universität innsbruck

Lecture 8. Object Recognition I: Categories

703142. Computer Vision

Assoz.Prof. Antonio Rodríguez-Sánchez, PhD.

Most slides thanks to Antonio Torralba and Rob Fergus

Outline

- Introduction
- Challenges
- Representation
- Learning
- Category recognition

Outline

- Introduction
- Challenges
- Representation
- Learning
- Category recognition

Introduction

ob·ject Pr…ion Key (ŏb′jĭkt, -jĕkt′)
n.

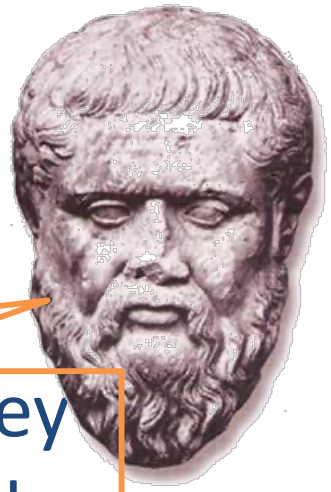
1. Somethi… one or more of the senses, especia… or touch; a…
2. A focus … g, thought, or action: *an object of co…*
3. The purpos… of a specific action or effort: *the object … game.*
4. *Grammar.*
   a. A noun, pronoun, … oun phrase that receives or is affected by the a…ion of a ve… within a sentence.
   b. A noun or substantive … verned by a preposition.
5. *Philosophy.* Something inte…gible 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

Introduction

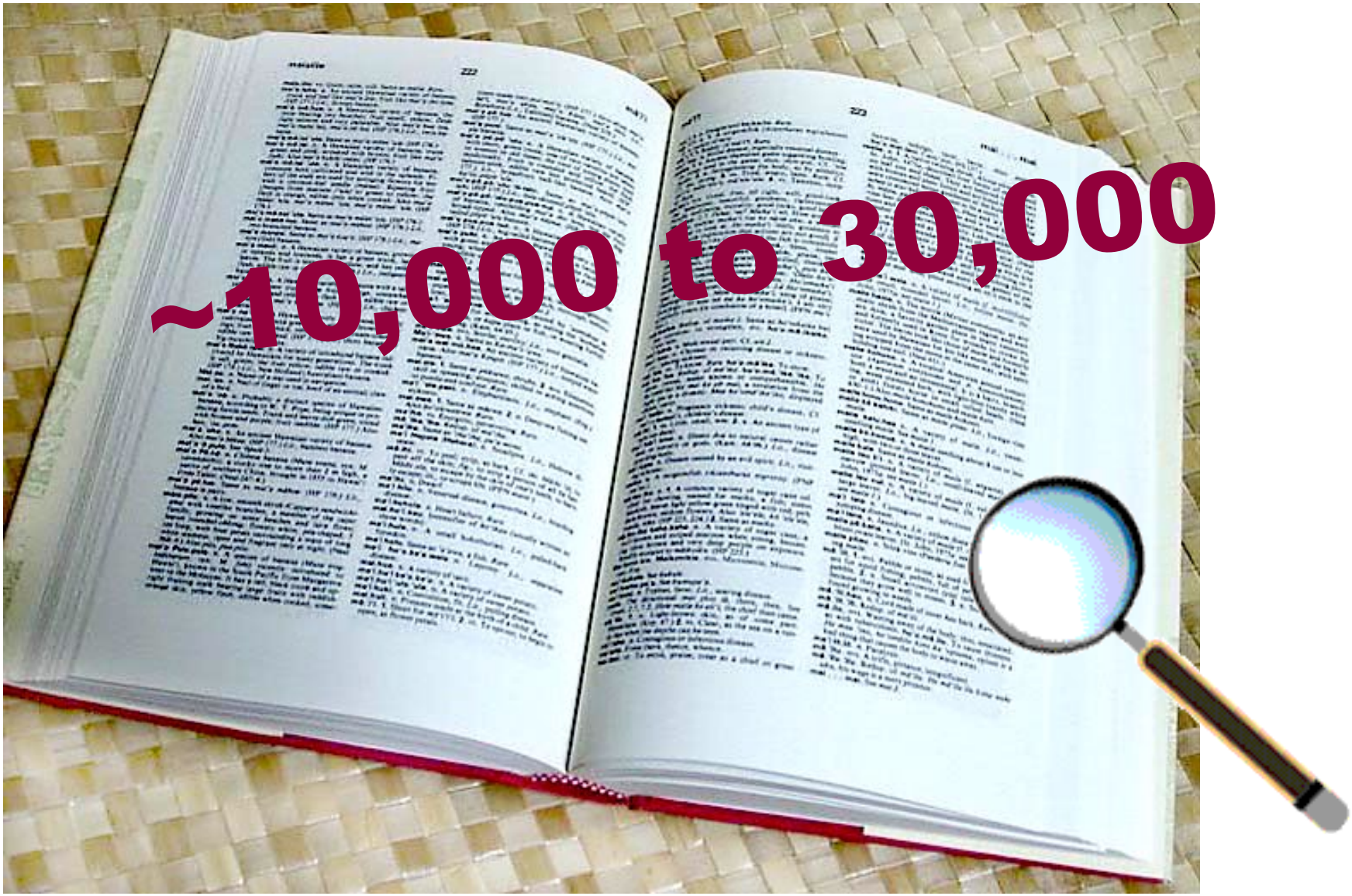


- Quoting Plato

Ordinary objects are classified together if they „participate" in the same **abstract** Form, such as the Form of a Human or the Form of Quartz.

  - Forms are proper subjects of philosophical investigation, for they have the highest degree of reality.
  - Ordinary objects, such as humans, trees, and stones, have a lower degree of reality than the Forms.
  - Fictions, shadows, and the like have a still lower degree of reality than ordinary objects and so are not proper subjects of philosophical enquiry.

How many object categories are there?



Biederman 1987

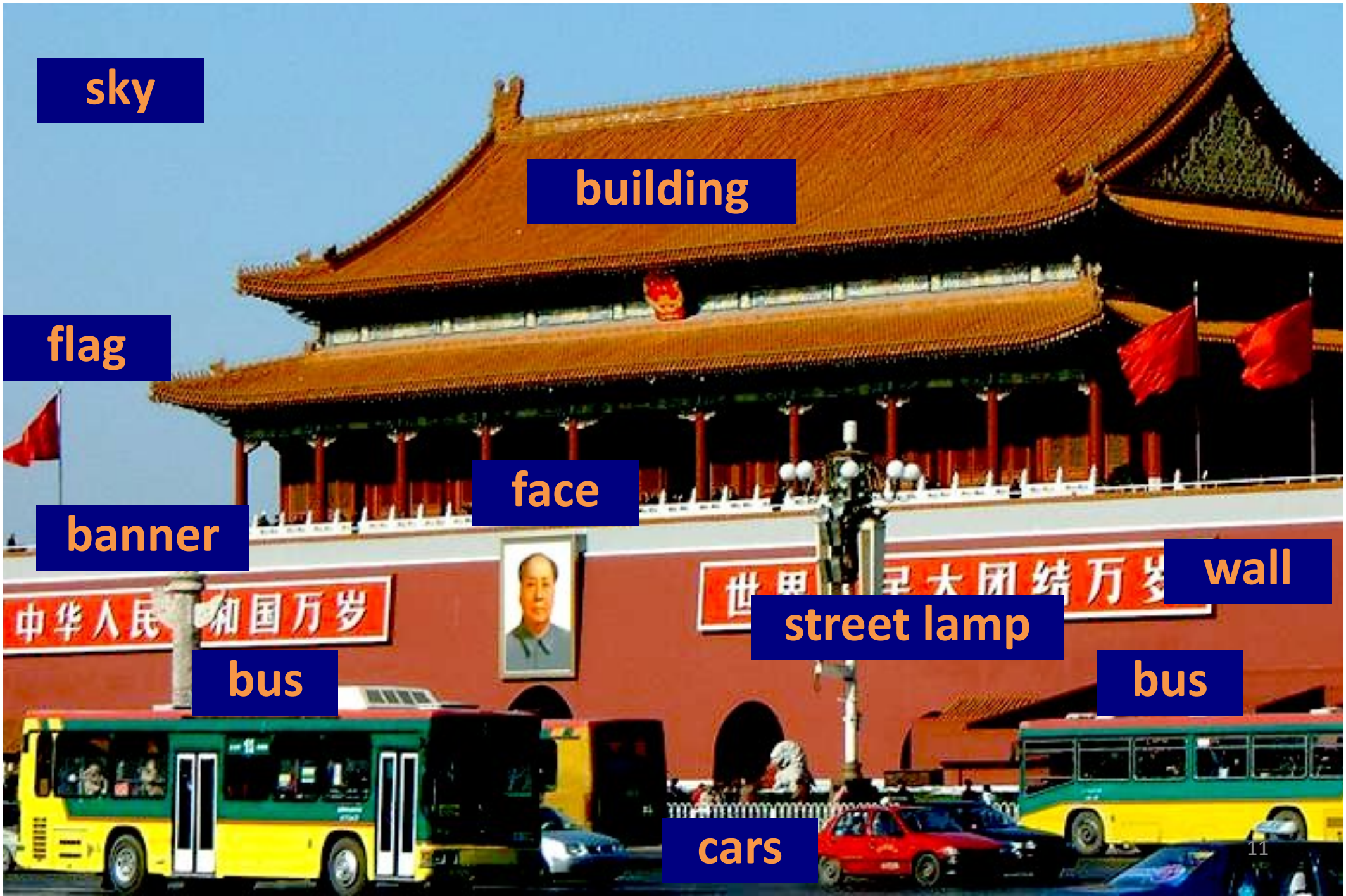
So what does object recognition involve?

Verification: is that a bus?

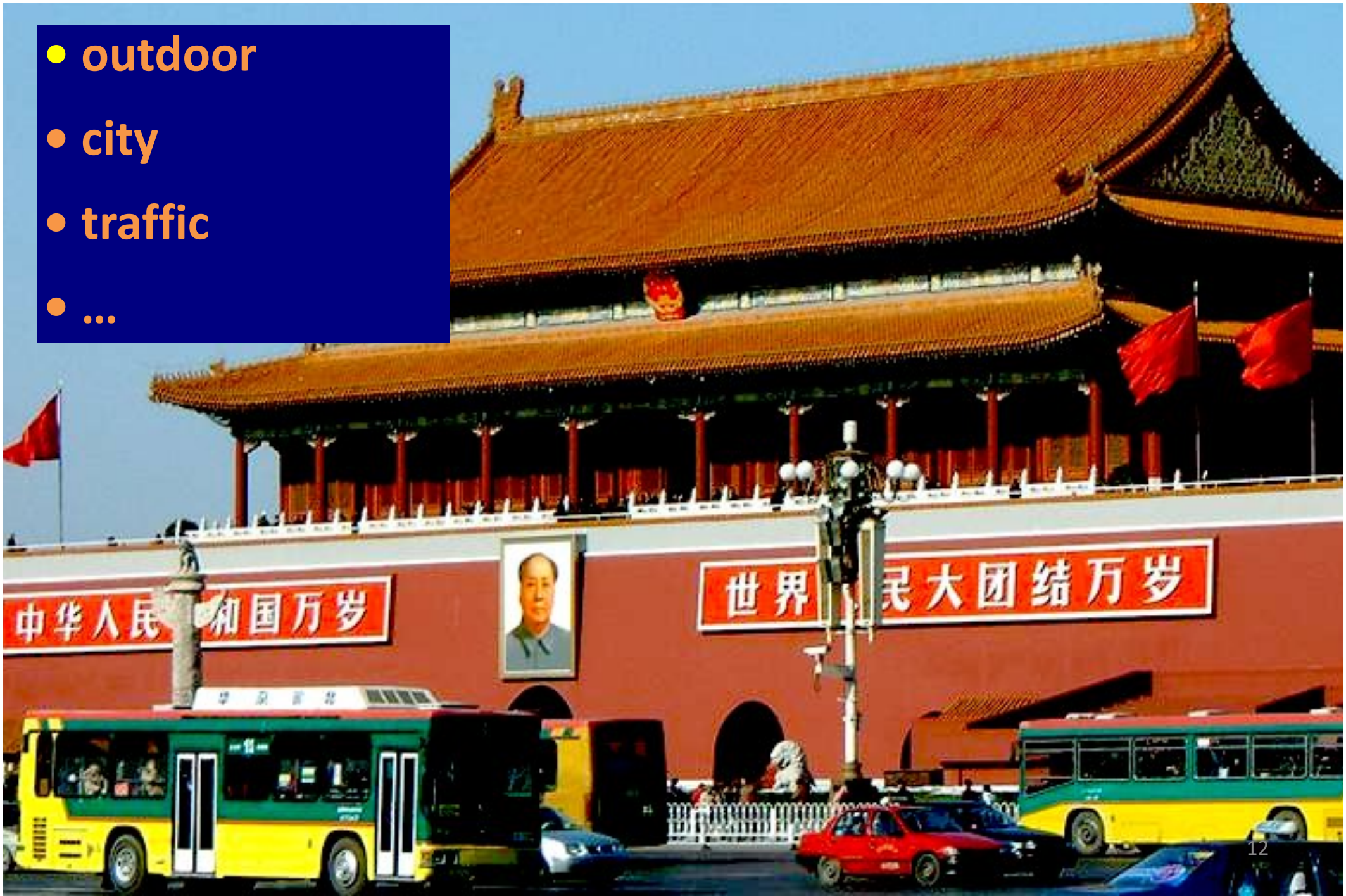
Detection: are there cars?

Identification: is that a picture of Mao?

Object categorization



Scene and context categorization



Introduction

- For Besl and Jain (1985), the problem of object recognition comprised the following steps:
  1. Given a set of objects, examine each object and label it.
  2. Given an array of pixels from a sensor and a list of objects, those questions arise:
     1. Is the object present in the scene?
     2. If so, how many times does it appear?
     3. For each occurrence find its location in the scene and determine its translation and rotation parameters referred to a known coordinate system.
  3. A third optional stage incorporates in the system any unknown objects in the scene (learn from experience).

