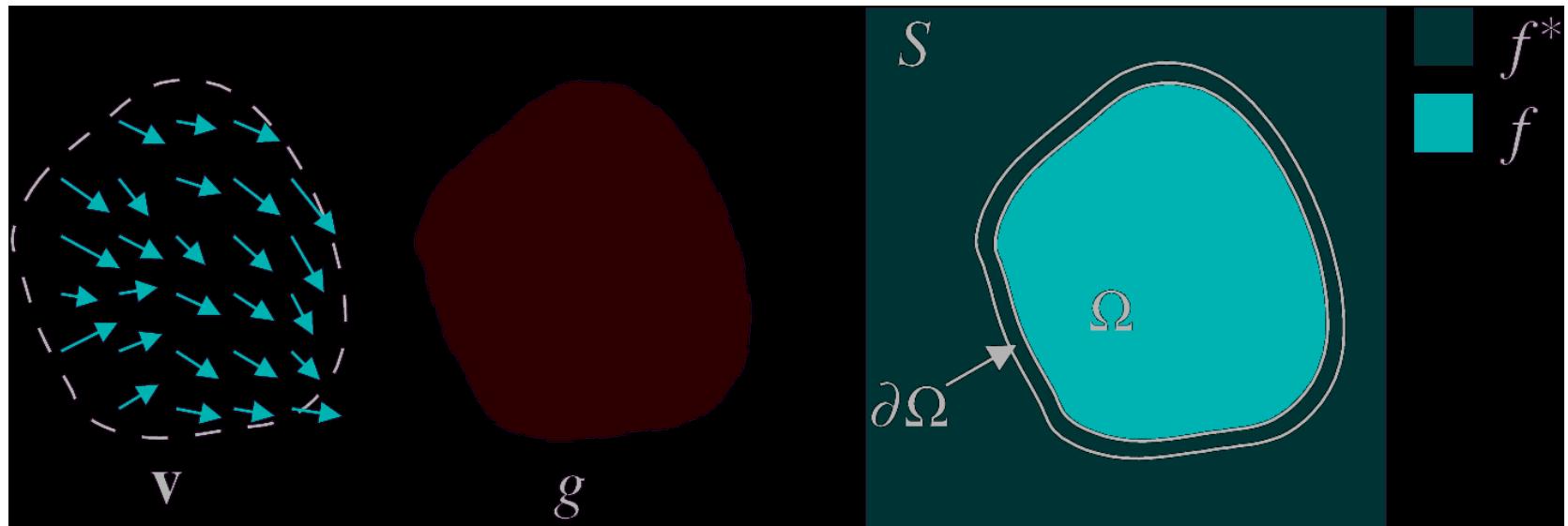
Minimum-Cost Cuts



Minimum-cost cut \equiv no blur

Poisson Image Blending

- Follow gradients of source subject to boundary conditions imposed by dest

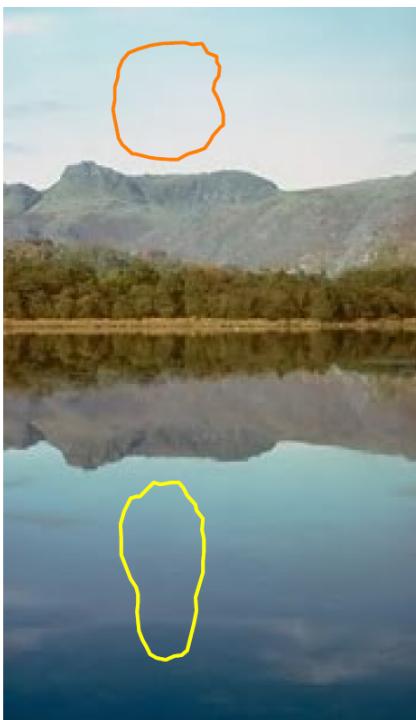


$$\begin{cases} \nabla^2 f = \nabla \cdot \mathbf{v} \\ f|_{\partial\Omega} = f^*|_{\partial\Omega} \end{cases}$$

Poisson Image Blending



sources



destinations

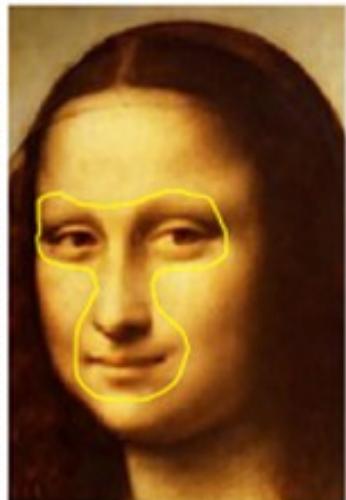


cloning



seamless cloning

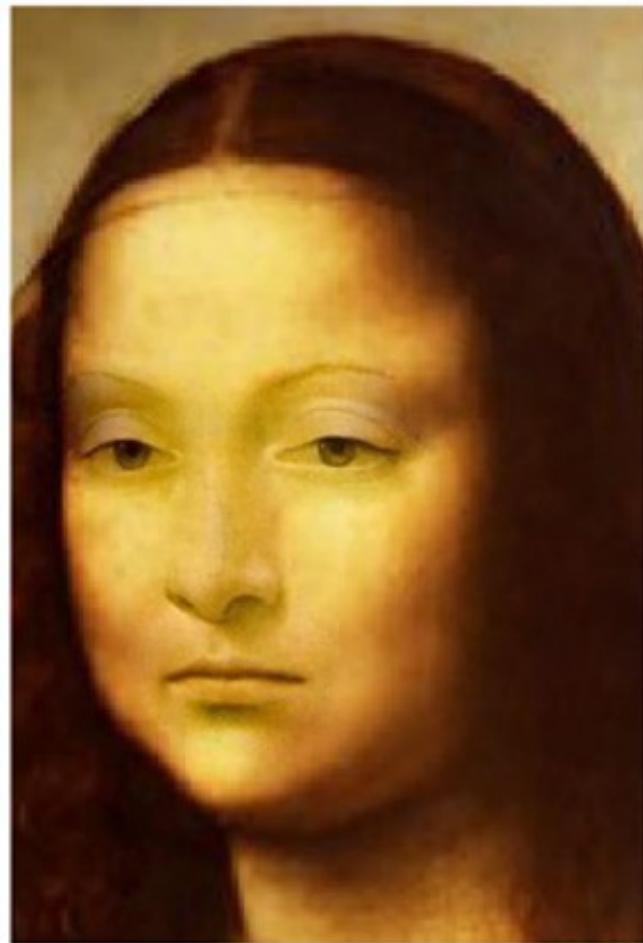
Poisson Image Blending



source/destination



cloning



seamless cloning

Recap: Feature-Based Alignment

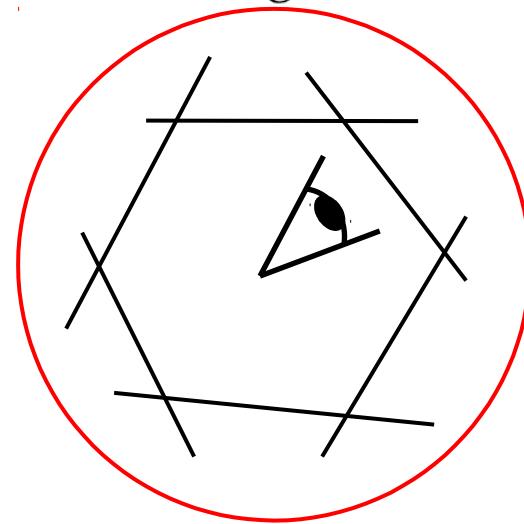
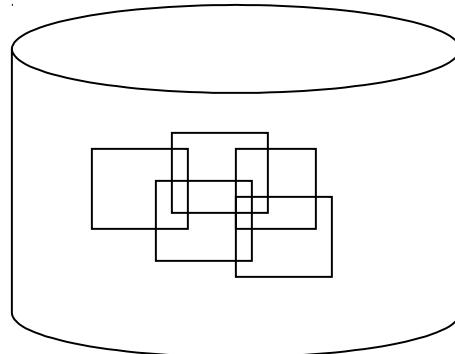
- Find keypoints; compute SIFT descriptors
- Generate candidate keypoint matches
- Use RANSAC to select a subset of matches
- Fit to find best image transformation
- Warp images according to transformation
- Blend images in overlapping regions
 - YouTube: search for “Interactive Digital Photomontage”

Real-World Panoramic Stitching

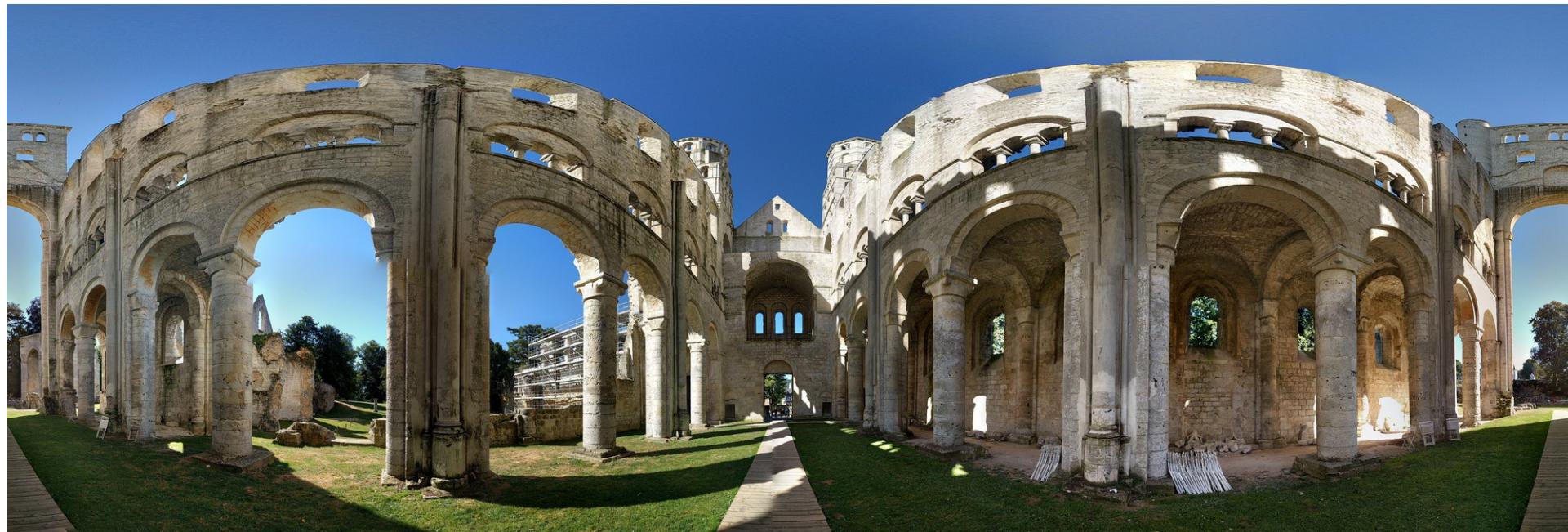
- How to handle more than 2 frames?
 - Align each frame to the previous: simple, but can lead to drift in alignment
 - Optimize for all transformations at once: “bundle adjustment”

Real-World Panoramic Stitching

- How to handle extremely wide total field of view?
 - Project onto cylinder – allows 360° viewing



