

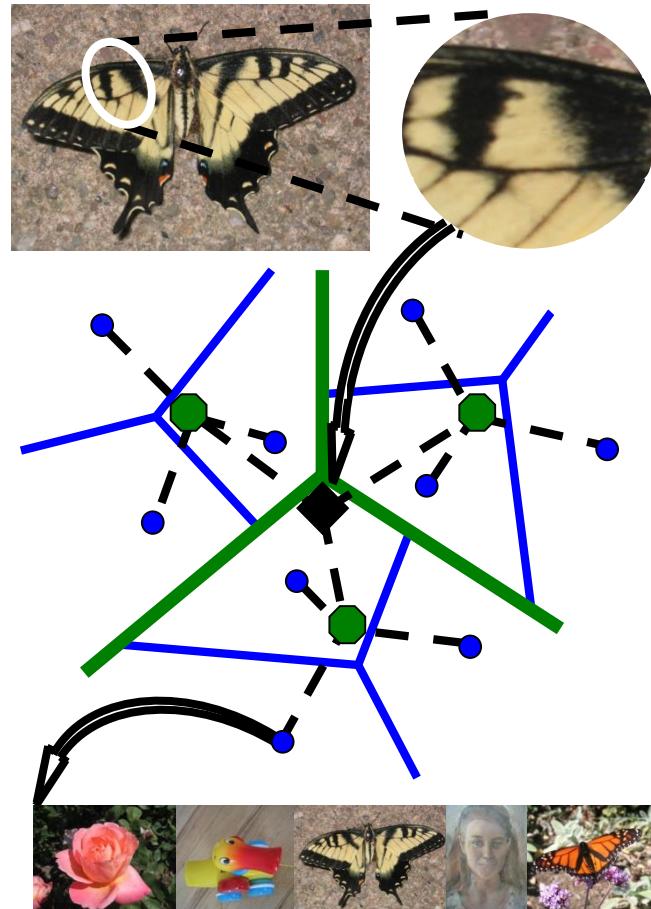
# Object Categorization

Vinay P. Namboodiri

Slide credit to Li Fei Fei, Rob  
Fergus and Antonio Torralba

# Last Class : Scalable Recognition with a Vocabulary Tree

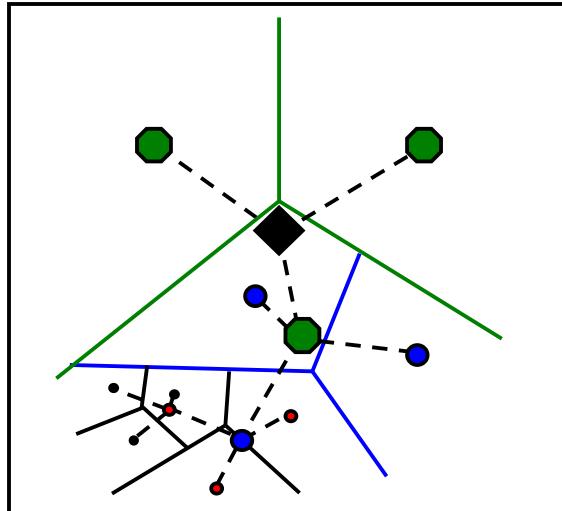
David Nistér, Henrik Stewénius



# Last Class: Take-Home Message

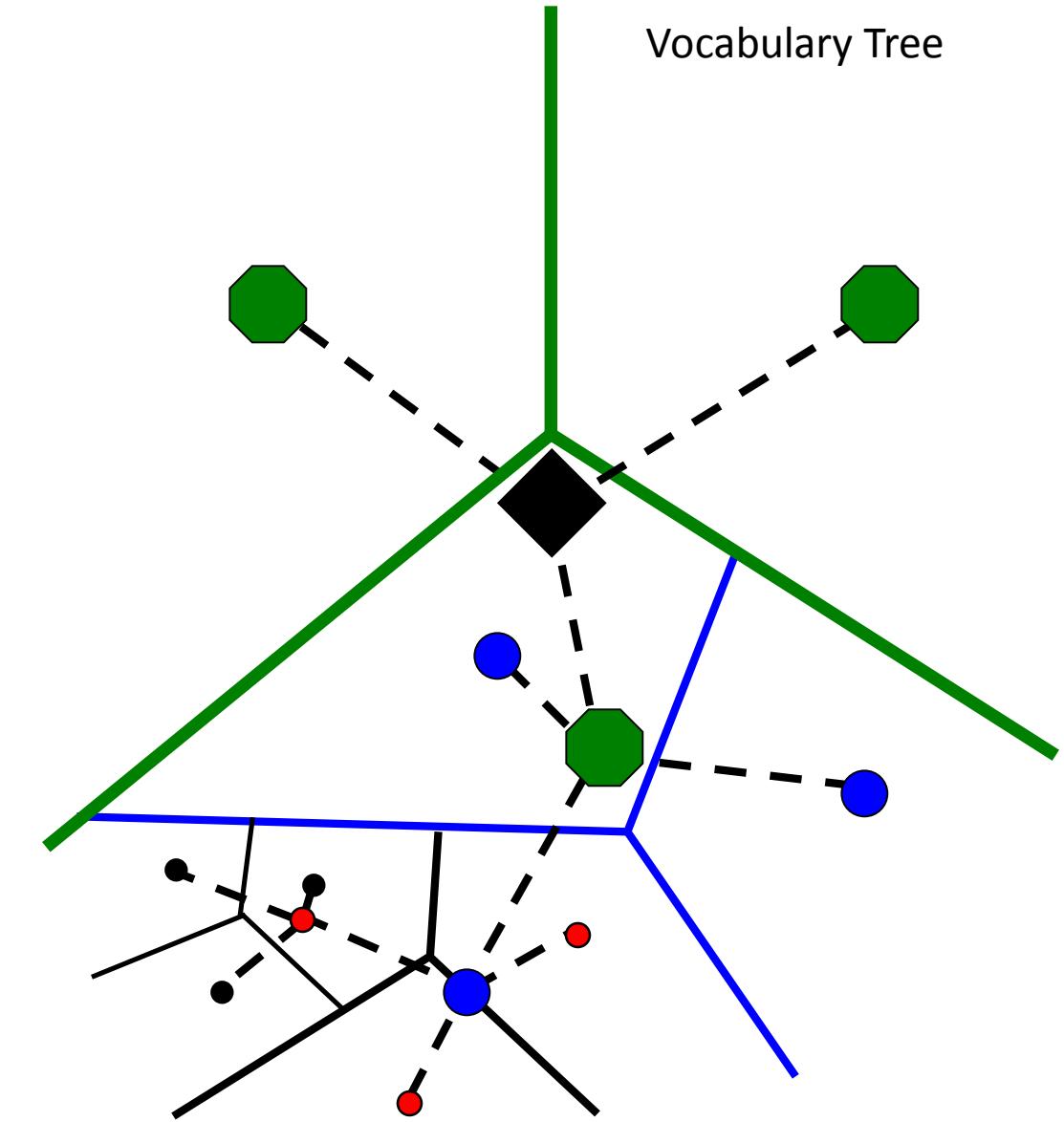
If we can get repeatable, discriminative features,

then recognition can scale to very large databases  
using the vocabulary tree and indexing approach  
described in Nistér & Stewénius CVPR 2006.