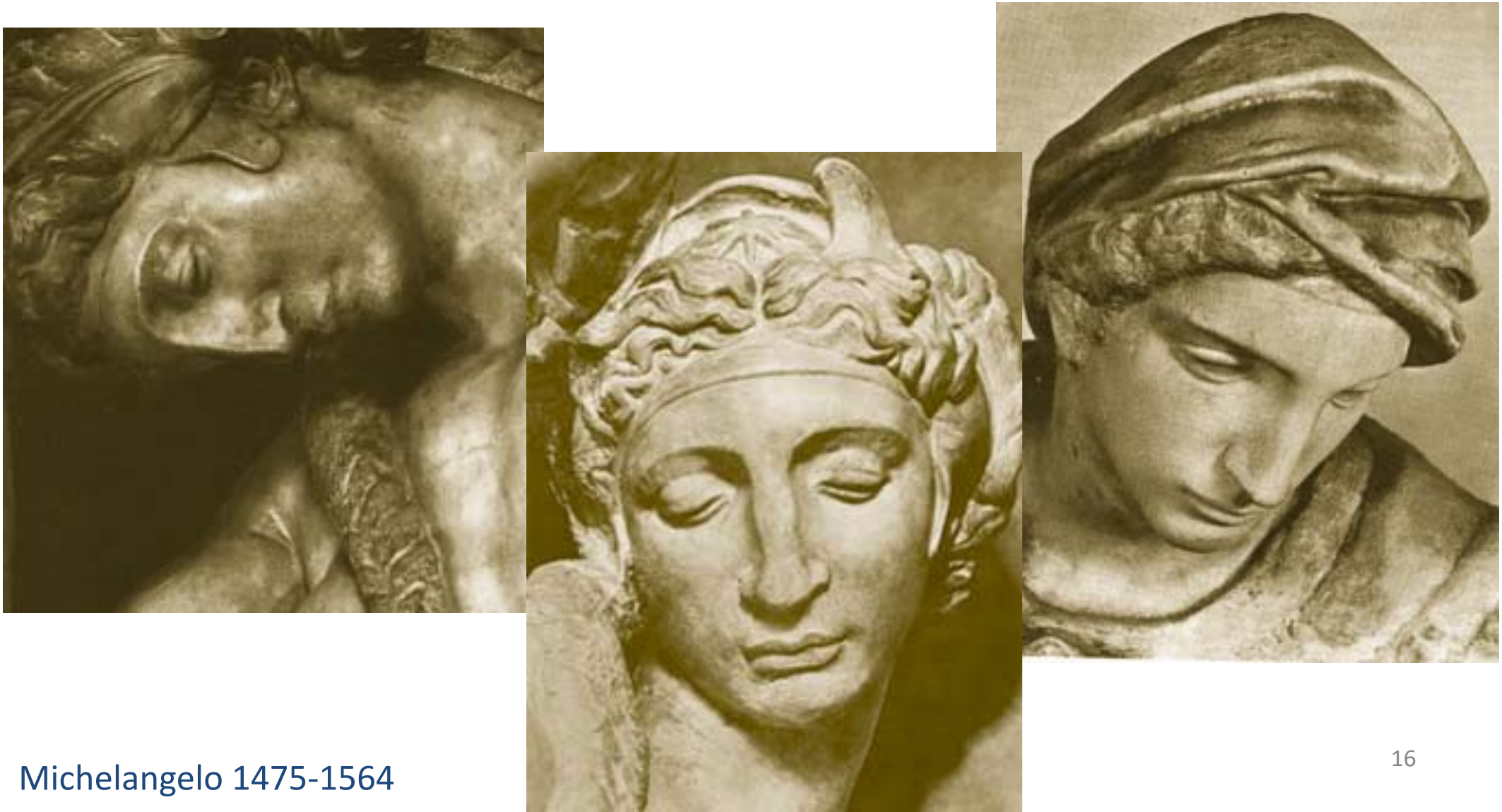
Outline

- Introduction
- Challenges
- Representation
- Learning
- Category recognition

Bruegel, 1564

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

Salvador Dali

Challenges

6. Background clutter

Klimt, 1913

Challenges

6. Background clutter

Challenges

6. Background clutter

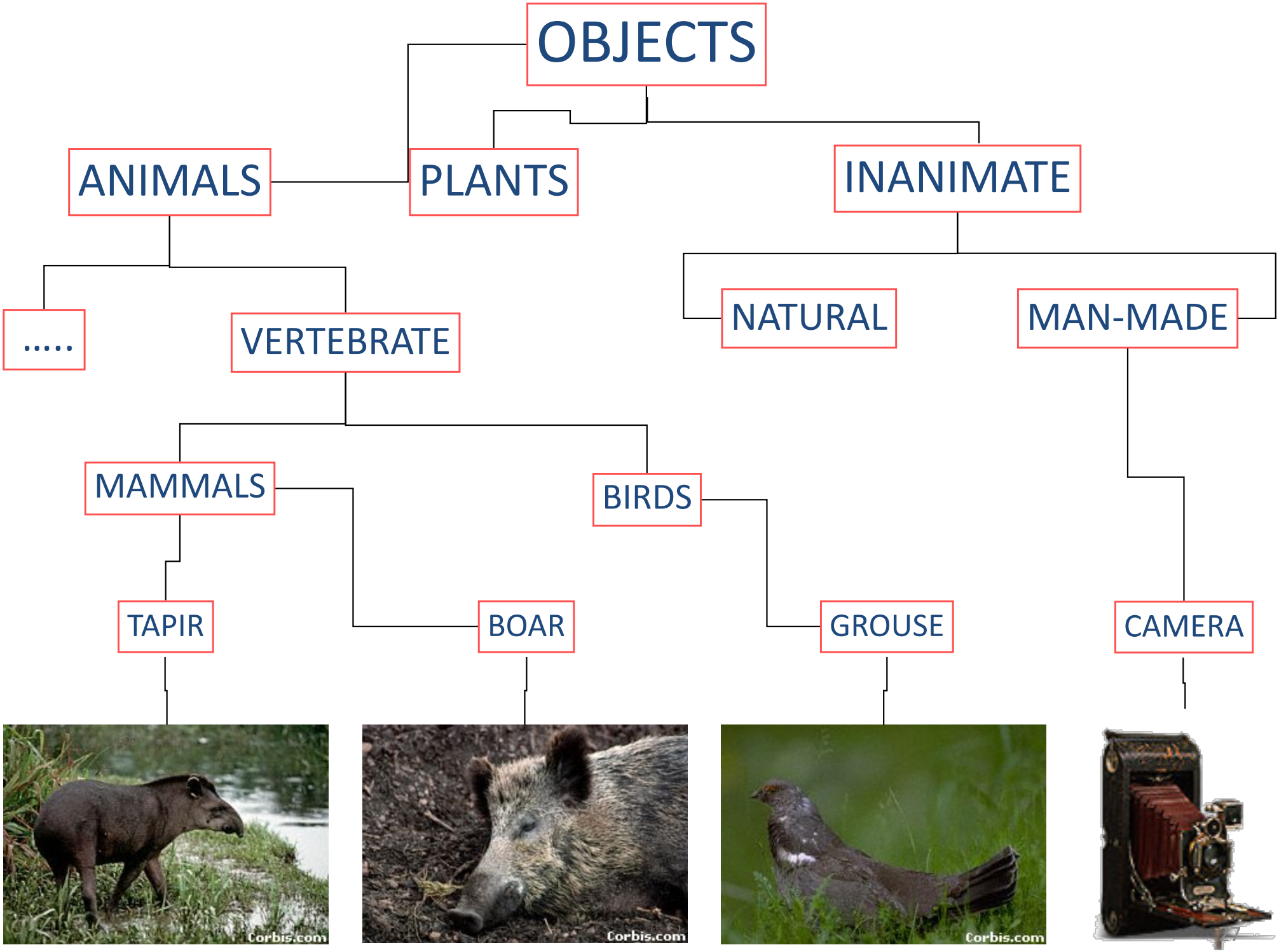
Challenges

7. Intra-class variation

Challenges

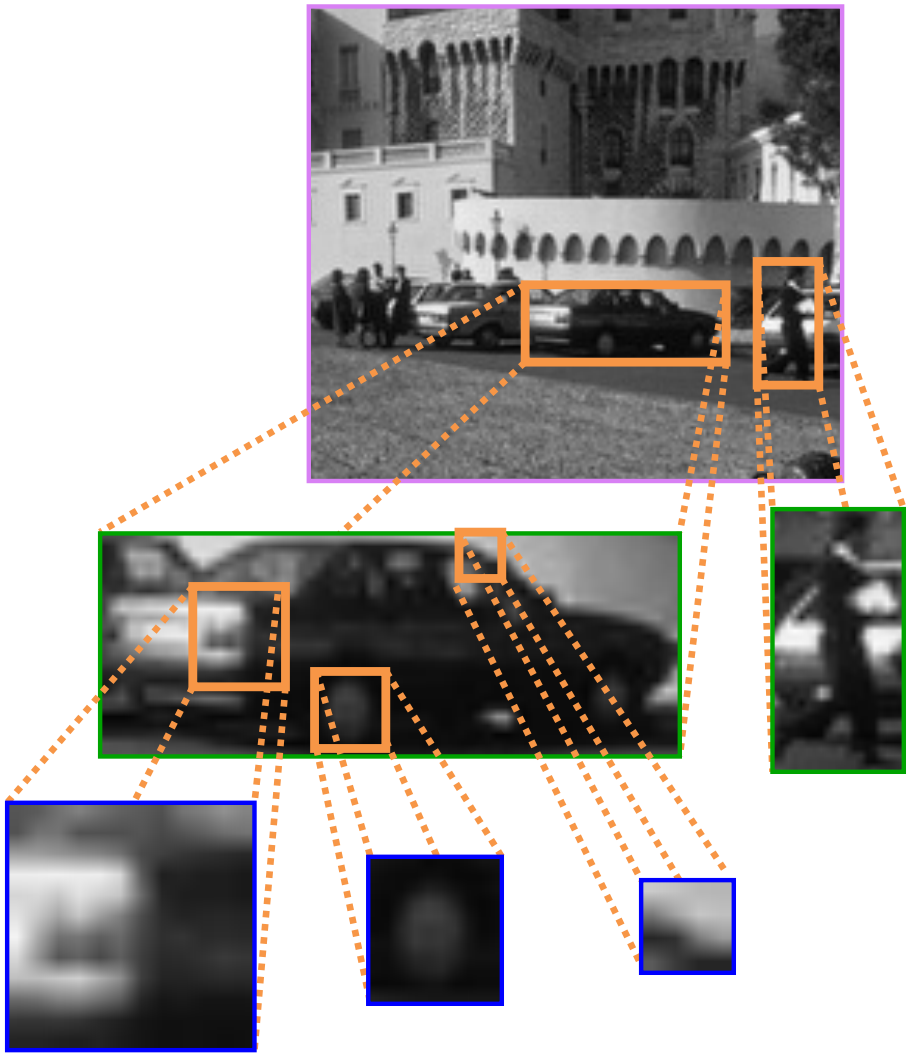
6. Intra-class variation


~10,000 to 30,000


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

Challenges

Scenes, Objects, and Parts



Scene

Objects

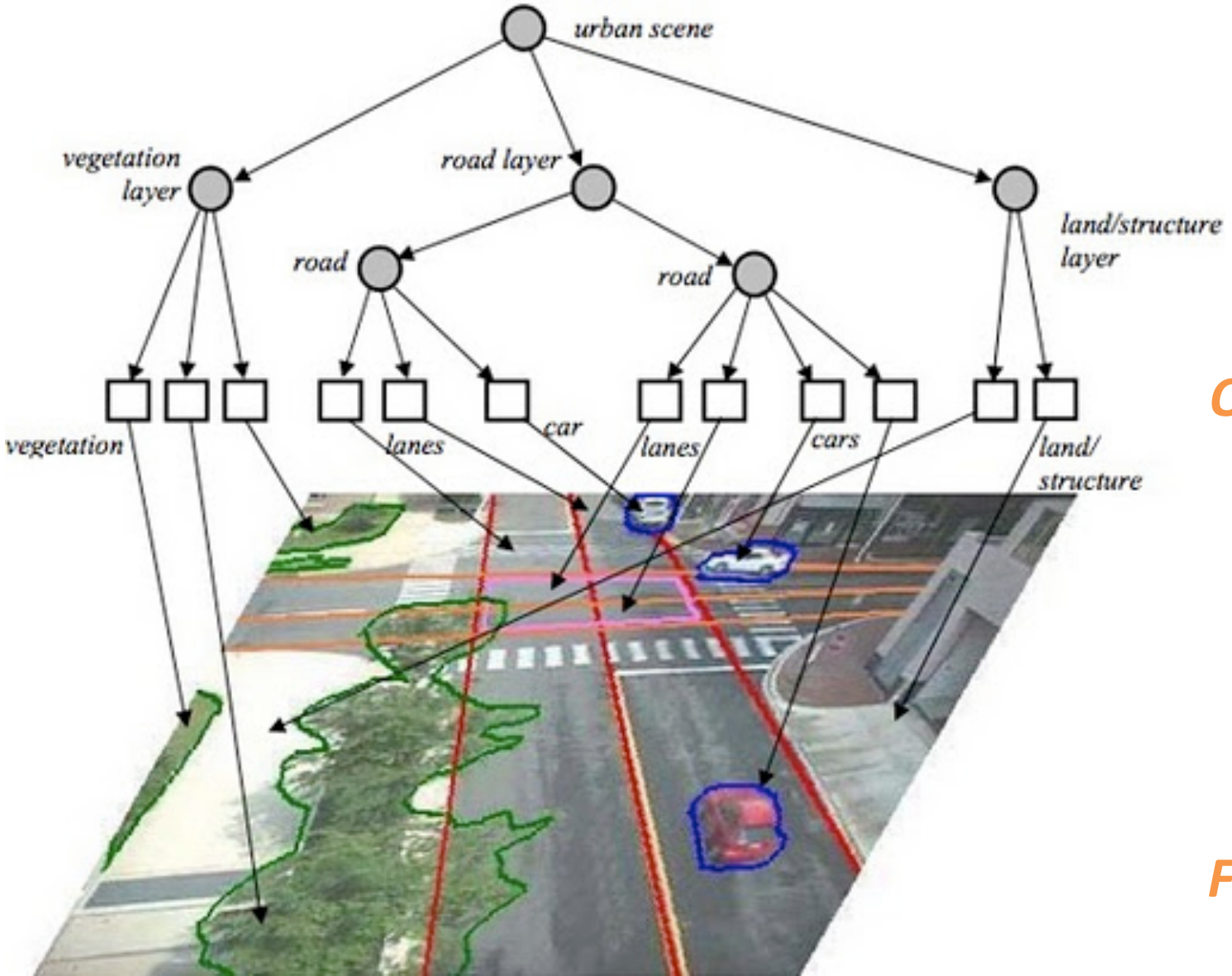
Parts

Features

E. Sudderth, A. Torralba, W. Freeman, A. Willsky. ICCV 2005.

Challenges

Scenes, Objects, and Parts



Scene

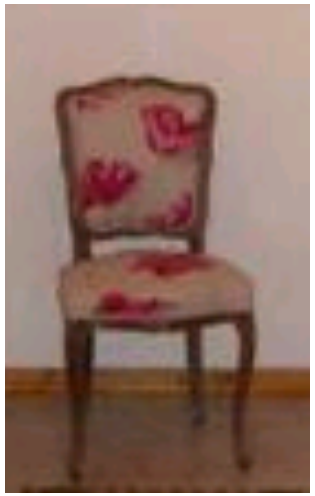
Objects

Parts

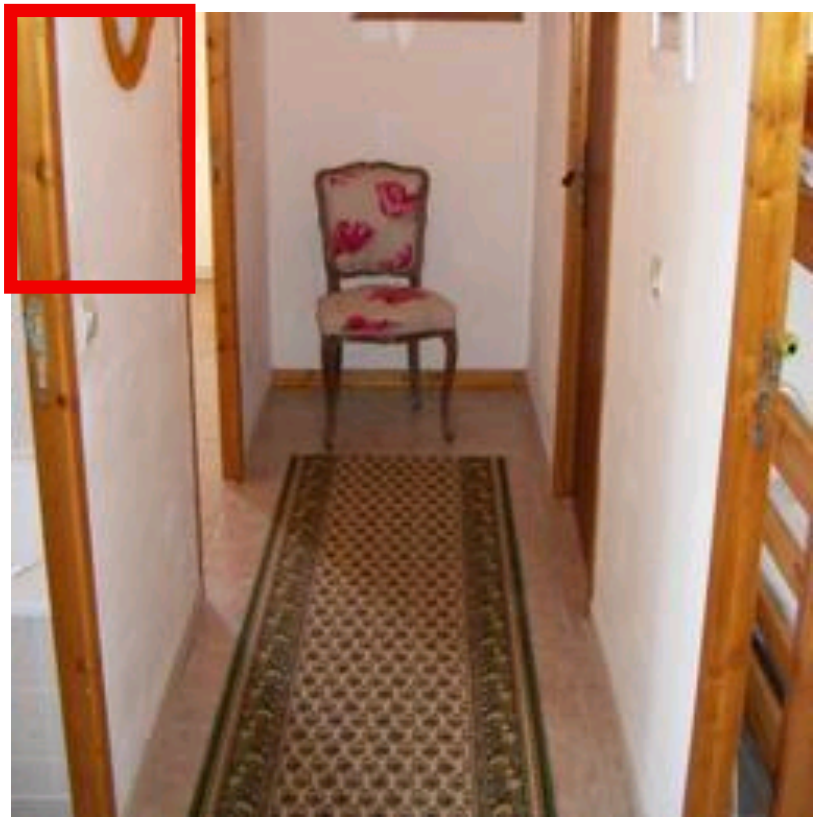
Features

Object recognition
Is it really so hard?

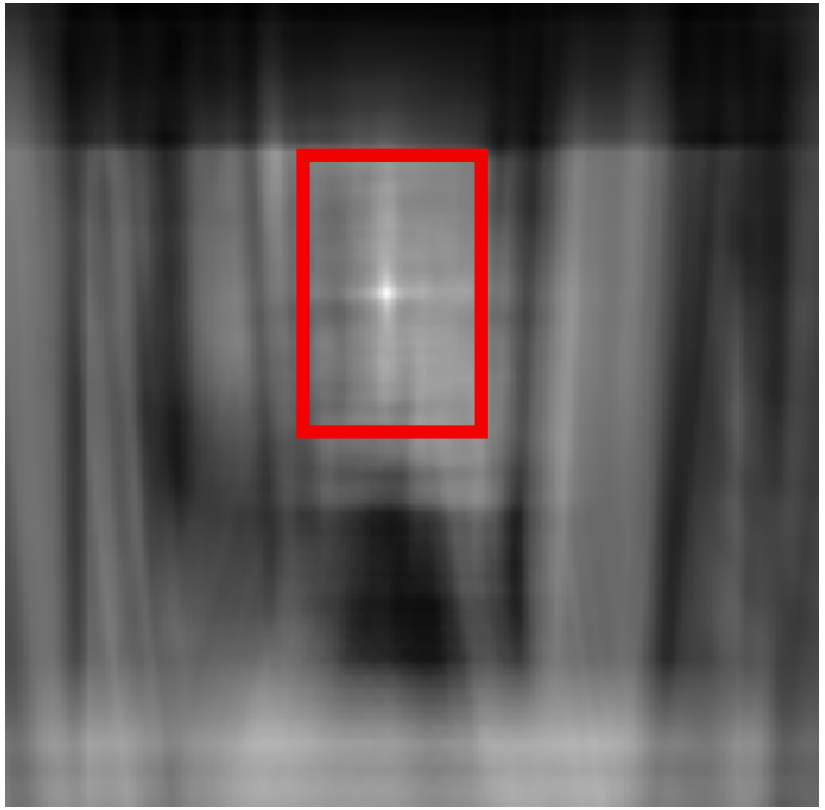
This is a chair

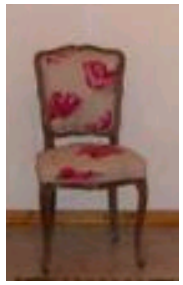


Find the chair in this image



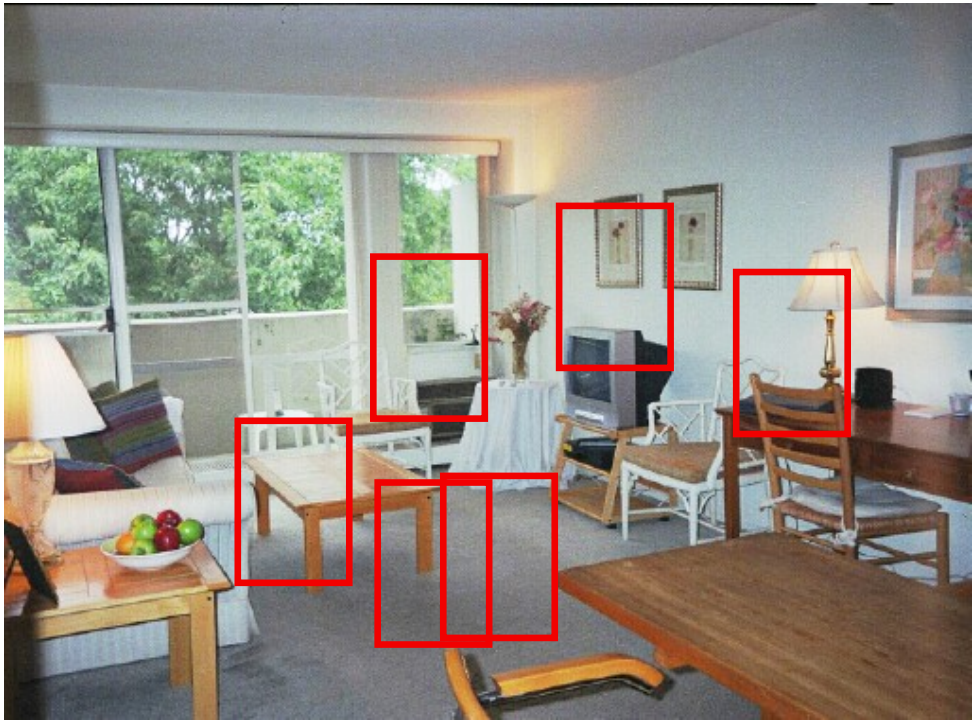
Output of normalized correlation

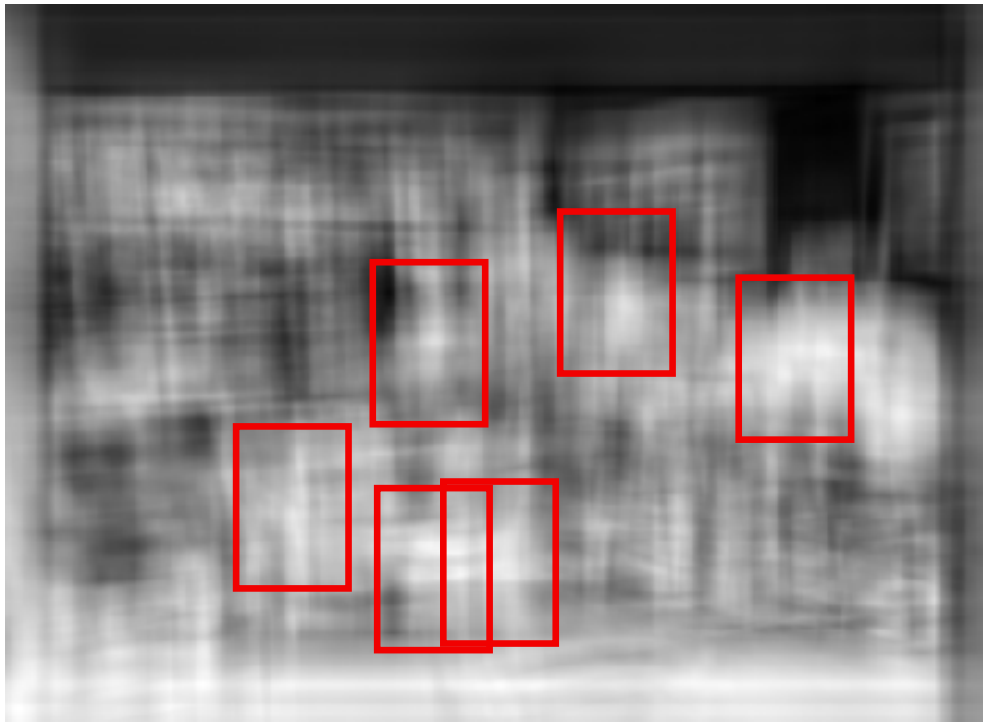




Object recognition
Is it really so hard?

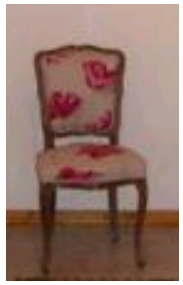
Find a chair in this image





Pretty much garbage

Simple template matching is not going to make it



Object recognition
Is it really so hard?

Find a chair in this image

A “popular method is that of template matching, by point to point correlation of a model pattern with the image pattern. These techniques are inadequate for three-dimensional scene analysis for many reasons, such as occlusion, changes in viewing angle, and articulation of parts.” (Nivatia & Binford, 1977).

Outline

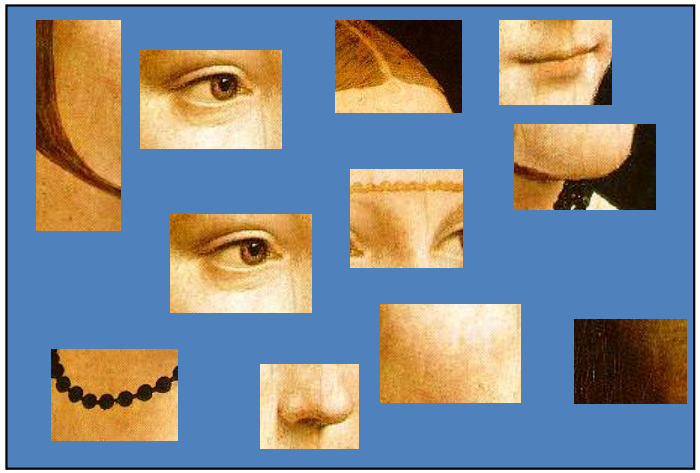
- Introduction
- Challenges
- Representation
- Learning
- Category recognition

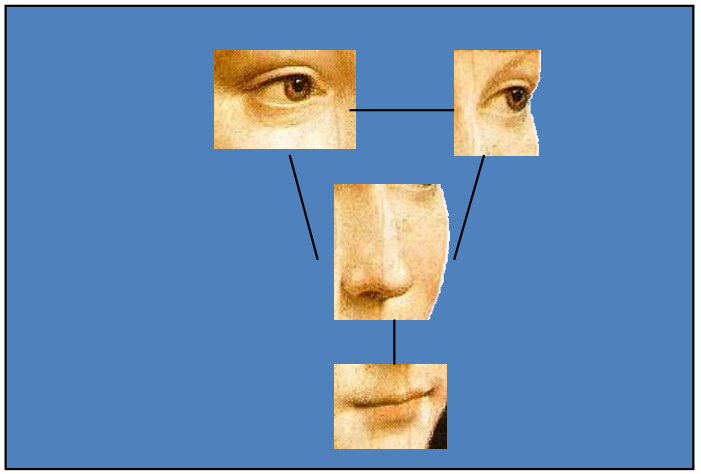
Three main issues

- Representation
  - How to represent an object category
- Learning
  - How to form the classifier, given training data
- Recognition
  - How the classifier is to be used on novel data

Representation

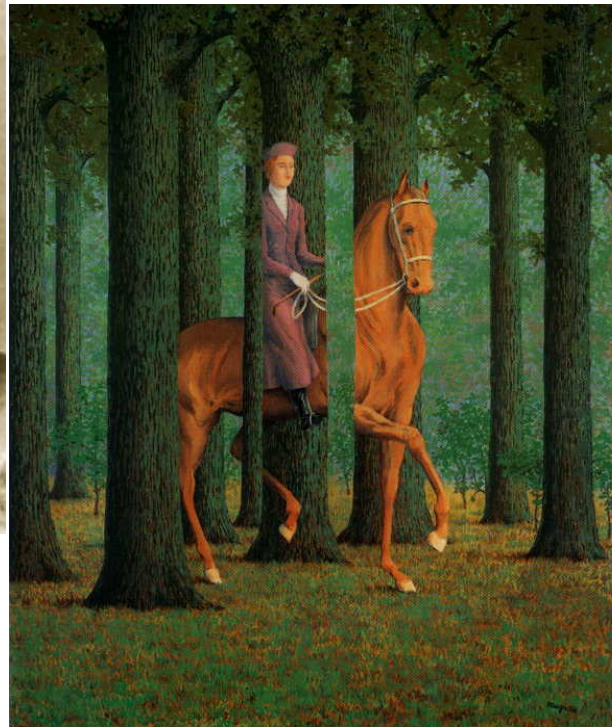
- Appearance only or location and appearance





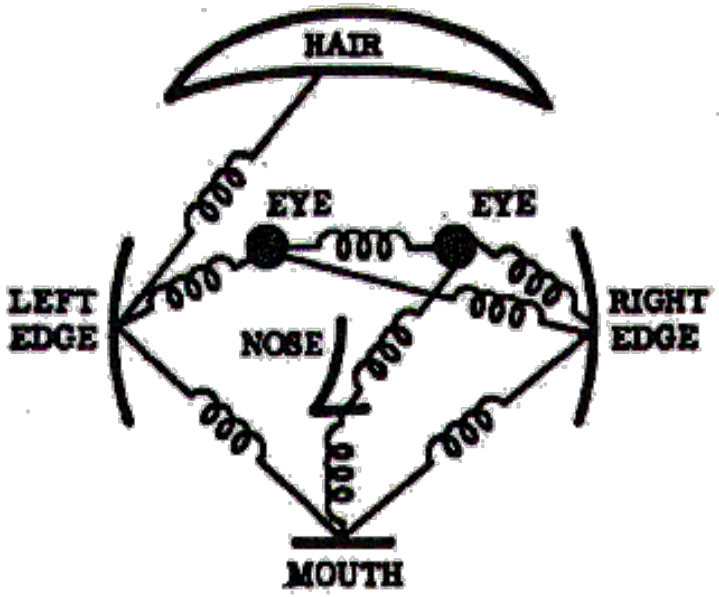
Representation

- Appearance only or location and appearance
- Invariances
  - View point
  - Illumination
  - Occlusion
  - Scale
  - Deformation
  - Clutter
  - etc.



