Object recognition

Jana Kosecka

Slides from D. Lowe, D. Forsythe and J. Ponce book, ICCV 2005 Tutorial Fei-Fei Li, Rob Fergus and A. Torralba, S. Lazebnik, T. Berg and many others

Dictionary.com

object | Search

Dictionary | Thesaurus | Encyclopedia | Web

ob·ject P Pr ion Key (ŏb′jĭkt, -jĕkt′)

n.

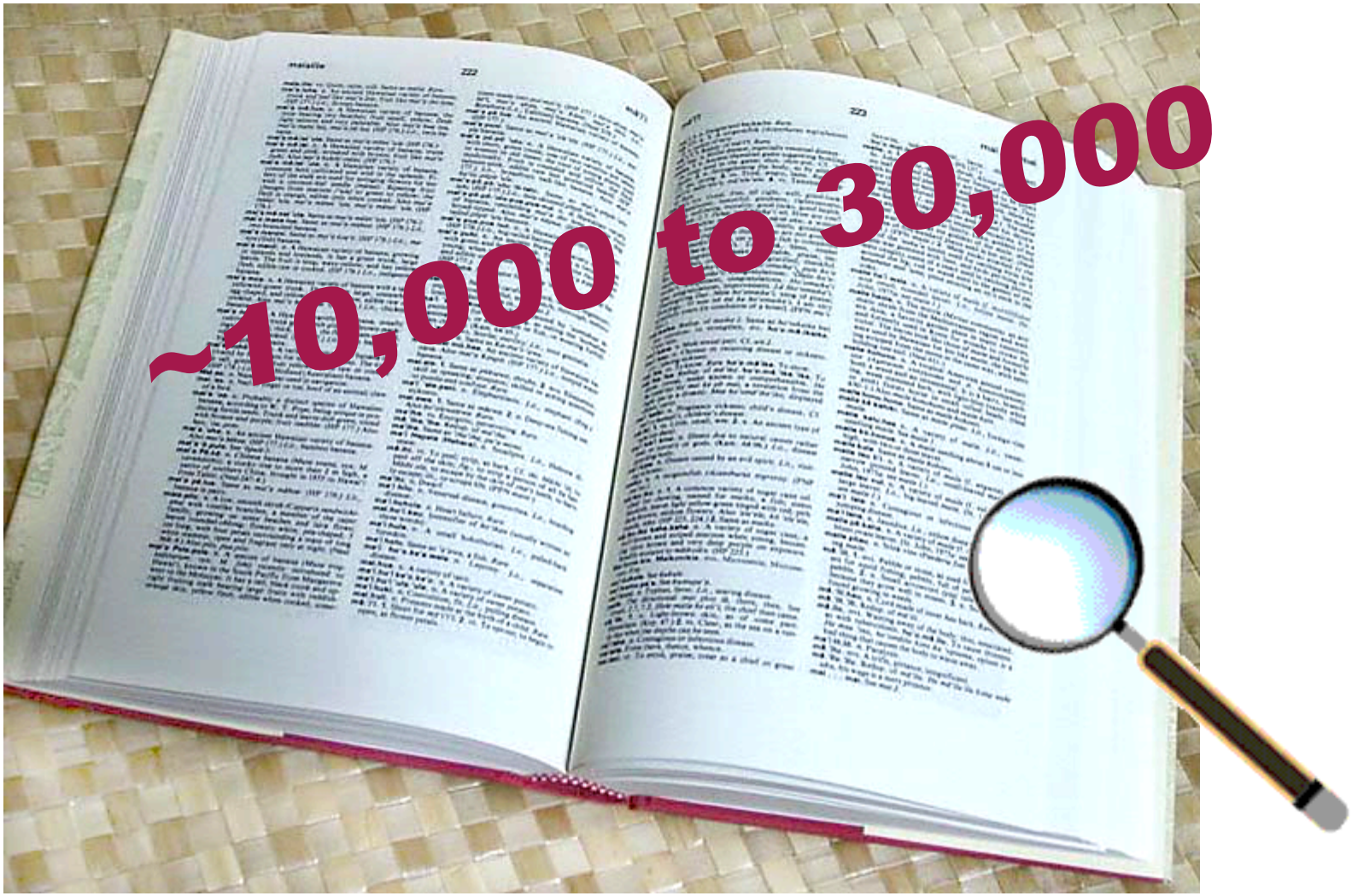
1. Somethi ne or more of the senses, especia or touch; a
2. A focus g, thought, or action: *an object of co*
3. The purpos of a specific action or effort: *the object game.*
4. *Grammar.*
   a. A noun, pronoun, oun phrase that receives or is affected by the ion of a ve within a sentence.
   b. A noun or substantive verned by a preposition.
5. *Philosophy.* Something inte ible or perceptible by the mind.
6. *Computer Science.* A discrete item that can be selected and maneuvered, such as an onscreen graphic. In object-oriented programming, objects include data and the procedures necessary to operate on that data.

perceptible

vision

material thing

How many object categories are there?



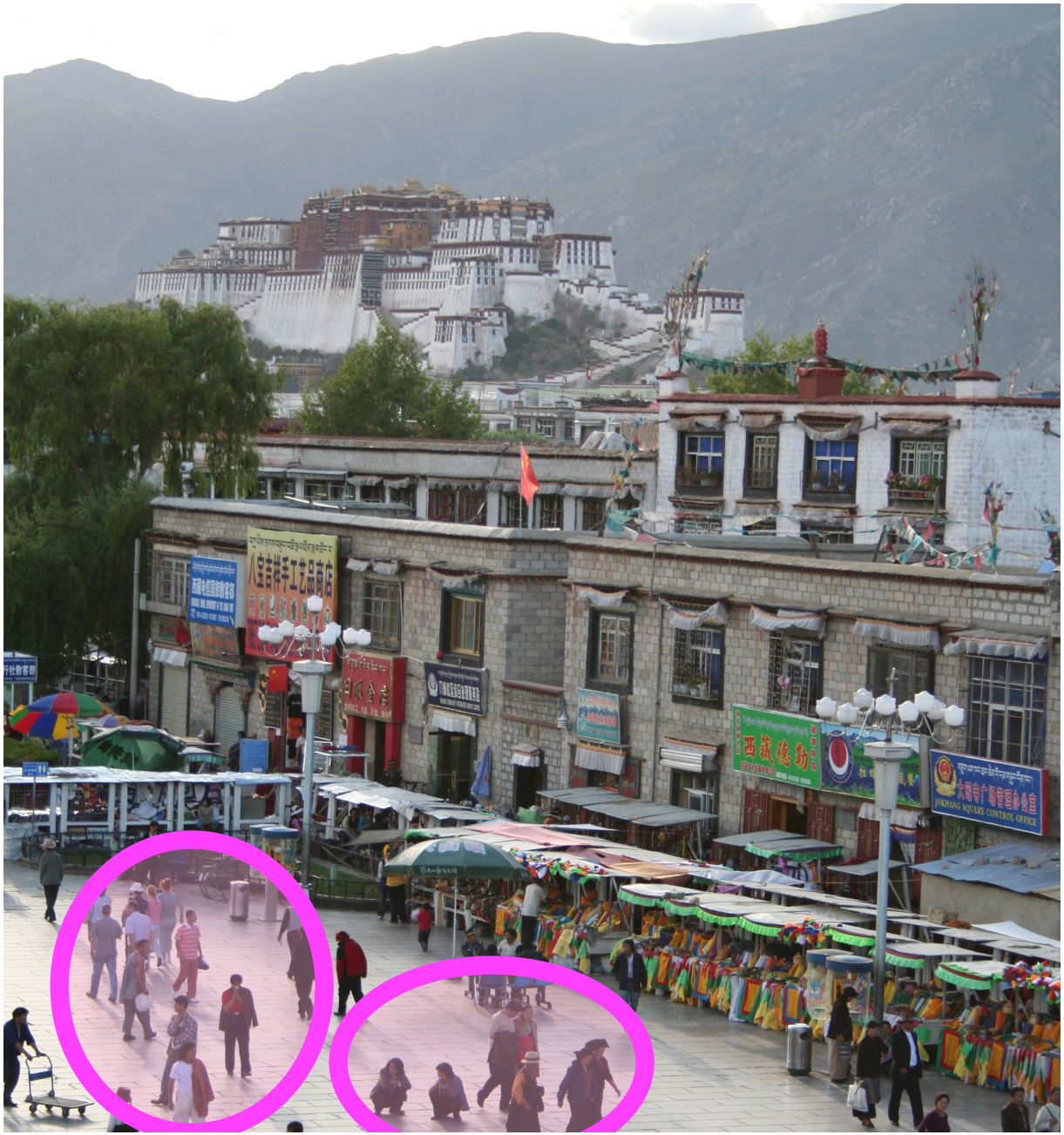
source: Svetlana Lazebnik Biederman 1987


~10,000 to 30,000

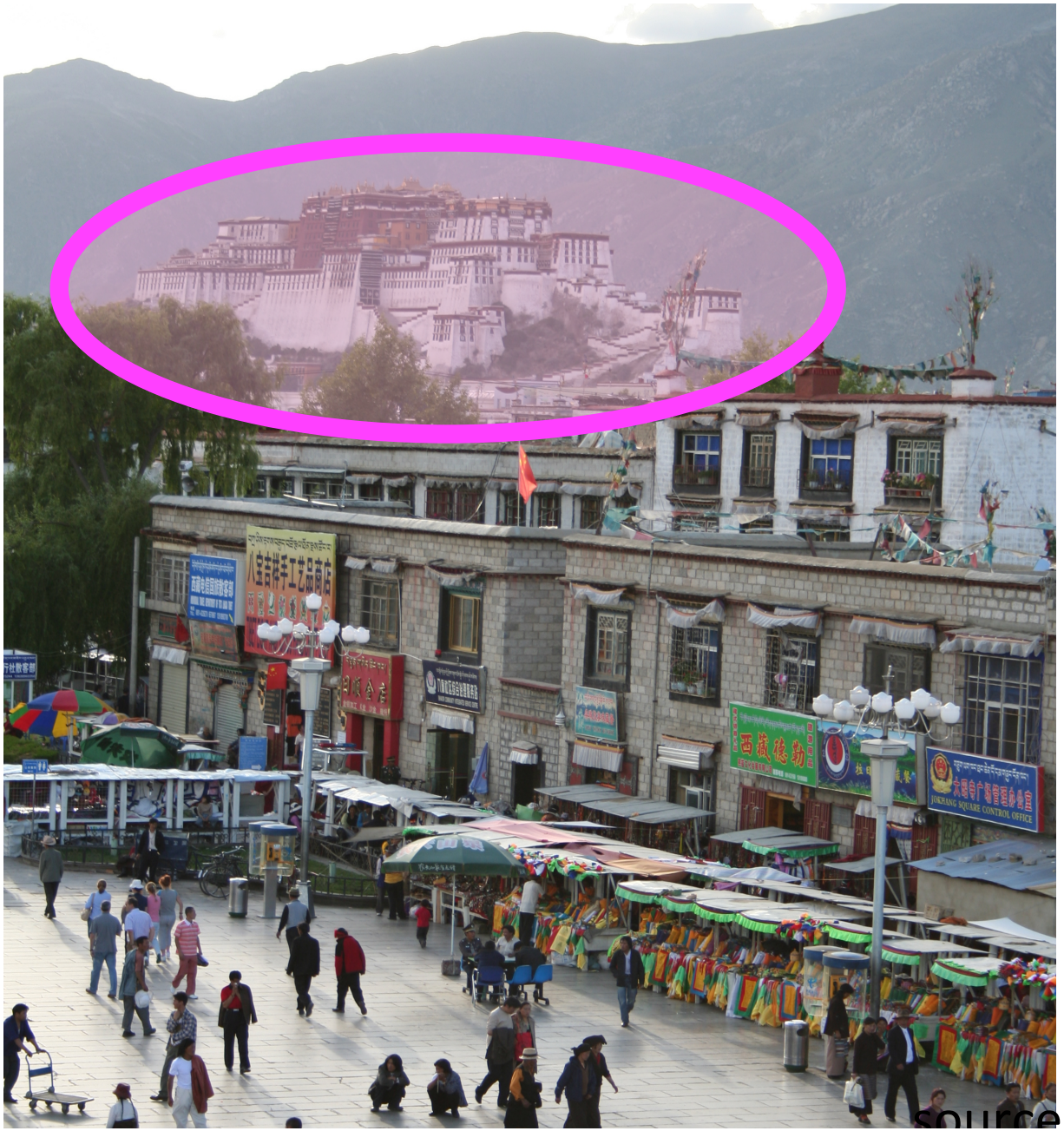
So what does object recognition involve?

Verification: is that a lamp?

Detection: are there people?