Panorama



Lecture 7

Introduction to Recognition

COS 429: Computer Vision



Object recognition: let's try something simple

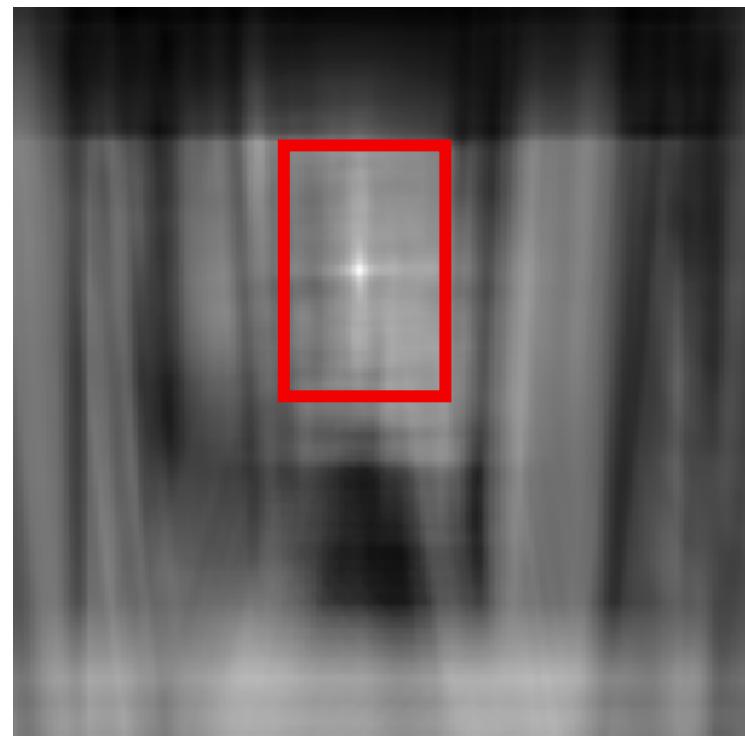
This is a chair



Find the chair in this image

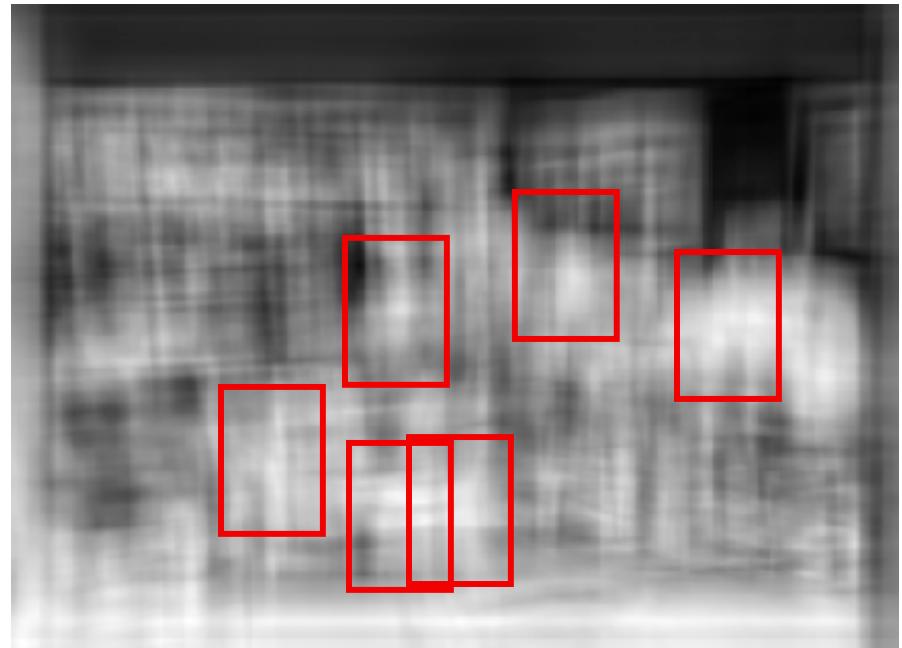
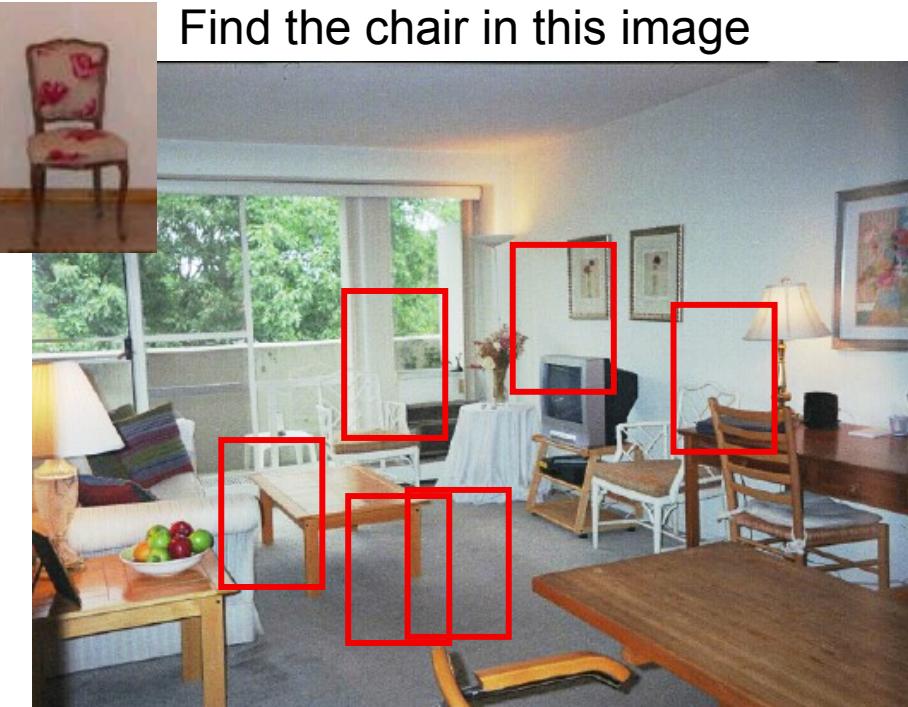


Output of normalized correlation



Object recognition: let's try something simple

Find the chair in this image



Simple template matching
is not going to be enough

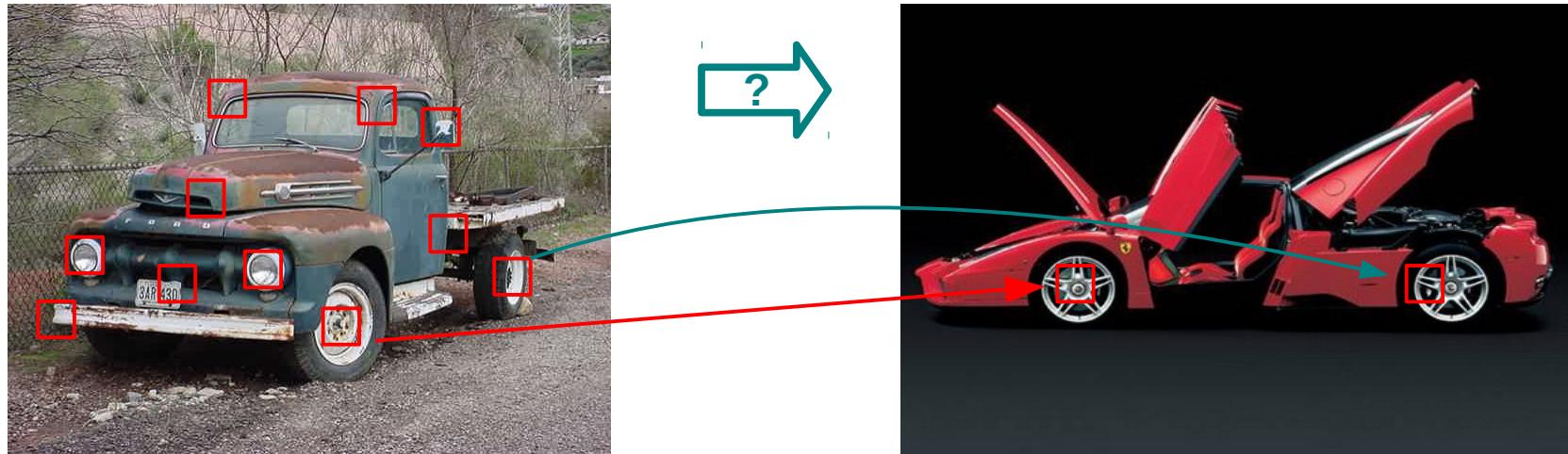
How about SIFT based alignment?



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!