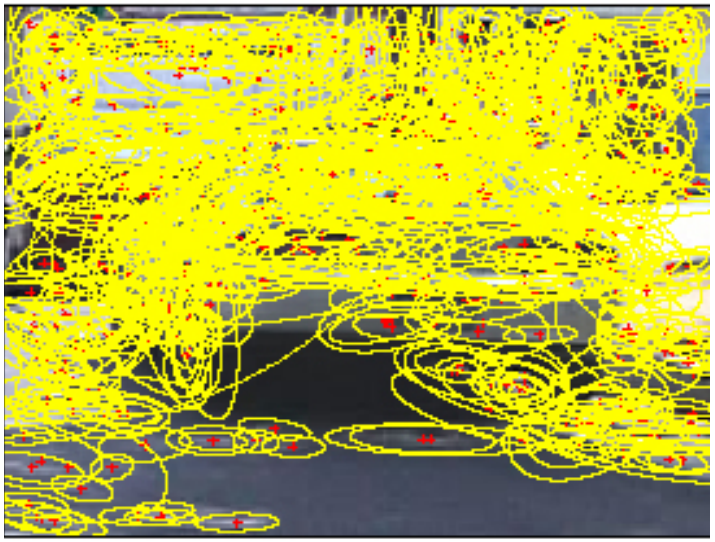
Representation

- Appearance only or location and appearance
- Invariances
- Part-based or global with sub-window



Representation

- Appearance only or location and appearance
- Invariances
- Part-based or global with sub-window
- Use set of features or each pixel in image



Three main issues

- Representation
  - How to represent an object category
- Learning
  - How to form the classifier, given training data
- Recognition
  - How the classifier is to be used on novel data

Learning

- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning

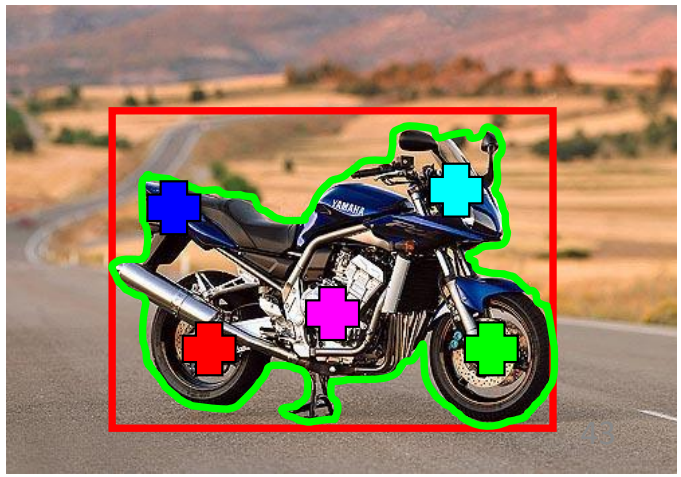
Learning

- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning.
- Methods of training: generative or discriminative
  - What are you maximizing? Likelihood (Generative) or performances on train/validation set (Discriminative)

Learning

- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning.
- Methods of training: generative or discriminative
  - What are you maximizing? Likelihood (Generative) or performances on train/validation set (Discriminative)
- Level of supervision
  - Manual segmentation
  - Bounding box
  - Image labels
  - Noisy labels

Contains a motorbike



Learning

- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning.
- Methods of training: generative or discriminative
  - What are you maximizing? Likelihood (Generative) or performances on train/validation set (Discriminative)
- Level of supervision
- Batch/incremental (on category and image level; user-feedback)

Learning

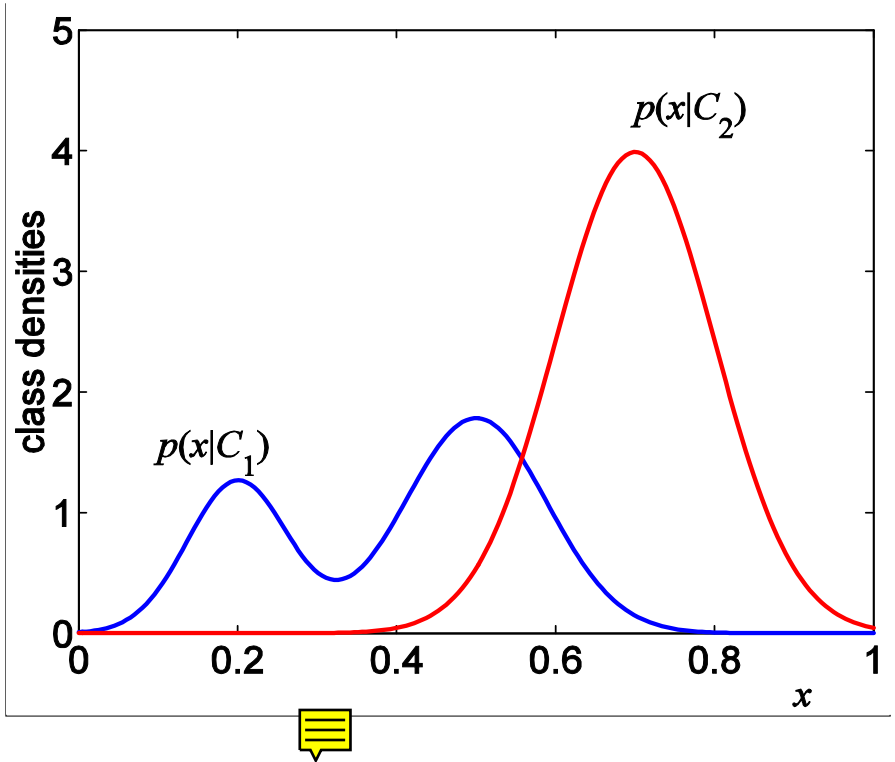
- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning.
- Methods of training: generative or discriminative
  - What are you maximizing? Likelihood (Generative) or performances on train/validation set (Discriminative)
- Level of supervision
- Batch/incremental (on category and image level; user-feedback)
- Training images:
  - Issue of overfitting
  - Negative images for discriminative methods

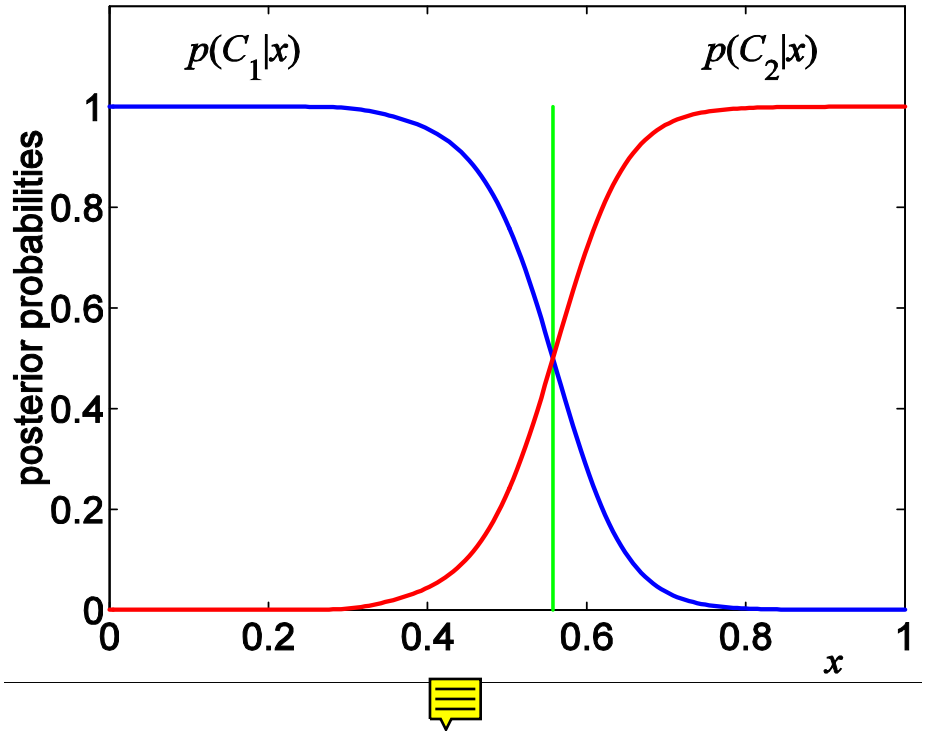
Learning

- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning.
- Methods of training: generative or discriminative
  - What are you maximizing? Likelihood (Generative) or performances on train/validation set (Discriminative)
- Level of supervision
- Batch/incremental (on category and image level; user-feedback)
- Training images
- Priors

Learning

- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning.
- Methods of training: generative or discriminative





Object categorization: the statistical viewpoint

$$p(zebra \mid image)$$

vs.

$$p(no\ zebra \mid image)$$

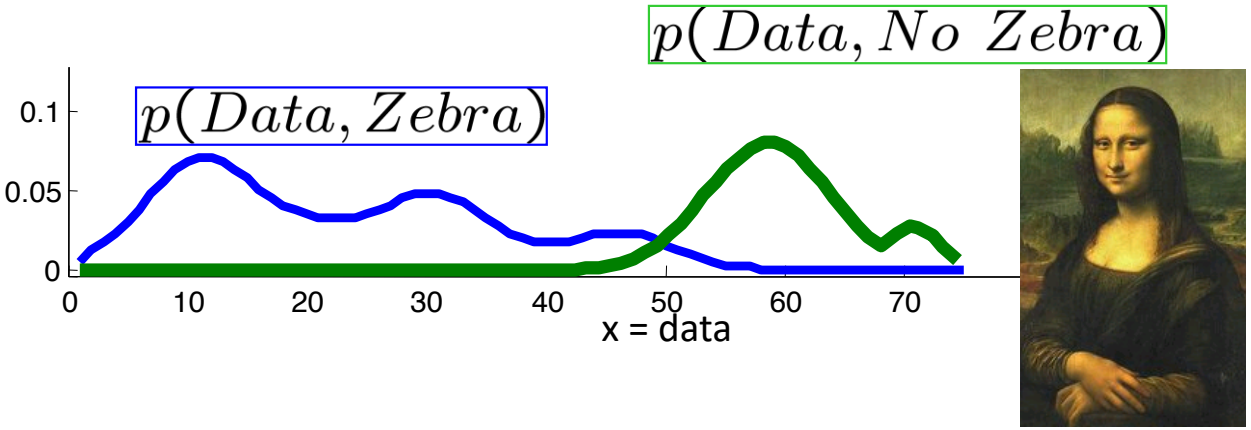
- Bayes rule:

$$\underbrace{\frac{p(zebra \mid image)}{p(no\ zebra \mid image)}}_{\text{posterior ratio}} = \underbrace{\frac{p(image \mid zebra)}{p(image \mid no\ zebra)}}_{\text{likelihood ratio}} \cdot \underbrace{\frac{p(zebra)}{p(no\ zebra)}}_{\text{prior ratio}}$$

Learning

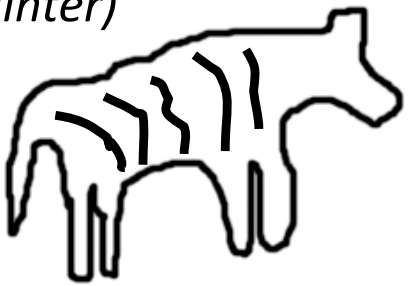
- Generative model

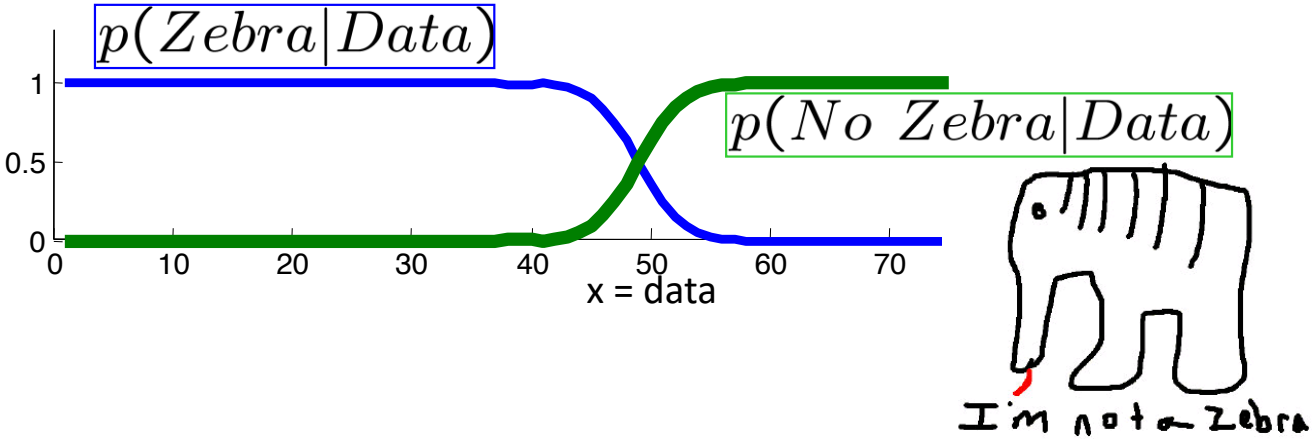
(*The artist*)



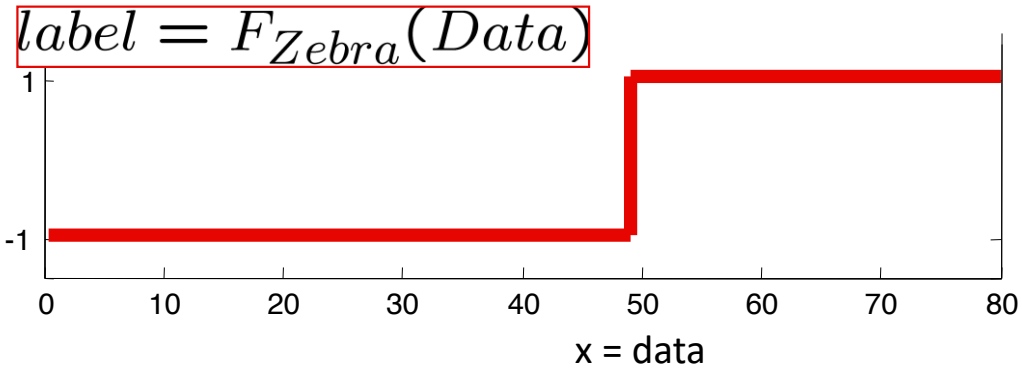
- Discriminative model

(The lousy painter)





- Classification function



Object categorization: the statistical viewpoint

$$\underbrace{\frac{p(zebra \mid image)}{p(no\ zebra \mid image)}}_{\text{posterior ratio}} = \underbrace{\frac{p(image \mid zebra)}{p(image \mid no\ zebra)}}_{\text{likelihood ratio}} \cdot \underbrace{\frac{p(zebra)}{p(no\ zebra)}}_{\text{prior ratio}}$$

- **Discriminative methods model posterior**

Object categorization: the statistical viewpoint

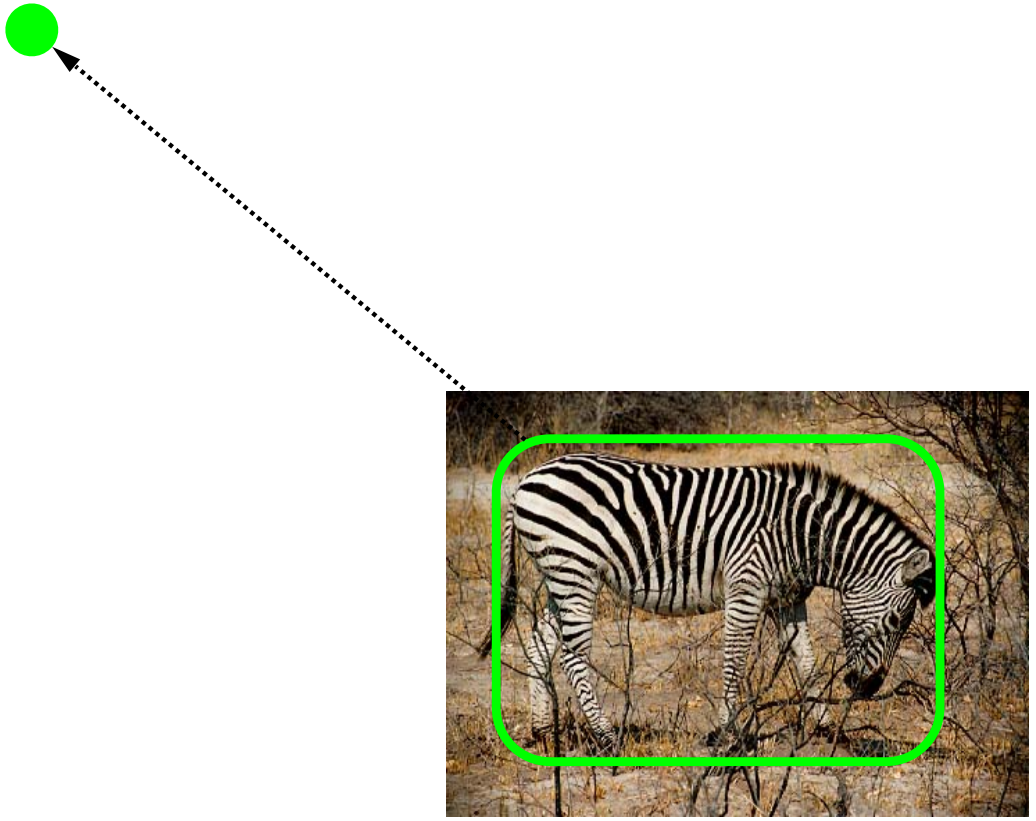
$$\underbrace{\frac{p(zebra \mid image)}{p(no\ zebra \mid image)}}_{\text{posterior ratio}} = \underbrace{\frac{p(image \mid zebra)}{p(image \mid no\ zebra)}}_{\text{likelihood ratio}} \cdot \underbrace{\frac{p(zebra)}{p(no\ zebra)}}_{\text{prior ratio}}$$

- **Discriminative methods model posterior**
- **Generative methods model likelihood and prior**