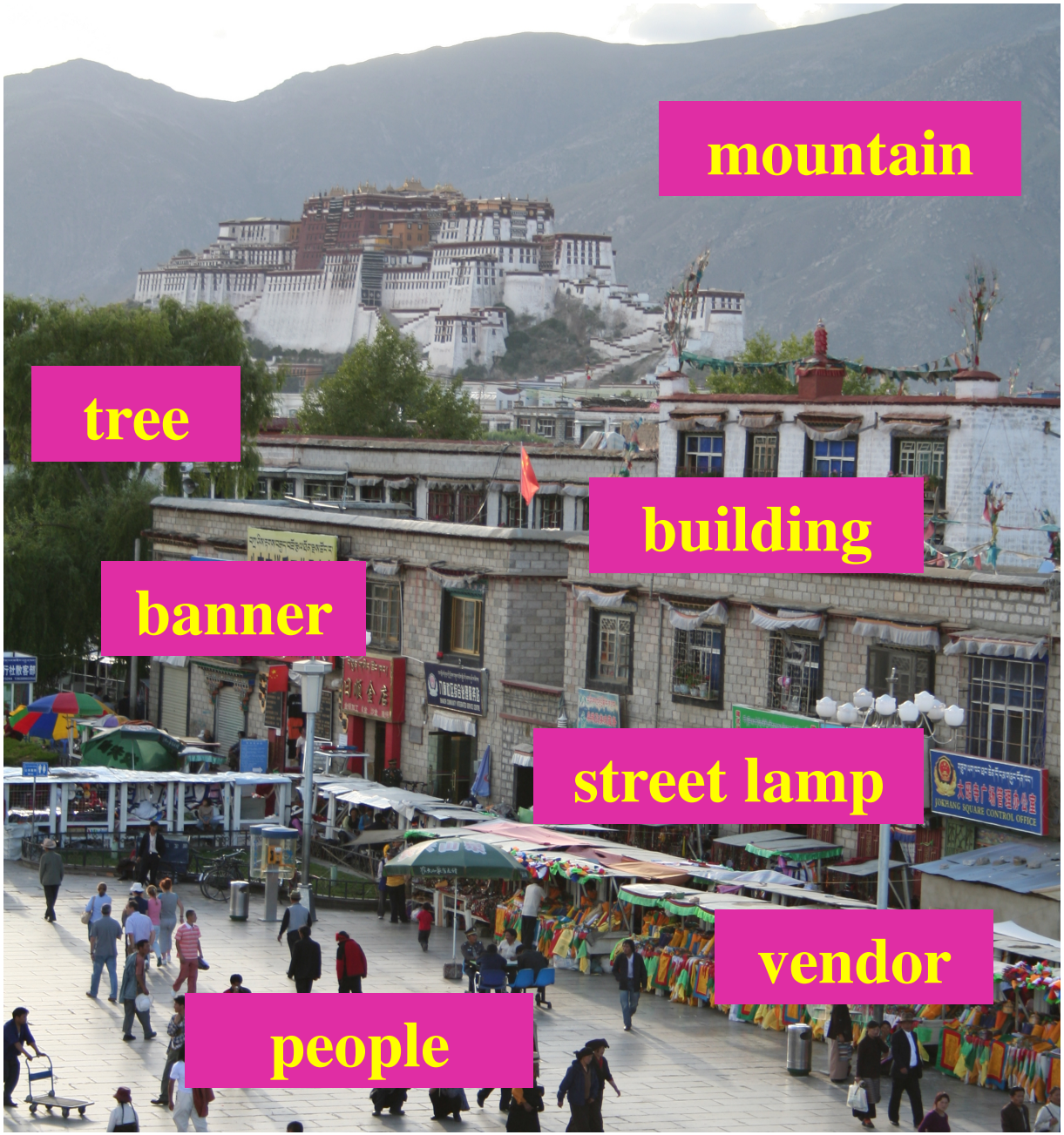


Identification: is that Potala Palace?

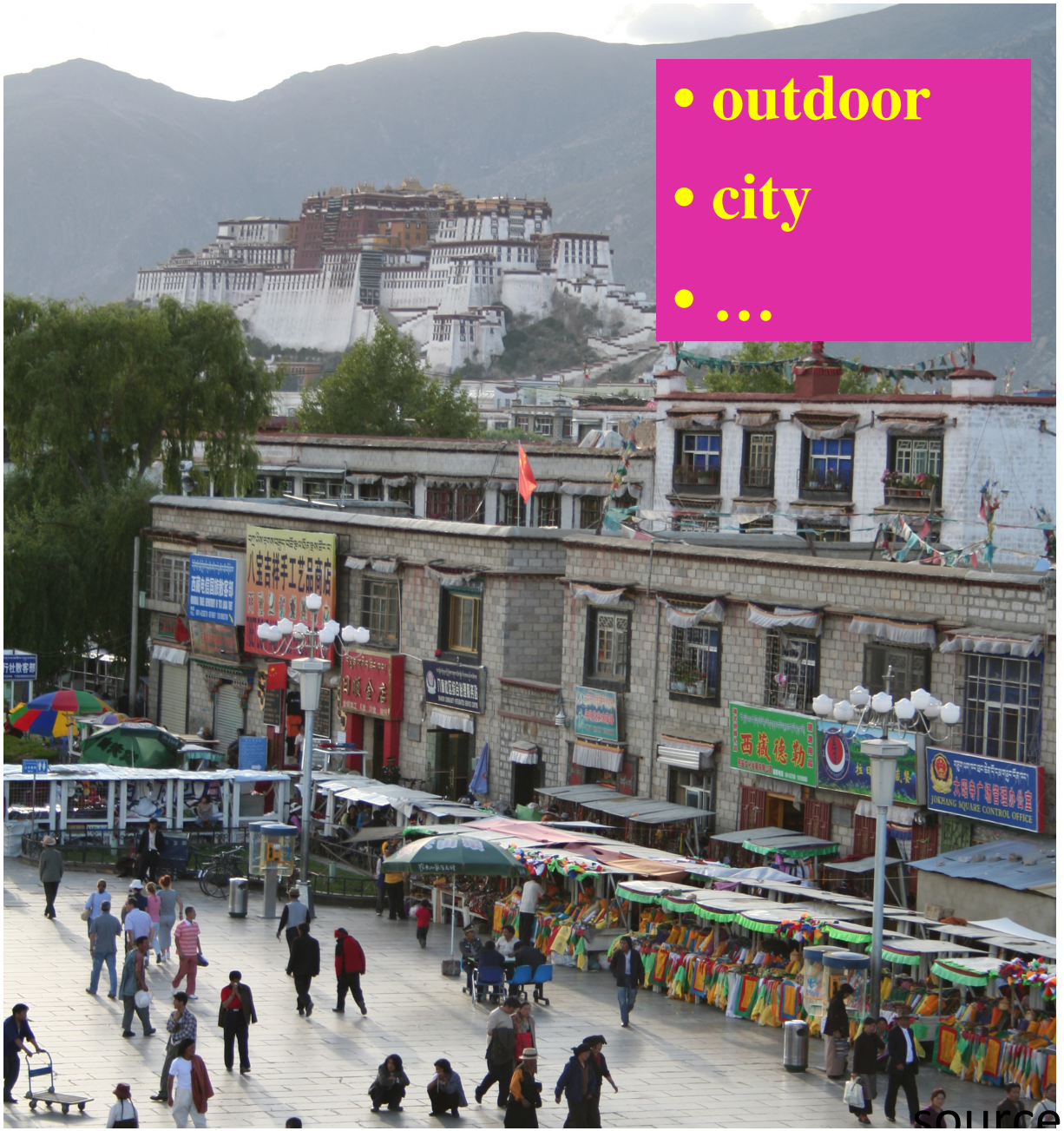


source: Svetlana La

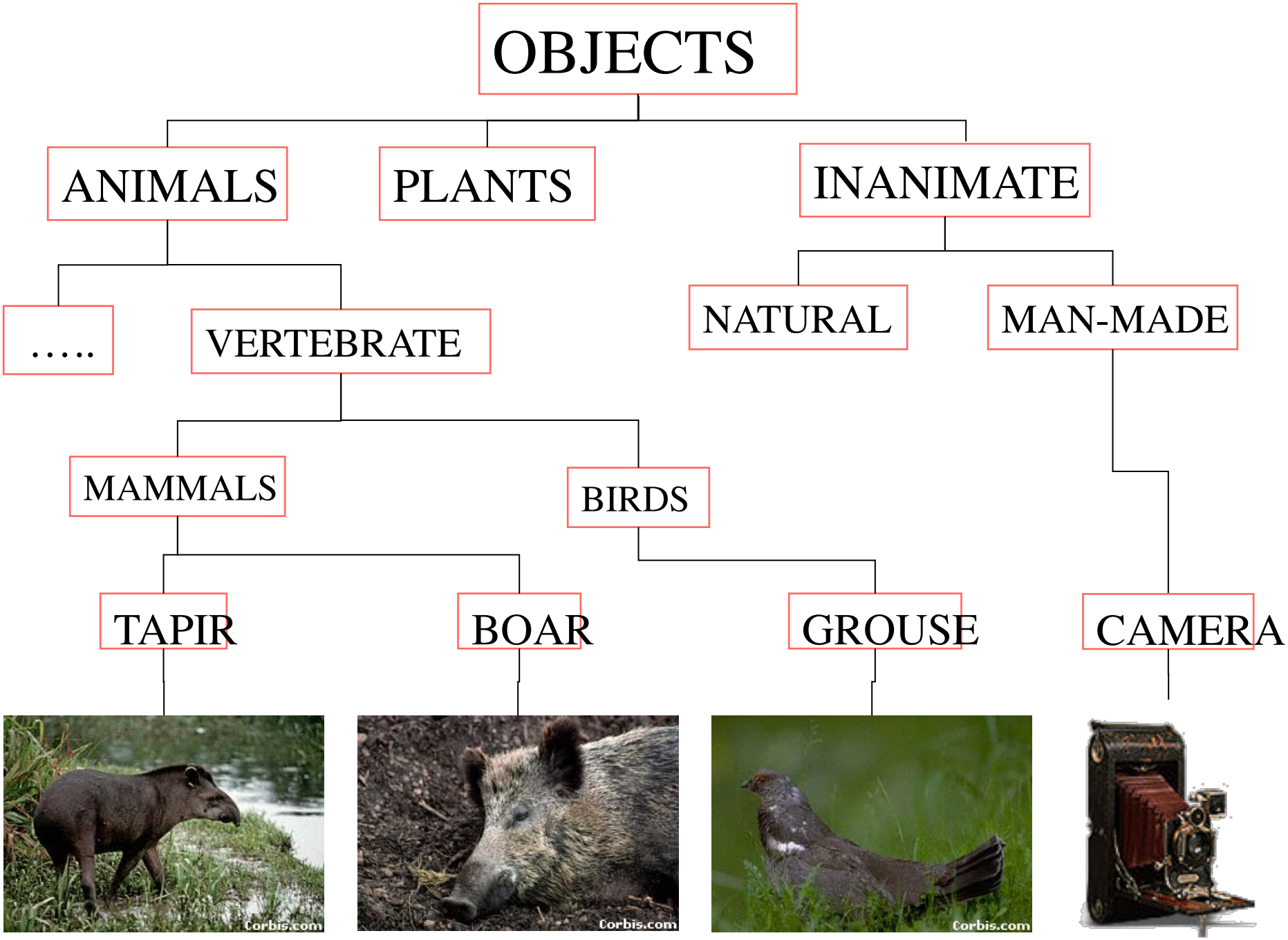
Object categorization



Scene and context categorization



source: Svetlana La


OBJECTS
ANIMALS
PLANTS
INANIMATE
…..
VERTEBRATE
NATURAL
MAN-MADE
MAMMALS
BIRDS
TAPIR
BOAR
GROUSE
CAMERA
Corbis.com
Corbis.com
Corbis.com