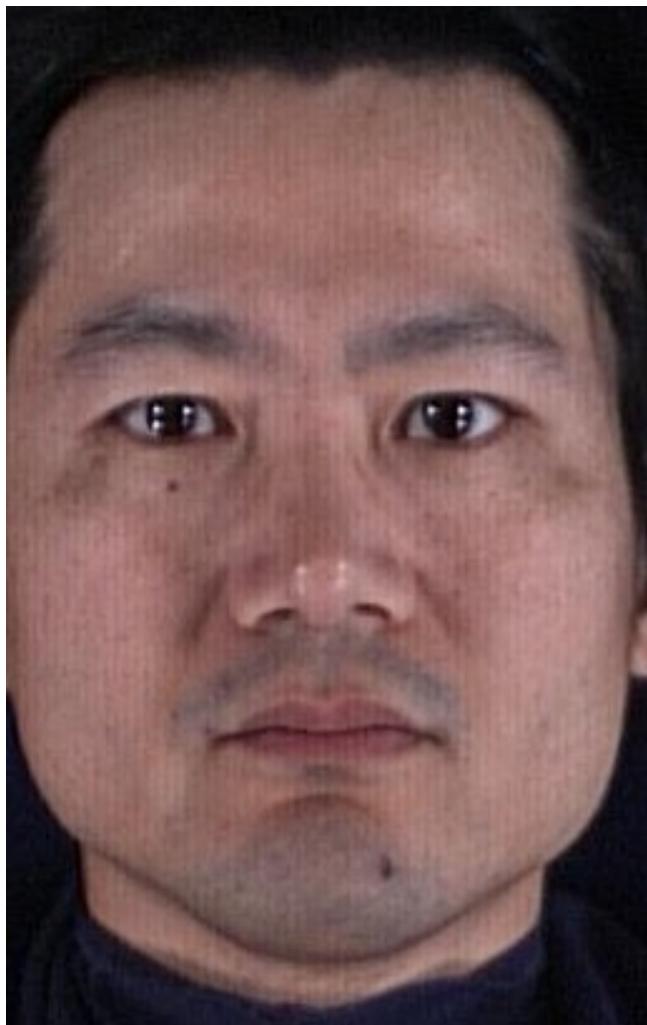
Challenges 1: view point variation



Michelangelo 1475-1564

Slides: course object recognition
ICCV 2005

Challenges 2: illumination



Challenges 3: occlusion



Magritte, 1957

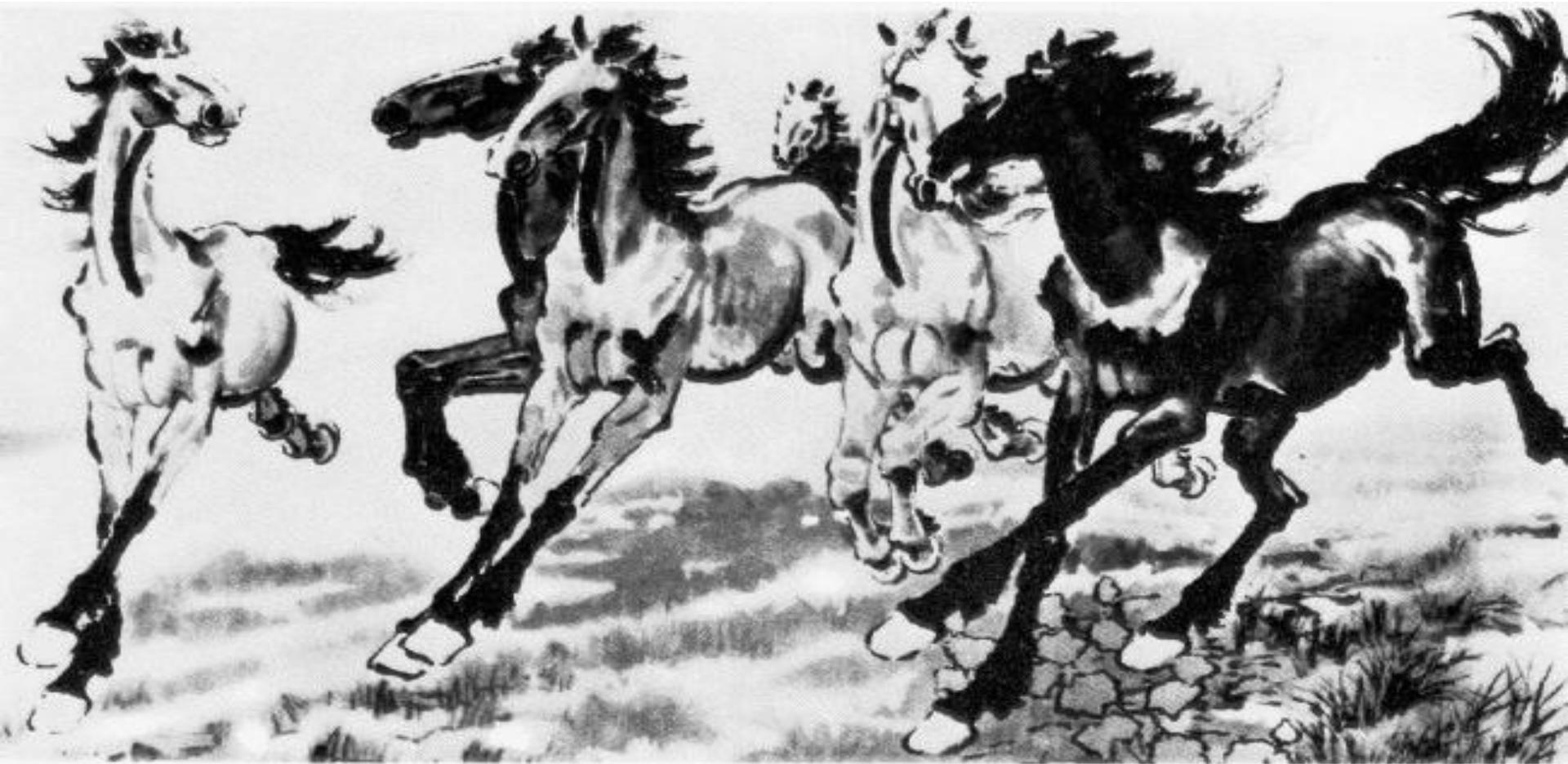
Slides: course object recognition
ICCV 2005

Challenges 4: scale



Slides: course object recognition
ICCV 2005

Challenges 5: deformation



Slides: course object recognition
ICCV 2005

Xu, Beihong 1943

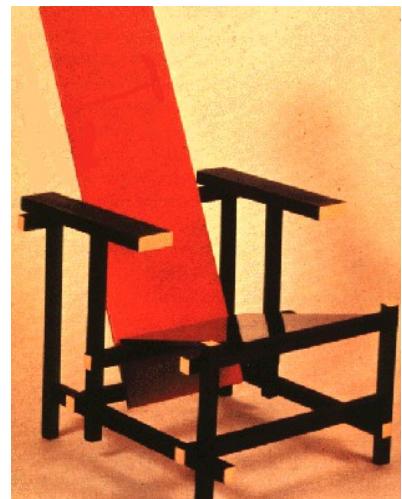
33 : COS429 : 05.10.17 : Andras Ferencz

Challenges 6: background clutter

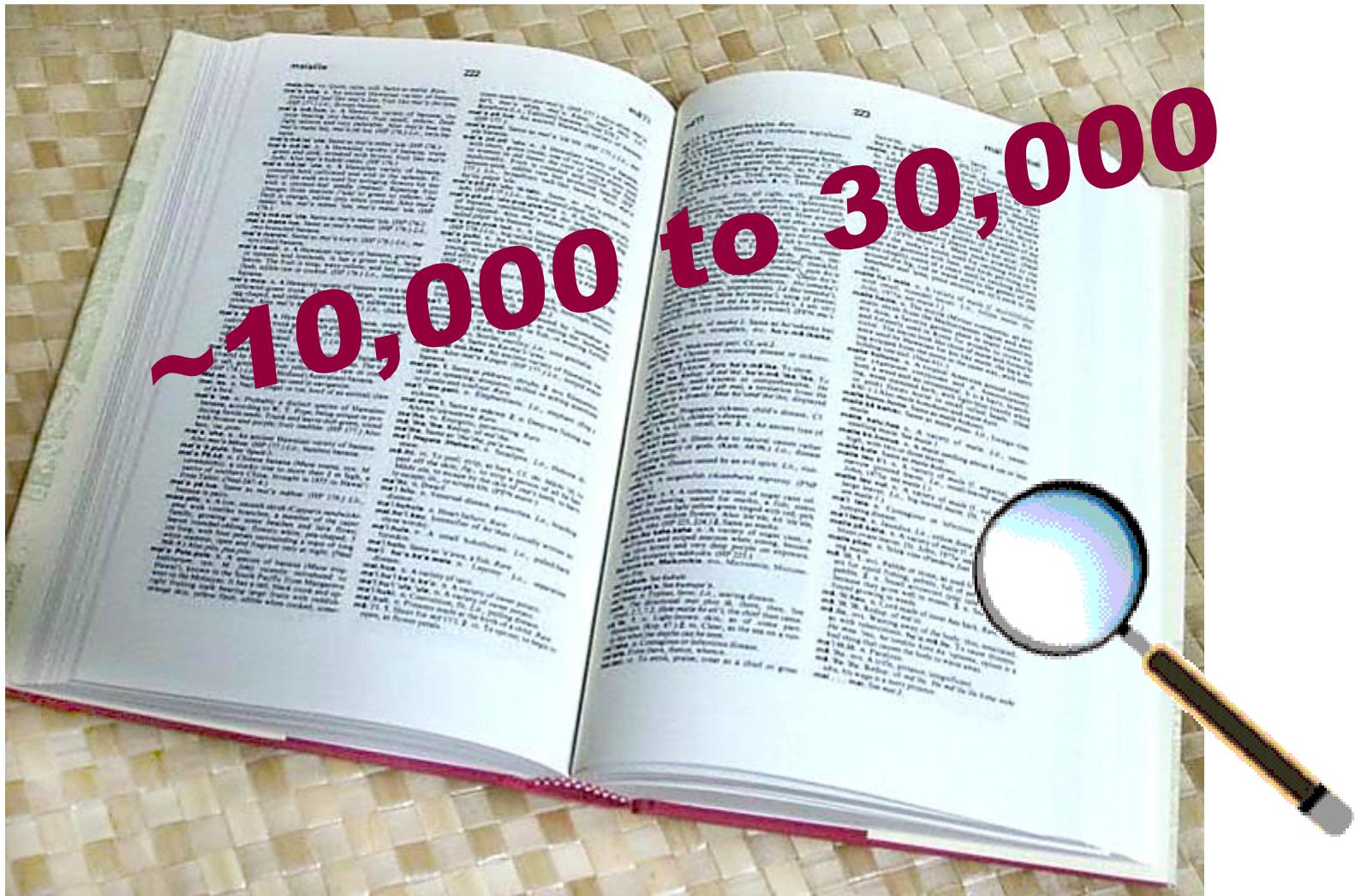


Brady, M. J., & Kersten, D. (2003). Bootstrapped learning of novel objects. *J Vis*, 3(6), 413-422

Challenges 7: intra-class variation



How many visual object categories are there?



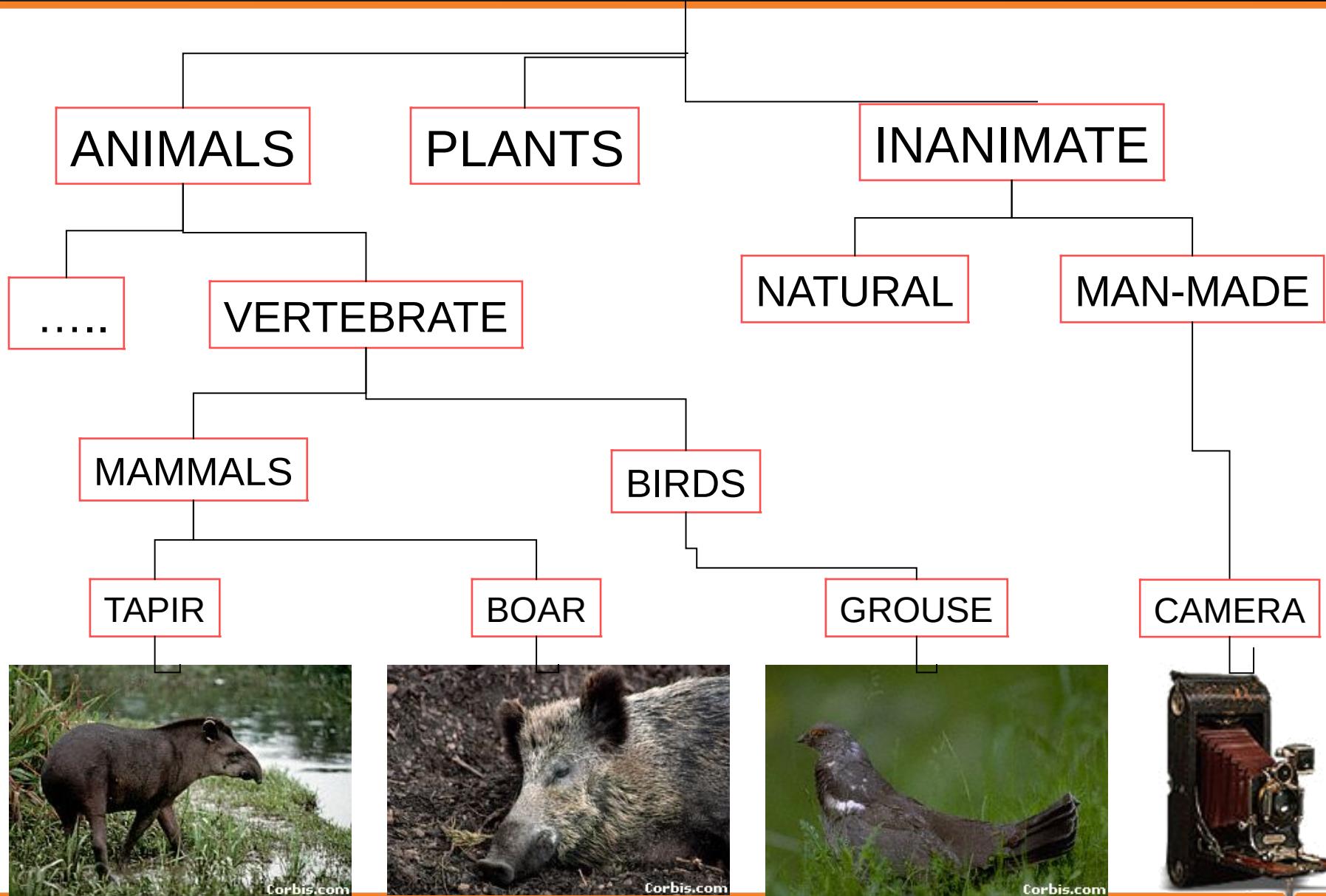
Biederman 1987



~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.