Last Class:

### Vocabulary Tree



# Scoring of Vocabulary tree

- 
- $$\bullet \quad q_i = n_i \cdot w_i$$
- Where  $n_i$  is the number of descriptors going to node  $i$
- $w_i = \ln N/N_i$  where
- $N$  is number of images in the database
- $N_i$  is the number of images in the database with atleast one descriptor vector path through node  $i$

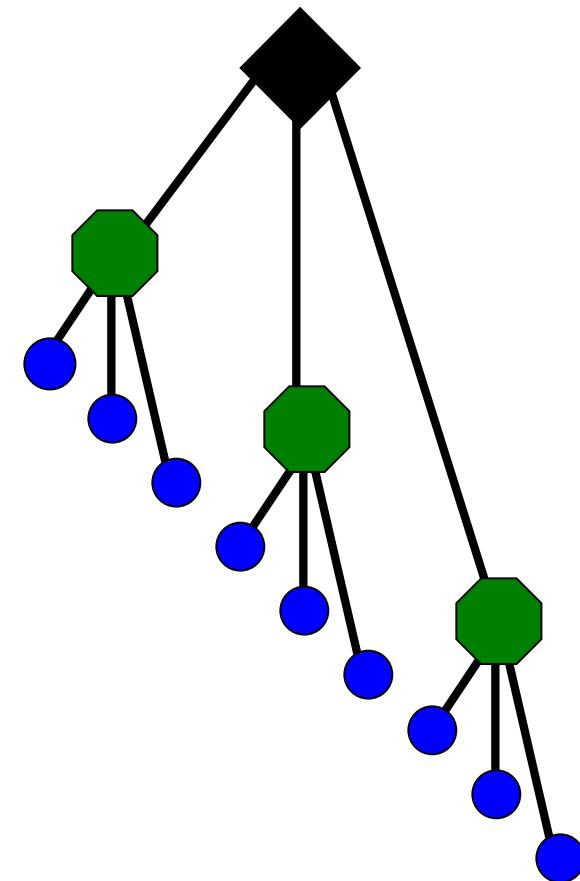
# Scoring

- Distance obtained by normalized distance between database and query vectors

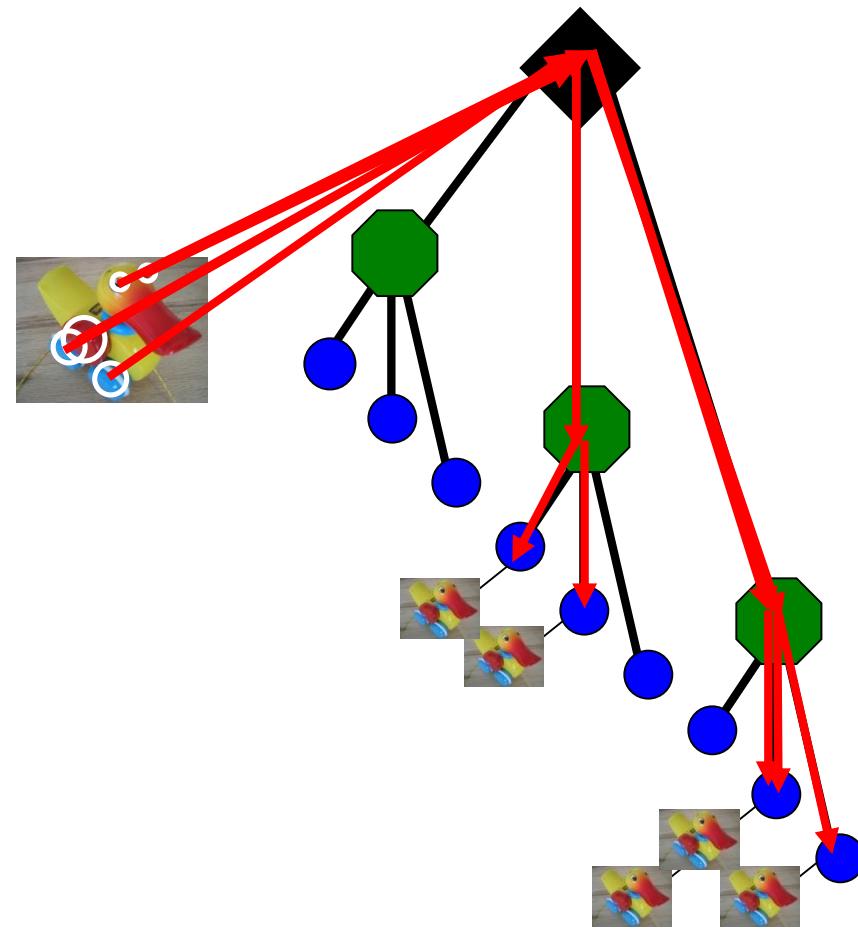
$$s(q, d) = \left\| \frac{d}{\|d\|} - \frac{q}{\|q\|} \right\|$$

Efficiently done by using inverted lists

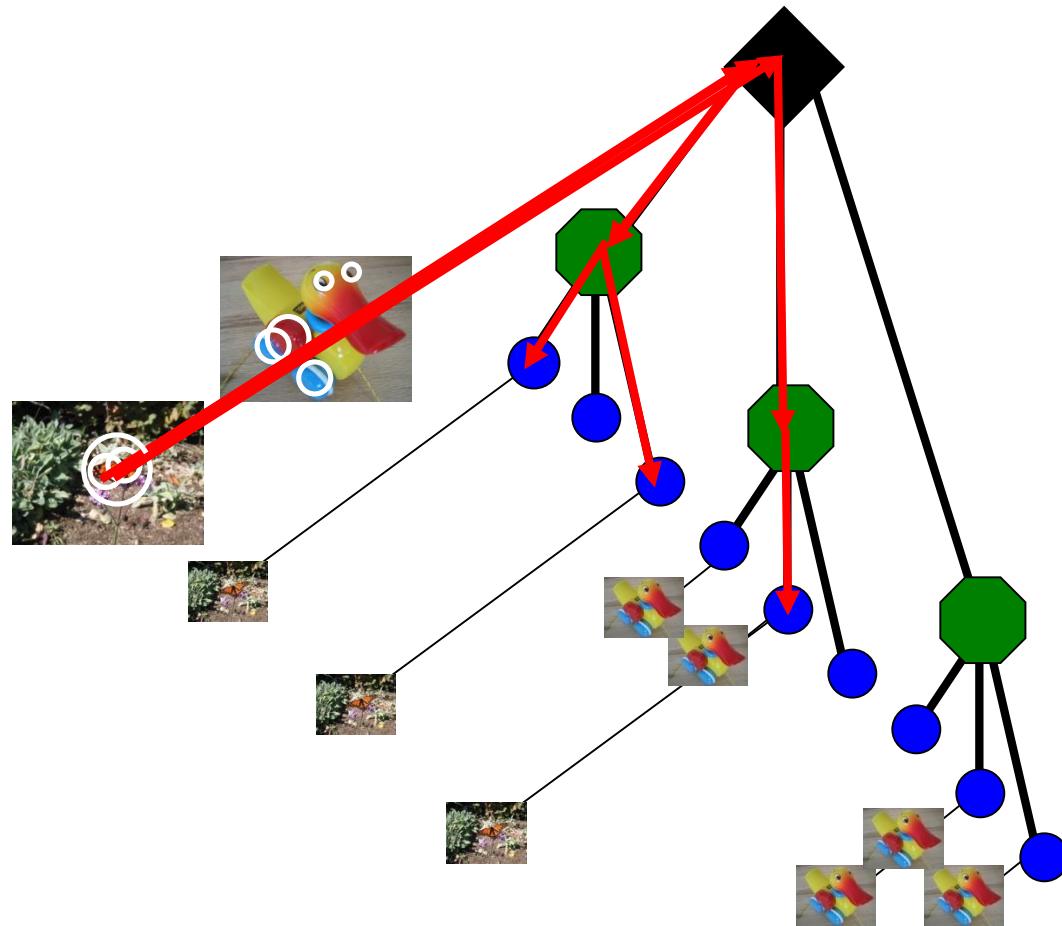
Last Class:



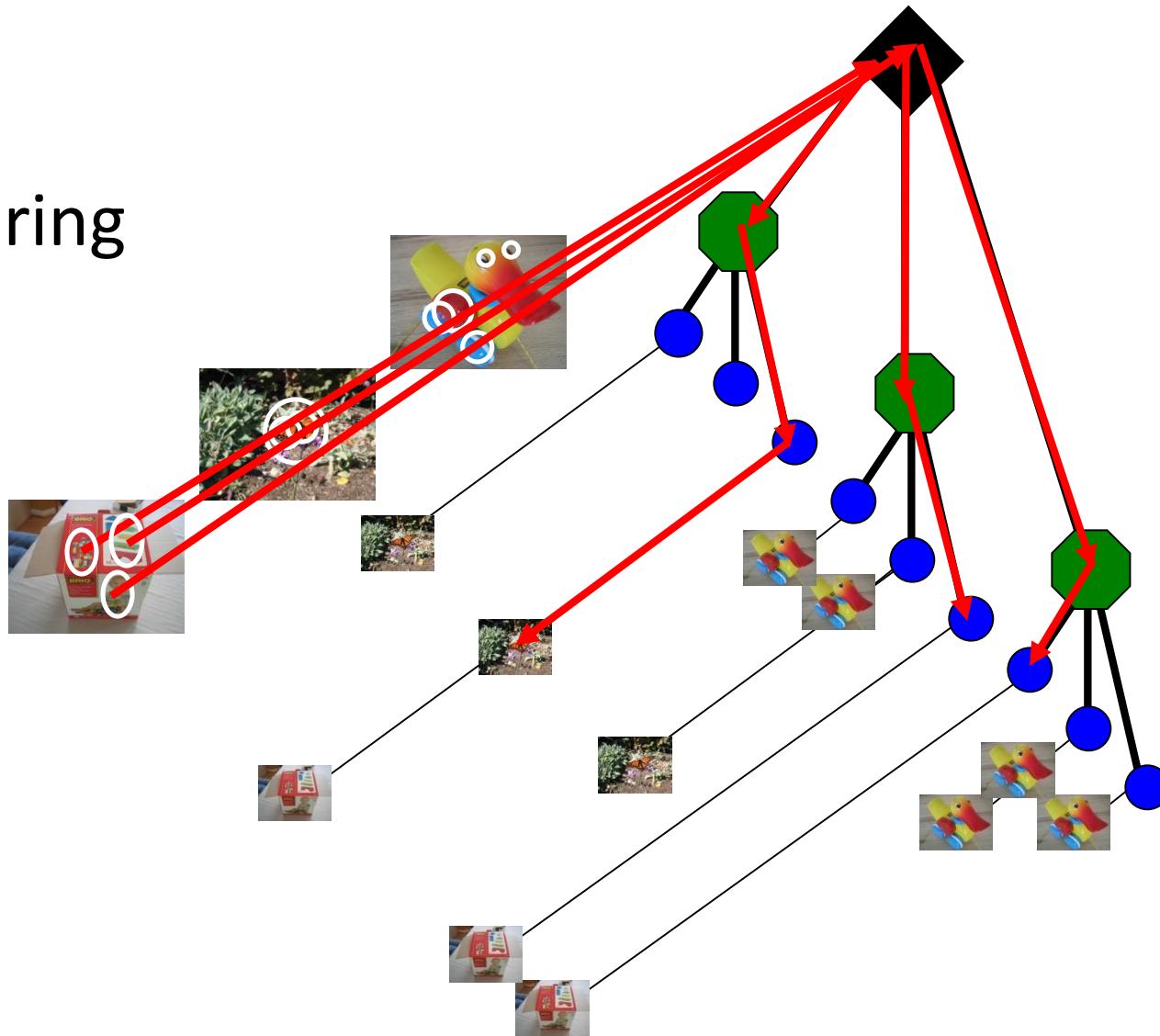
# Scoring



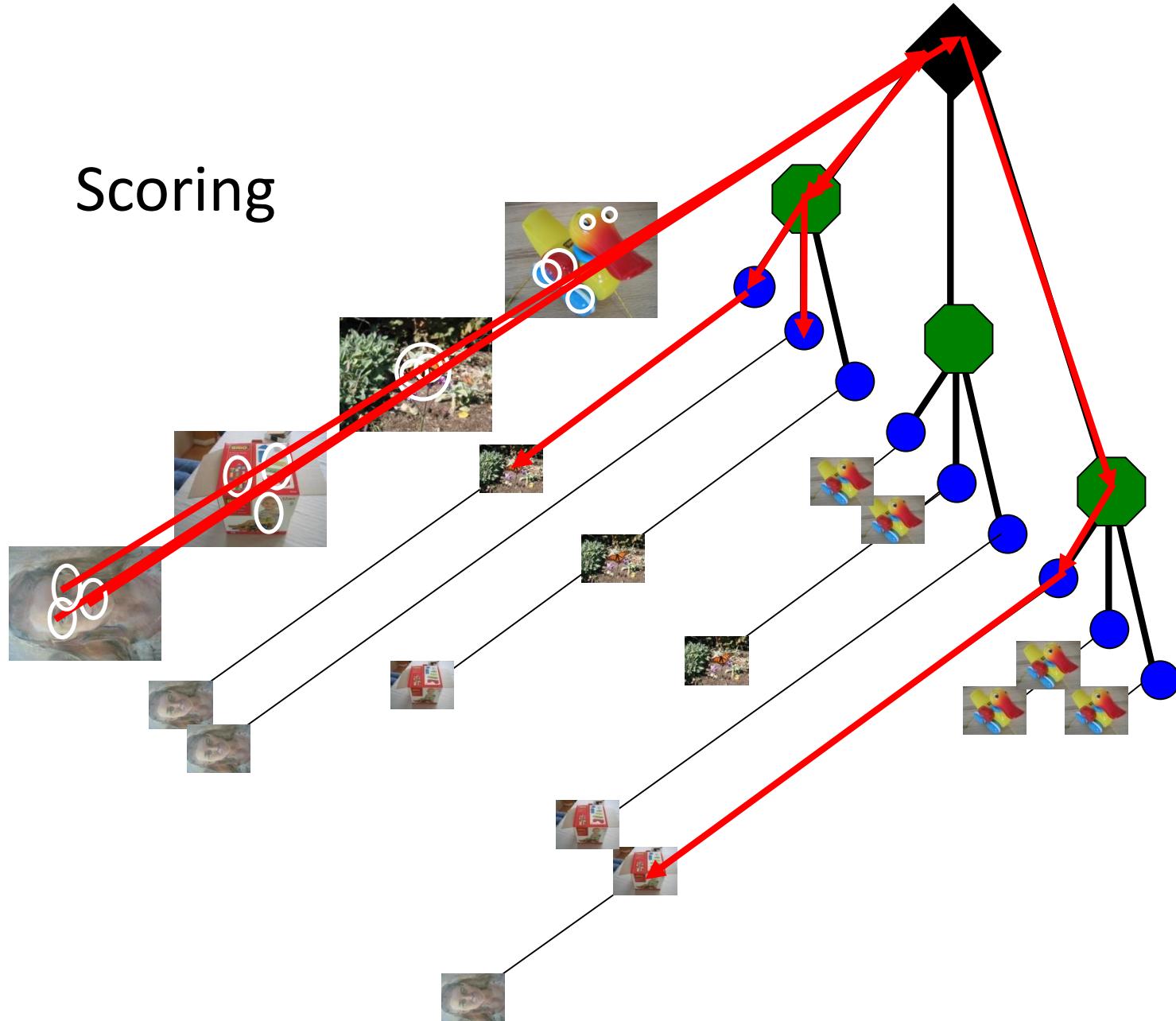
# Scoring



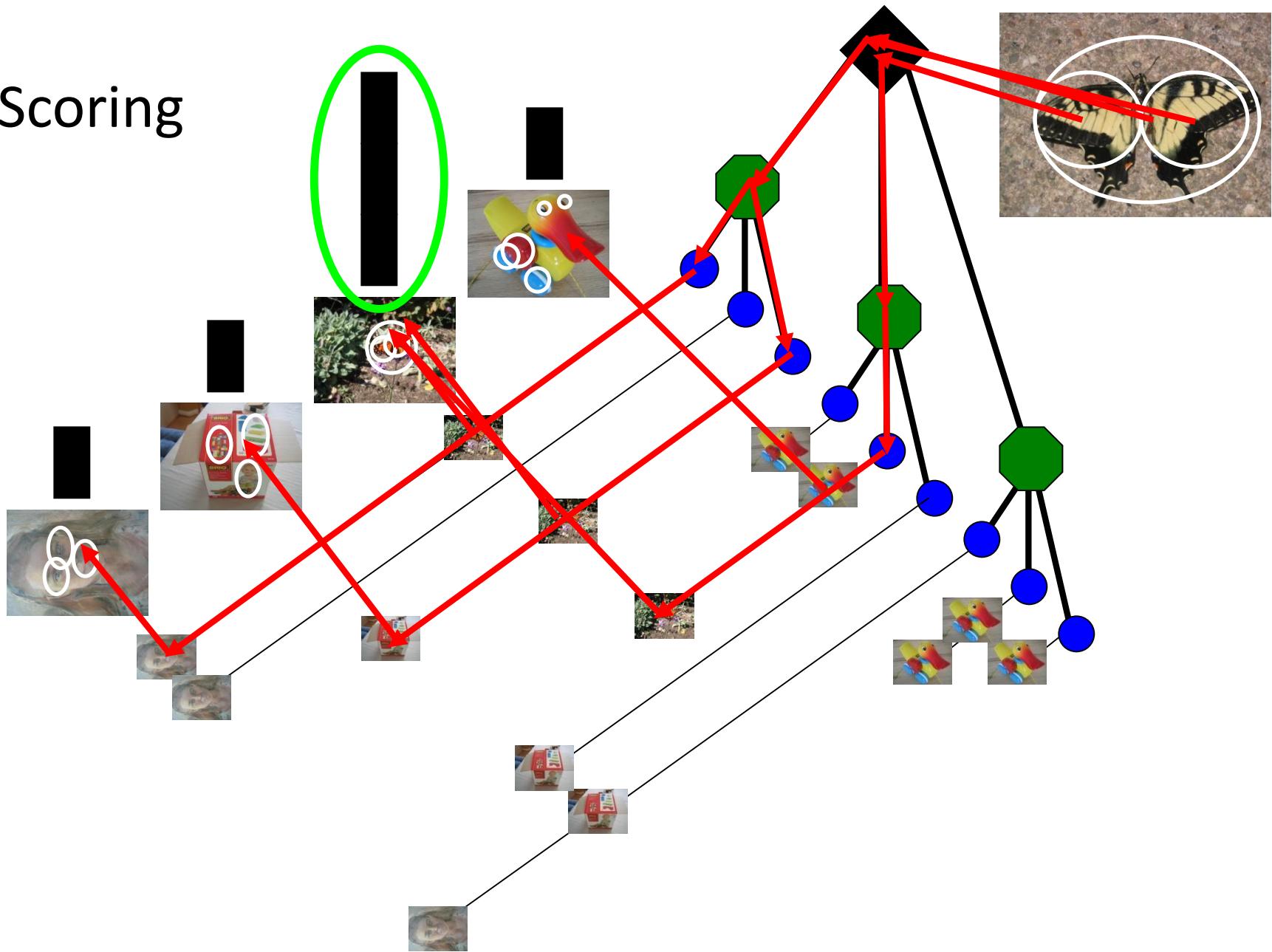
# Scoring



# Scoring

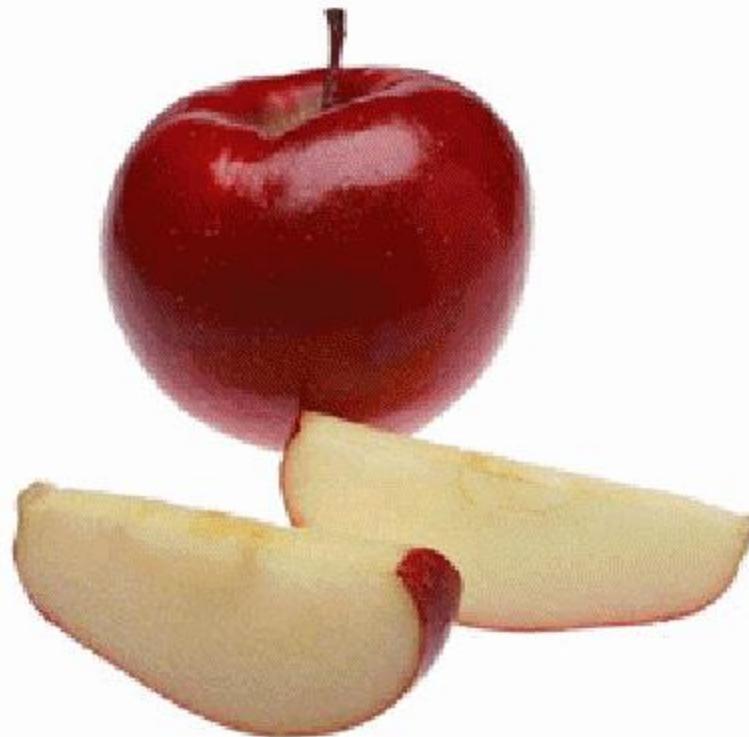


# Scoring

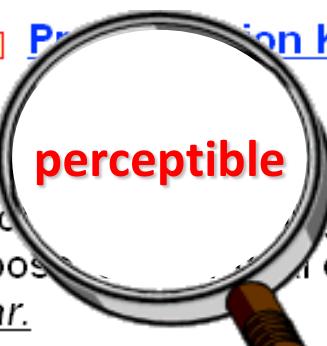


This Class

# **OBJECT CATEGORIZATION**



**obj·ect**   [Prunciation Key](#) (əbj'ikt, -jĕkt')*n.*

- 
1. Something **perceptible** that can be detected by one or more of the senses, especially sight or touch; a **material thing**.
- 
2. A focus of attention, interest, thought, or action: *an object of concern*.
- 
3. The purpose or goal of a specific action or effort: *the object of a game*.
4. Grammar.
- A noun, pronoun, or noun phrase that receives or is affected by the action of a verb within a sentence.
  - A noun or substantive governed by a preposition.
5. Philosophy. Something intelligible 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.