$$p(X, Y) = p(Y \mid X)p(X)$$

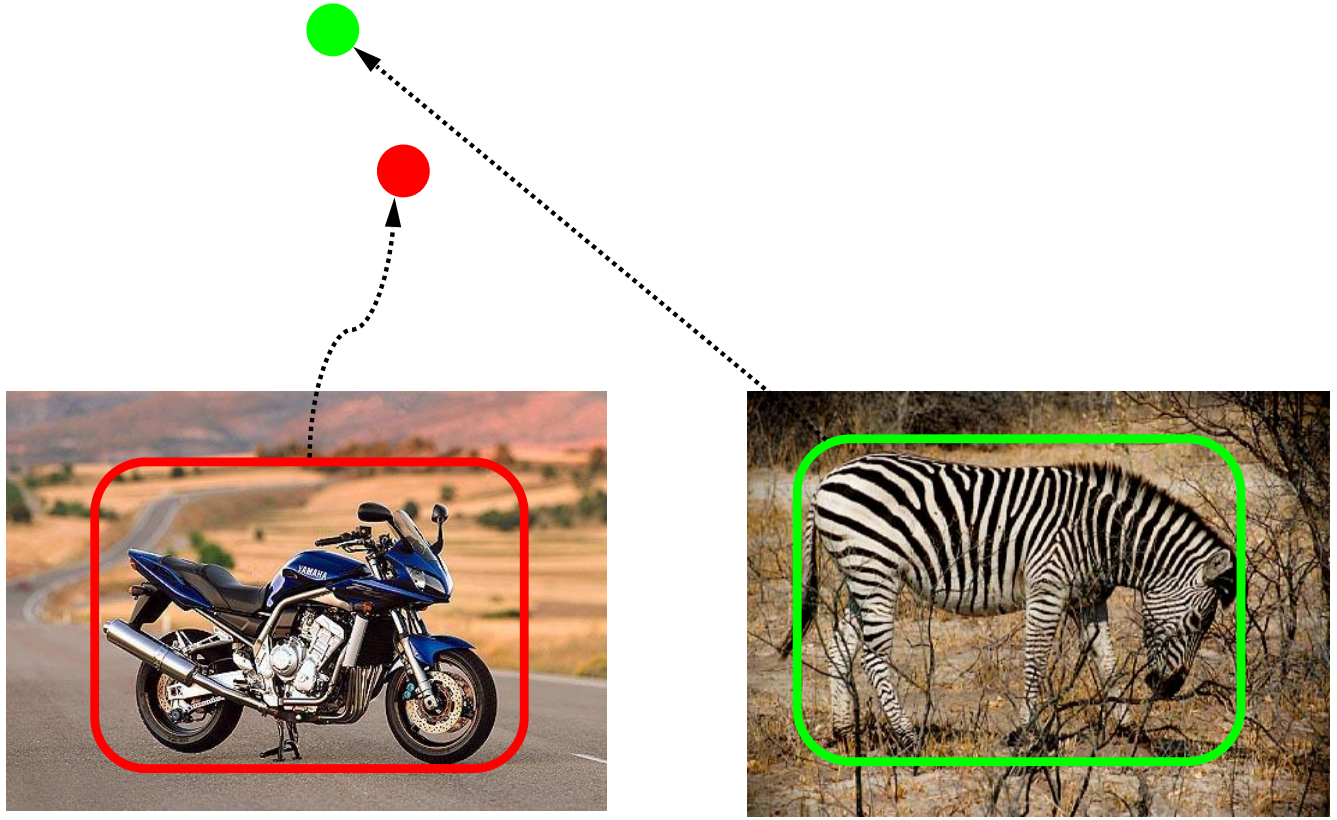
Discriminative

- Direct modeling of $\frac{p(zebra \mid image)}{p(no\ zebra \mid image)}$



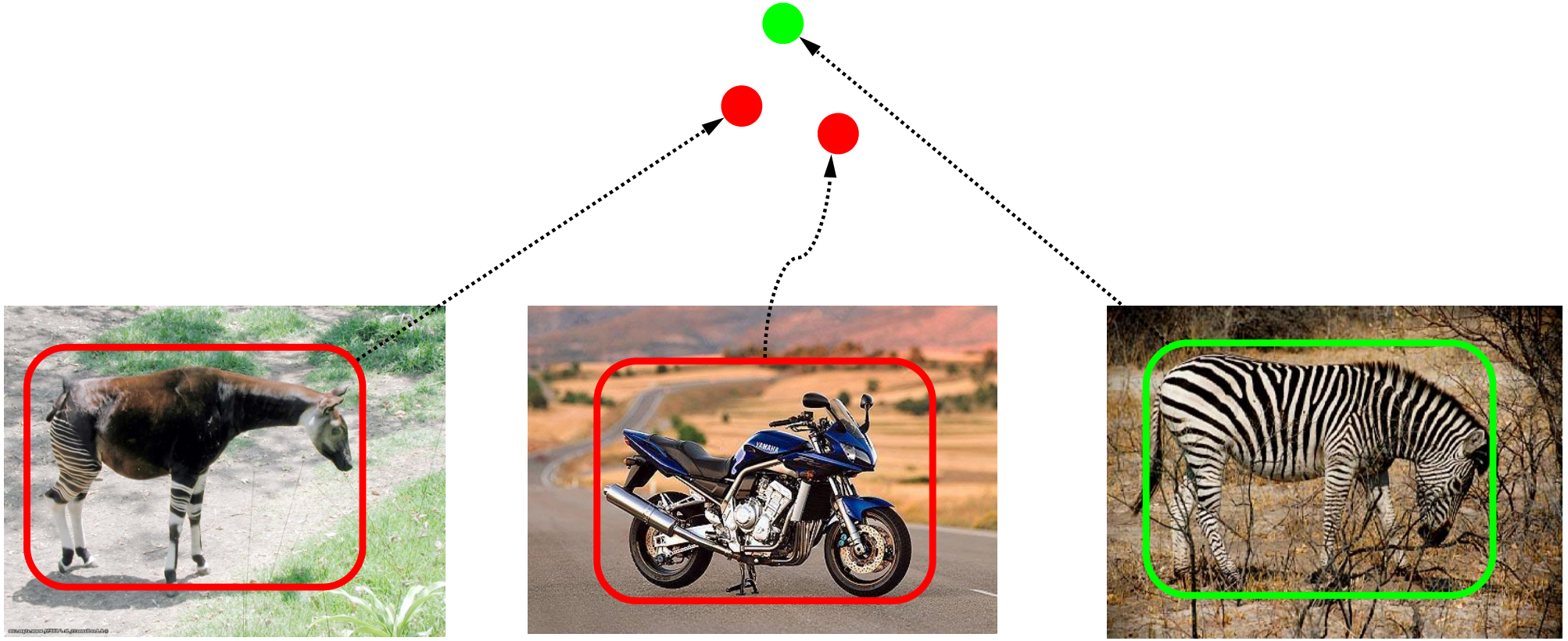
Discriminative

- Direct modeling of $\frac{p(zebra \mid image)}{p(no\ zebra \mid image)}$



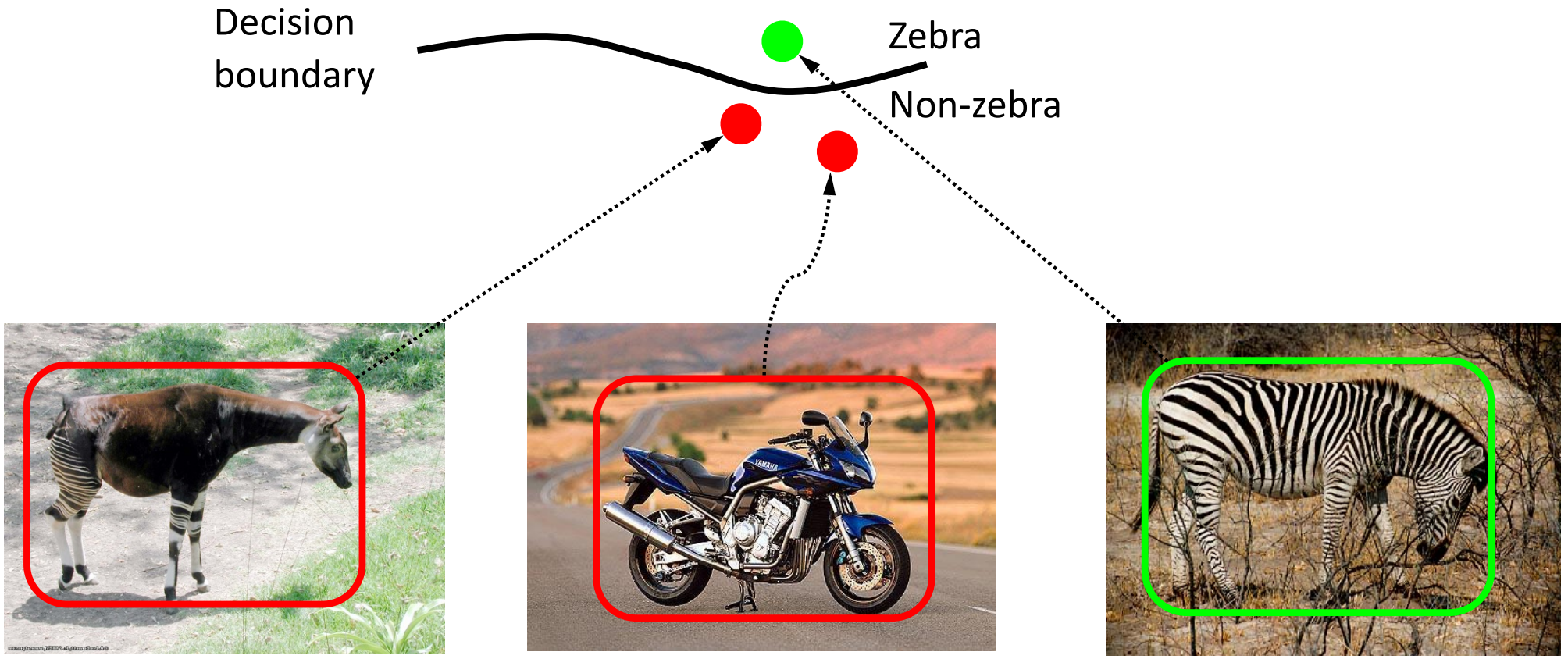
Discriminative

- Direct modeling of $\frac{p(zebra \mid image)}{p(no\ zebra \mid image)}$



Discriminative

- Direct modeling of $\frac{p(zebra \mid image)}{p(no\ zebra \mid image)}$



Discriminative

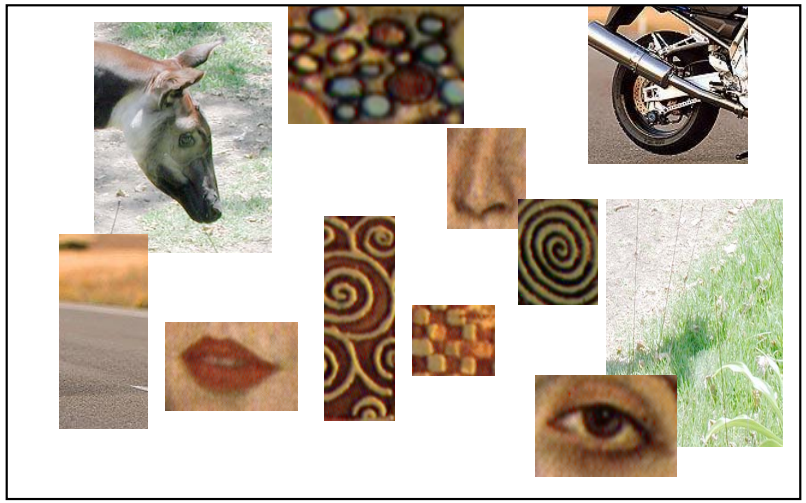
- Directly estimate posterior probabilities
- No attempt to model underlying probability distributions
- Focus computational resources on a given task – better performance
- Popular models
  - Logistic regression, SVMs
  - Traditional neural networks, Nearest neighbors
  - Conditional Random Fields (CRF)

Discriminative

- Disadvantages
  - Lack of elegance of generative
  - Alternative notions of penalty functions, regularization, kernel functions
  - Feel like black-boxes
    - Relationships between variables are not explicit and visualizable

Generative

- Model $p(image \mid zebra)$ and $p(image \mid no\ zebra)$



$p(image \mid zebra)$	$p(image \mid no\ zebra)$
Low	Middle
High	Middle→Low

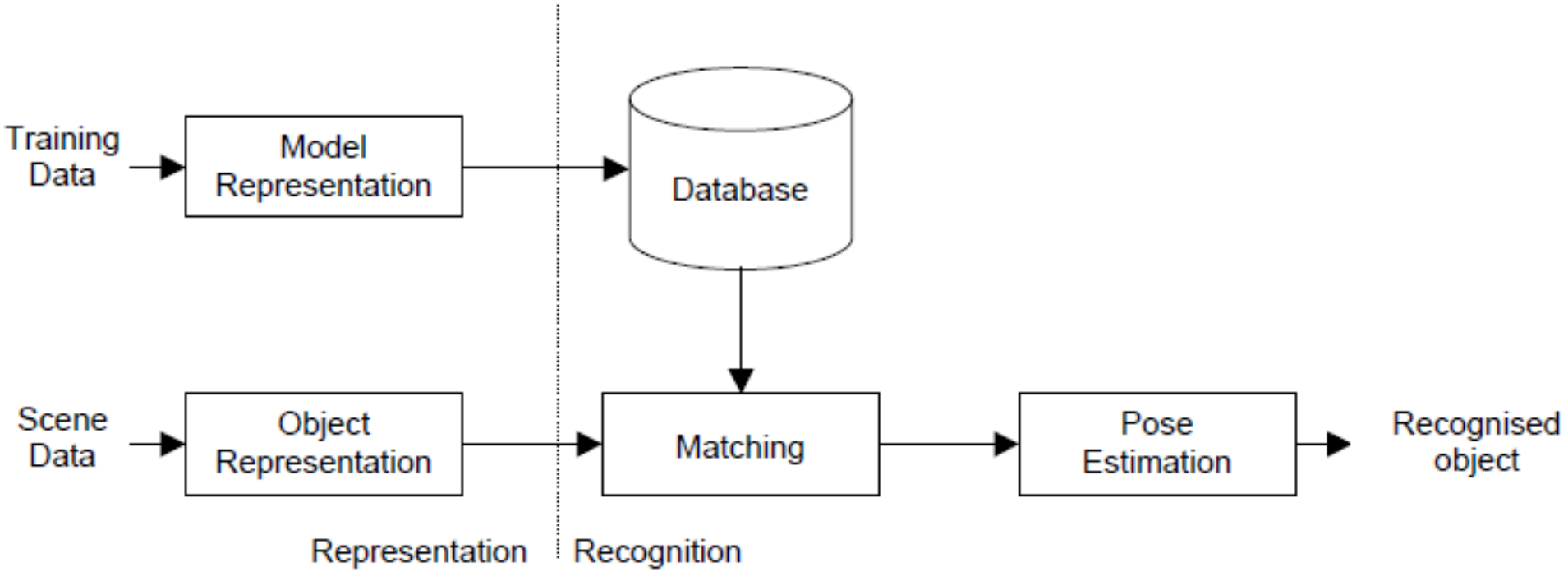
Generative

- Model class-conditional probability distribution functions and prior probabilities
- Generative since they can generate synthetic data points
- Popular models
  - Gaussians, Naïve Bayes, Mixtures of multinomials
  - Mixtures of Gaussians, Hidden Markov Models (HMM)
  - Sigmoidal belief networks, Bayesian networks, Markov random fields (MRF)

Three main issues

- Representation
  - How to represent an object category
- Learning
  - How to form the classifier, given training data
- Recognition
  - How the classifier is to be used on novel data

Recognition



Mamic and Benamoun, 2002

Outline

- Introduction
- Challenges
- Representation
- Learning
- Category recognition

Classical Methods

1. Bag of words approaches

2. Parts and structure approaches

3. Recognition with segmentation

Classical Methods

1. **Treat a feature vectors for standard classifier (e.g. SVM)**
   - **Bag of words approaches**

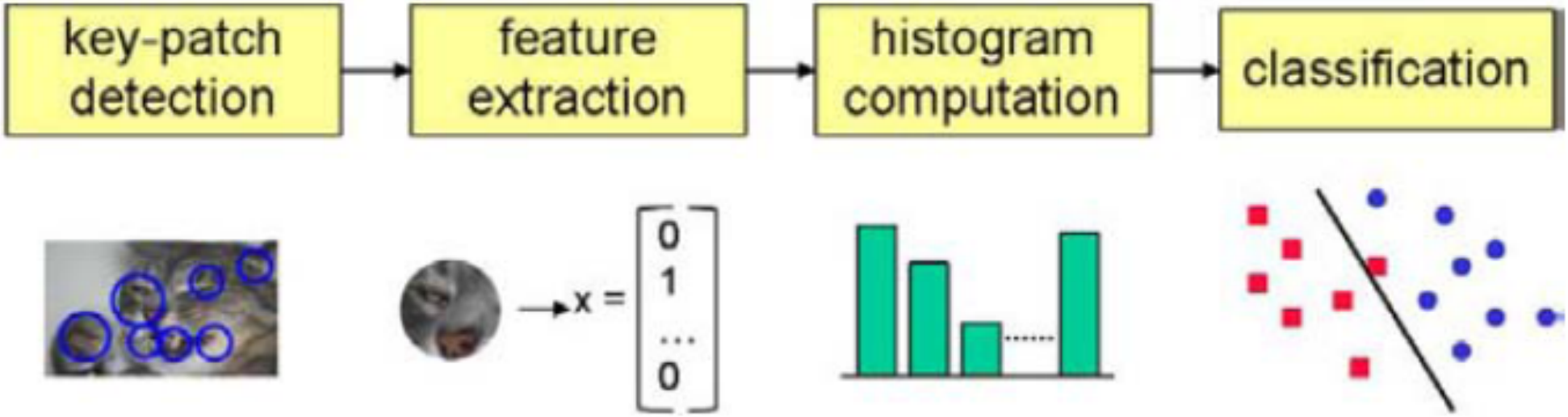
2. Parts-based representation
   - Hierarchical models
   - Decompose scene/object

3. Scene

Bag of Words

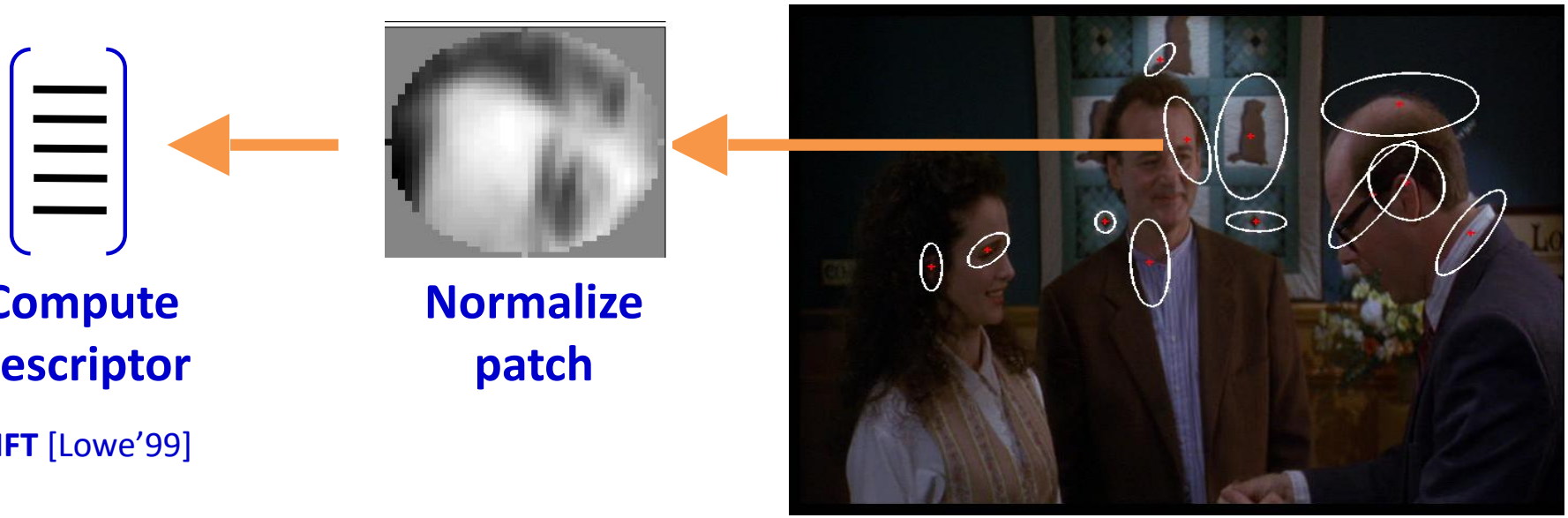
Object → **Bag of 'words'**

Bag of Words


key-patch detection
feature extraction
histogram computation
classification
x =
0
1
…
0

Bag of Words

1.Feature detection and representation



Compute descriptor

e.g. **SIFT** [Lowe'99]

Normalize patch

Detect patches:

Local interest operator

or

Regular grid

[Mikojaczyk and Schmid '02]

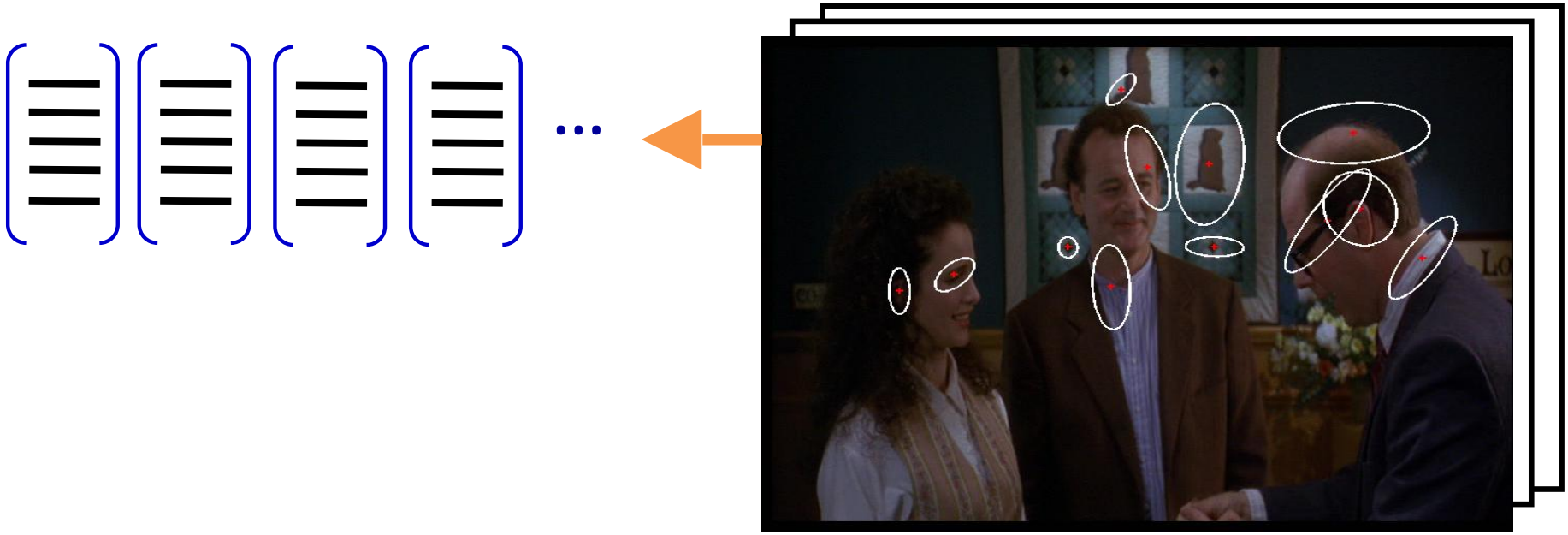
[Mata, Chum, Urban & Pajdla, '02]

[Sivic & Zisserman, '03]