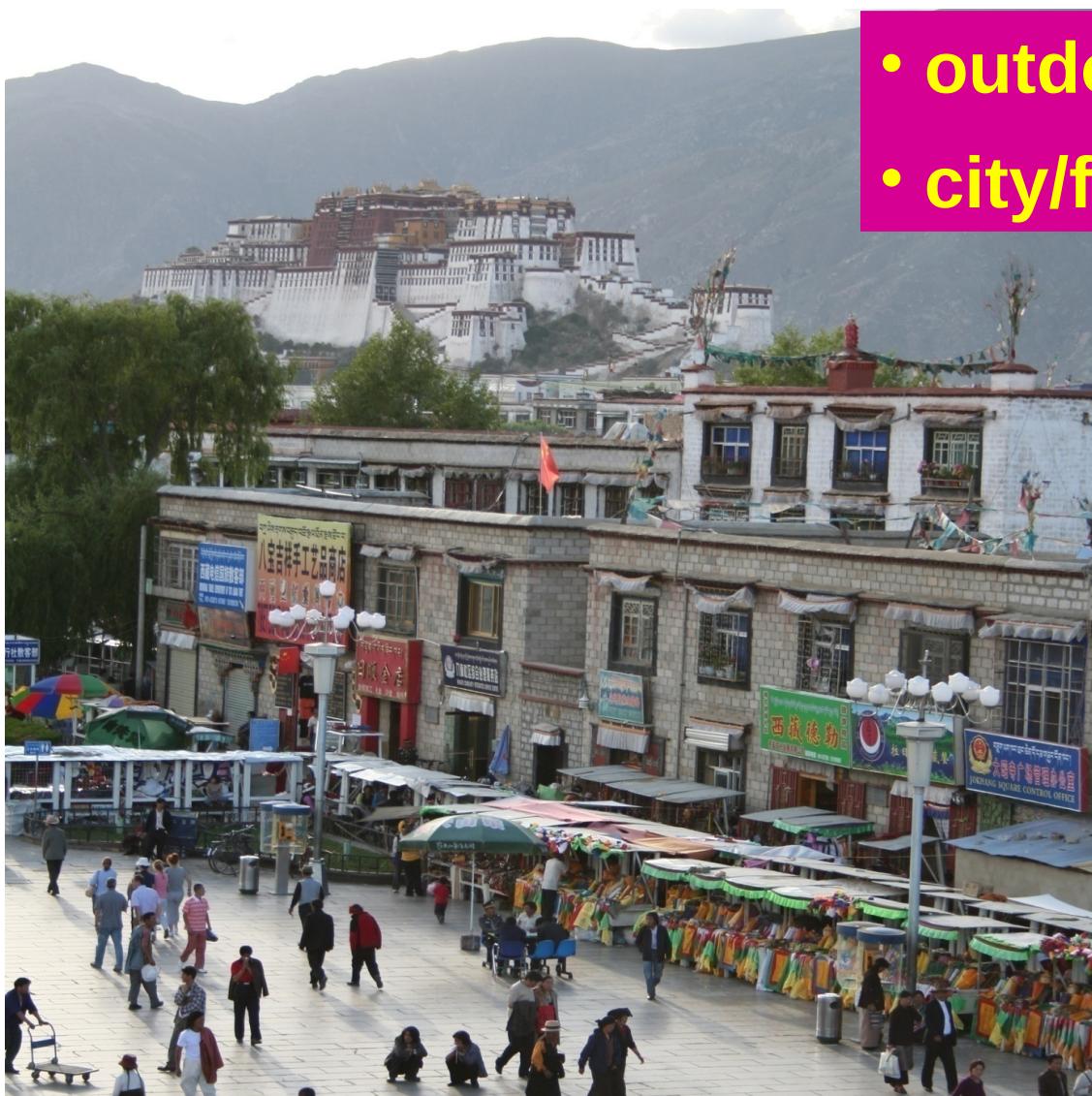


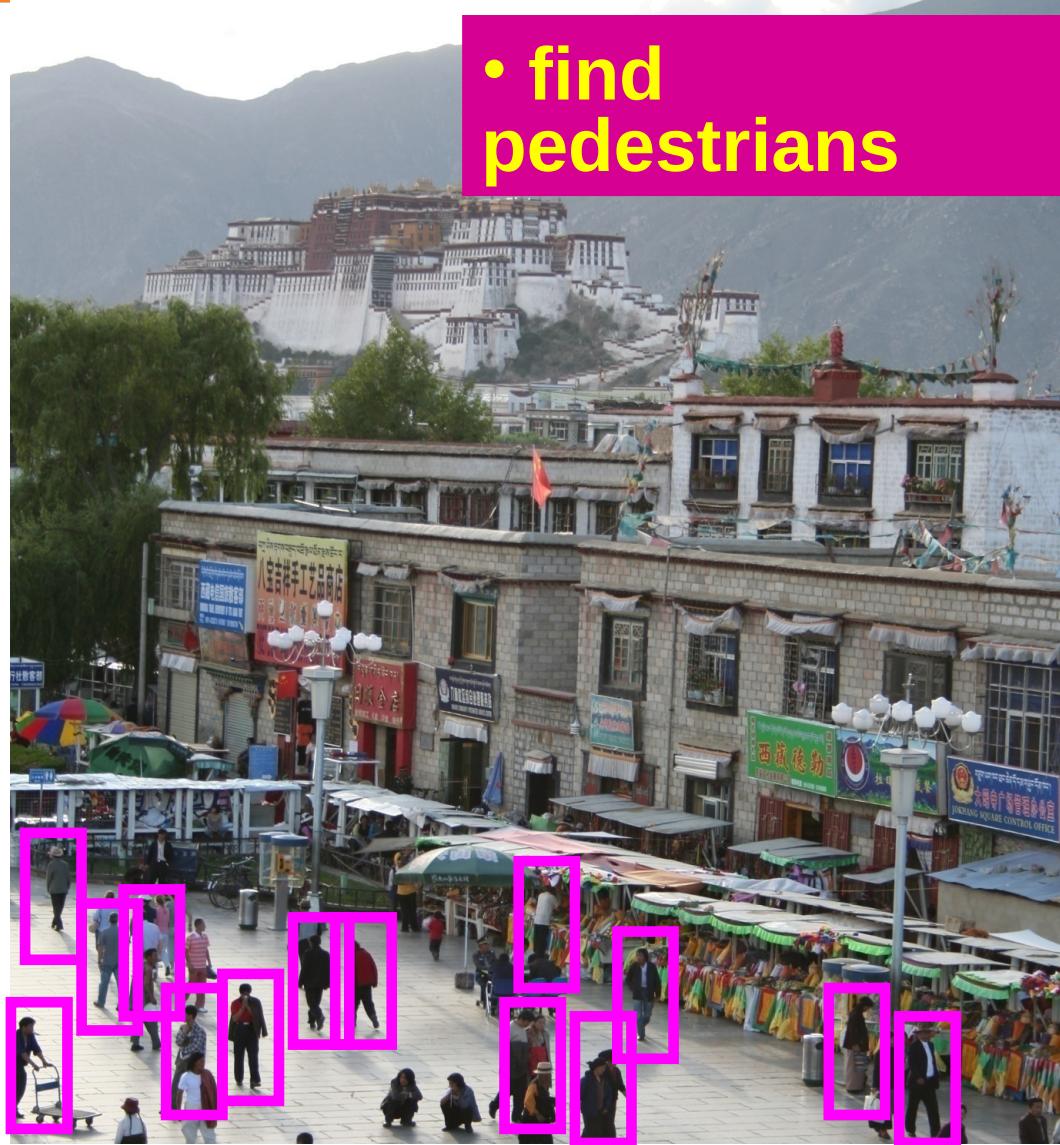
Image annotation / tagging / attributes



- 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

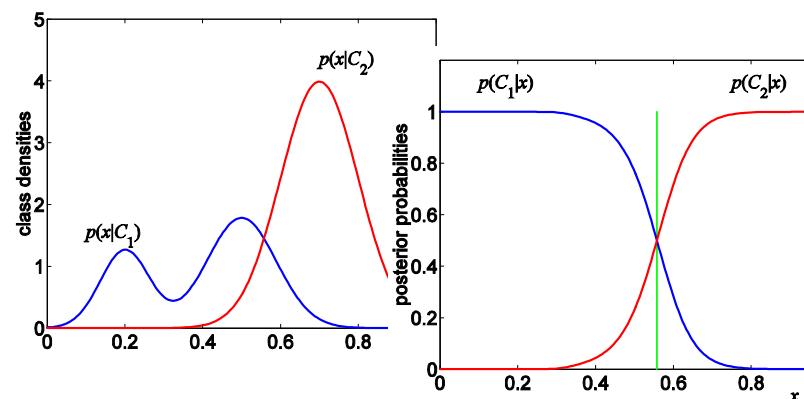
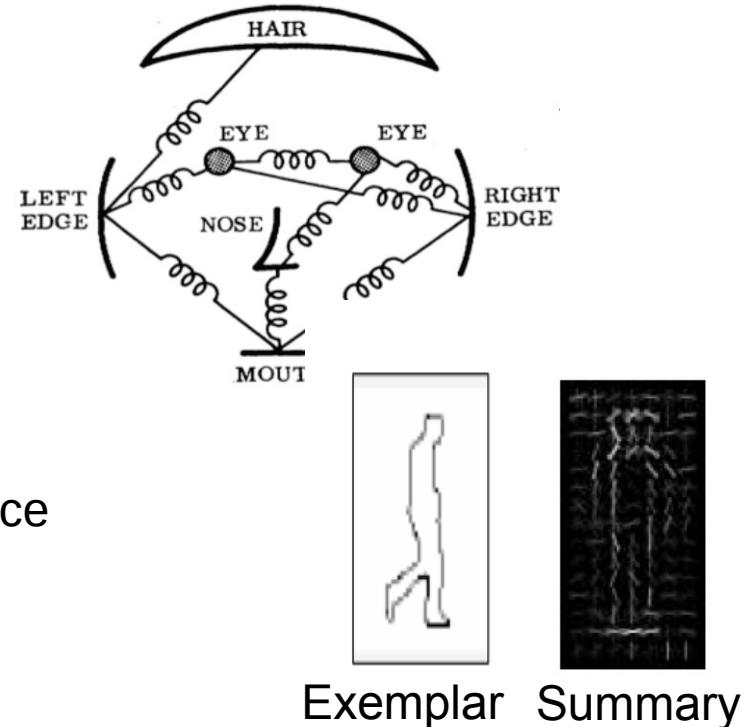
Scene understanding?