Address http://search.yahoo.com/search?p=object&amp;fr=FP-tab-web-t&amp;toggle=1&amp;cop=&amp;ei=UTF-8

Google object Search 48 blocked ABC Check AutoLink AutoFill Options object

Yahoo! My Yahoo! Mail Welcome, Guest [Sign In]

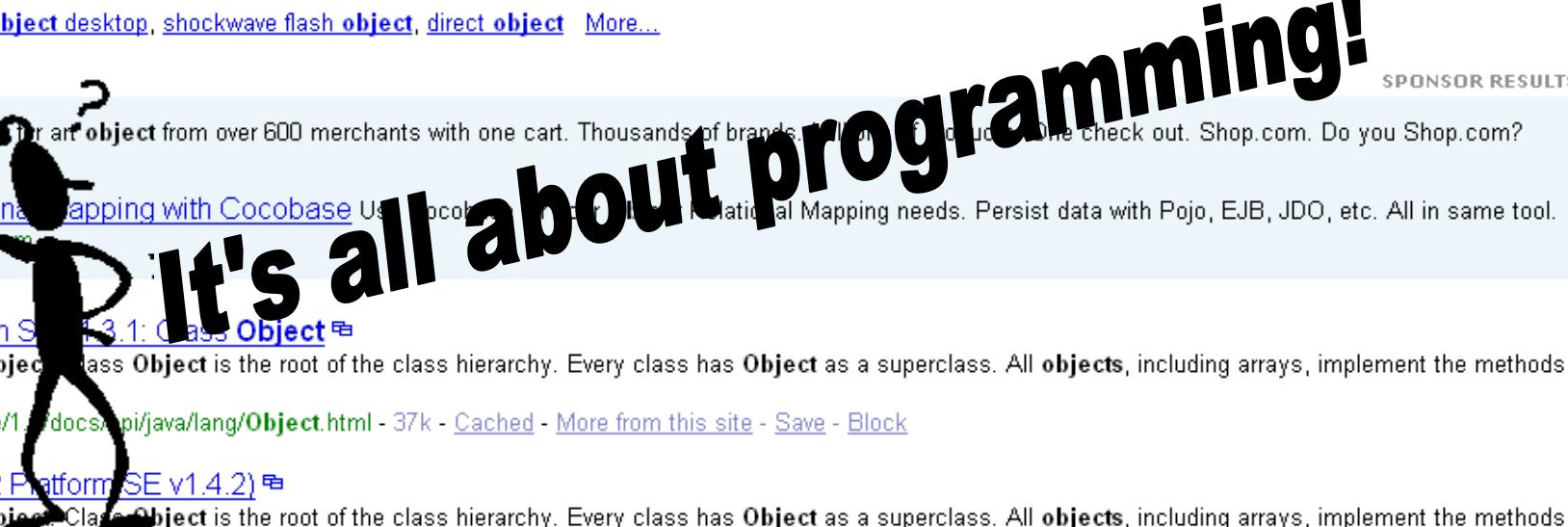


My Web BETA

Search Services

## Search Results

Results 1 - 10 of about 201,000,000 fo

Also try: [object dock](#), [object desktop](#), [shockwave flash object](#), [direct object](#) More...

- [Art Object](#) Shop for an object from over 600 merchants with one cart. Thousands of brands. Millions of products. One check out. Shop.com. Do you Shop.com?  
www.shop.com
- [Object](#) Relational Mapping with Cocobase Use Cocobase to fulfill your Relational Mapping needs. Persist data with Pojo, EJB, JDO, etc. All in same tool.  
www.thoughtinc.com

1. [Java 2 Platform SE v1.3.1: Class Object](#)

... public class **Object**. Class **Object** is the root of the class hierarchy. Every class has **Object** as a superclass. All **objects**, including arrays, implement the methods of this class ...

[java.sun.com/j2se/1.3.1/docs/api/java/lang/Object.html](http://java.sun.com/j2se/1.3.1/docs/api/java/lang/Object.html) - 37k - [Cached](#) - [More from this site](#) - [Save](#) - [Block](#)

2. [Object \(Java 2 Platform SE v1.4.2\)](#)

... public class **Object**. Class **Object** is the root of the class hierarchy. Every class has **Object** as a superclass. All **objects**, including arrays, implement the methods of this class ...

[java.sun.com/j2se/1.4.2/docs/api/java/lang/Object.html](http://java.sun.com/j2se/1.4.2/docs/api/java/lang/Object.html) - 40k - [Cached](#) - [More from this site](#) - [Save](#) - [Block](#)

3. [Objects, Images, and Applets in HTML documents](#)

... address these issues, HTML 4 introduces the **OBJECT** element, which offers an all-purpose solution to generic **object** inclusion. The **OBJECT** element allows HTML ...