Slide credit: Josef Sivic

Bag of Words

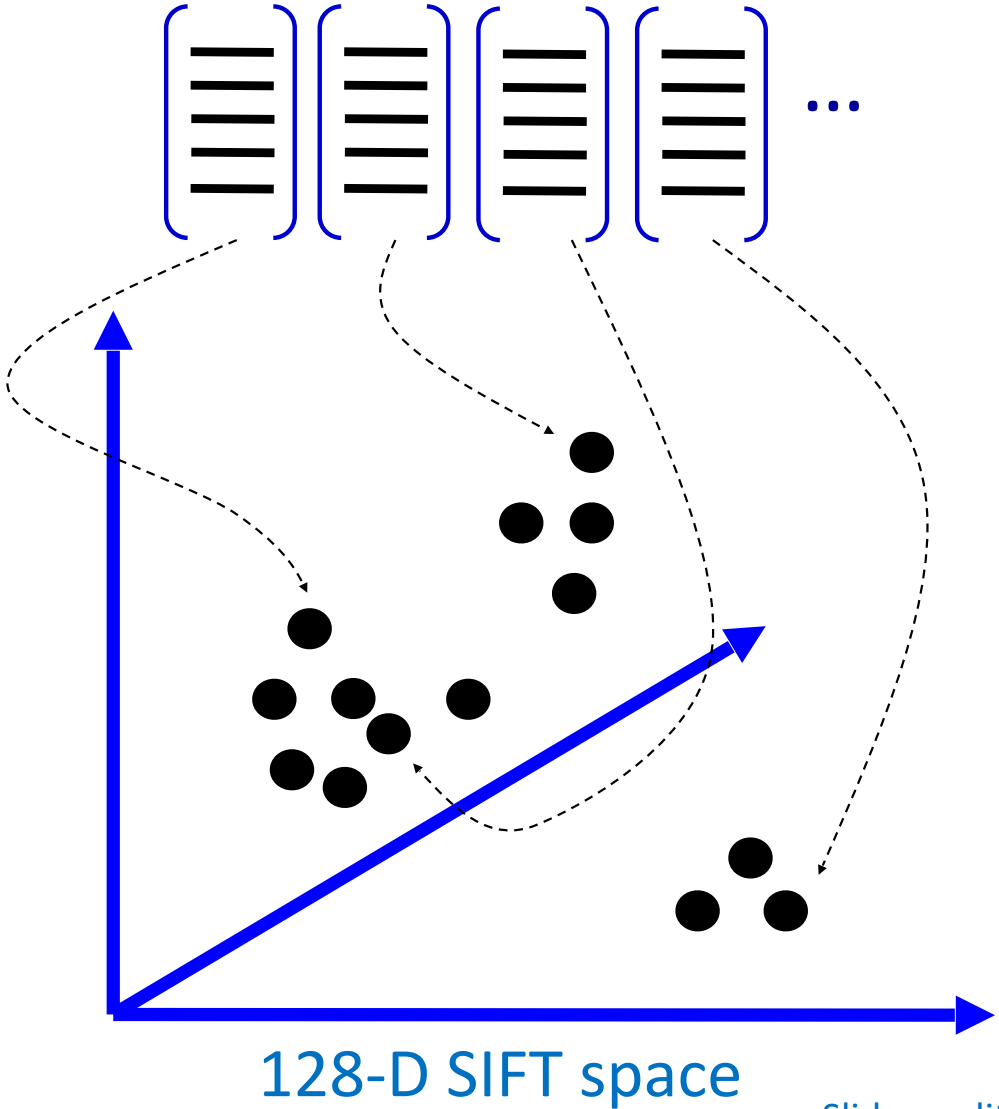
1.Feature detection and representation



Slide credit: Josef Sivic

Bag of Words

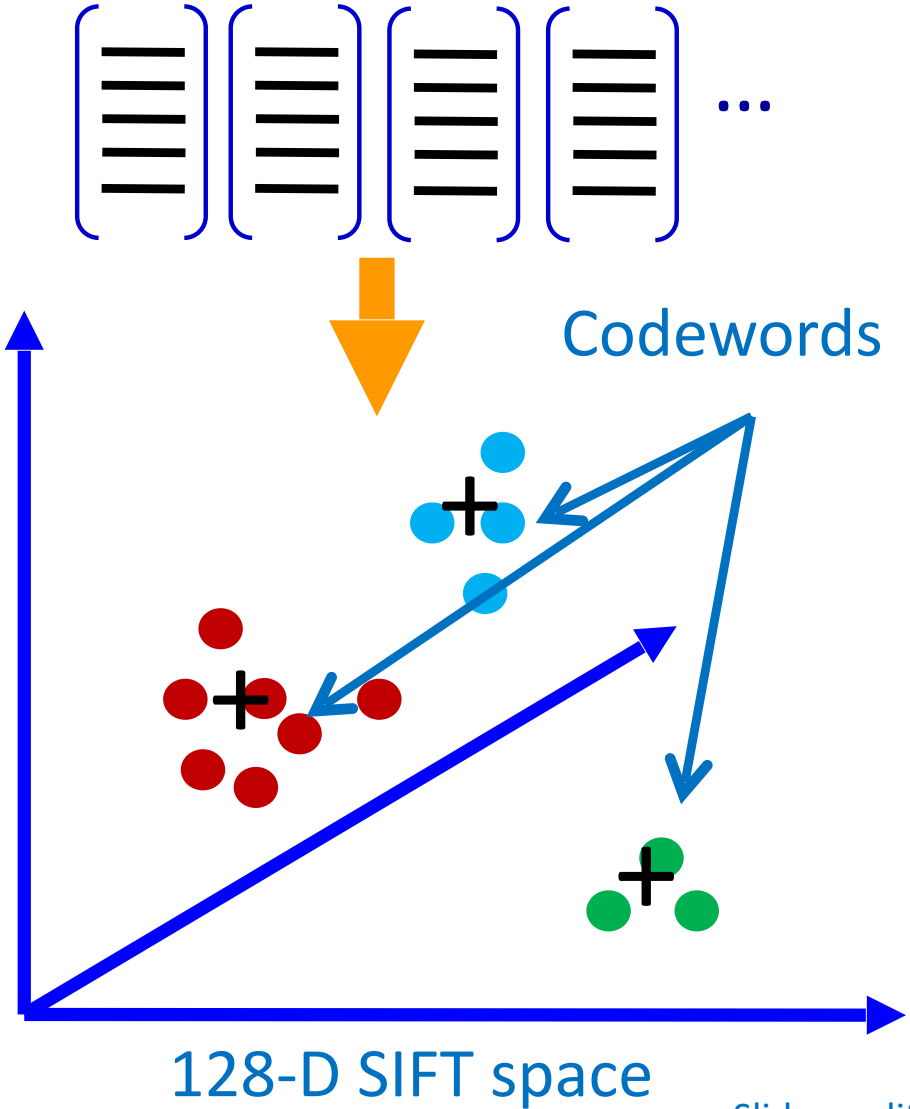
2. Codewords dictionary formation



Slide credit: Josef Sivic

Bag of Words

2. Codewords dictionary formation

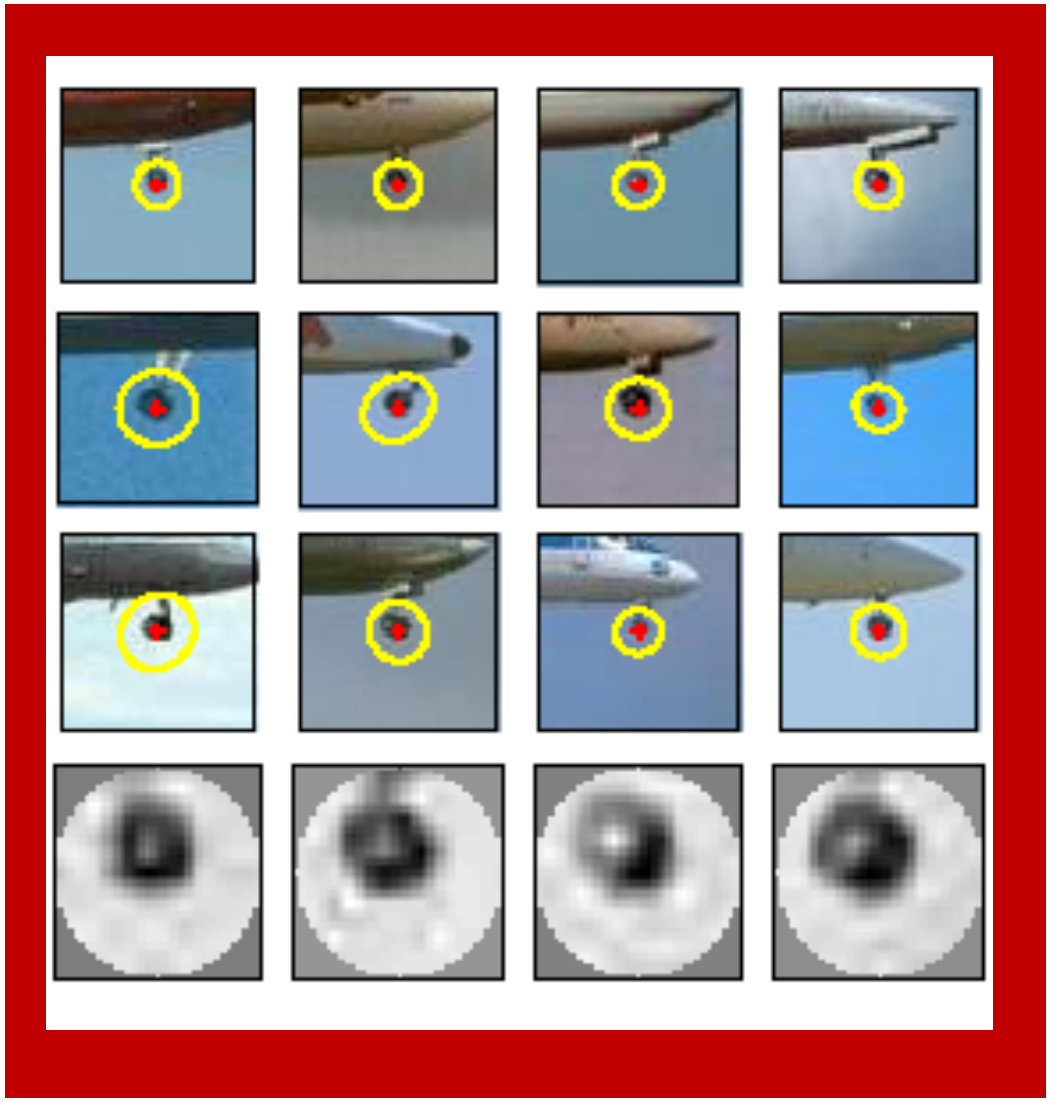


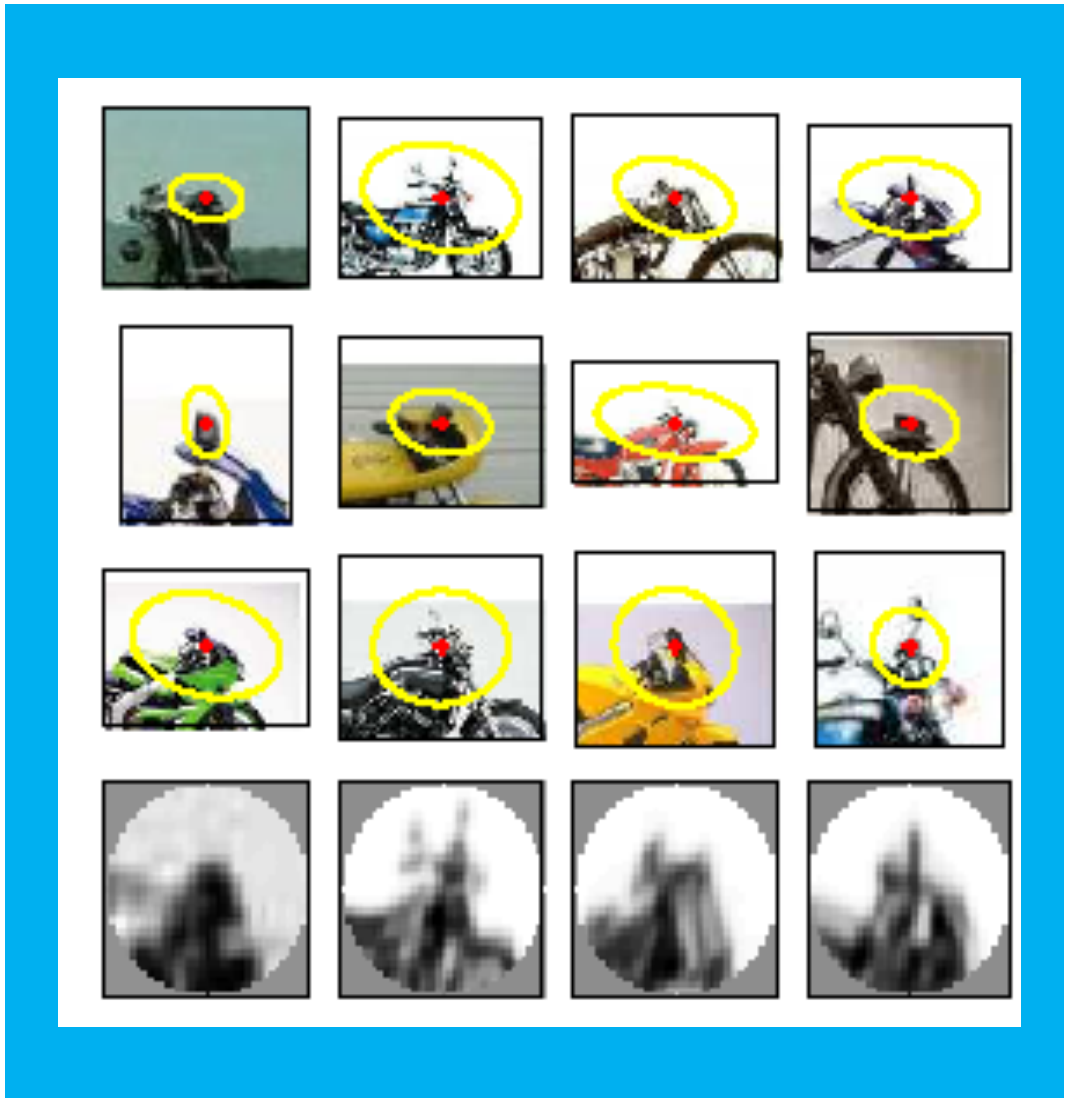
Vector quantization

Slide credit: Josef Sivic

Bag of Words

Image patch examples of codewords



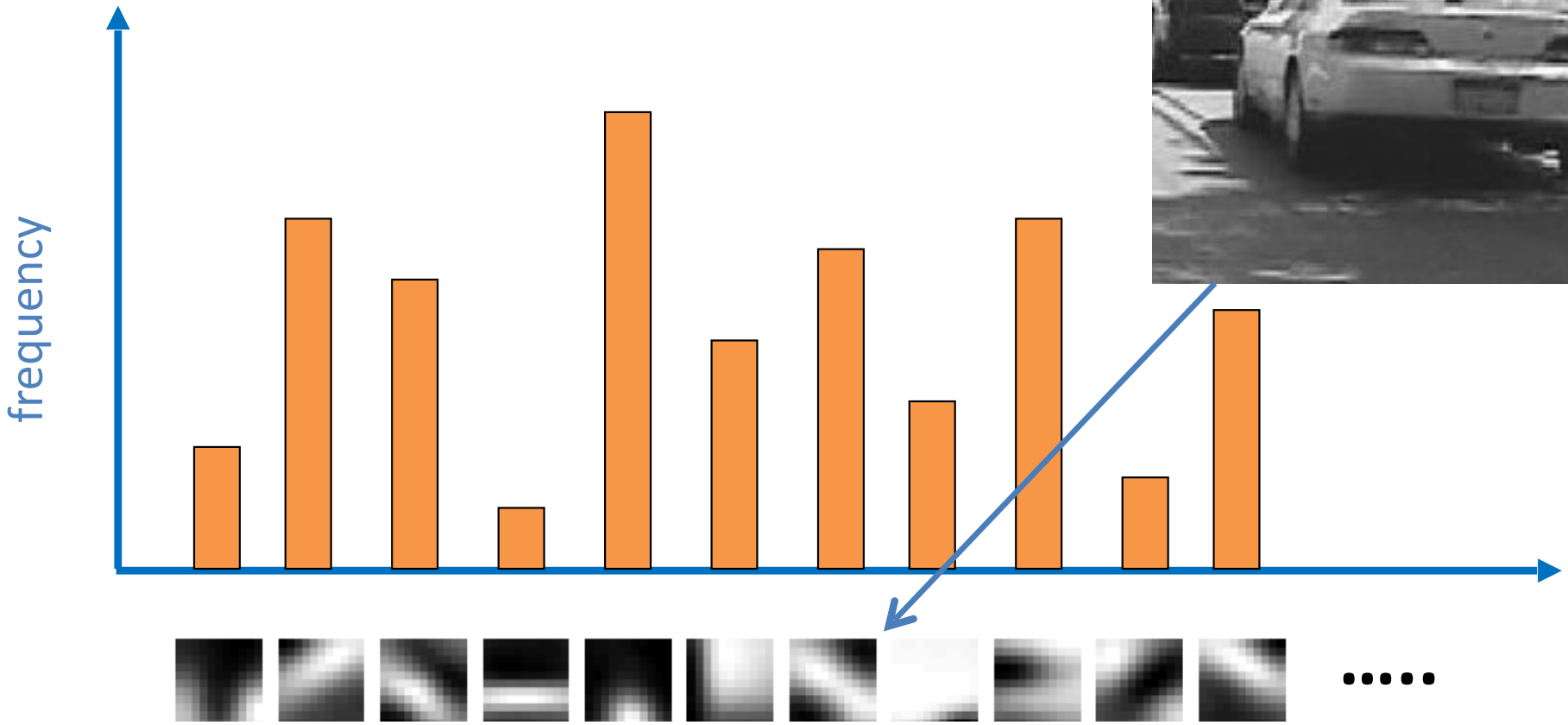


Sivic et al. 2005

Bag of Words

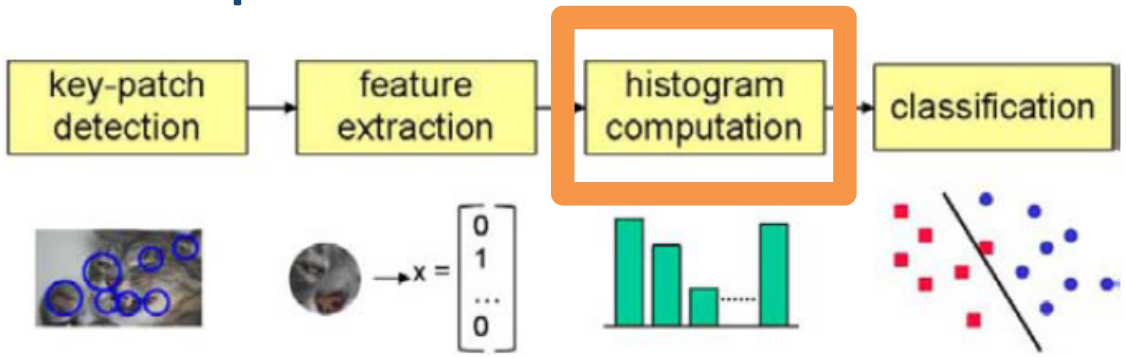
3. Image representation

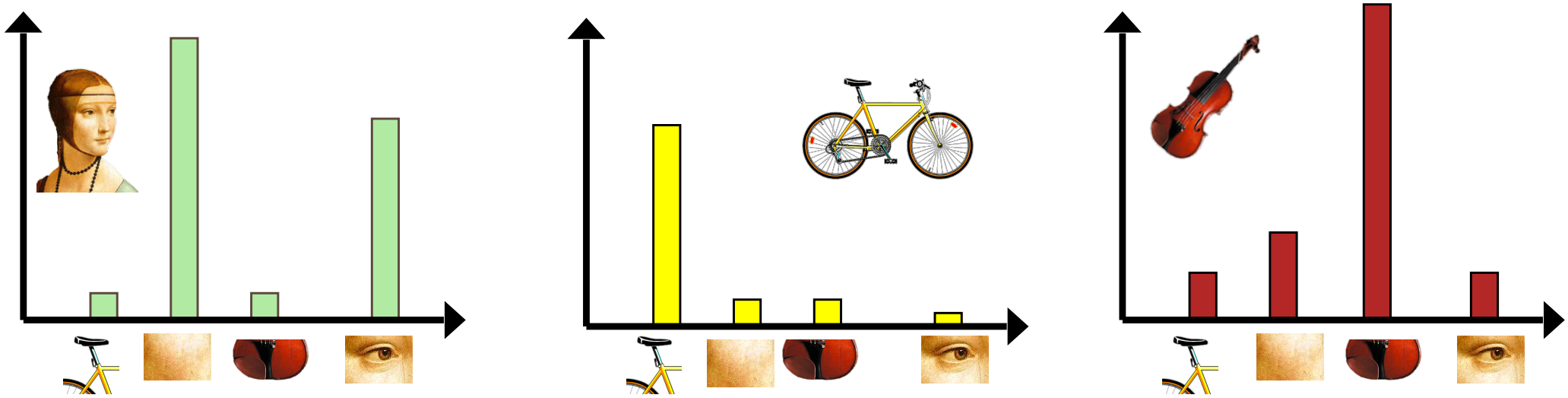
Histogram of features
assigned to each cluster



Bag of Words

- Independent features
- Histogram representation





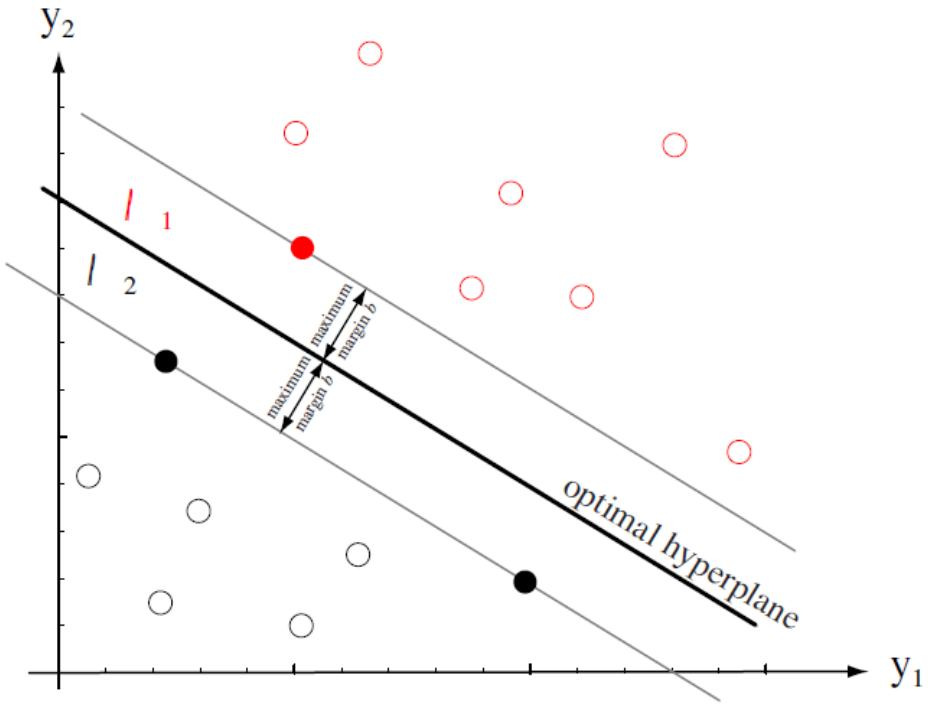
BoW as input to classifier (Discriminative or Generative)

- SVM for object classification
  - Csurka, Bray, Dance & Fan, 2004

- Naïve Bayes

BoW as input to classifier (Discriminative)

- Support Vector Machines (SVM)
  - Represent patterns in a dimension higher than the original feature space
  - Categories are separated by a “hyperplane”