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



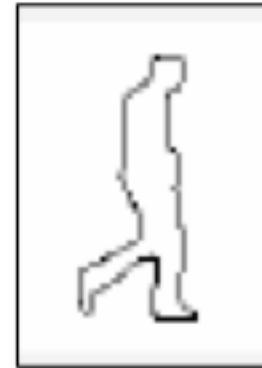
Example 1: Chamfer matching (Pedestrian Detection)



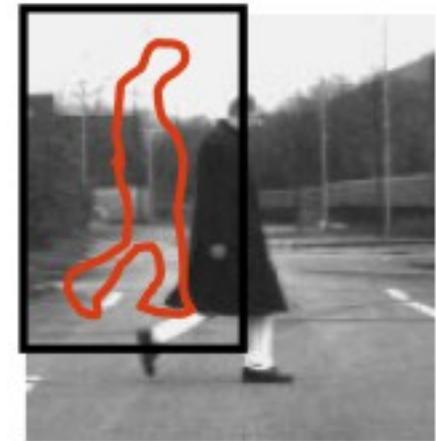
Input Image



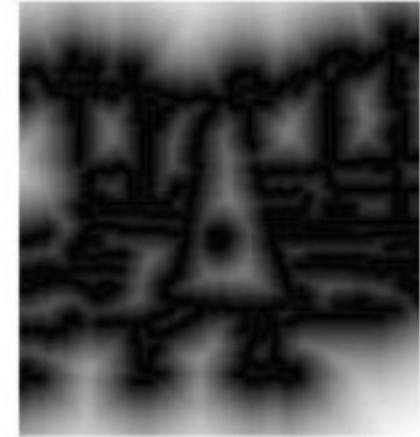
Edge Detection



Template

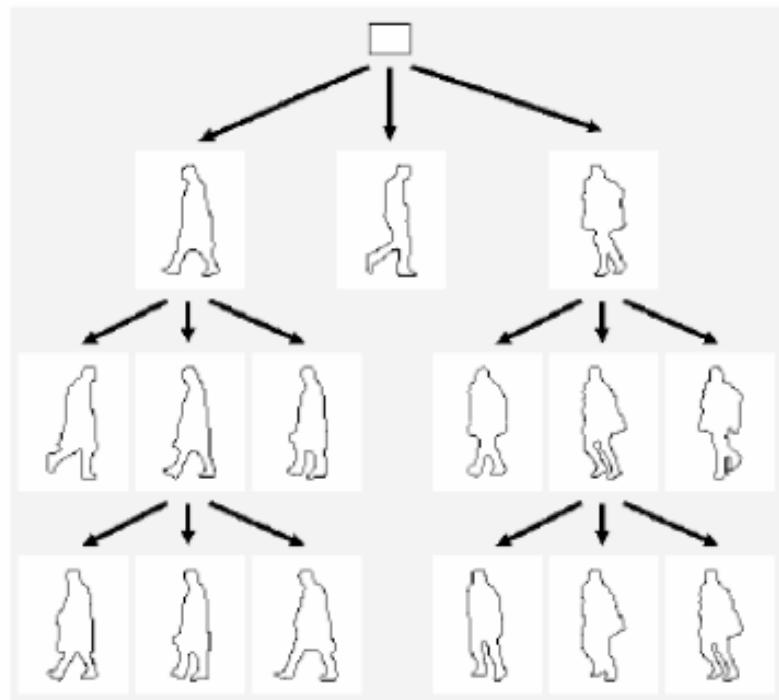


Best Match

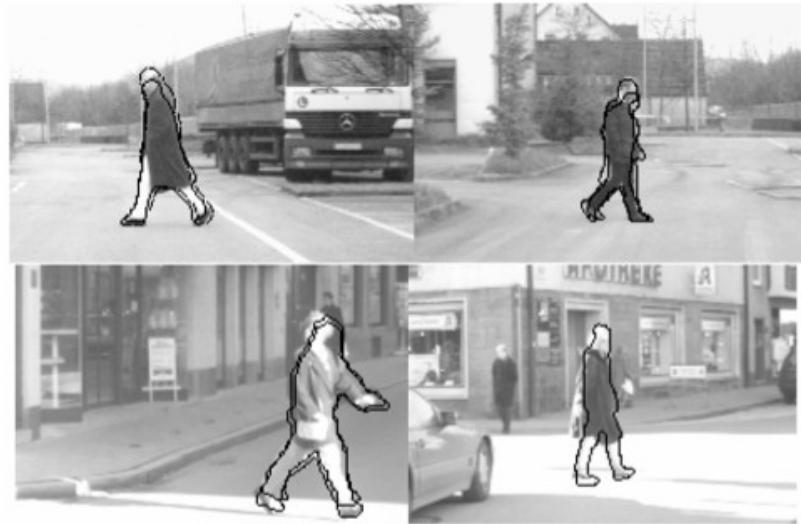


Distance Transform

Example 1: Chamfer matching (Pedestrian Detection)



Hierarchy of templates



Example 2: Viola/Jones (Face Detection)

Robust Realtime Face Detection, IJCV 2004, Viola and Jones

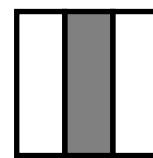
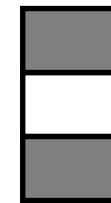
Features: “Haar-like Rectangle filters”

- Differences between sums of pixels in adjacent rectangles

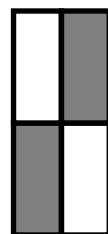
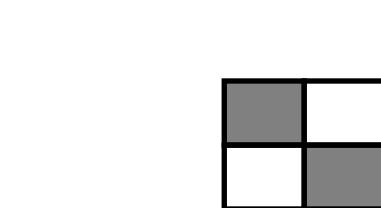
-1 +1



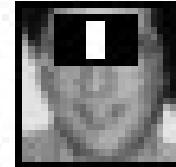
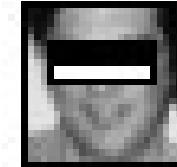
2-rectangle features



3-rectangle features



4-rectangle features

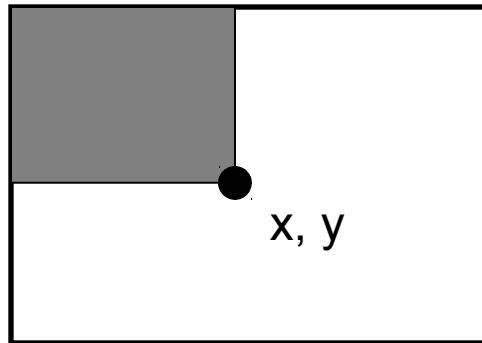


$60,000 \times 100 = 6,000,000$
Unique Features

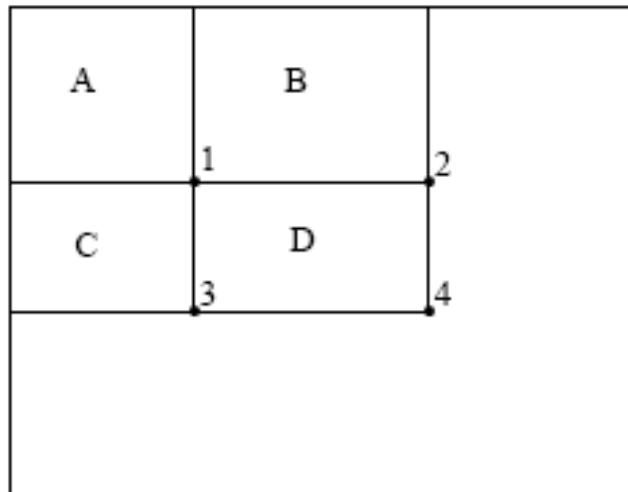
Slide from: Derek Hoiem

Example 2: Viola/Jones - Integral Images

- `ii = cumsum(cumsum(im, 1), 2)`



$ii(x,y) = \text{Sum of the values in the grey region}$



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

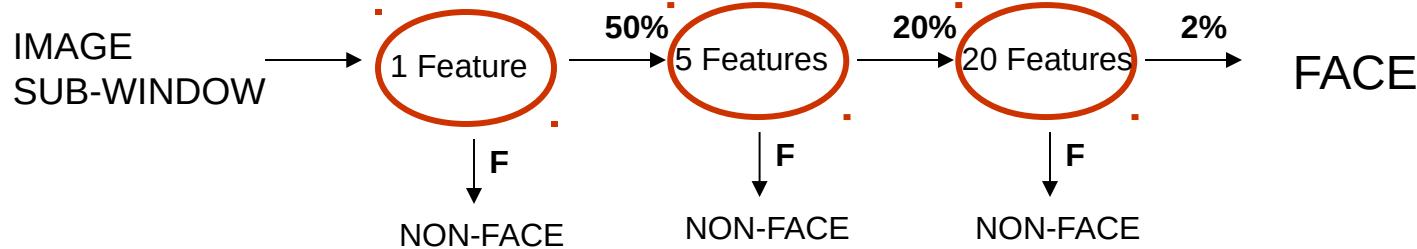
Trained Classifier
(for each stage of cascade)

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

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

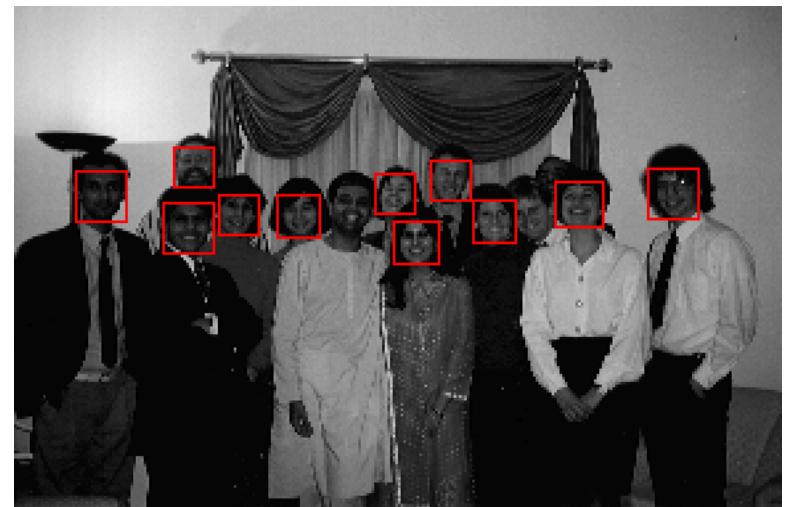
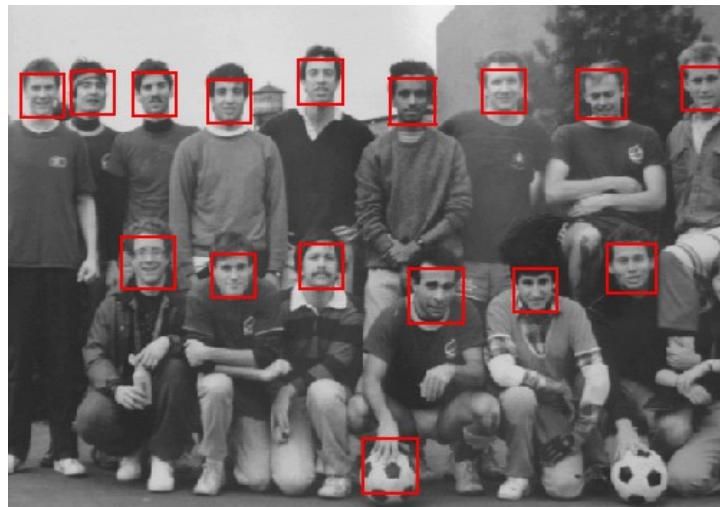
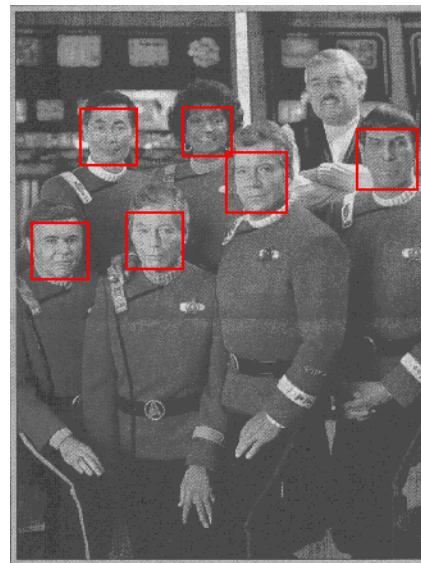
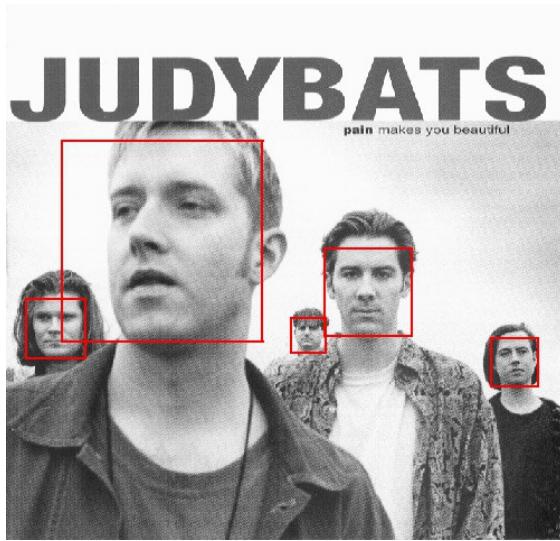
$$\text{Decision} = \begin{cases} \text{face,} & \text{if } Y(x) > 0 \\ \text{non-face,} & \text{otherwise} \end{cases}$$