Within-class variations






Challenges 1: view point variation

Michelangelo 1475-1564

Challenges 2: illumination

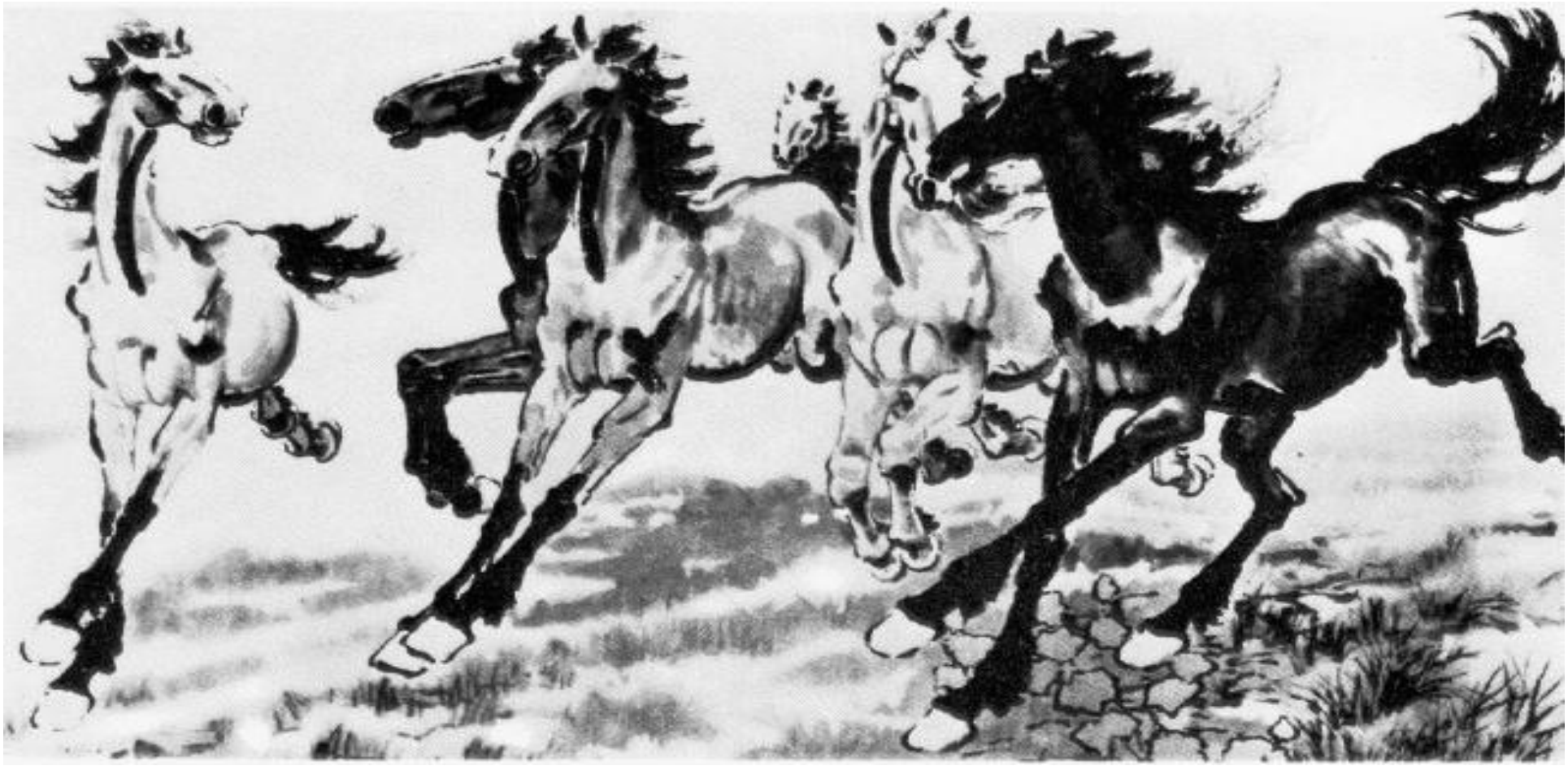
slide credit: S. Ullman

Challenges 3: occlusion

Magritte, 1957

Challenges 4: scale

Challenges 5: deformation

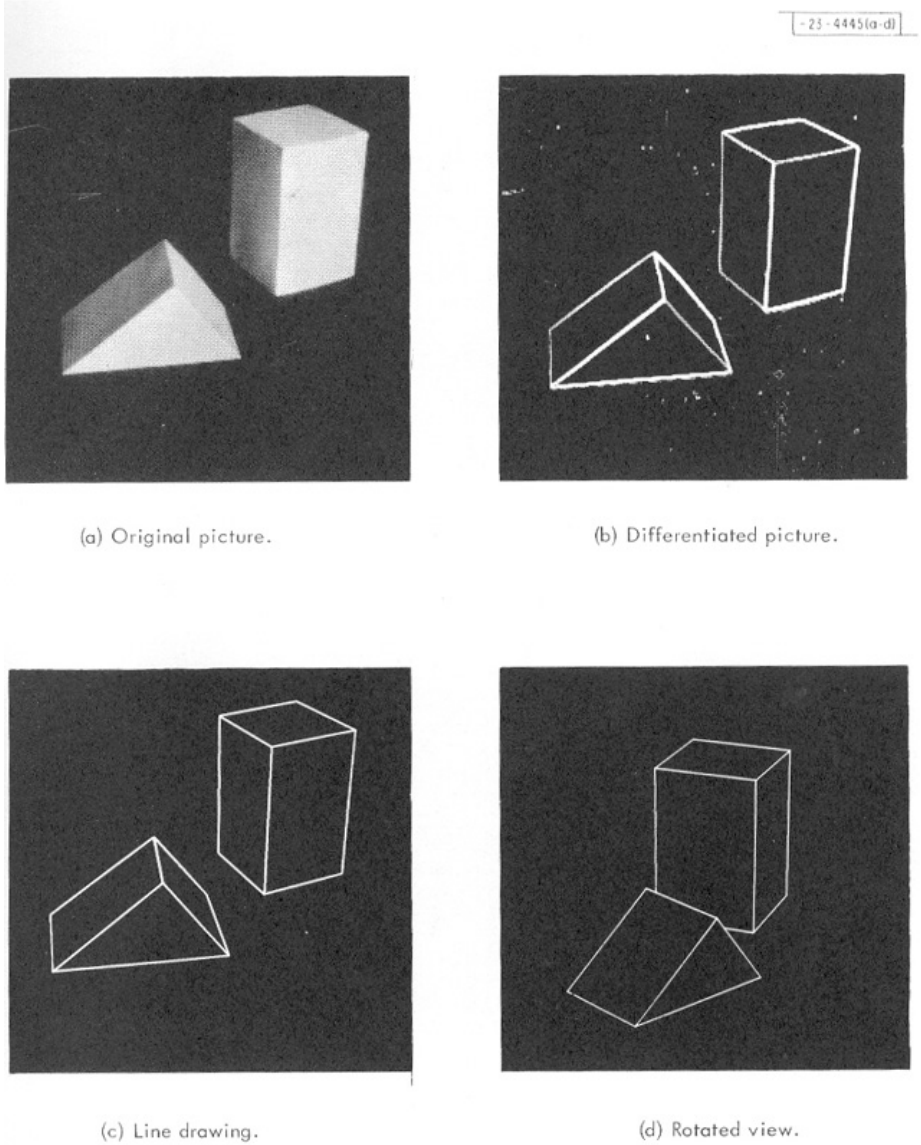


Xu, Beihong 1943

Challenges 6: background clutter

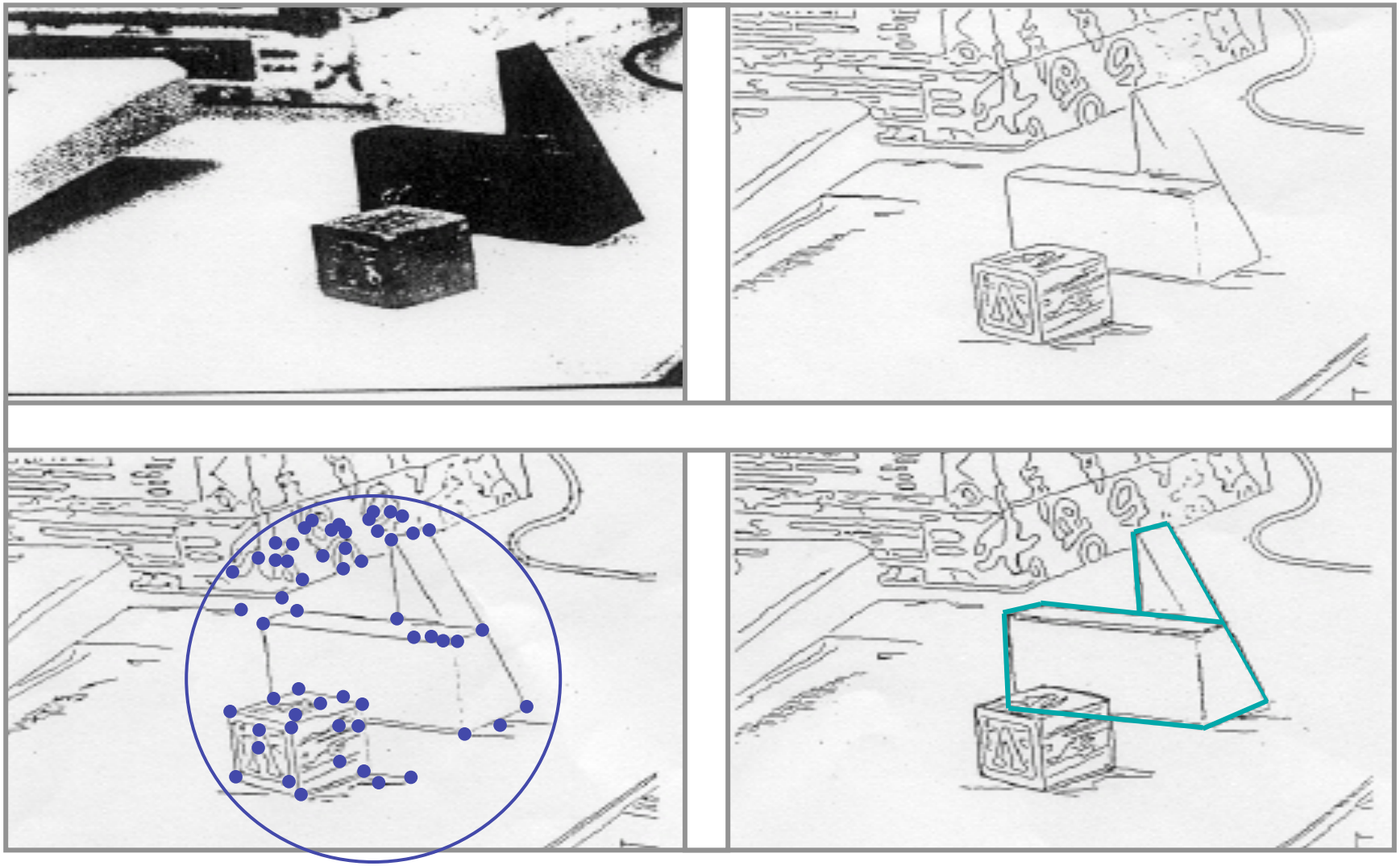
Klimt, 1913

Recall: Origins of computer vision (and AI)



L. G. Roberts, *Machine Perception of Three Dimensional Solids*, Ph.D. thesis, MIT Department of Electrical Engineering, 1963.

Alignment: Huttenlocher & Ullman (1987)