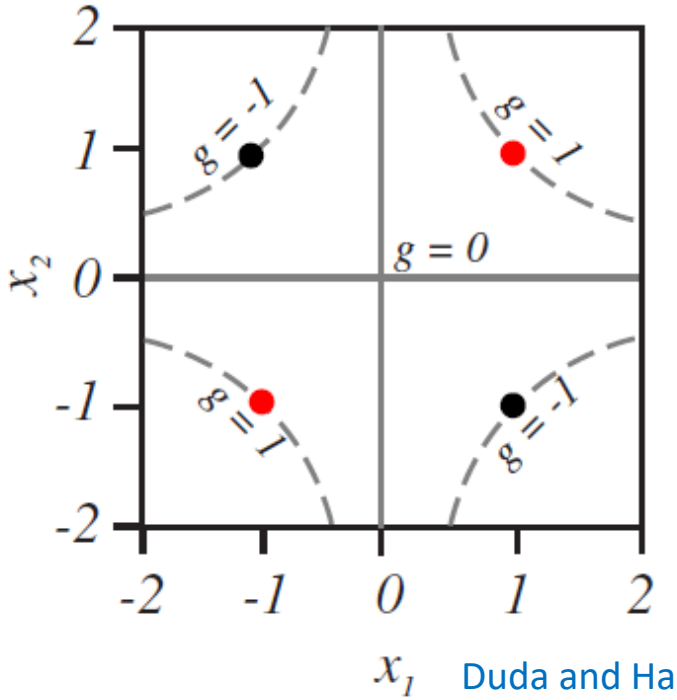


Duda and Hart, 2004

BoW as input to classifier (Discriminative)

- Support Vector Machines (SVM)
  - Represent patterns in a dimension higher than the original feature space
  - Categories are separated by a “hyperplane”

XOR problem

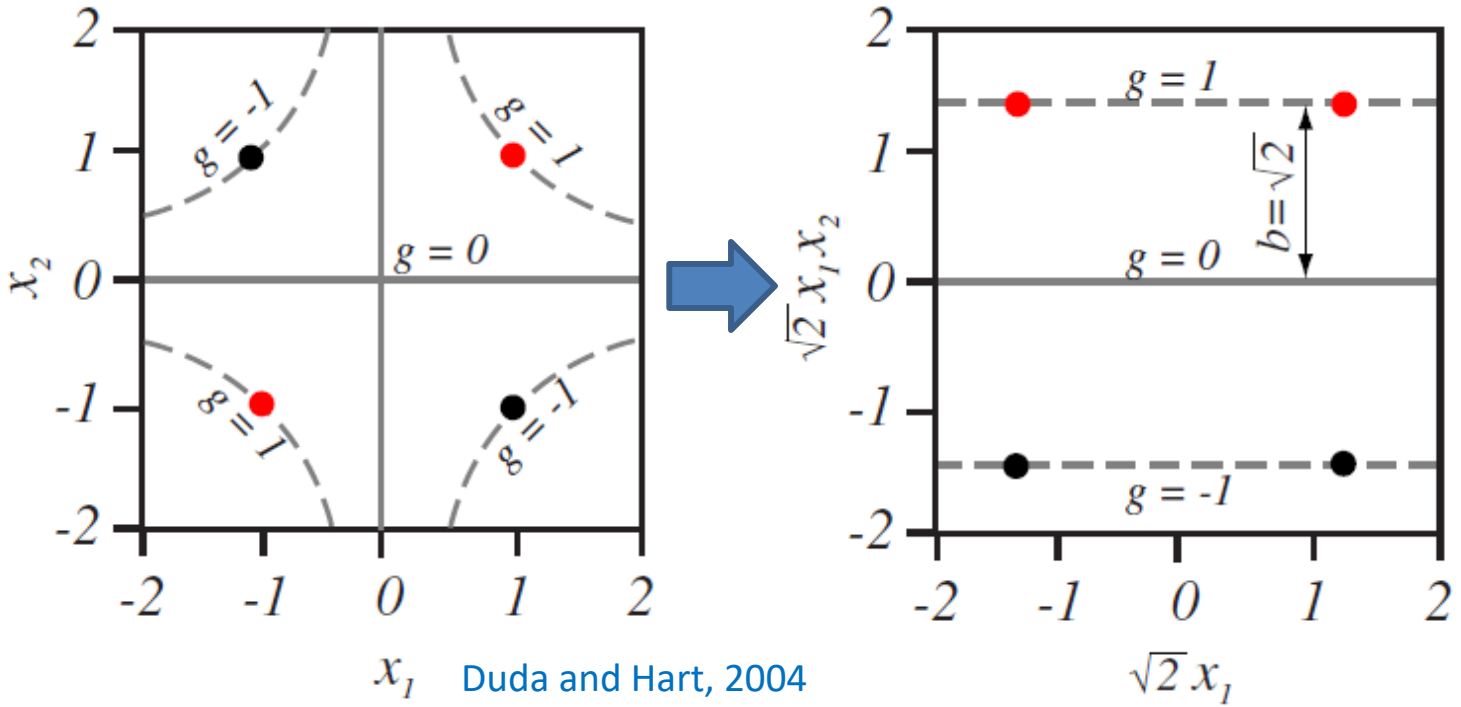


Duda and Hart, 2004

BoW as input to classifier (Discriminative)

- Support Vector Machines (SVM)
  - Represent patterns in a dimension higher than the original feature space
  - Categories are separated by a “hyperplane”

XOR problem



Duda and Hart, 2004

BoW as input to classifier (Generative)

- Naïve Bayes
  - Assumption: Features are independent of one another
  - Training: Estimate parameters of probability distribution
  - Prediction: Compute the posterior probability

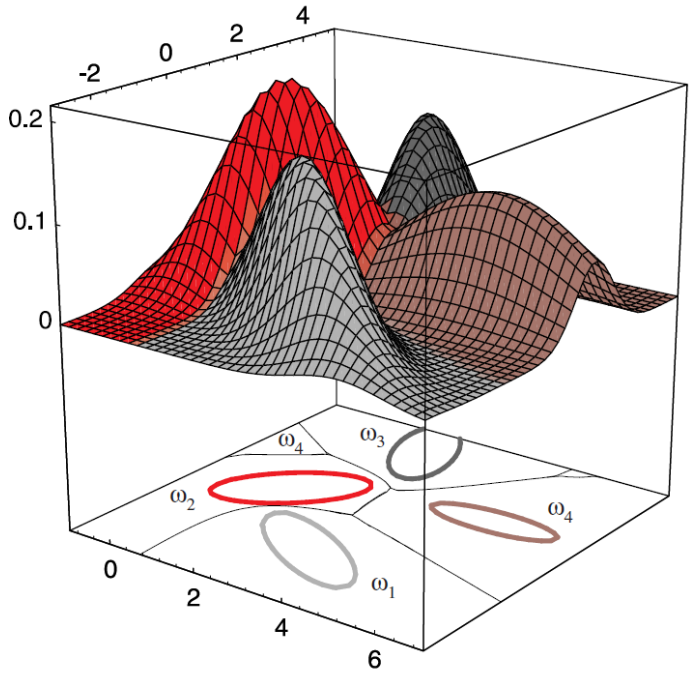
$$\frac{p(zebra \mid image)}{p(no\ zebra \mid image)} = \frac{p(image \mid zebra)}{p(image \mid no\ zebra)} \cdot \frac{p(zebra)}{p(no\ zebra)}$$

Duda and Hart, 2004

BoW as input to classifier (Generative)

- Naïve Bayes
  - Assumption: Features are independent of one another
  - Training: Estimate parameters of probability distribution
  - Prediction: Compute the posterior probability

$$\frac{p(zebra \mid image)}{p(no\ zebra \mid image)} = \frac{p(image \mid zebra)}{p(image \mid no\ zebra)} \cdot \frac{p(zebra)}{p(no\ zebra)}$$



Duda and Hart, 2004

BoW as input to classifier (Generative)

- Naïve Bayes: Example. Training set.

sex	height (feet)	weight (lbs)	foot size(inches)
male	6	180	12
male	5.92 (5'11")	190	11
male	5.58 (5'7")	170	12
male	5.92 (5'11")	165	10
female	5	100	6
female	5.5 (5'6")	150	8
female	5.42 (5'5")	130	7
female	5.75 (5'9")	150	9

Duda and Hart, 2004

BoW as input to classifier (Generative)

- Naïve Bayes: Example. Parameters.

sex	mean (height)	variance (height)	mean (weight)	variance (weight)	mean (foot size)	variance (foot size)
male	5.855	3.5033e-02	176.25	1.2292e+02	11.25	9.1667e-01
female	5.4175	9.7225e-02	132.5	5.5833e+02	7.5	1.6667e+00

Duda and Hart, 2004

BoW as input to classifier (Generative)

- Naïve Bayes: Example. Test.

sex	height (feet)	weight (lbs)	foot size(inches)
sample	6	130	8

Duda and Hart, 2004

BoW as input to classifier (Generative)

- Naïve Bayes: Example. Test.

sex	height (feet)	weight (lbs)	foot size(inches)
sample	6	130	8

$$posterior(male) = \frac{P(male)\,p(height \mid male)\,p(weight \mid male)\,p(footsize \mid male)}{evidence}$$

$$evidence = P(male)\,p(height \mid male)\,p(weight \mid male)\,p(footsize \mid male)$$
$$+\,P(female)\,p(height \mid female)\,p(weight \mid female)\,p(footsize \mid female)$$

Duda and Hart, 2004

BoW as input to classifier (Generative)

- Naïve Bayes: Example. Test.

sex	height (feet)	weight (lbs)	foot size(inches)
sample	6	130	8

$$posterior(male) = \frac{P(male)\,p(height \mid male)\,p(weight \mid male)\,p(footsize \mid male)}{evidence}$$

$$p(height \mid male) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(6-\mu)^2}{2\sigma^2}\right) \approx 1.5789$$

Duda and Hart, 2004

BoW as input to classifier (Generative)

- Naïve Bayes: Example. Test.

sex	height (feet)	weight (lbs)	foot size(inches)
sample	6	130	8

$$posterior(male) = \frac{P(male)\,p(height \mid male)\,p(weight \mid male)\,p(footsize \mid male)}{evidence}$$

$$p(height \mid male) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(6-\mu)^2}{2\sigma^2}\right) \approx 1.5789$$

- Same for the rest, posterior(male)=6.2e-09
- posterior(female)=5.4e-4

Duda and Hart, 2004

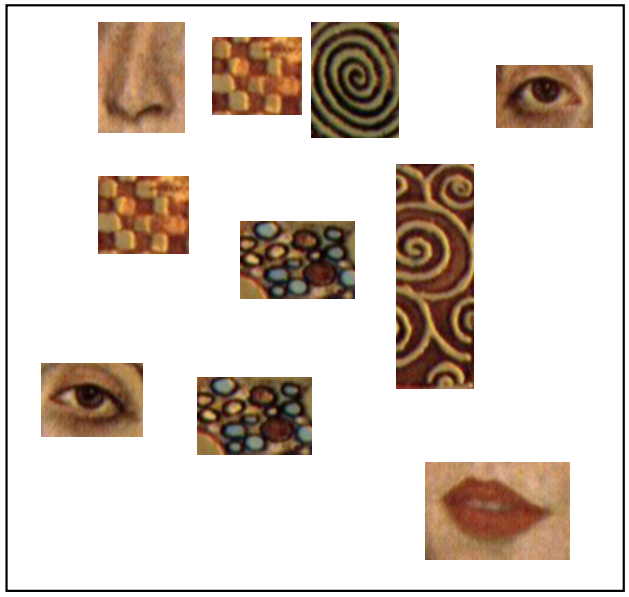
Classical Methods

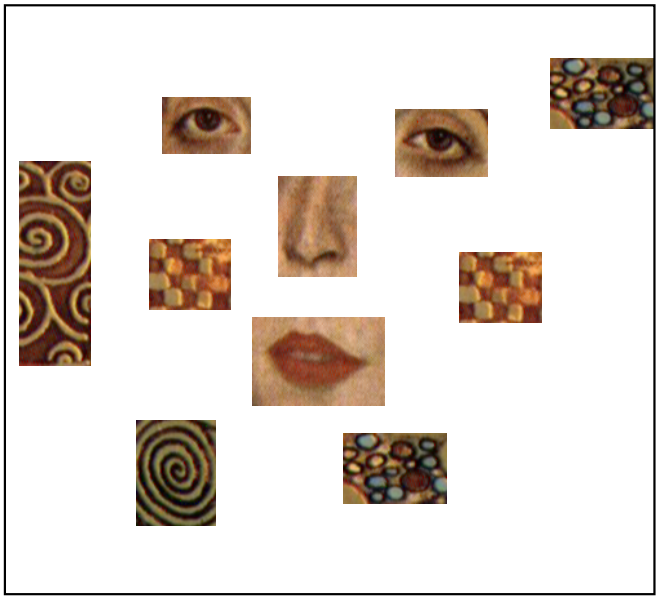
1. Treat a feature vectors for standard classifier (e.g. SVM)
   - Bag of words approaches

2. **Parts-based representation**
   - **Decompose scene/object**
   - Hierarchical models

3. Scene

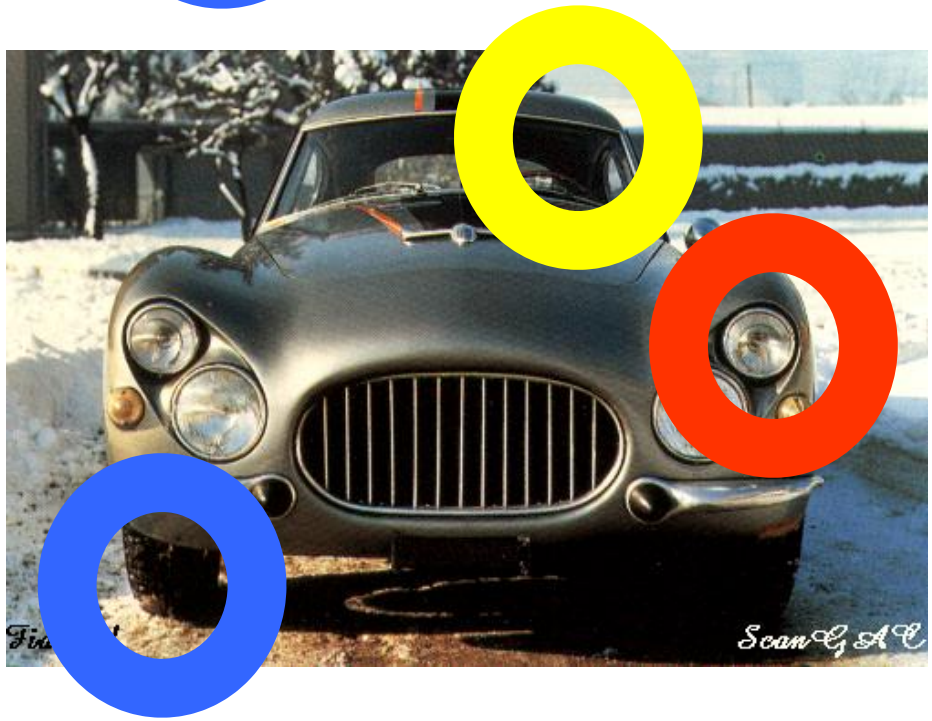
Problem with bag-of-words





- All have equal probability for bag-of-words methods
- Location information is important
- BoW + location still doesn’t give correspondence

Model: Parts and Structure



Representation

- Object as set of parts
  - Generative representation
- Model:
  - Relative locations between parts
  - Appearance of part
- Issues:
  - How to model location
  - How to represent appearance
  - How to handle occlusion/clutter

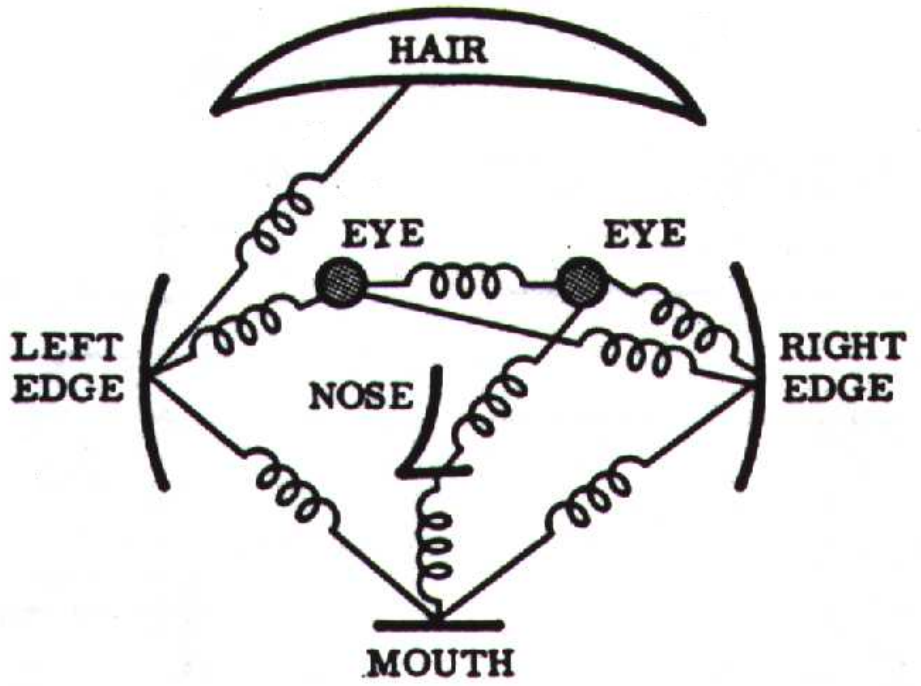
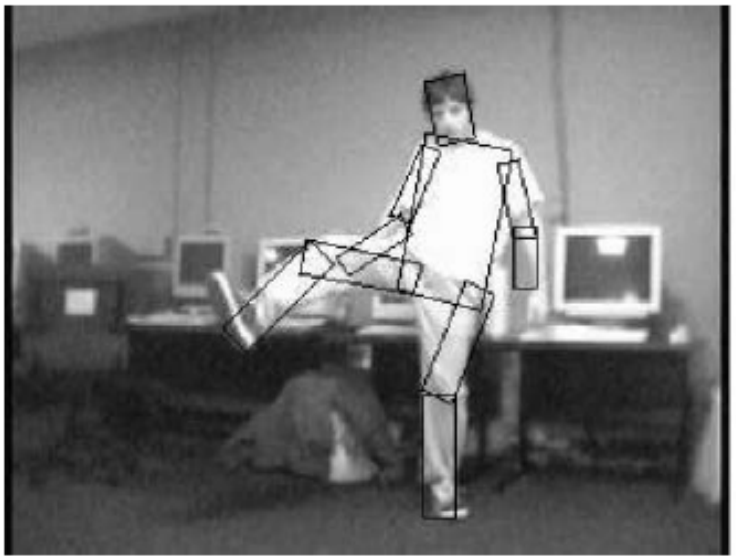
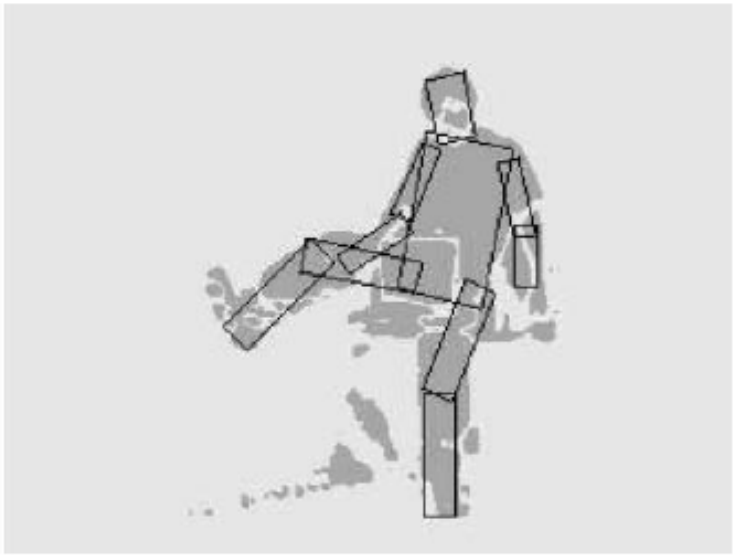


Figure from [Fischler & Elschlager 73]

Representation

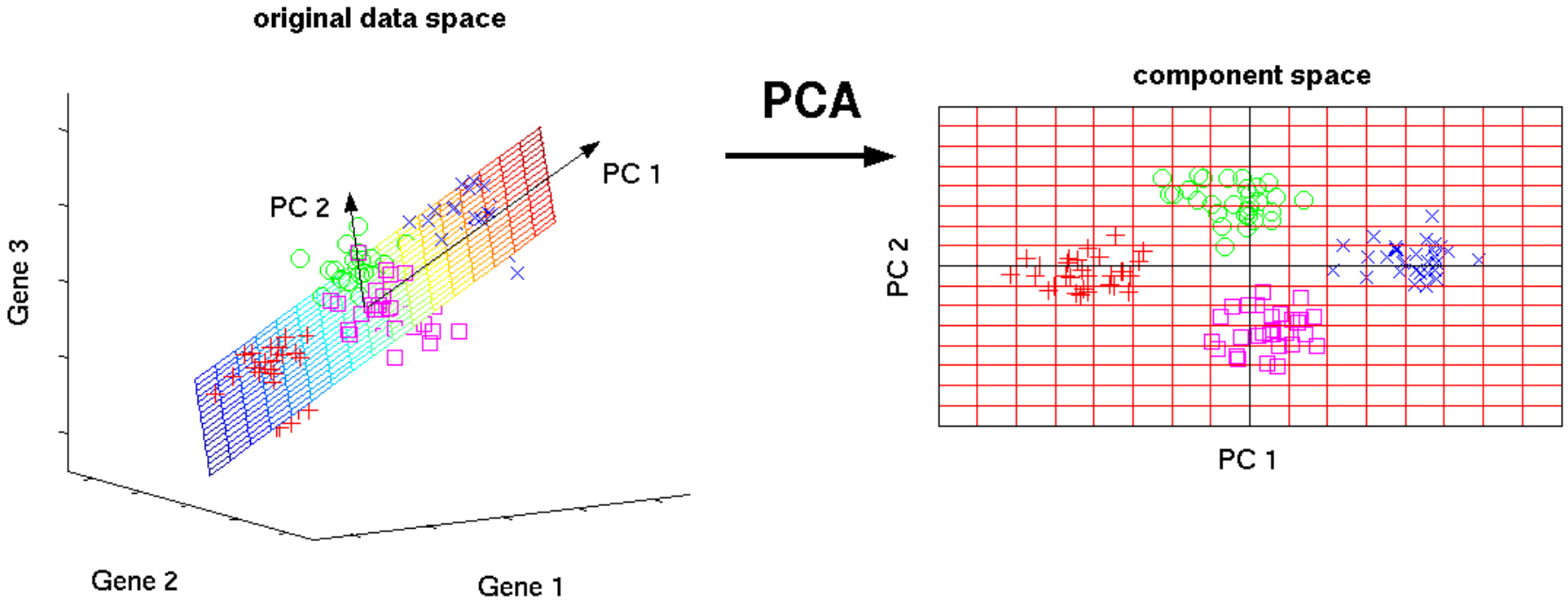
- Object as set of parts
  - Generative representation
- Model:
  - Relative locations between parts
  - Appearance of part
- Issues:
  - How to model location
  - How to represent appearance
  - How to handle occlusion/clutter