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

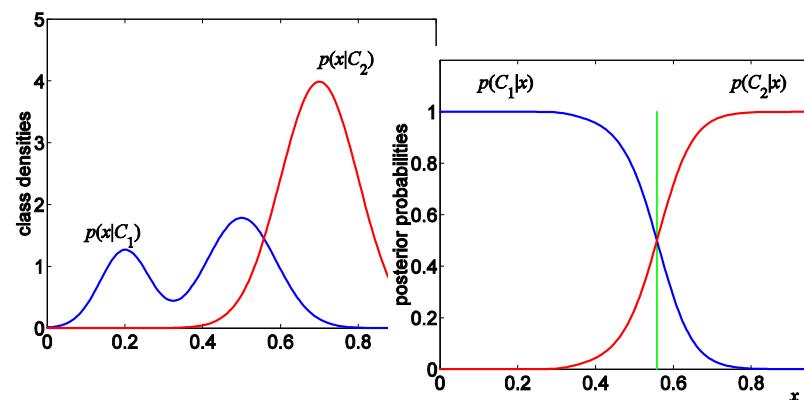
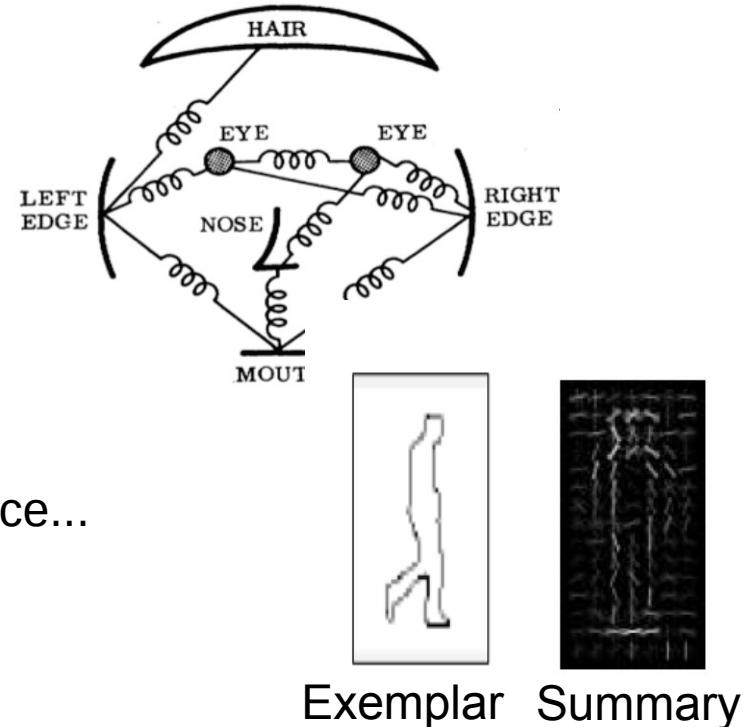
Example 2: Viola/Jones results



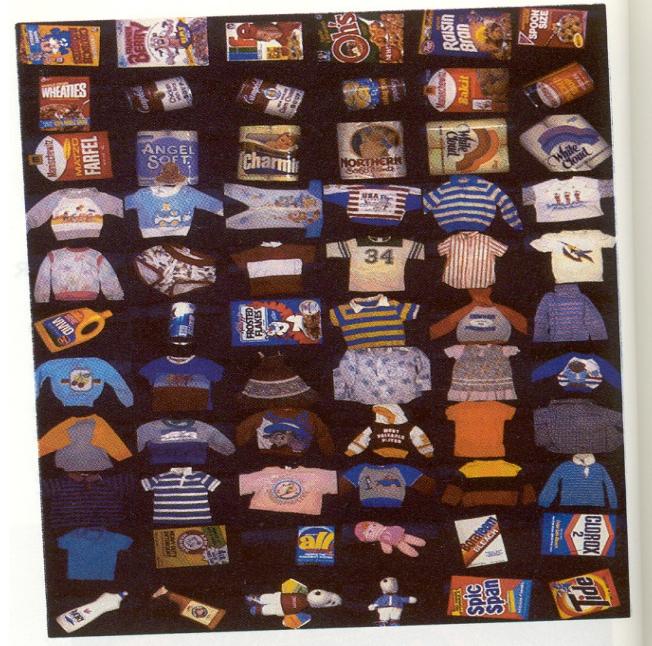
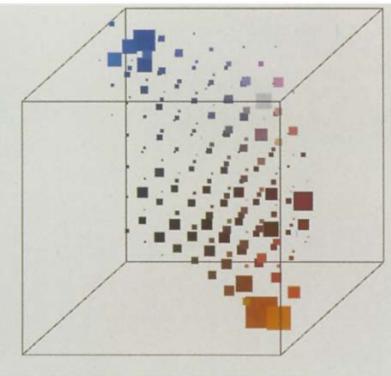
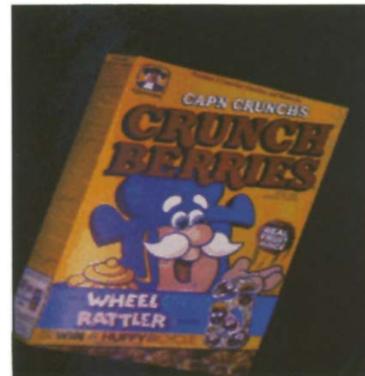
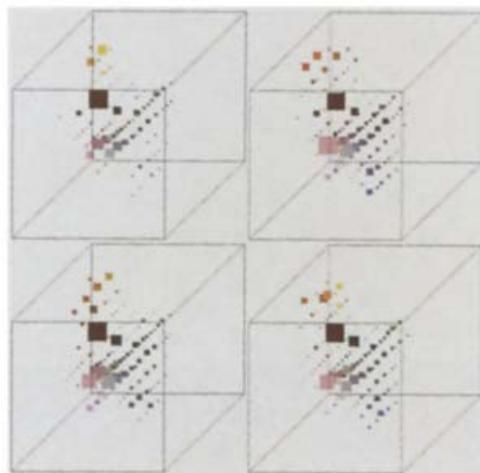
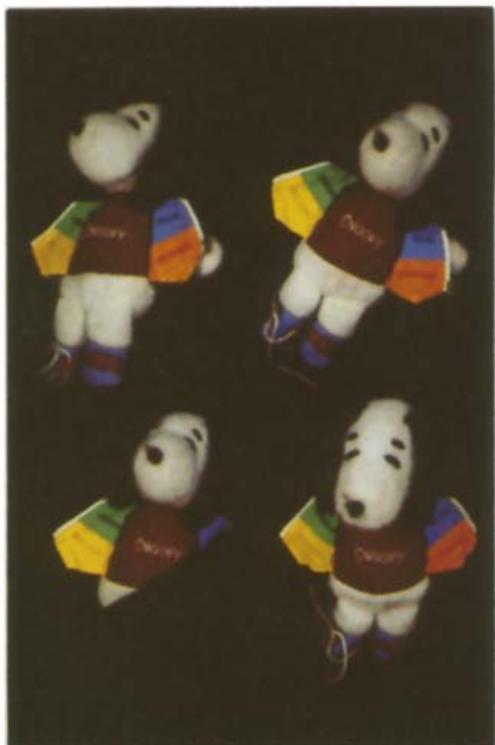
Run-time: 15fps (384x288 pixel image on a 700 Mhz Pentium III)

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
 - Blur, Histogram, Tree/Star, Pairwise/Covariance...
 - **Geometry is Hard. let's ignore it...**
- **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