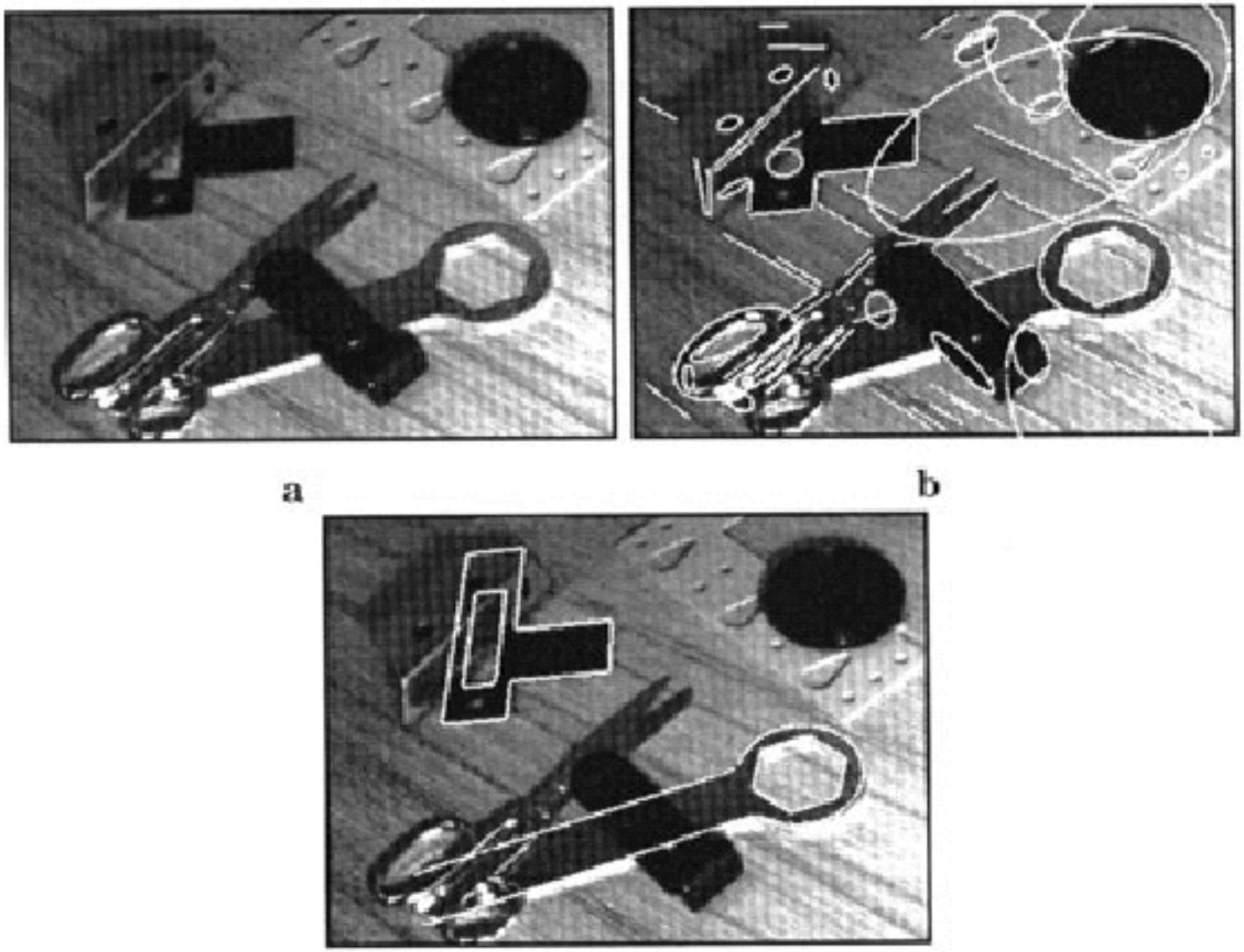
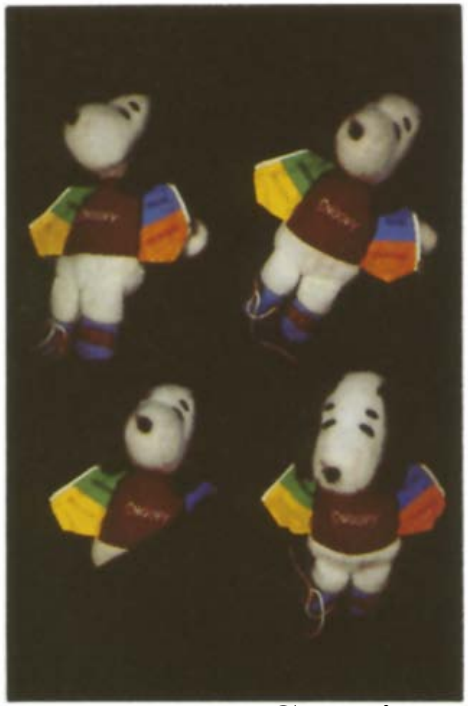
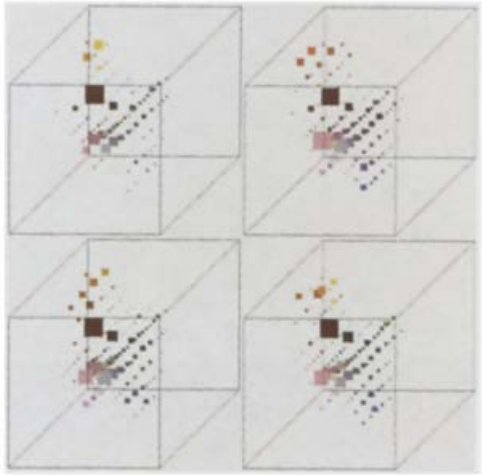


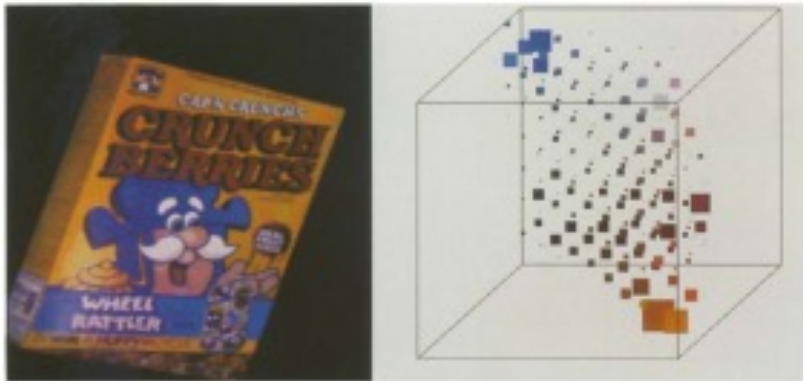
Figure from "Efficient model library access by projectively invariant indexing functions," by C.A. Rothwell et al., Proc. Computer Vision and Pattern Recognition, 1992, copyright 1992, IEEE - (courtesy Forsythe, Ponce CV, Prentice Hall)

Color Histograms

- Represent the objects by its color histogram
- In training and testing - objects on black background
- Some robustness wrt to clutter, partial occlusions, some view point change
- Not sufficiently discriminative (sometimes still used for large image based retrieval tasks)

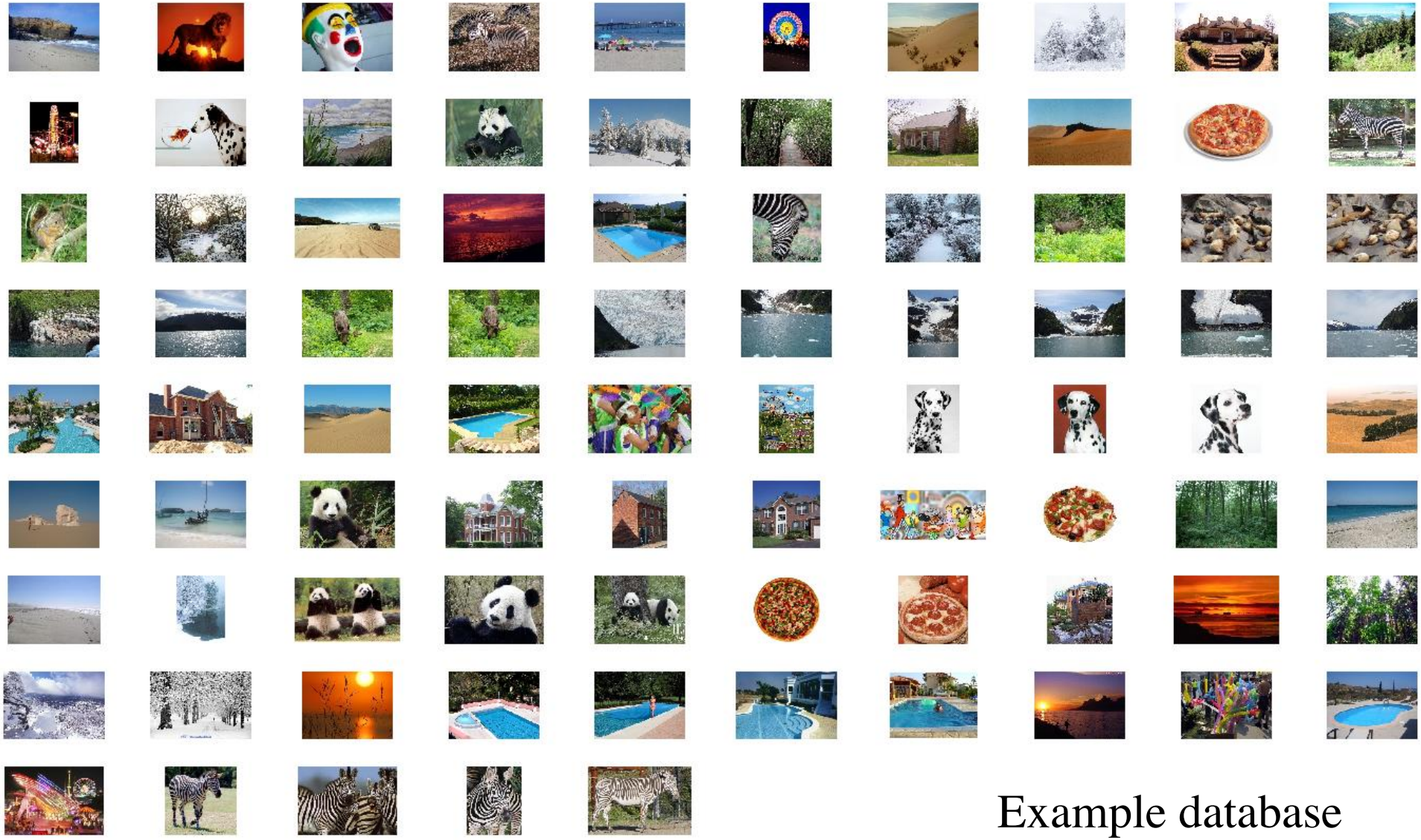






Swain and Ballard, Color Indexing, IJCV 1991.