Representation

- PCA (Principal Component Analysis)
  - Seeks to represent the *d-dimensional* data in a lower-dimensional space
  - The goal is to represent data in a space that best describes the variation in a sum-squared error sense



Sparse representation

- Computationally tractable (10^5 pixels → 10^1 -- 10^2 parts)
- Generative representation of class
- Avoid modeling global variability
- Success in specific object recognition

- Throw away most image information
- Parts need to be distinctive to separate from other classes

Learn Appearance

- Generative models of appearance
  - Can learn with little supervision
  - E.g. Fergus et al' 03
- Discriminative training of part appearance model
  - SVM part detectors
  - Felzenszwalb, Mcallester, Ramanan, CVPR 2008
  - Much better performance

Different connectivity structures

O(N⁶) O(N²) O(N³) O(N²)

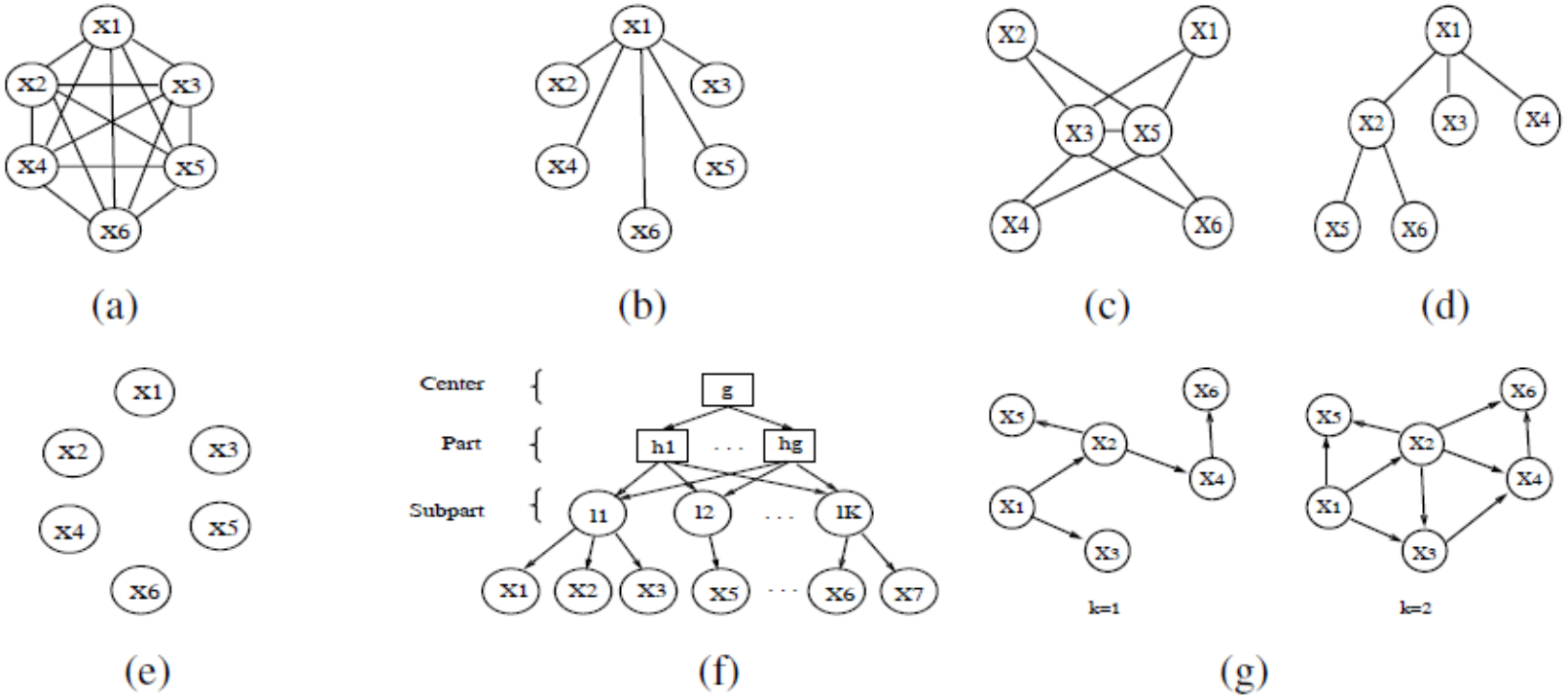
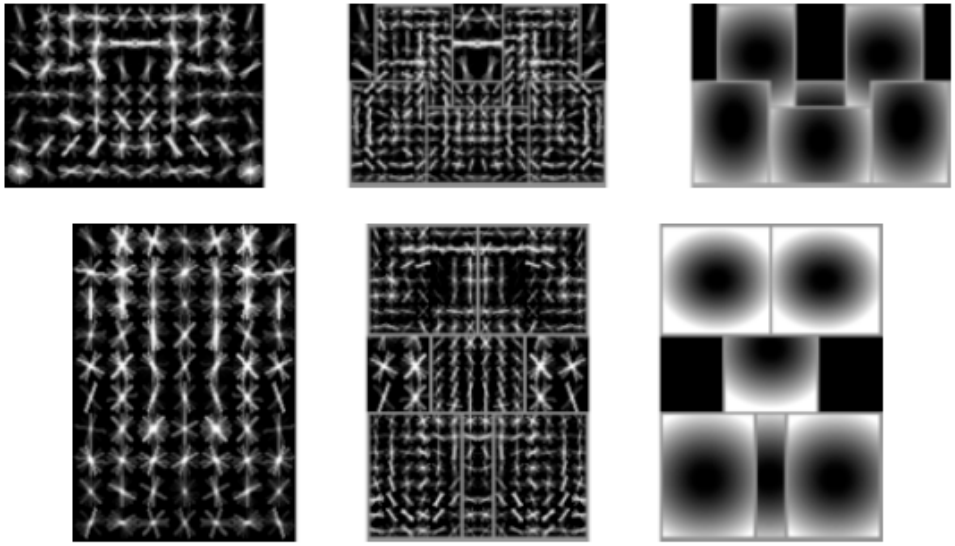


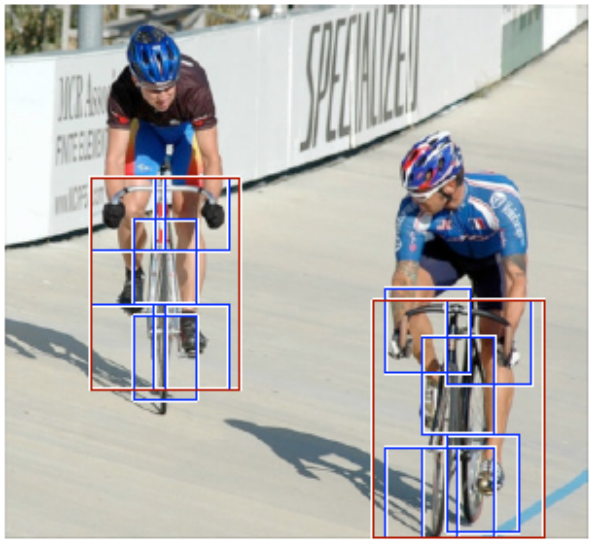
Figure 14.41 Graphical models for geometric spatial priors (Carneiro and Lowe 2006) © 2006 Springer: (a) constellation (Fergus, Perona, and Zisserman 2007); (b) star (Crandall, Felzenszwalb, and Huttenlocher 2005; Fergus, Perona, and Zisserman 2005); (c) k-fan ($k = 2$) (Crandall, Felzenszwalb, and Huttenlocher 2005); (d) tree (Felzenszwalb and Huttenlocher 2005); (e) bag of features (Csurka, Dance, Fan *et al.* 2004); (f) hierarchy (Bouchard and Triggs 2005); (g) sparse flexible model (Carneiro and Lowe 2006).

Felzenszwalb, Mcallester, Ramanan, CVPR 2008

- 2-scale model
  - Whole object
  - Parts
- HOG representation + SVM training to obtain robust part detectors
- Distance transforms allow examination of every location in the image

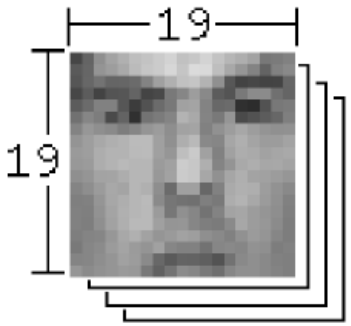


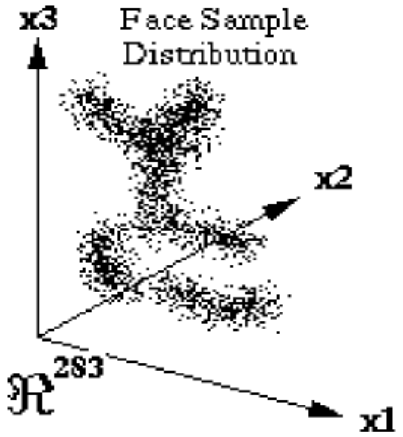
root filters coarse resolution | part filters finer resolution | deformation models

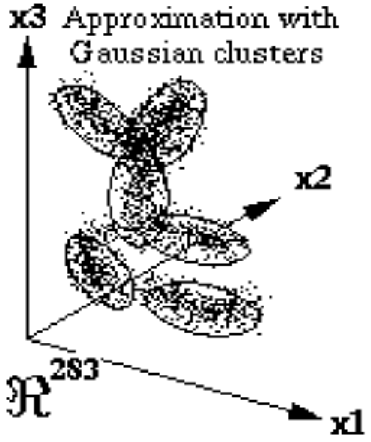


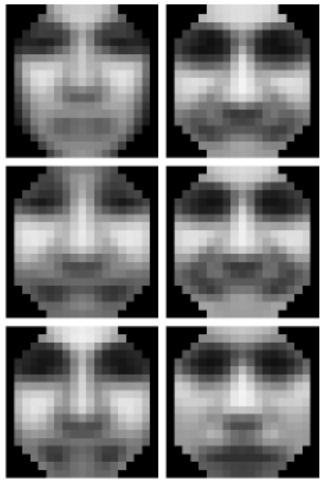
Clustering parts-based

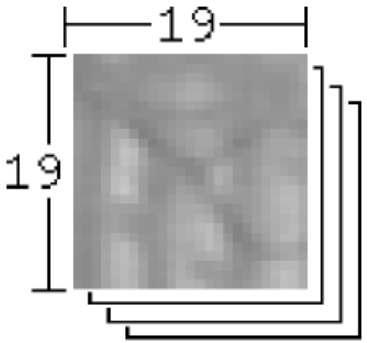
- Mixture of Gaussians and PCA
  - Sung and Poggio (1998)

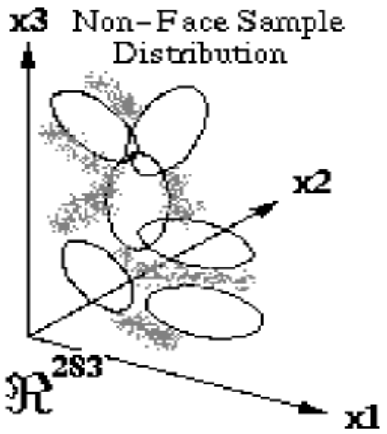


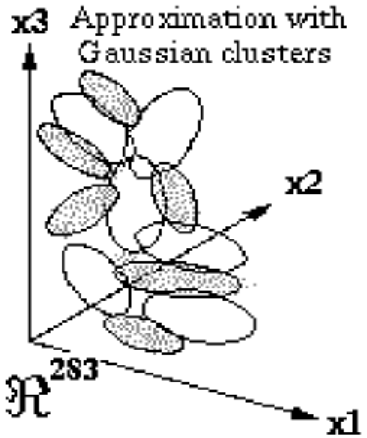


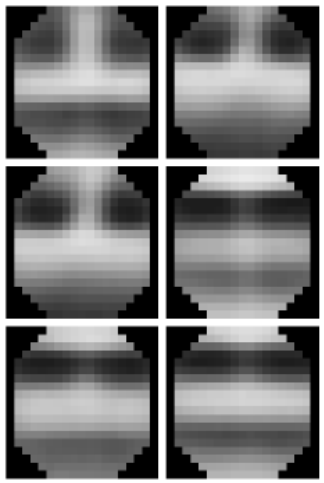



Face Centroids








Non-Face Centroids

Classical Methods

1. Treat a feature vectors for standard classifier (e.g. SVM)
   - Bag of words approaches

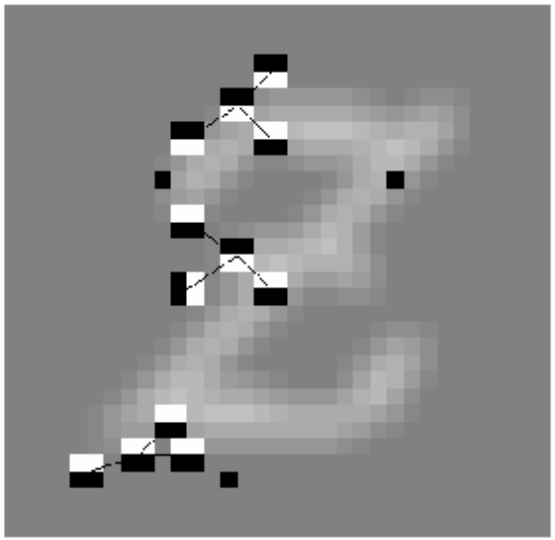
2. **Parts-based representation**
   - Decompose scene/object
   - **Hierarchical models**

3. Scene

Hierarchical Representations

- Pixels → Pixel groupings → Parts → Object

- Multi-scale approach increases number of low-level features

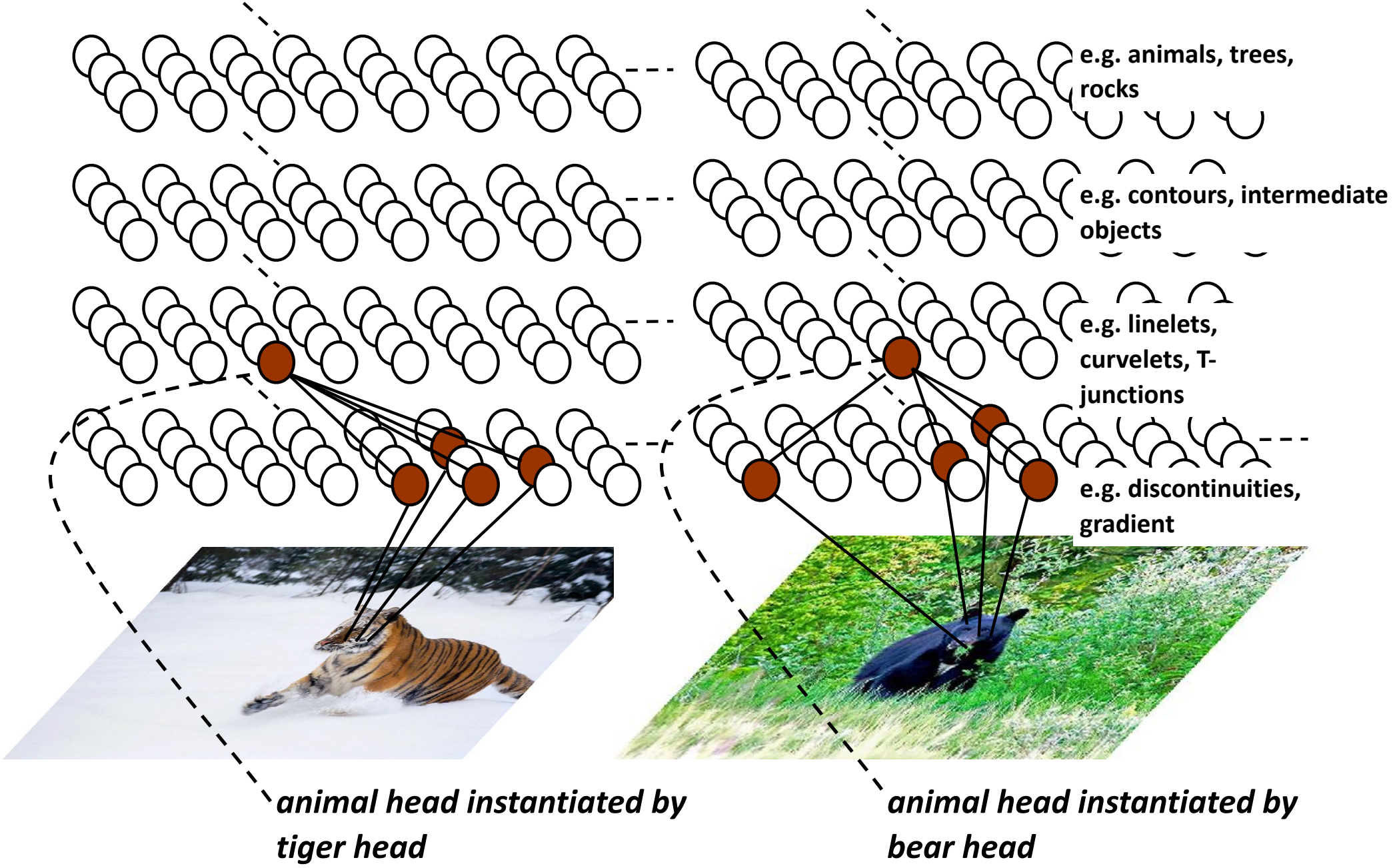


- Amit and Geman ’98
- Ullman et al.
- Bouchard & Triggs ’05
- Zhu and Mumford
- Jin & Geman ‘06
- Zhu & Yuille ’07
- Fidler & Leonardis ‘07

Images from [Amit98]

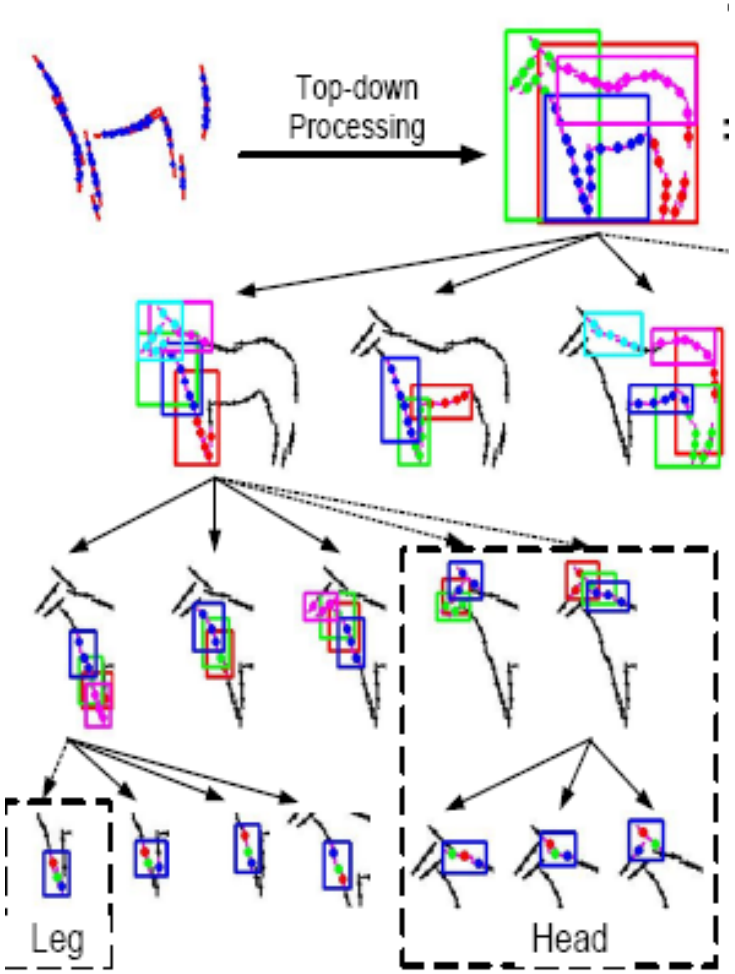
Context and Hierarchy in a Probabilistic Image Model

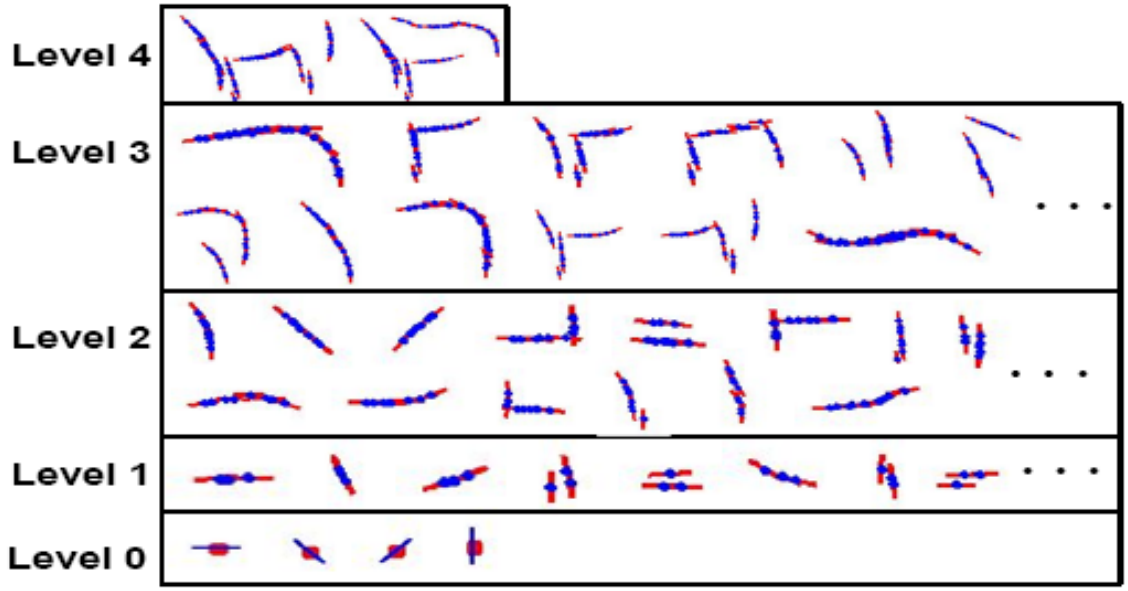
Jin & Geman (2006)



A Hierarchical Compositional System for Rapid Object Detection

Long Zhu, Alan L. Yuille, 2007.