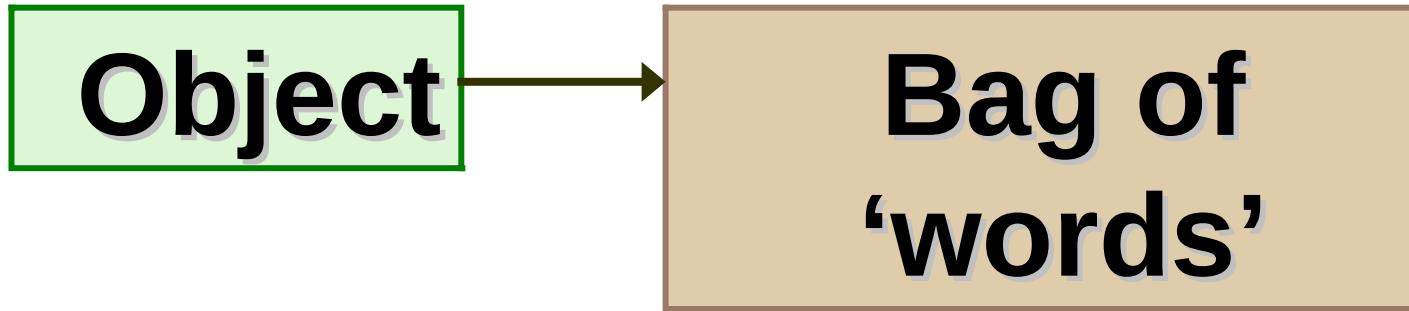


(No Geometry) Example: Color Histograms



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

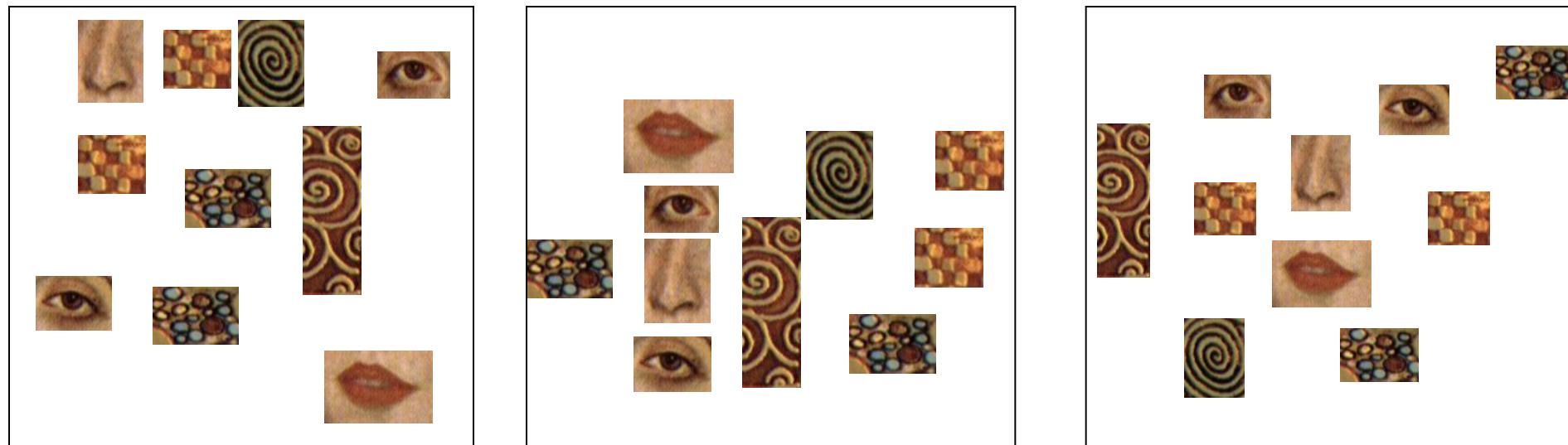
(No Geometry) Example: Bag of Words



Slide from: Svetlana Lazebnik

Objects as texture

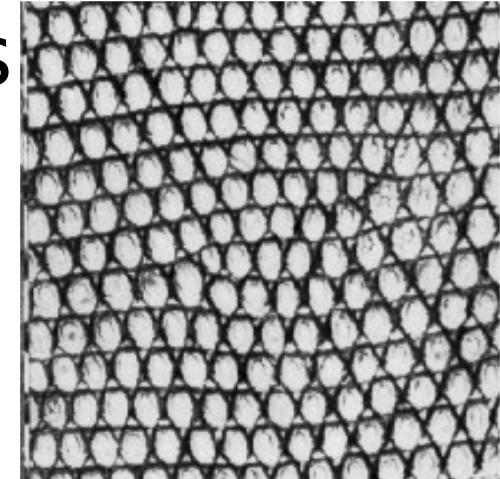
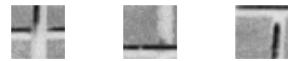
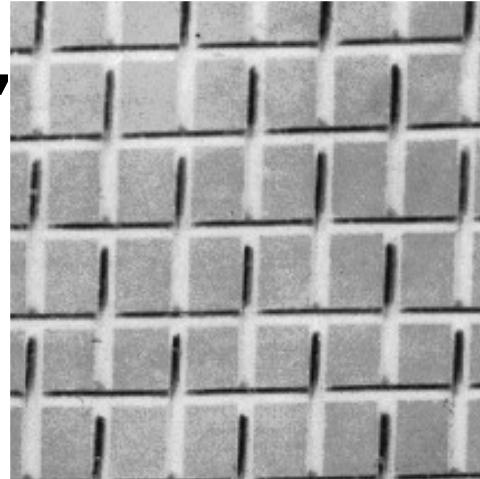
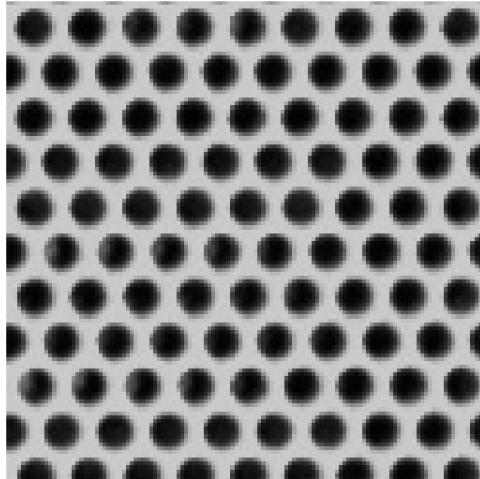
- All of these are treated as being the same



- 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 matters



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**

insurgents **ira** palestinian pay

september **sh** violence viol

1962-10-22: Soviet Missiles in Cuba

John F. Kennedy (1961-63)

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

declined **defensiv** elimination emerge halt hazards **hem** modernization neglec

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

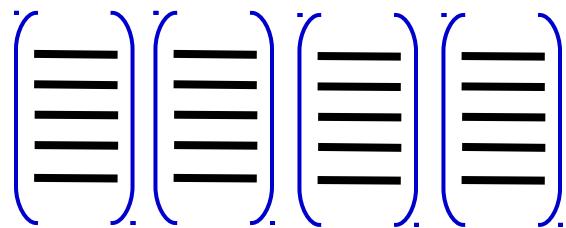
treachery true tyranny undertaken victory **war** wartime washington

- Orderless document representation:
frequencies of words from a dictionary

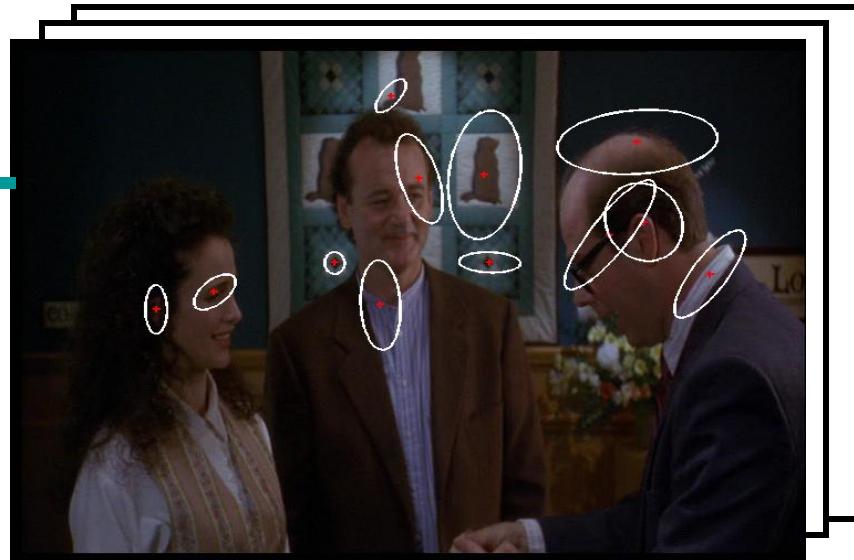
McGill (1983)

Salton &

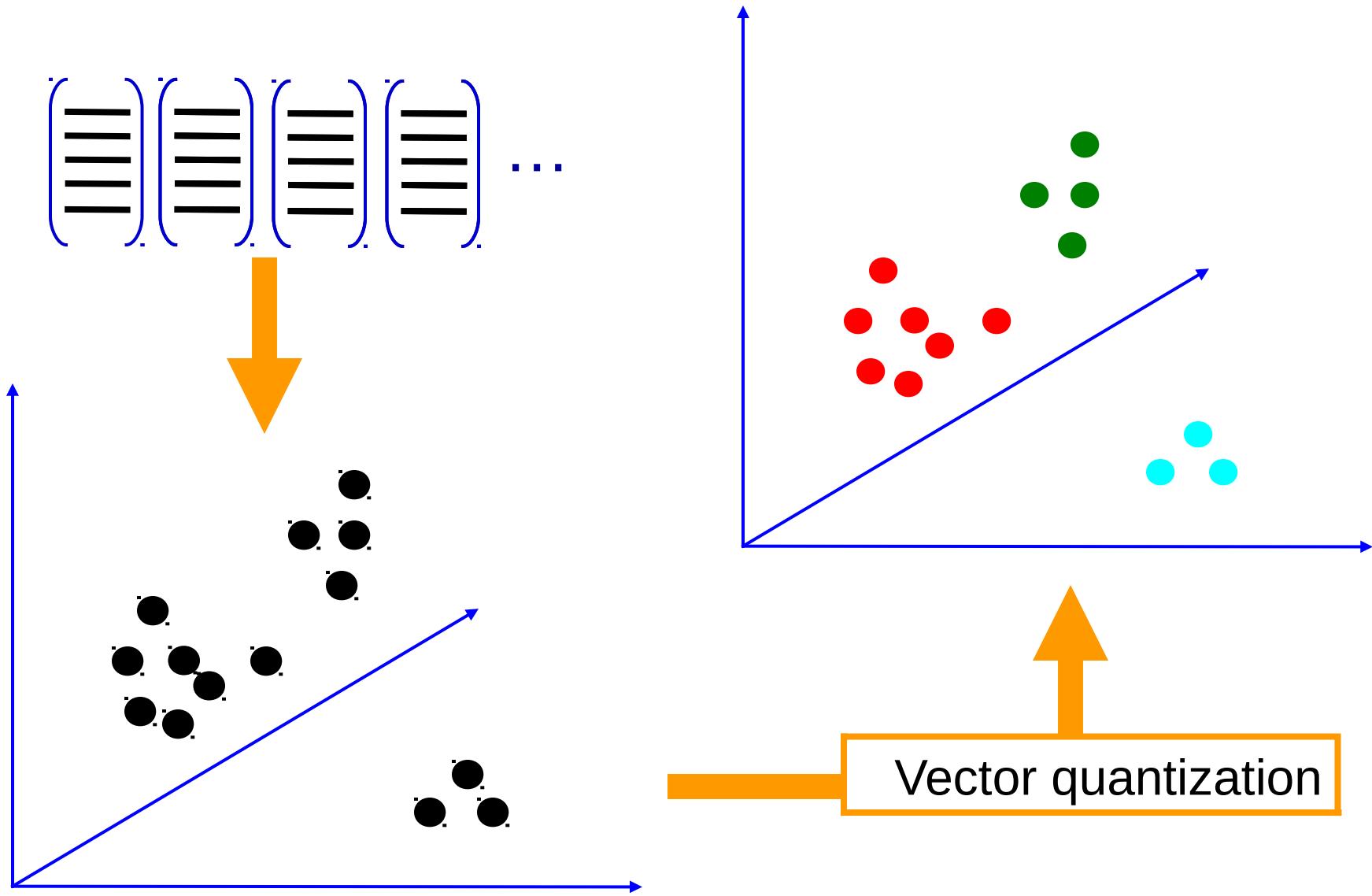
Interest Point Features



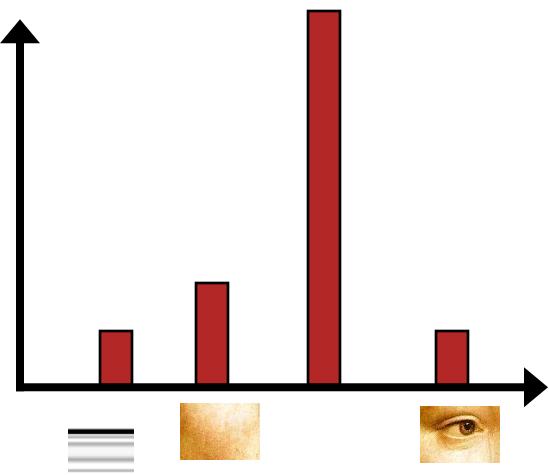
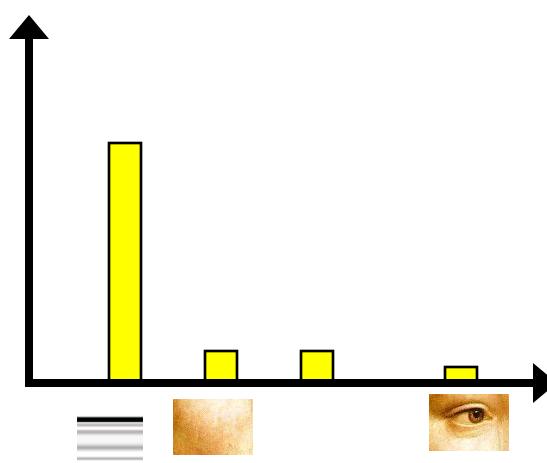
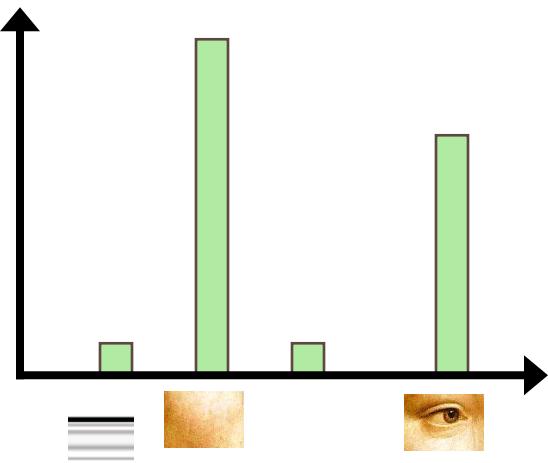
... ←