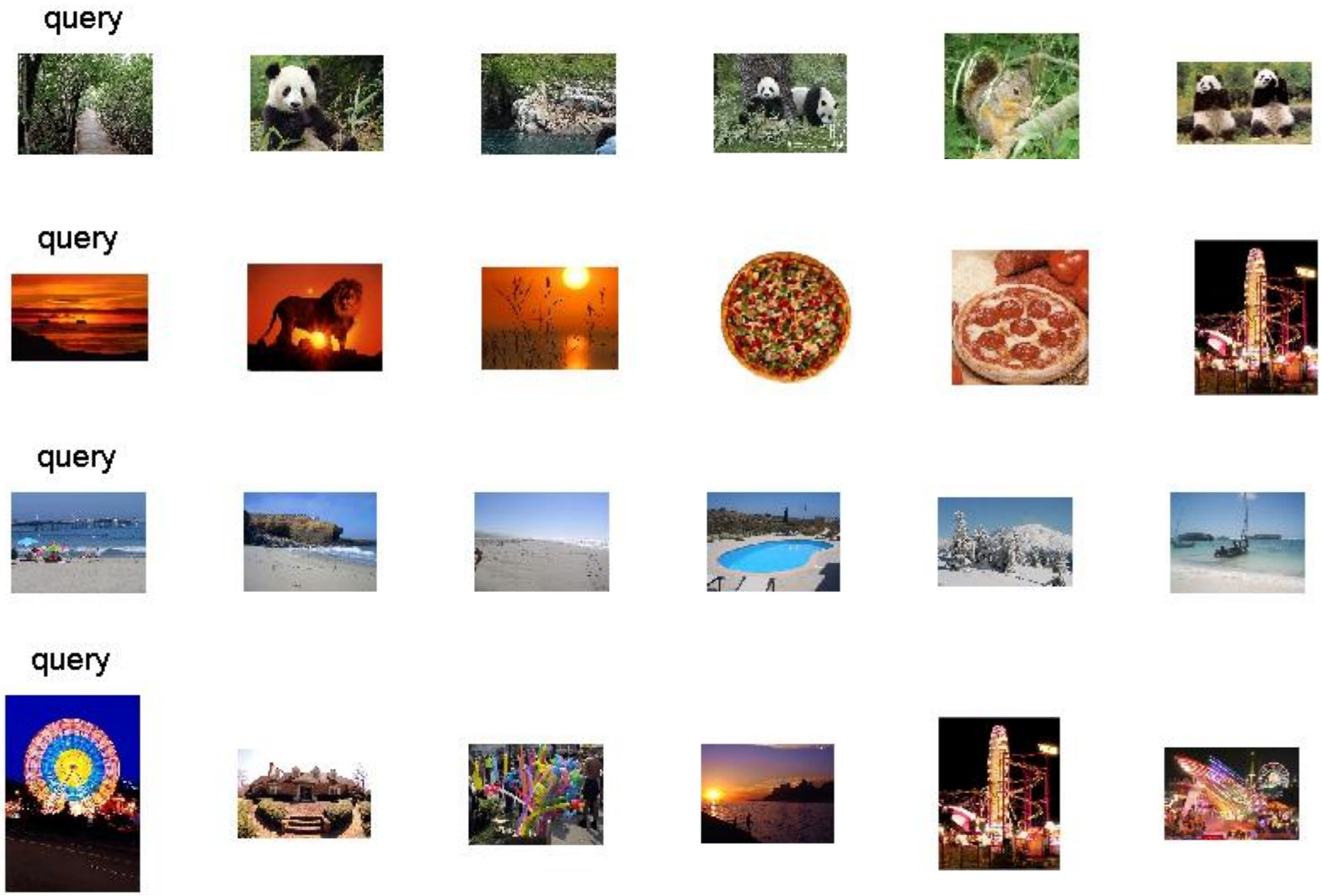
Color-based image retrieval



Example database

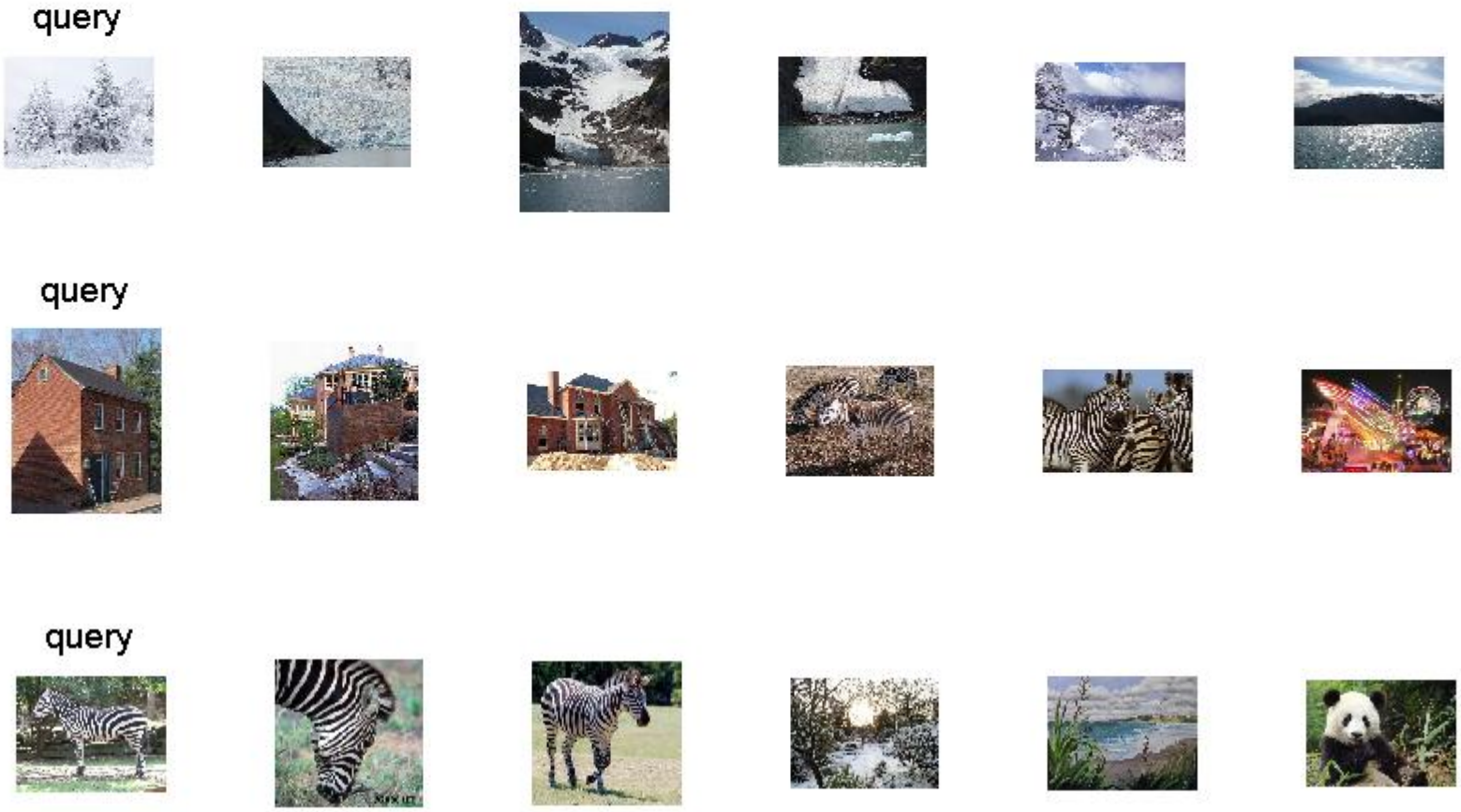
Kristen Grauman

Color-based image retrieval



Example retrievals

Color-based image retrieval



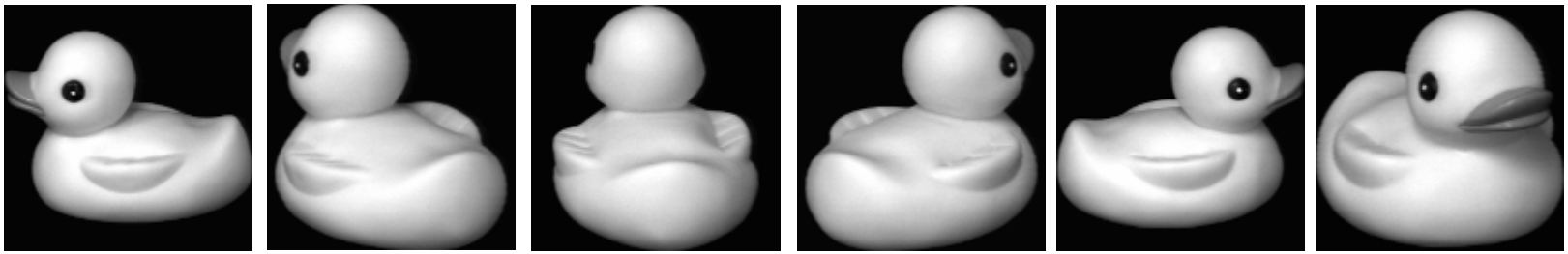
Example retrievals

See More: color and light lecture

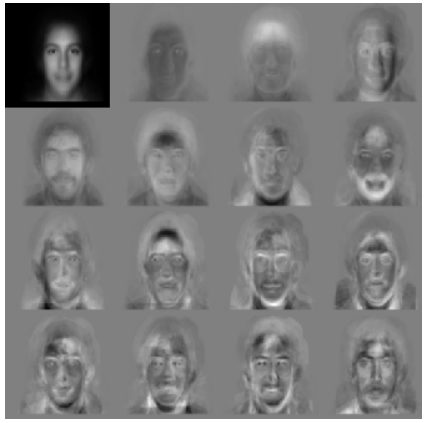
Kristen Grauman

Appearance based Object Recognition

- Holistic representation of images
- Instead of features, use the whole image



- PCA based techniques, Eigen-faces [Turk 91]
- Not robust with respect to occlusions, clutter changes of viewpoint
- Global representation
- No features



The space of all face images

- When viewed as vectors of pixel values, face images are extremely high-dimensional
  - 100x100 image = 10,000 dimensions
- However, relatively few 10,000-dimensional vectors correspond to valid face images
- We want to effectively model the subspace of face images

Principal Component Analysis -- PCA

(also called Karhunen-Loeve transformation)

- **PCA** transforms the original input space into a lower dimensional space, by constructing dimensions that are linear combinations of the given features;

- The objective is to consider independent dimensions along which data have largest variance (i.e., greatest variability);

Geometric view

- Given set of datapoints in D dimensional space – find some transformation which will transform the points to lower dimensional space. U_d is *D x d* matrix with *d* orthonormal column vectors

$$x_i = x_0 + U_d y_i$$

- y_i are the new coordinates of x_i in *d*-dimensional space
- Derivation on the board – see handout for more details

Statistical view

- Given multivariate random variable *x* and set of sample points x_i, find *d* uncorrelated linear components of *x* such that variance of the components is maximized

$$y_i = u_i^T x$$

- Such that

$$u_i^T u_i = 1 \quad and \quad Var(y_1) \geq Var(y_2) \cdots$$

- Derivation on the board

Principal Component Analysis -- PCA

- **PCA** enables transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components;

- The first principal component accounts for as much of the variability in the data as possible;

- Each succeeding component (orthogonal to the previous ones) accounts for as much of the remaining variability as possible.

Principal Component Analysis

- PCA is the most commonly used dimension reduction technique.
- (Also called the Karhunen-Loeve transform).
- Data samples $x_1, \cdots, x_N$
- Compute the mean $\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$
- Computer the covariance:

$$\Sigma_x = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})^T$$

Principal Component Analysis

- Compute the eigenvalues λ and eigenvectors e of the matrix Σ_x
- Solve $\Sigma_x x = \lambda x$
- Order them by magnitude:

$$\lambda_1 \geq \lambda_2 \geq .\lambda_N.$$

- PCA reduces the dimension by keeping direction e such that $\lambda < T$.

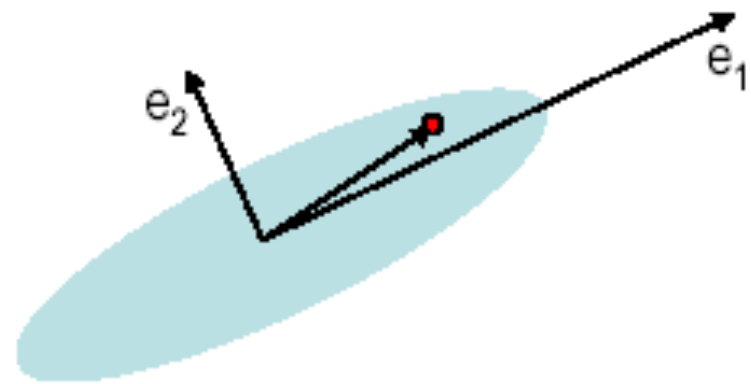
Principal Component Analysis

- For many datasets, most of the eigenvalues are negligible and can be discarded.

The eigenvalue λ measures the variation
In the direction of corresponding eigenvector

Example:

$$\lambda_1 \neq 0, \lambda_2 = 0.$$



Principal Component Analysis

- How to get uncorrelated components which Capture most of the variance
- Project the data onto the selected eigenvectors:

$$y_i = e_i^T (x_i - \bar{x})$$

- If we consider first M eigenvectors we get new lower dimensional representation

$$[y_1, \cdots, y_M]$$

- Proportion covered by first M eigenvalues

$$\frac{\sum_{i=1}^{M} \lambda_i}{\sum_{i=1}^{N} \lambda_i}$$

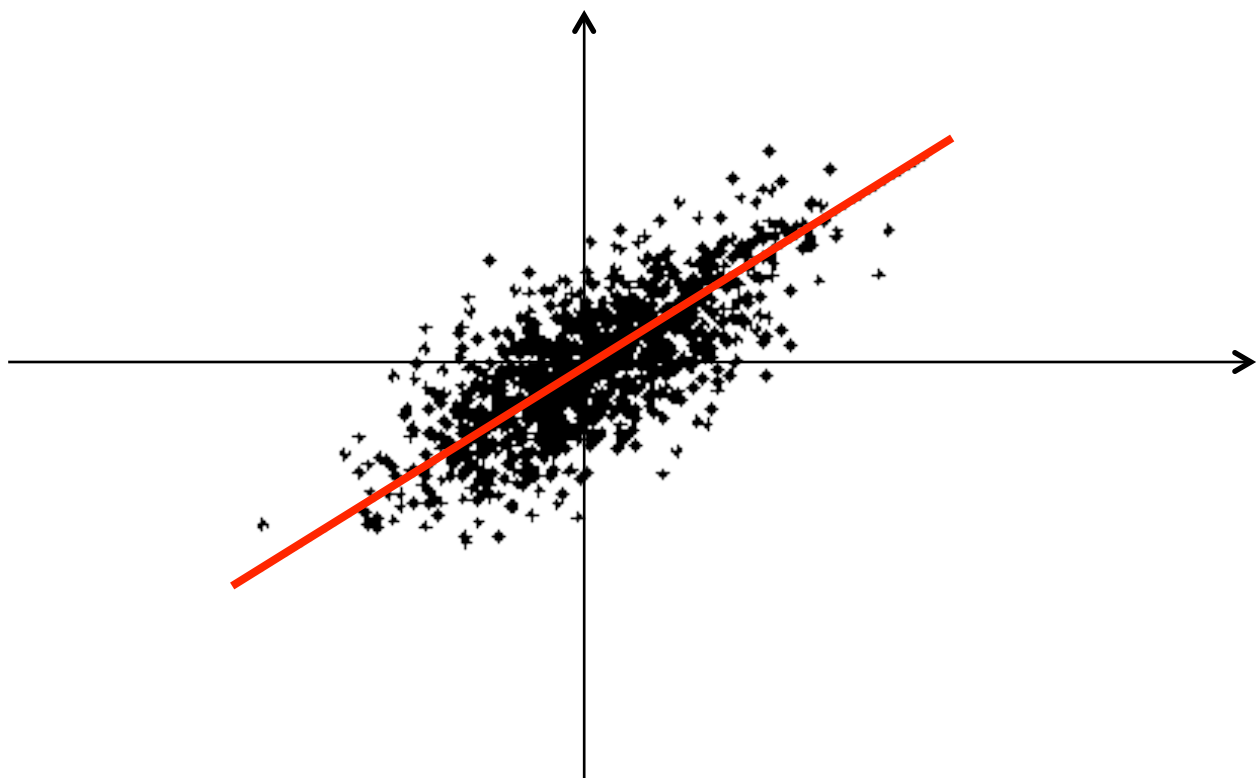
PCA Example

- The images of an object under different lighting lie in a low-dimensional space.
- The original images are 256x 256. But the data lies mostly in 3-5 dimensions.
- First we show the PCA for a face under a range of lighting conditions. The PCA components have simple interpretations.
- Then we plot $\frac{\sum_{i=1}^{M} \lambda_i}{\sum_{i=1}^{N} \lambda_i}$ as a function of M for several objects under a range of lighting.