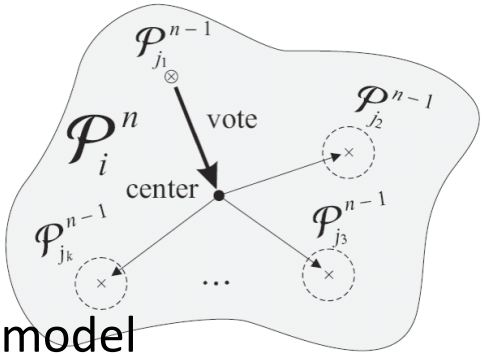




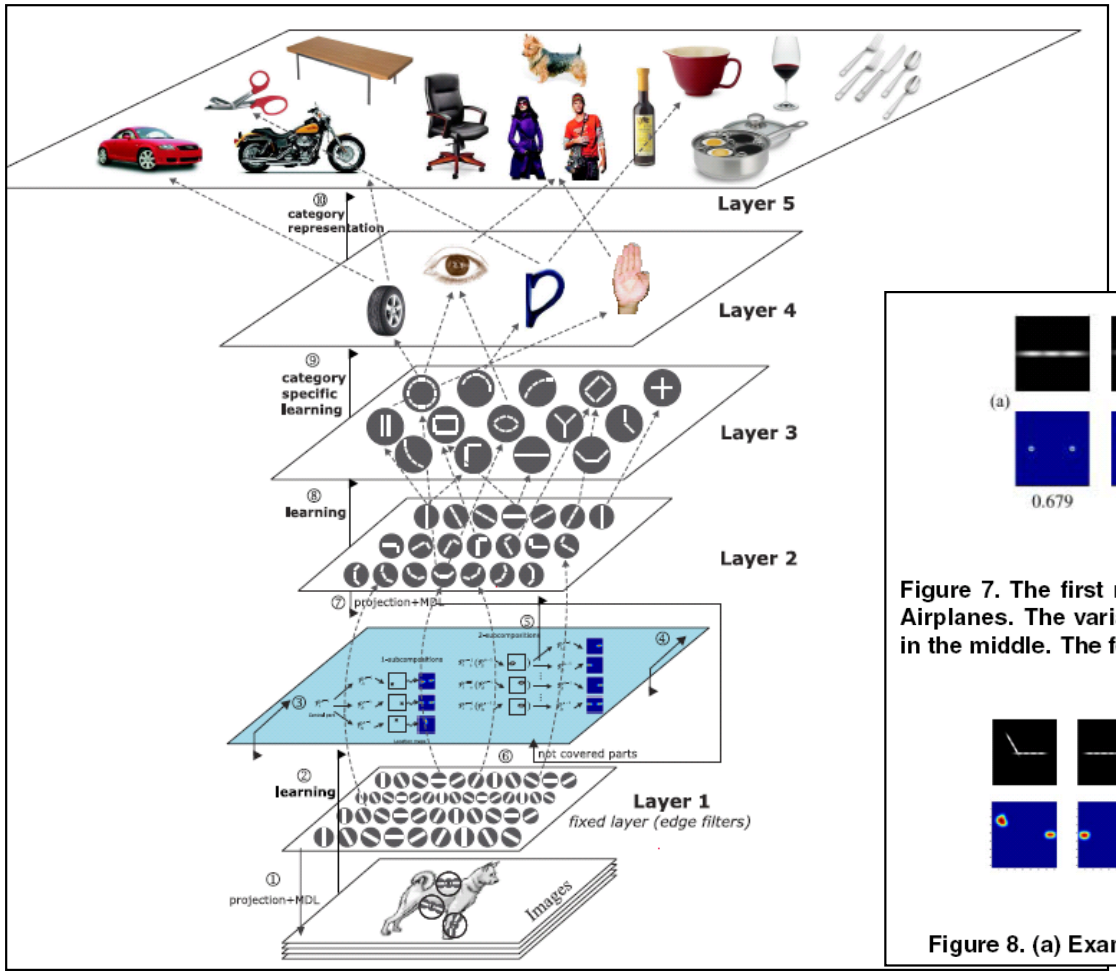
Able to learn #parts at each level

Learning a Compositional Hierarchy of Object Structure

Fidler & Leonardis, CVPR'07; Fidler, Boben & Leonardis, CVPR 2008



Parts model



The architecture

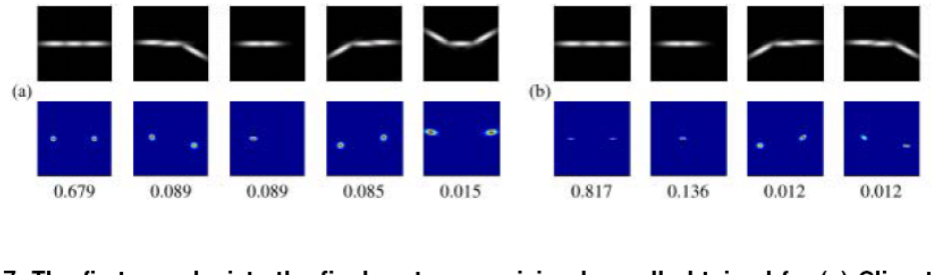


Figure 7. The first row depicts the final parts comprising Layer II obtained for (a) Cliparts and (b) Airplanes. The variances of position distributions of parts, relative to the central part, are depicted in the middle. The feature probabilities are listed in the last row.

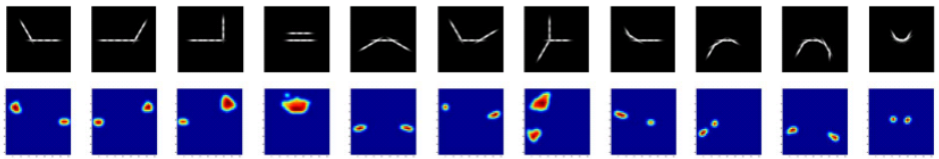


Figure 8. (a) Examples of Layer 3 parts, (b) variances of positions of the surrounding subparts

Learned parts

Classical Methods

1. Treat a feature vectors for standard classifier (e.g. SVM)
   - Bag of words approaches

2. Parts-based representation
   - Decompose scene/object
   - Hierarchical models

3. Scene

Context and Scene understanding

- Related to part-based models

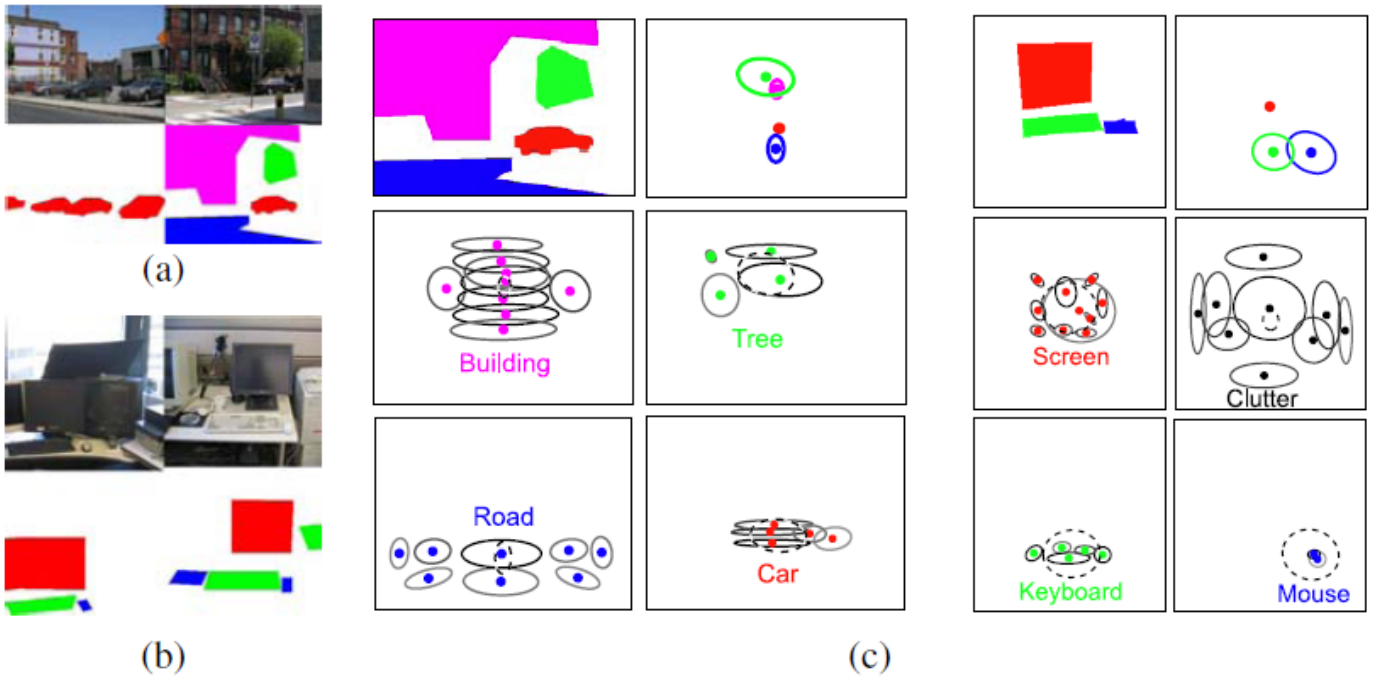


Figure 14.50 Contextual scene models for object recognition (Sudderth, Torralba, Freeman *et al.* 2008) © 2008 Springer: (a) some street scenes and their corresponding labels (magenta = buildings, red = cars, green = trees, blue = road); (b) some office scenes (red = computer screen, green = keyboard, blue = mouse); (c) learned contextual models built from these labeled scenes. The top row shows a sample label image and the distribution of the objects relative to the center red (car or screen) object. The bottom rows show the distributions of parts that make up each object.

Parts and Structure models: Summary

- Explicit notion of correspondence between image and model
- Efficient methods for large number of parts and number of positions in image
- With powerful part detectors, can get state-of-the-art performance
- Hierarchical models allow for more parts

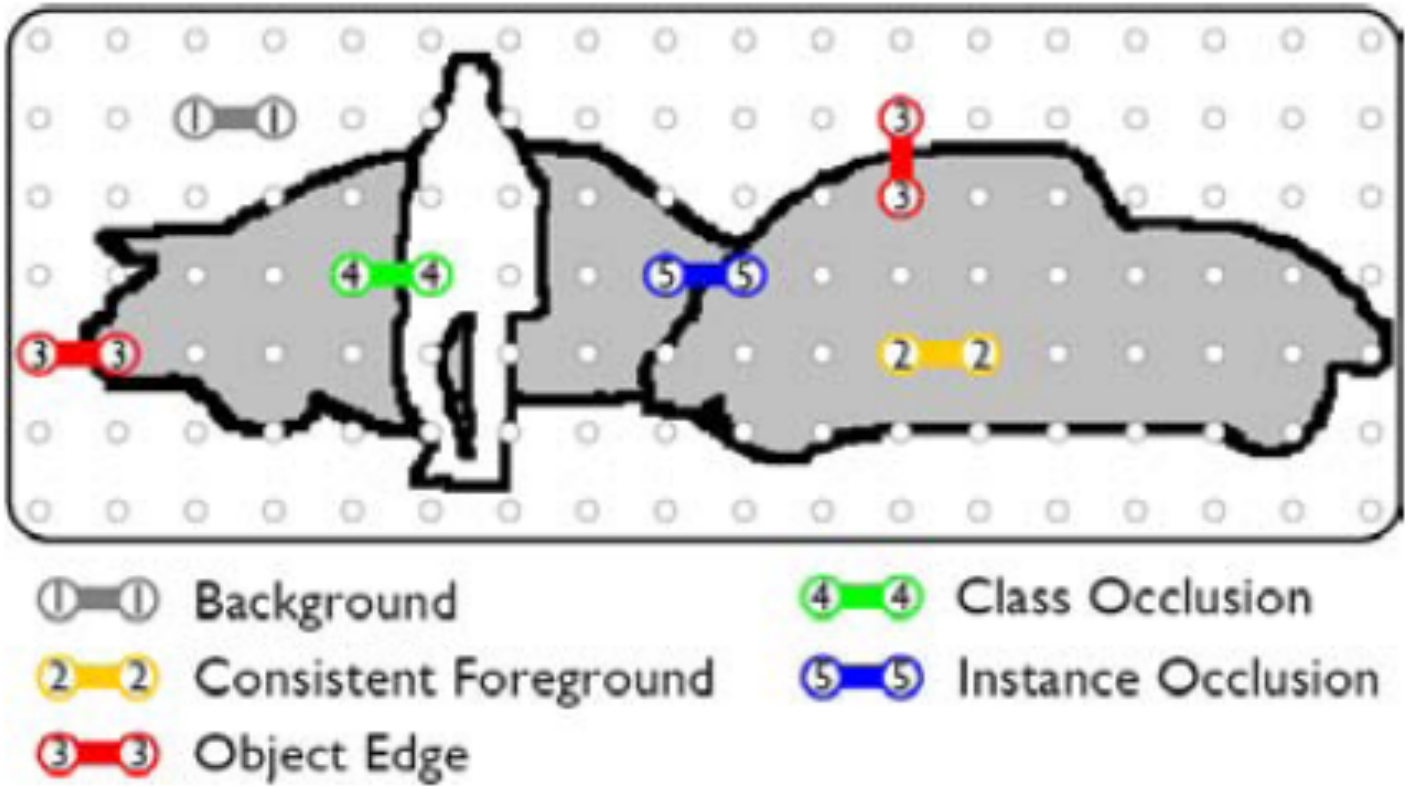
Classical Methods

1. Bag of words approaches

2. Parts and structure approaches

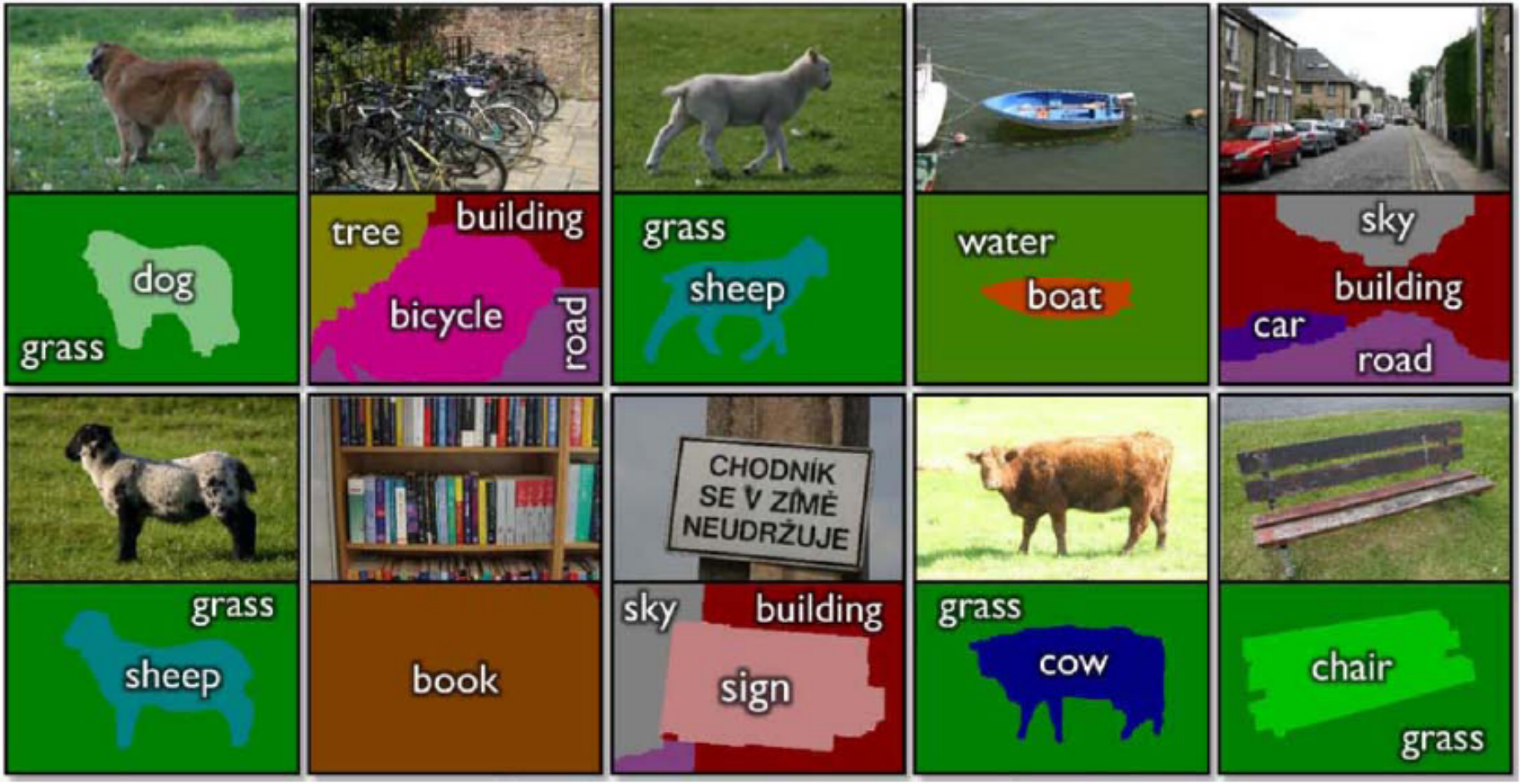
3. Recognition with segmentation

Recognition with Segmentation



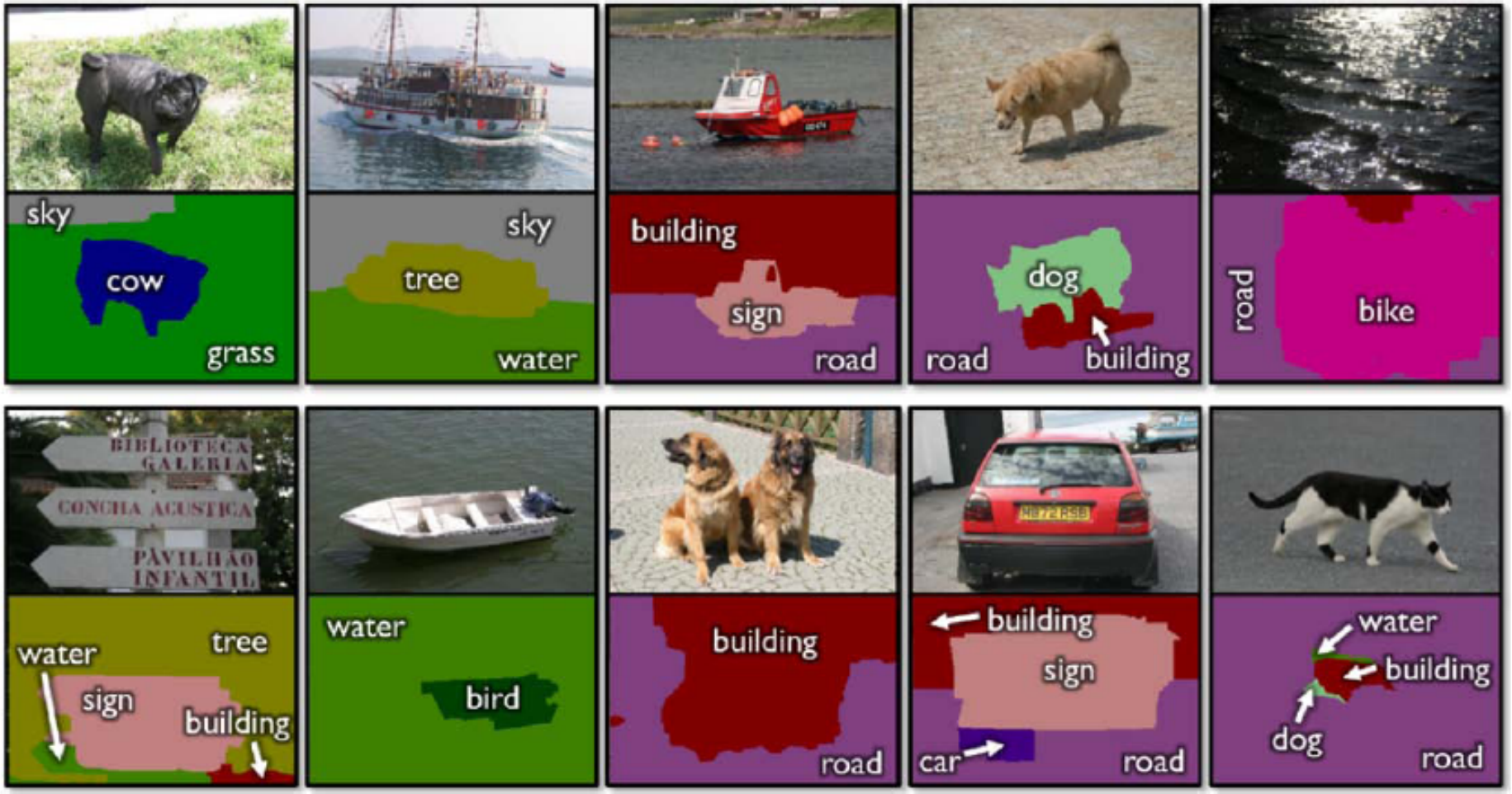
Recognition with Segmentation

- TextonBoost



Recognition with Segmentation

- TextonBoost



Summary

- Introduction
- Challenges
- Representation
- Learning
- Category recognition