[www.w3.org/TR/REC-html40/struct/objects.html](http://www.w3.org/TR/REC-html40/struct/objects.html) - 105k - [Cached](#) - [More from this site](#) - [Save](#) - [Block](#)

4. [OBJECT - Embedded Object](#)



File Edit View Favorites Tools Help



Address <http://images.google.com/images?svnum=10&hl=en&lr=&q=object>

Go Links

Google object Search 48 blocked Check AutoLink AutoFill Options object



Web Images Groups News Froogle Local more »  
[Advanced Image Search](#)  
[Preferences](#)

Moderate SafeSearch is on

## Images

Results 1 - 20 of about 1,030,000 for [object \[definition\]](#). (1.20 seconds)

Show: All sizes - Large - Medium - Small



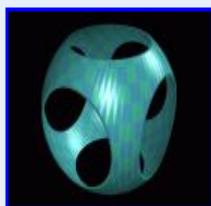
**object.jpg**

360 x 504 pixels - 24k  
[pantransit.reptiles.org/.../object.jpg](http://pantransit.reptiles.org/.../object.jpg)



**OBJECT.GIF**

650 x 410 pixels - 216k  
[www.hawaii.edu/lruby/art400/OBJECT.GIF](http://www.hawaii.edu/lruby/art400/OBJECT.GIF)



**object.gif**

640 x 480 pixels - 73k  
[www.am.ics.keio.ac.jp/proj/sed/images/object.gif](http://www.am.ics.keio.ac.jp/proj/sed/images/object.gif)



**object docl.JPG**

1600 x 1200 pixels - 160k  
[www.4peeps.com/pain/object%20docl.JPG](http://www.4peeps.com/pain/object%20docl.JPG)



**front.jpg**

388 x 392 pixels - 27k  
[www.renewal.org.au/object/](http://www.renewal.org.au/object/)



**AH\_1.jpg**

380 x 385 pixels - 11k  
[www.pbs.org/art21/artists/hamilton/card1.html](http://www.pbs.org/art21/artists/hamilton/card1.html)



**Img0001492.jpg**

650 x 432 pixels - 184k  
[yoshi.txt-nifty.com/photo/images/Img0001492.jpg](http://yoshi.txt-nifty.com/photo/images/Img0001492.jpg)

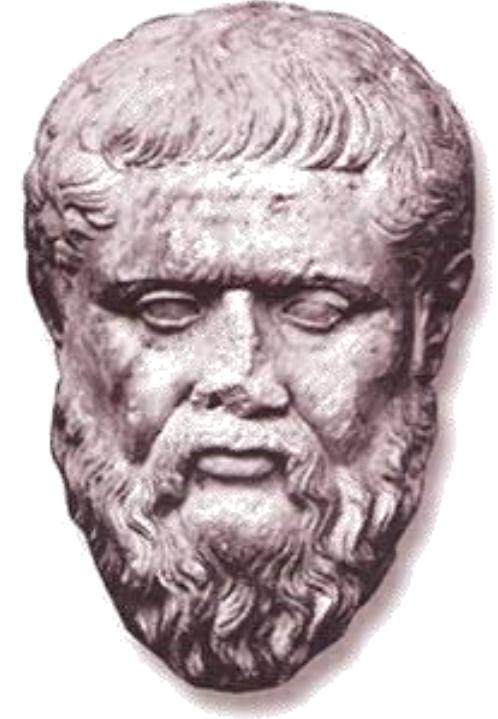


**Don120 Amorphium Object.jpg**

320 x 240 pixels - 7k  
[www.computer-music.com/images/amorphium/Don12...](http://www.computer-music.com/images/amorphium/Don12...)

# Plato said...

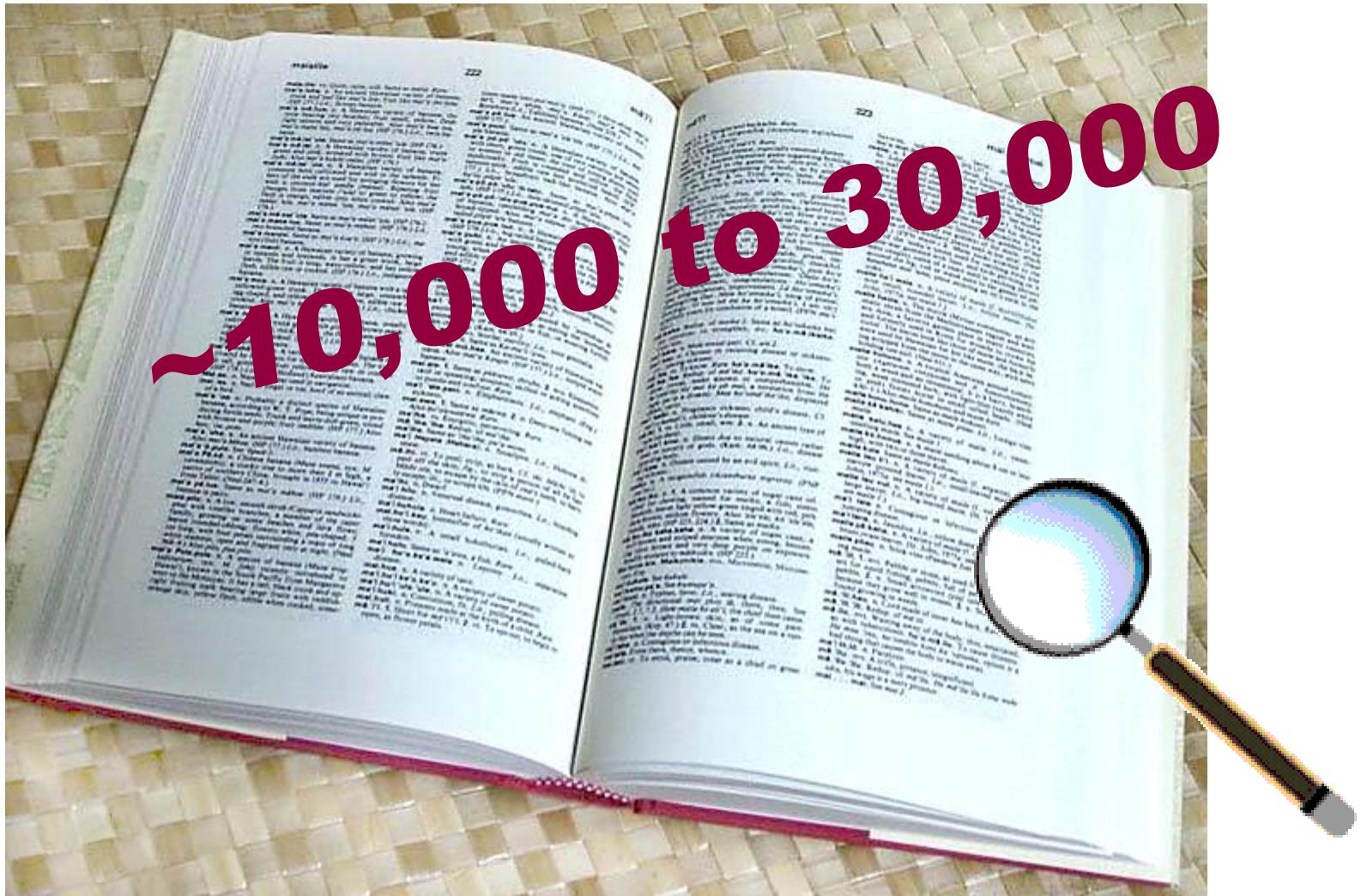
- Ordinary objects are classified together if they 'participate' in the same abstract Form, such as the Form of a Human or the Form of Quartz.
- Forms are proper subjects of philosophical investigation, for they have the highest degree of reality.
- Ordinary objects, such as humans, trees, and stones, 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.