Limitations of PCA

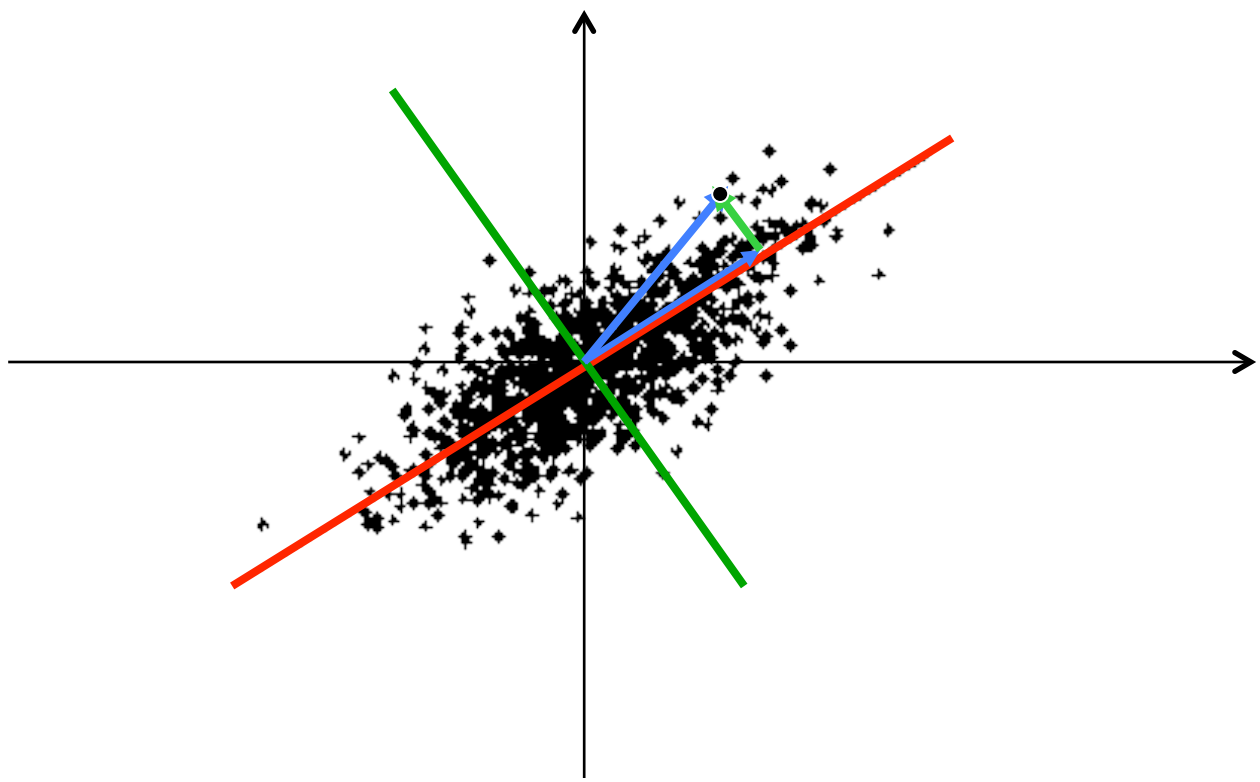
- PCA is not effective for some datasets.
- For example, if the data is a set of strings
- (1,0,0,0,…), (0,1,0,0…),…,(0,0,0,…,1) then the eigenvalues do not fall off as PCA requires.

Simple illustration of PCA



First principal component of a two-dimensional data set.

Simple illustration of PCA



Second principal component of a two-dimensional data set.

Determining the number of components

- Plot the eigenvalues – each eigenvalue is related to the amount of variation explained by the corresponding axis (eigenvector);
- If the points on the graph tend to level out (show an “elbow” shape), these eigenvalues are usually close enough to zero that they can be ignored.

- In general: Limit the variance accounted for.

Critical information lies in low dimensional subspaces

- A typical eigenvalue spectrum and its division into two orthogonal subspaces

Applications

- Need to analyze large amounts multivariate data.
  - Human Faces.
  - Speech Waveforms.
  - Global Climate patterns.
  - Gene Distributions.

- Difficult to visualize data in dimensions just greater than three.

- Discover compact representations of high dimensional data.
  - Visualization.
  - Compression.
  - Better Recognition.
  - Probably meaningful dimensions.

Eigenfaces: Key idea

- Assume that most face images lie on a low-dimensional subspace determined by the first k ($k<d$) directions of maximum variance

- Use PCA to determine the vectors or "eigenfaces" $\mathbf{u}_1,\ldots \mathbf{u}_k$ that span that subspace

- Represent all face images in the dataset as linear combinations of eigenfaces

M. Turk and A. Pentland, Face Recognition using Eigenfaces, CVPR 1991

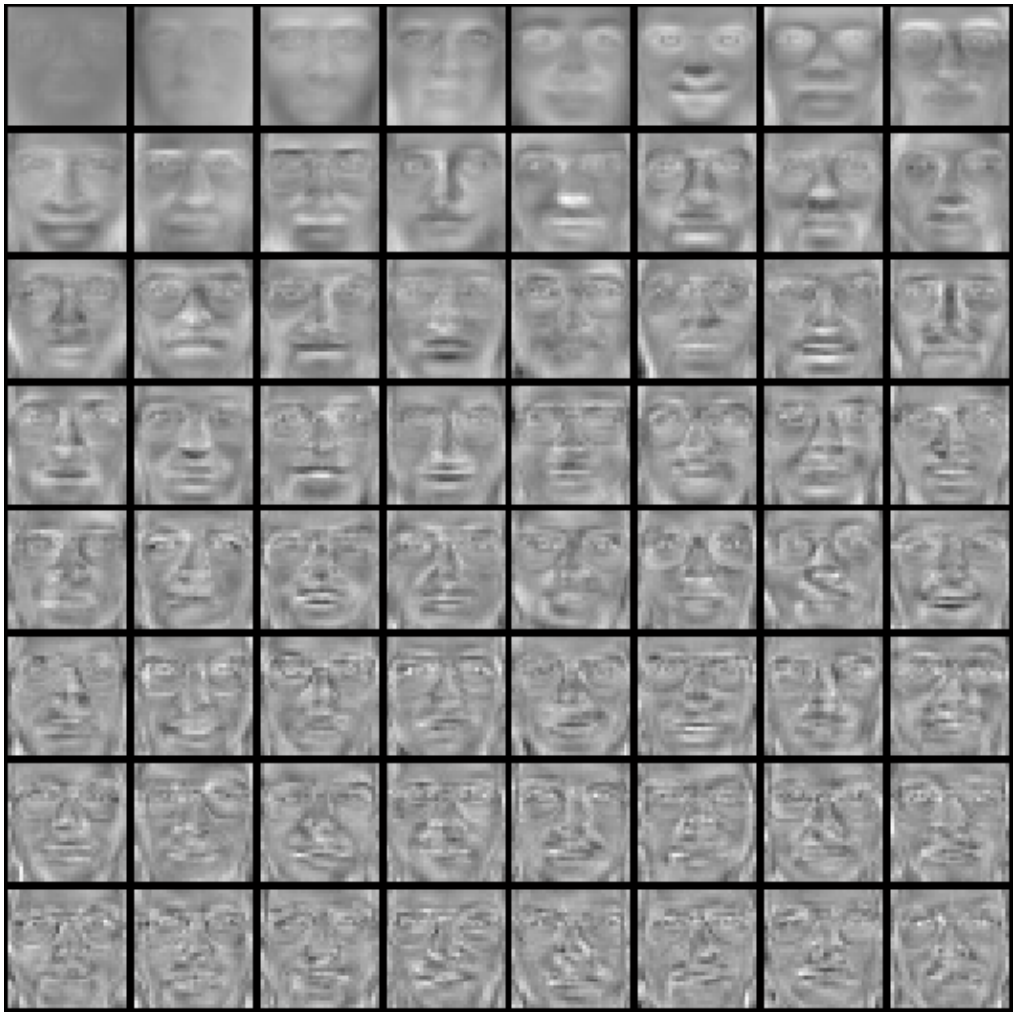
Eigenfaces example

- Training images
- $\mathbf{x}_1, \ldots, \mathbf{x}_N$

Eigenfaces example

Mean: μ

Top eigenvectors: $u_1, \ldots u_k$



Eigenfaces example

Principal component (eigenvector) u_k

$\mu + 3\sigma_k u_k$

$\mu - 3\sigma_k u_k$

Eigenfaces example

- Representation

$$(w_{i1},\ldots,w_{ik}) = (\mathbf{u}_1^T(\mathbf{x}_i - \boldsymbol{\mu}), \ldots, \mathbf{u}_k^T(\mathbf{x}_i - \boldsymbol{\mu}))$$

- Reconstruction

$$\hat{x} = \mu + w_1u_1 + w_2u_2 + w_3u_3 + w_4u_4 + \ldots$$

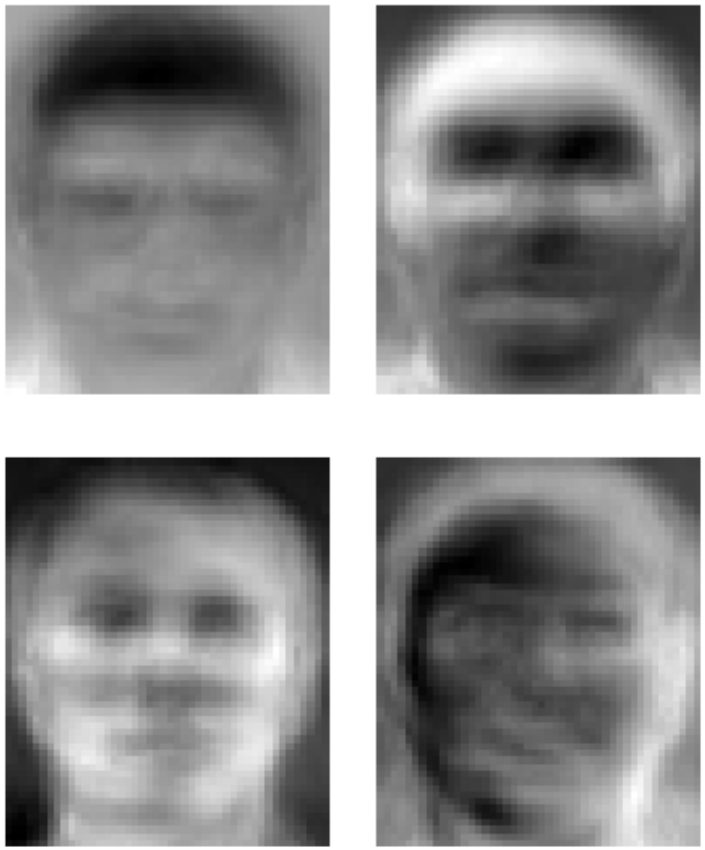
Recognition with eigenfaces

- Process labeled training images:
- Find mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$
- Find k principal components (eigenvectors of $\boldsymbol{\Sigma}$) $\mathbf{u}_1,\ldots\mathbf{u}_k$
- Project each training image $\mathbf{x}_i$ onto subspace spanned by principal components:
  $(w_{i1},\ldots,w_{ik}) = (\mathbf{u}_1^T(\mathbf{x}_i - \boldsymbol{\mu}), \ldots , \mathbf{u}_k^T(\mathbf{x}_i - \boldsymbol{\mu}))$

- Given novel image $\mathbf{x}$:
- Project onto subspace:
  $(w_1,\ldots,w_k) = (\mathbf{u}_1^T(\mathbf{x} - \boldsymbol{\mu}), \ldots , \mathbf{u}_k^T(\mathbf{x} - \boldsymbol{\mu}))$
- Optional: check reconstruction error $\mathbf{x} - \hat{\mathbf{x}}$ to determine whether image is really a face
- Classify as closest training face in k-dimensional subspace

M. Turk and A. Pentland, Face Recognition using Eigenfaces, CVPR 1991

Eigenfaces (Turk & Pentland, 1991)



Experimental	Correct/Unknown Recognition Percentage		
Condition	Lighting	Orientation	Scale
Forced classification	96/0	85/0	64/0
Forced 100% accuracy	100/19	100/39	100/60
Forced 20% unknown rate	100/20	94/20	74/20