Clustering (usually k-means)



Slide credit: Josef Sivic



learning



feature detection
& representation

codewords dictionary

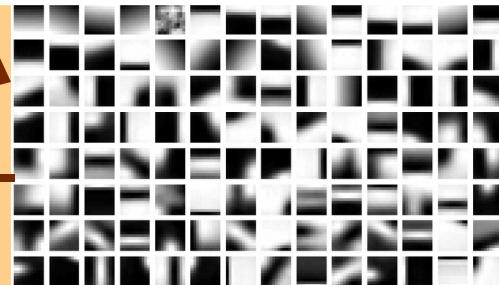
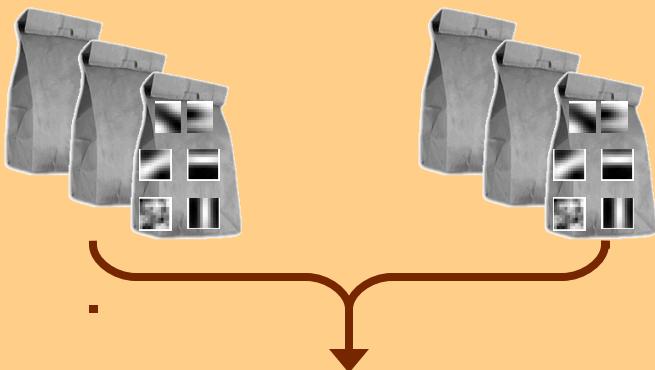


image representation



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

recognition

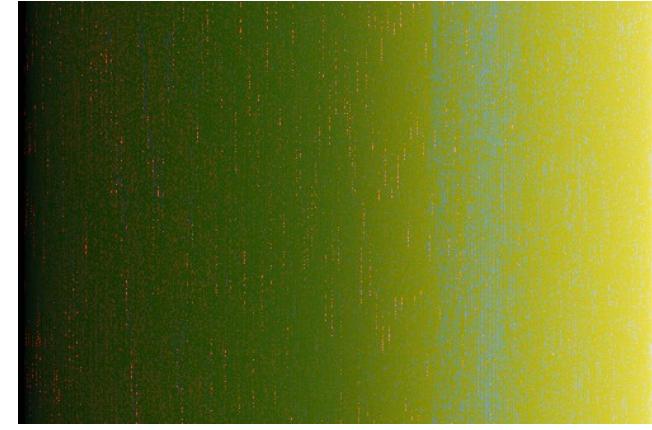
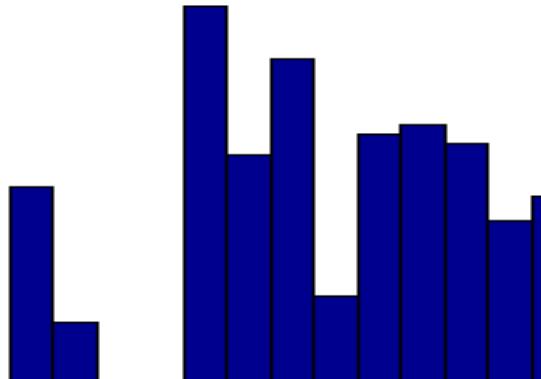


feature detection
& representation



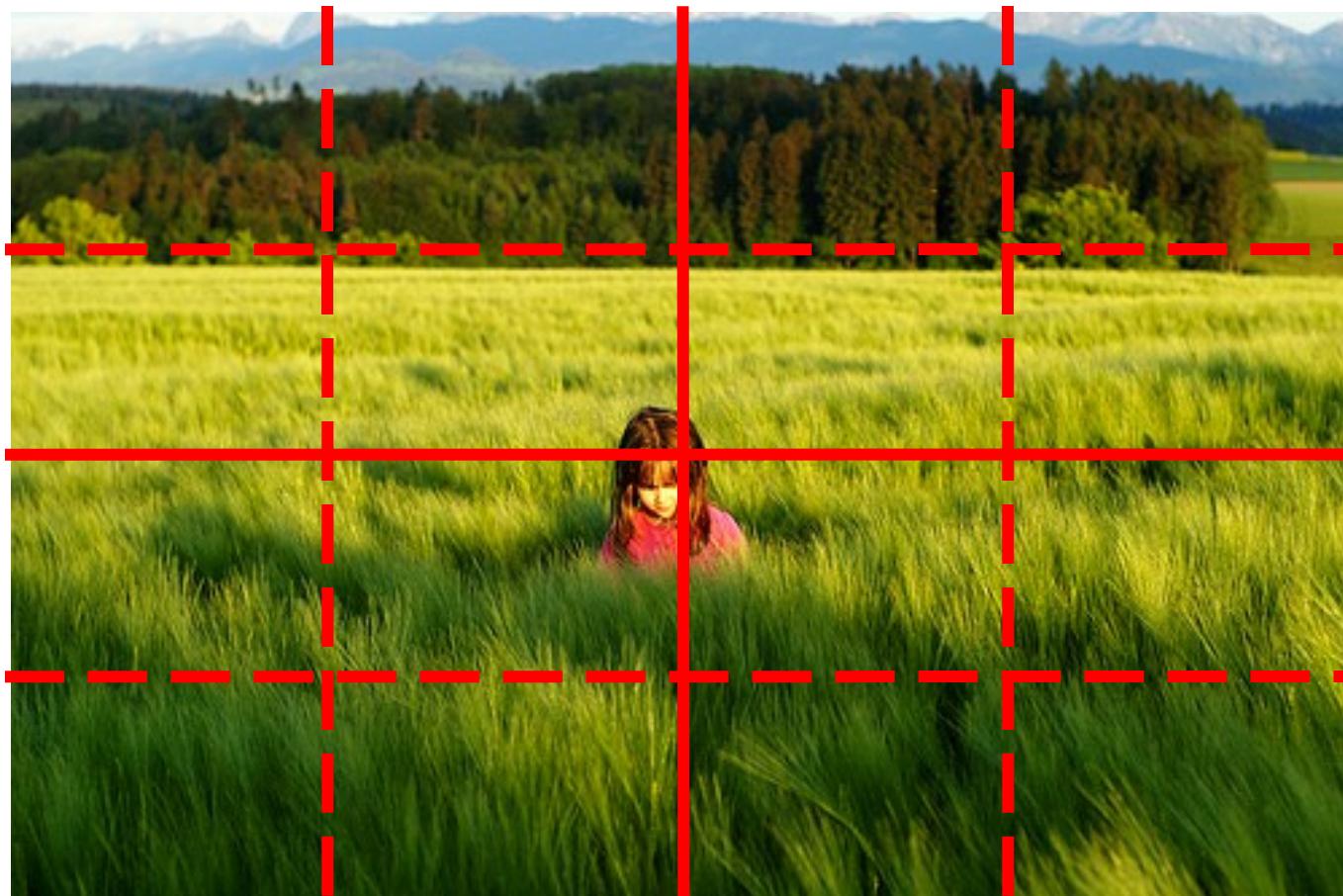
**category
decision**

The (obvious) problem with ignoring Geometry



All of these images have the same color histogram

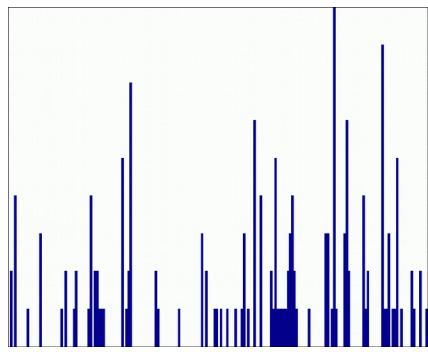
Adding Geometry back: Spatial pyramid



Compute histogram in each spatial bin

Spatial pyramid representation

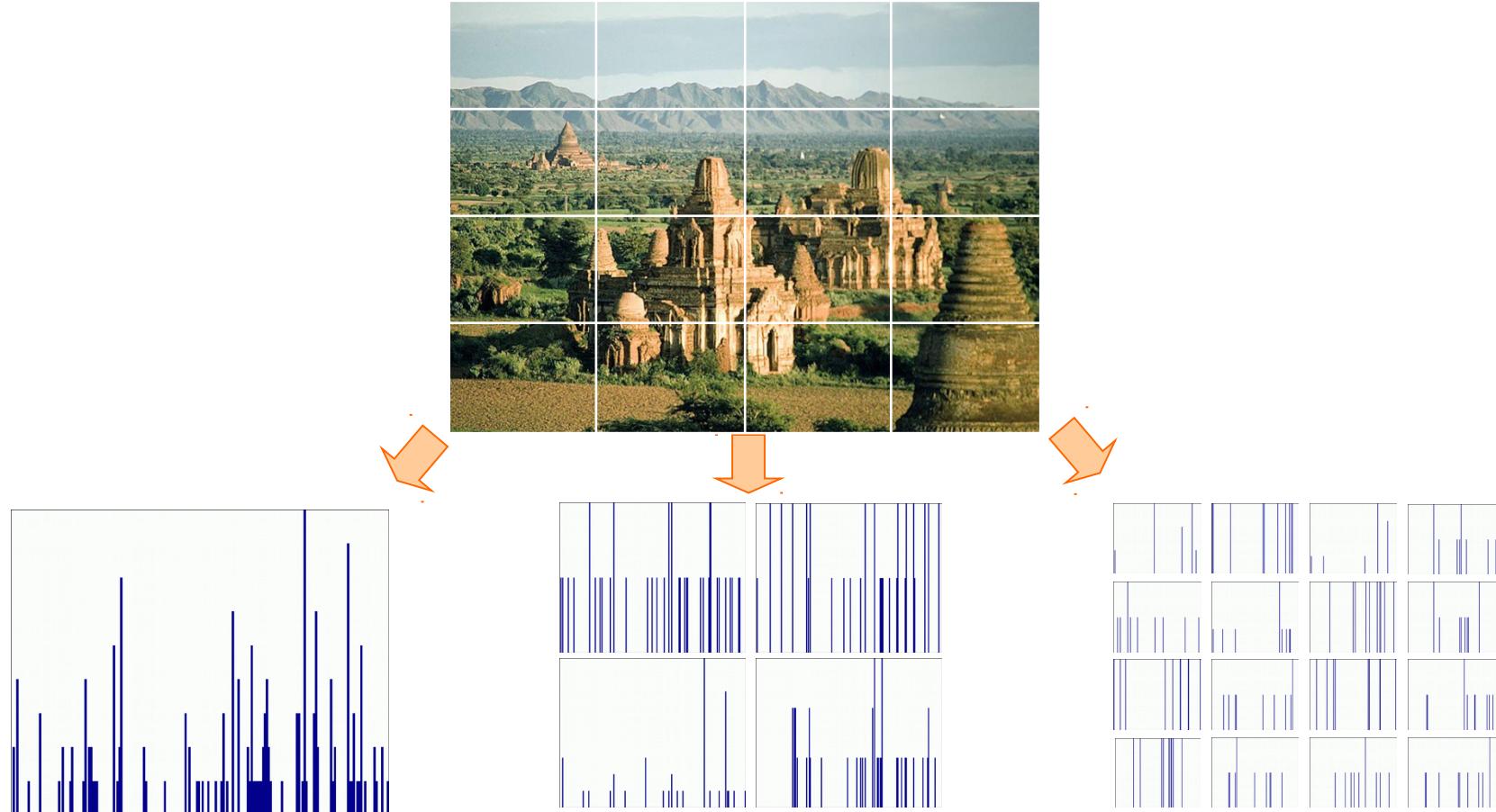
- Extension of a bag of features
- Locally orderless representation at several levels of resolution



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