Bruegel, 1564

# How many object categories are there?



# So what does object recognition involve?



# Verification: is that a lamp?



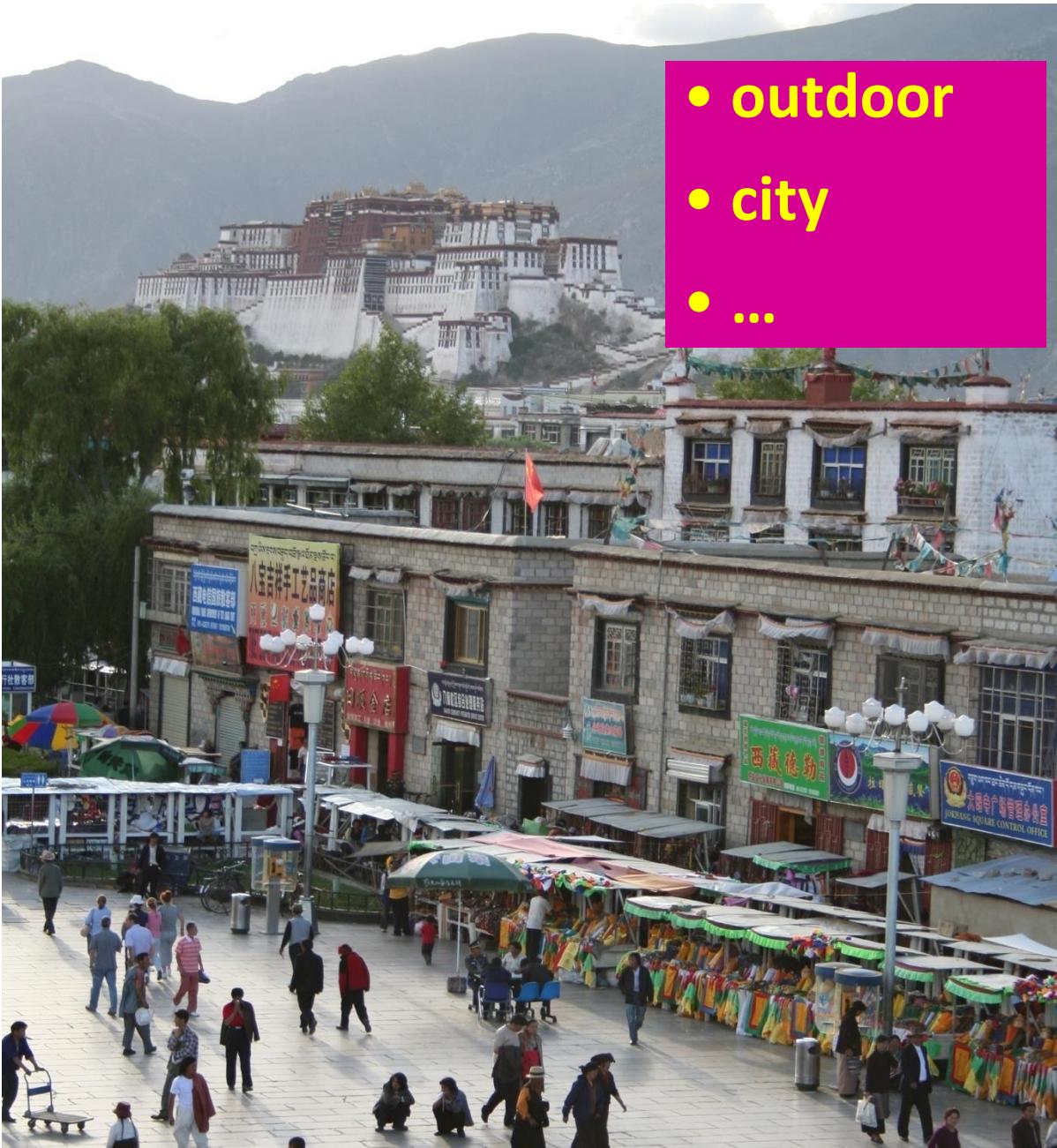
# Detection: are there people?



# Object categorization



# Scene and context categorization



- **outdoor**
- **city**
- ...