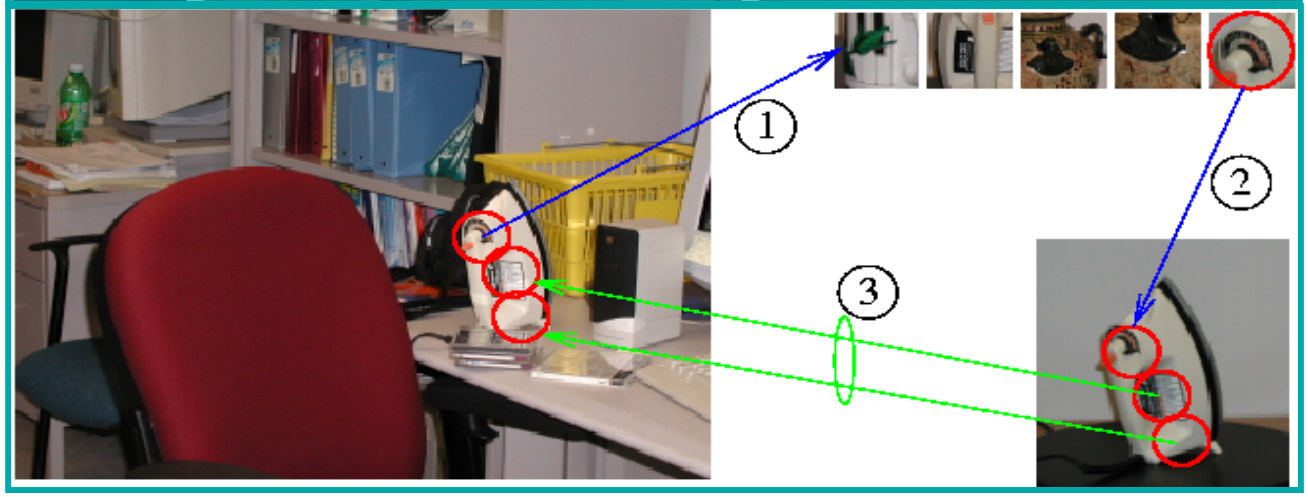
Limitations

Requires good alignment, no background, no outliers and/or partial occlusions

Local Features

Combining *local* appearance, spatial constraints, invariants, and classification techniques from machine learning.

Schmid & Mohr' 97

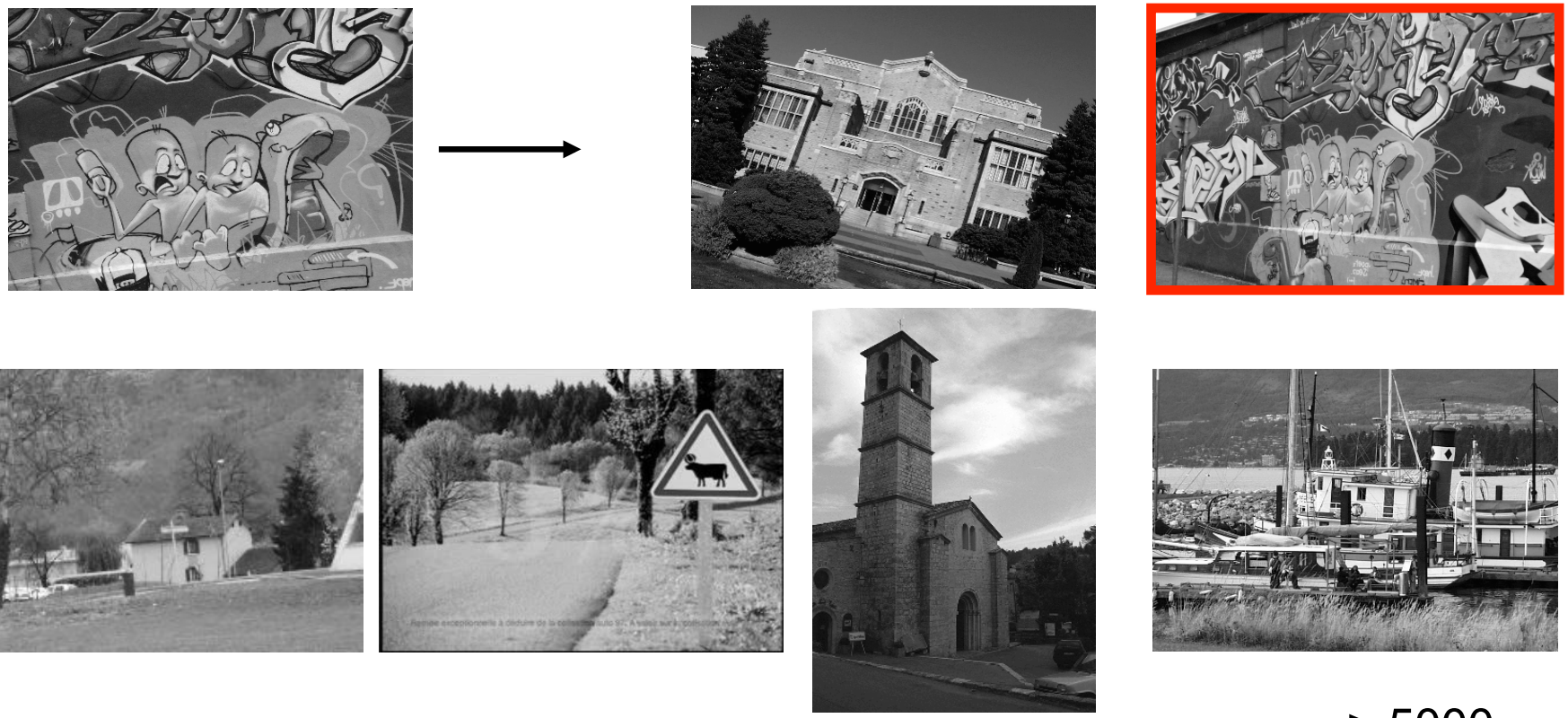


Mahamud & Hebert' 03

Recognition using local features

- Define a set of local feature templates
  - could find these with filters, etc., SIFT features
  - corner detector+filters, feature detectors and descriptors
- Each feature votes for all patterns that contain it
- Pattern with the most votes wins

Recognition with local features



> 5000 images

[Local greyvalue invariants for image retrieval,
C. Schmid and R. Mohr, PAMI 1997]

Semi-local constraints, neighboring points should match, angles, length ratios should be similar

Recognition with local features

- For each feature in the query
- Find the nearest neighbour in the database of features
- Check which object/model it belongs to
- That object/model gets one vote
- Repeat for all detected features in the query
- Model with largest number of votes wins

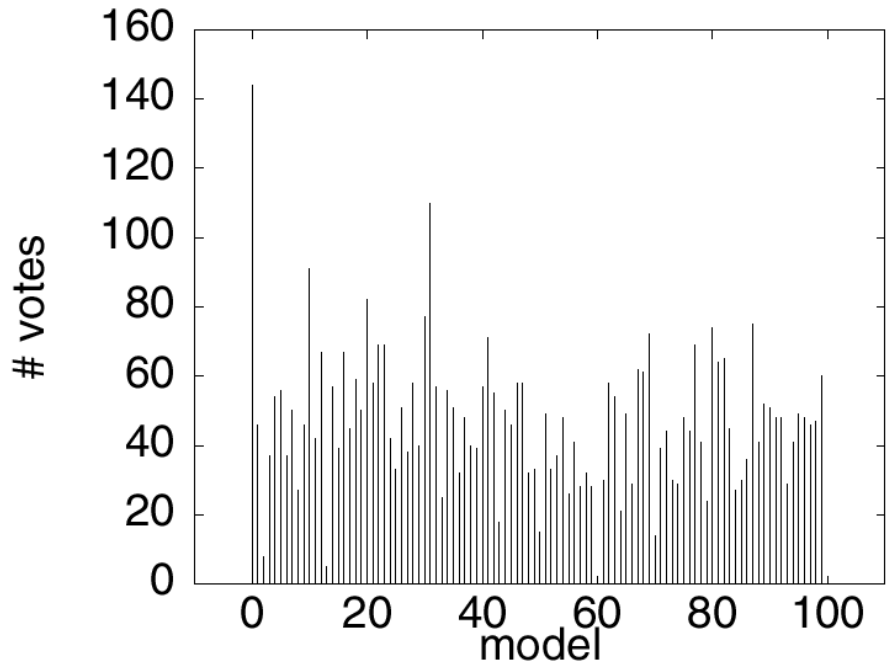


Figure from "Local grayvalue invariants for image retrieval," by C. Schmid and R. Mohr, IEEE Trans. Pattern Analysis and Machine Intelligence, 1997 copyright 1997, IEEE

Employ spatial relations

Instead of matching each feature separately, match the feature and its spatial neighbours to and verify that the match and its neighbours have similar spatial relationship

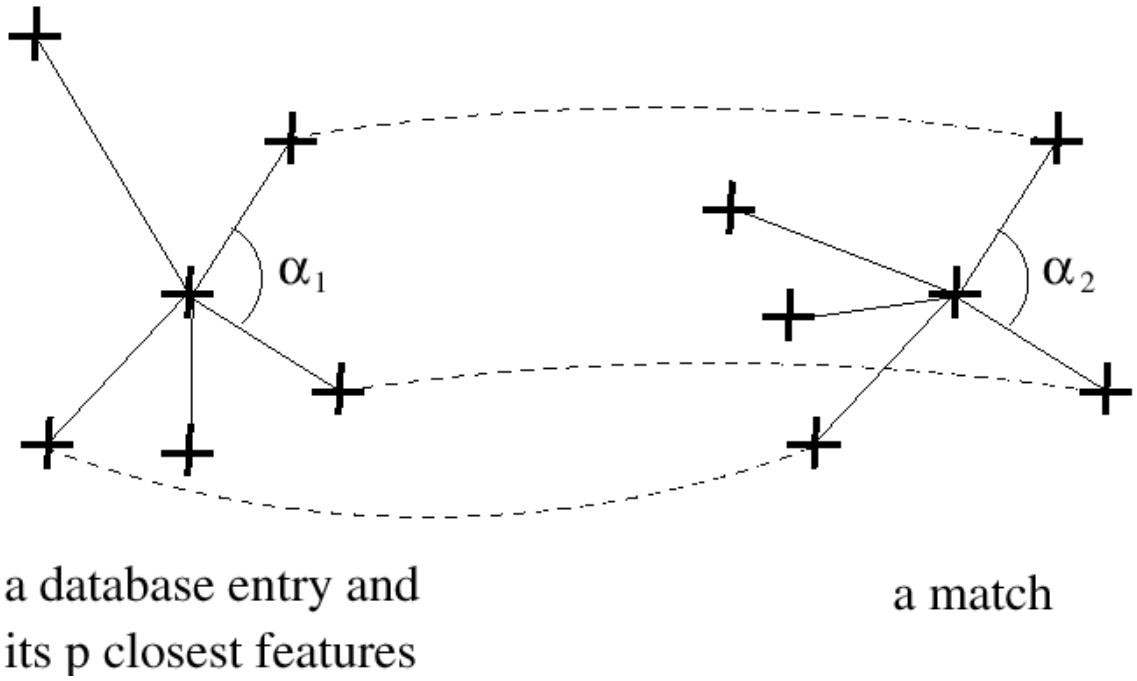
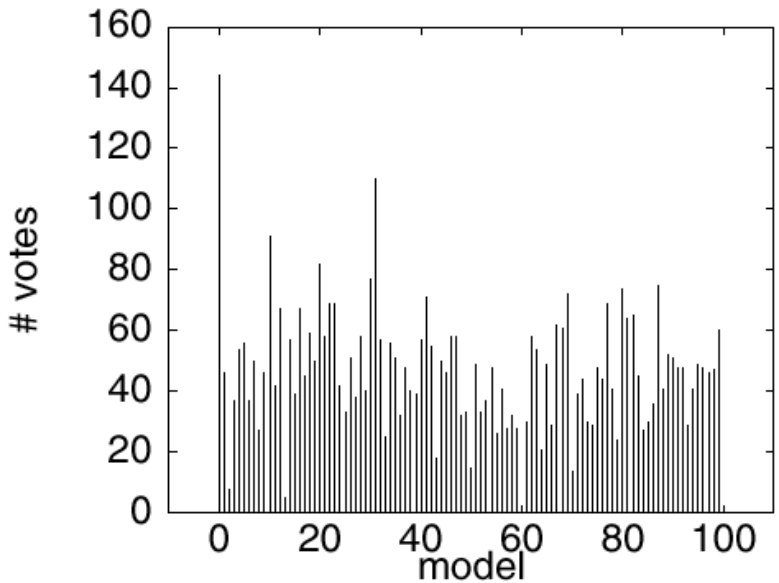


Figure from "Local grayvalue invariants for image retrieval," by C. Schmid and R. Mohr, IEEE Trans. Pattern Analysis and Machine Intelligence, 1997 copyright 1997, IEEE

Recognition with local features

- Improved performance, with spatial relationships

- Baseline method to image based retrieval using local features



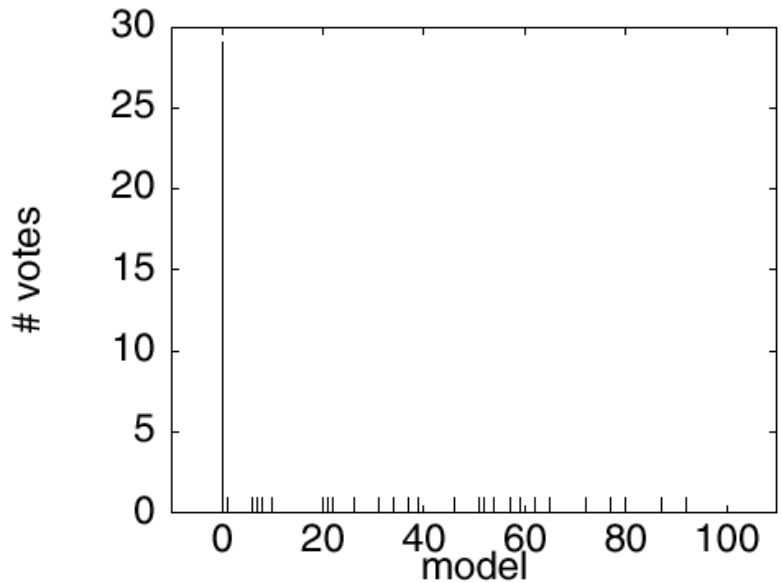


Figure from "Local grayvalue invariants for image retrieval," by C. Schmid and R. Mohr, IEEE Trans. Pattern Analysis and Machine Intelligence, 1997 copyright 1997, IEEE

Local features for recognition of object instances

Local features for recognition of object instances

- Lowe, et al. 1999, 2003
- Mahamud and Hebert, 2000
- Ferrari, Tuytelaars, and Van Gool, 2004
- Rothganger, Lazebnik, and Ponce, 2004
- Moreels and Perona, 2005
- …

Representing categories: Parts and Structure

Weber, Welling & Perona (2000), Fergus, Perona & Zisserman (2003)

Parts-and-shape representation

- Model:
  - Object as a set of parts
  - Relative locations between parts
  - Appearance of part

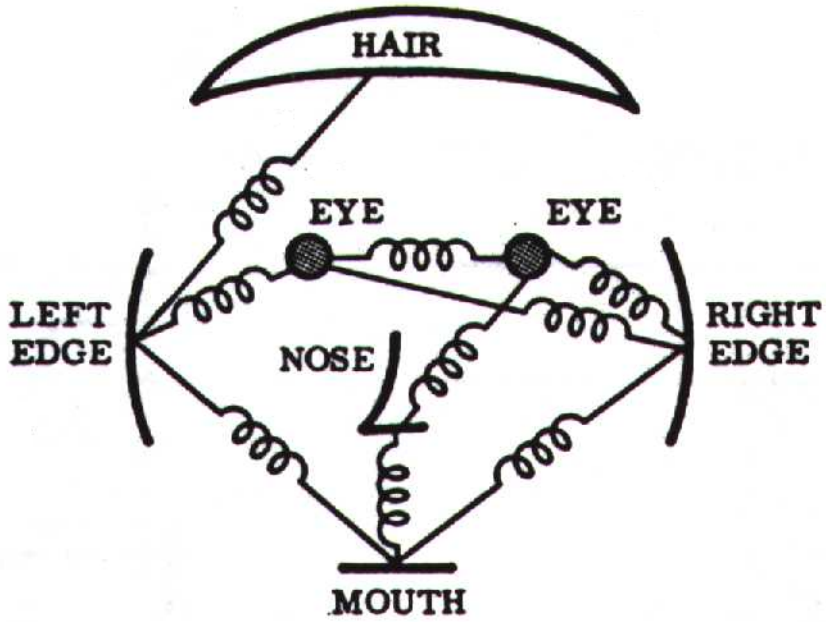


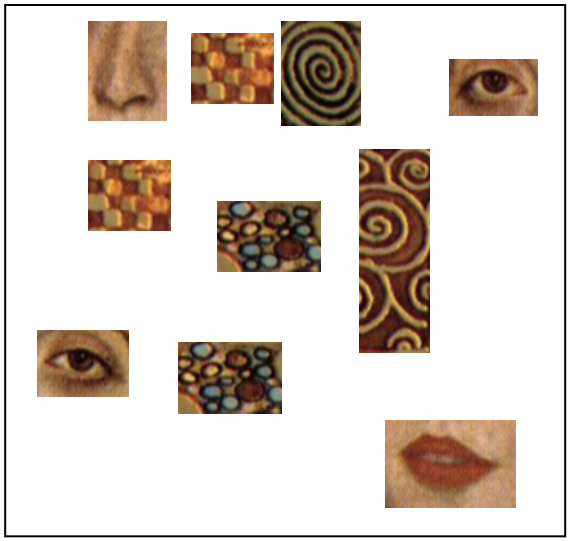
Figure from [Fischler & Elschlager 73]

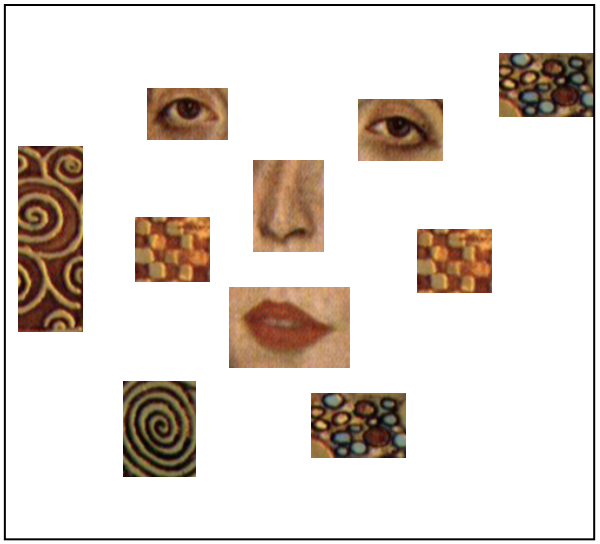
Bag-of-features models

Object → **Bag of 'words'**

Objects as texture

- All of these are treated as being the same





- No distinction between foreground and background: scene recognition?

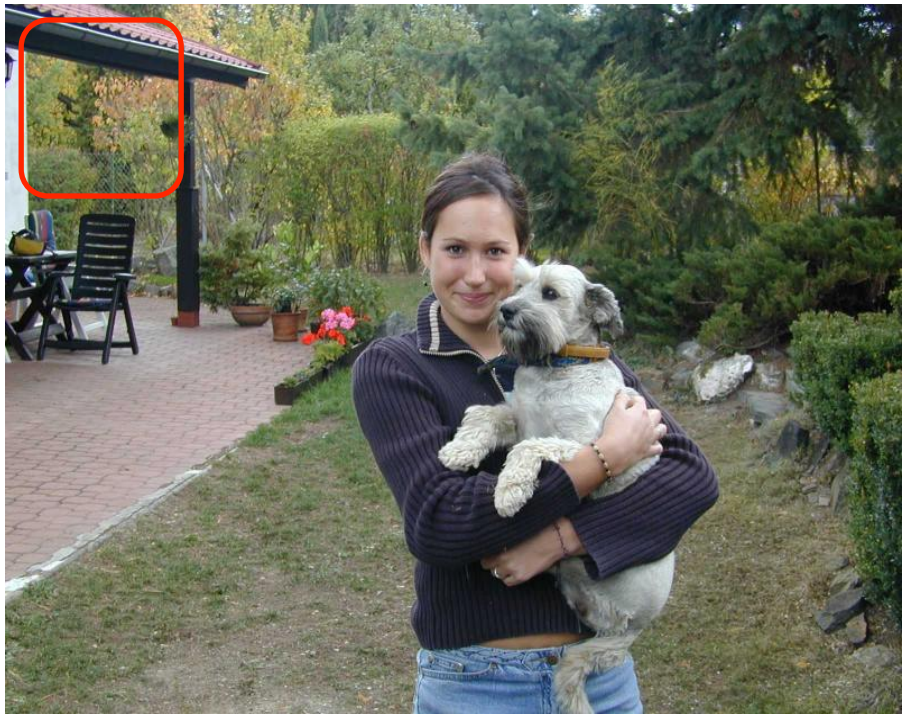
Object Detection – Sliding window approaches

Face detection
Car detection
Pedestrian detection

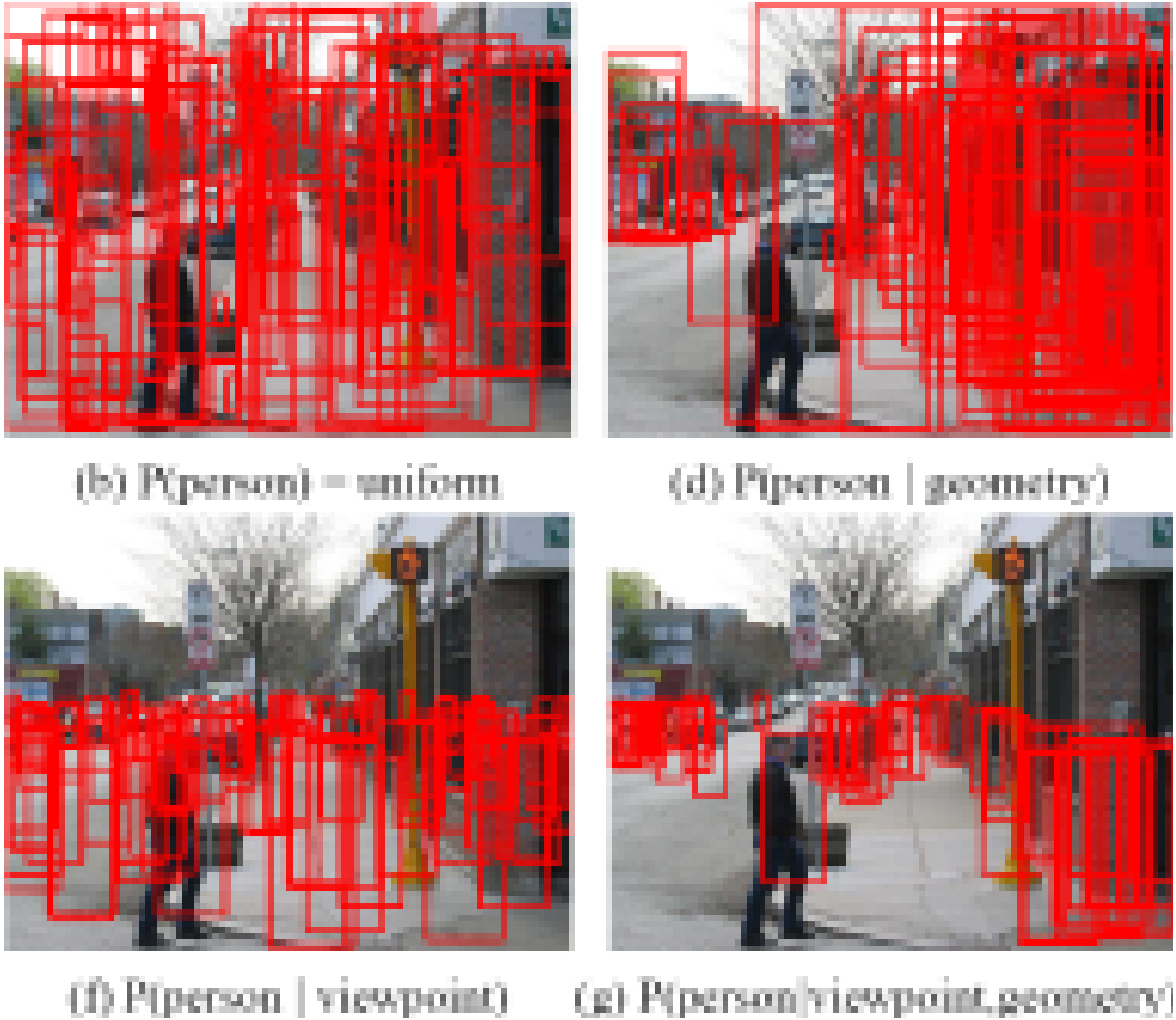
- At each image location (pixel)
- Evaluate how likely is that pixel
- To contain object of interest at
- Particular size and pose

Sliding window approaches

- Scale / orientation range to search over
- Speed
- Context



Context



Hoiem, Efros, Herbert, 2006

Recognition Today

- Integration of multiple view models (Complex 3D objects)
- Generative vs Discriminative Models
- Scaling issues > 10000 object
- Recognition of object categories
- Alternative models of context
- intra-object-within-class variations (chairs)
- Enable models with large number of parts
- Image based retrieval – annotating by semantic context
- Associating words with pictures