# Challenges 1: view point variation



Michelangelo 1475-1564

## Challenges 2: illumination



slide credit: S. Ullman

# Challenges 3: occlusion

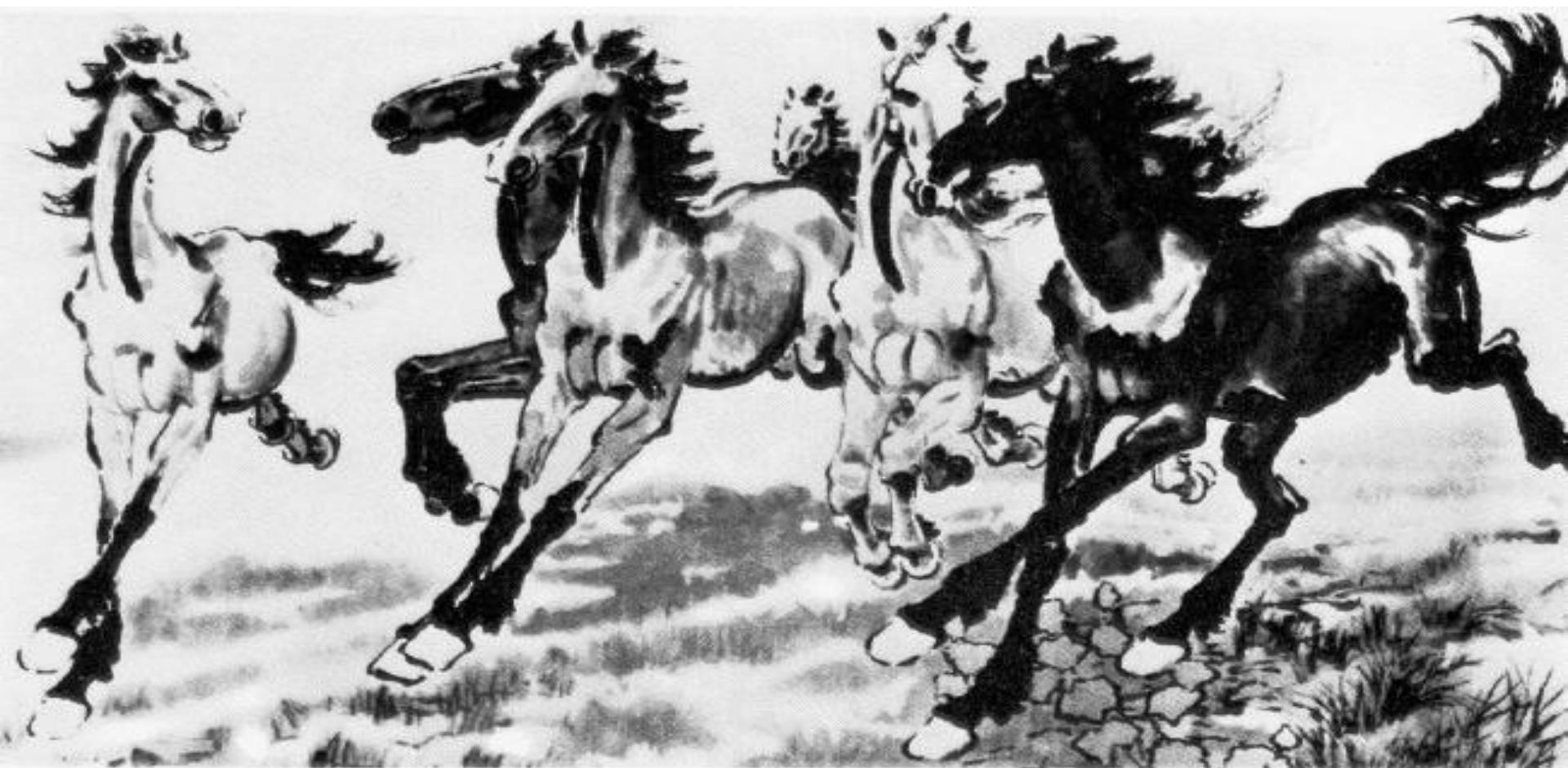


Magritte, 1957

## Challenges 4: scale



## Challenges 5: deformation



Xu, Beihong 1943

## Challenges 6: background clutter







# Challenges 7: intra-class variation



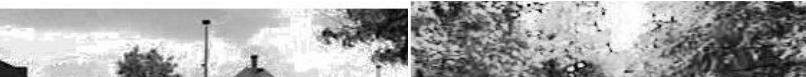


1 7 9 6  
7 8 6 3  
2 1 7 9 7 1 2  
4 8 1 9 0 1 8  
7 6 1 8 6 4 1 3 0 0  
7 5 9 2 6 5 8 1 9 7  
2 2 2 2 2 3 4 4 8 0  
0 2 3 8 0 7 3 8 5 7  
0 1 4 6 4 6 0 2 4 3  
7 1 2 8 7 6 9 8 6 1





7 6 1 8 6 4 1 5 6 0  
7 5 9 2 6 5 8 1 9 7  
2 2 2 2 2 3 4 4 8 0  
0 2 3 8 0 7 3 8 5 7  
0 1 4 6 4 6 0 2 4 3  
7 1 2 8 7 6 9 8 6 1

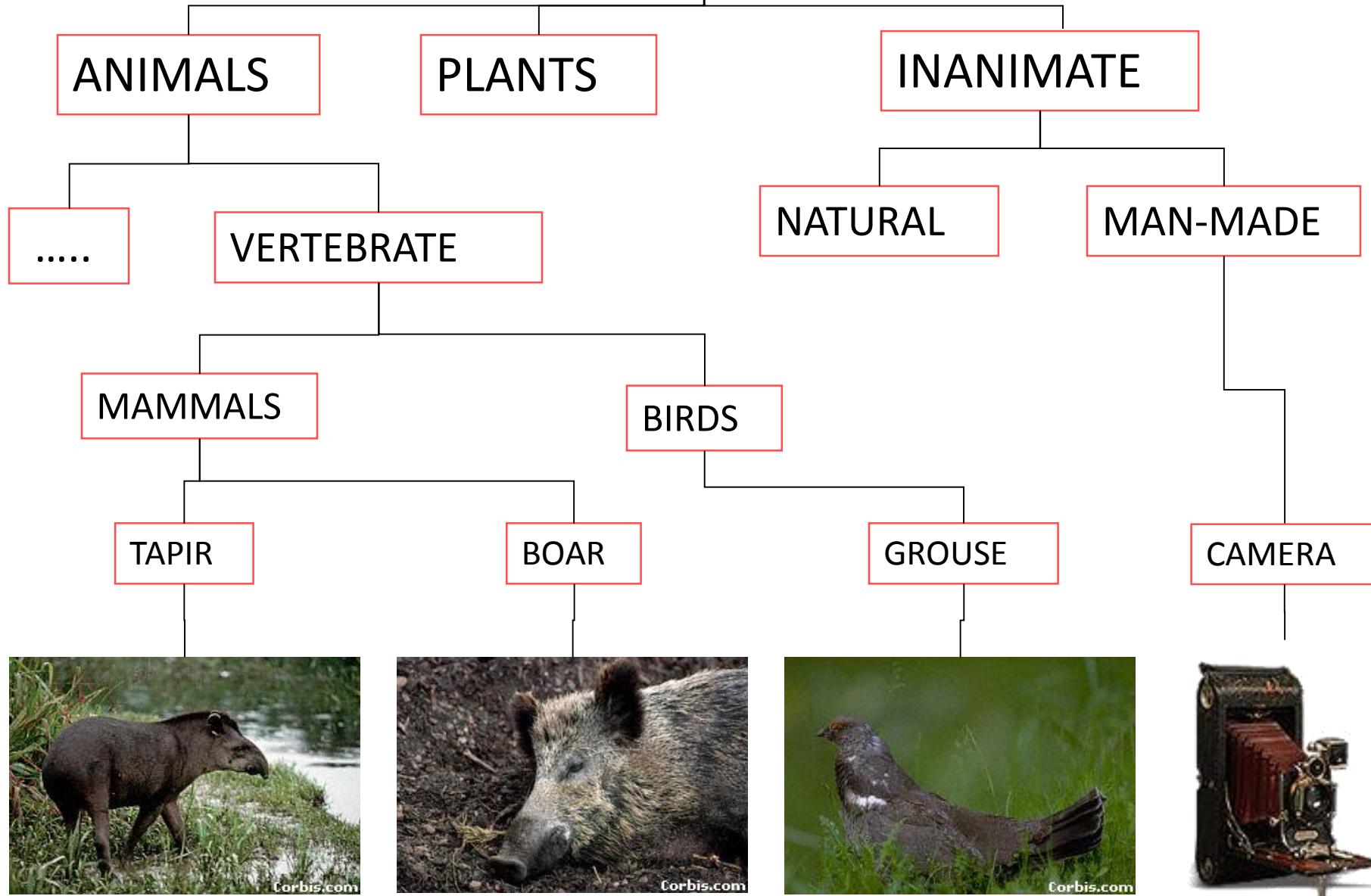


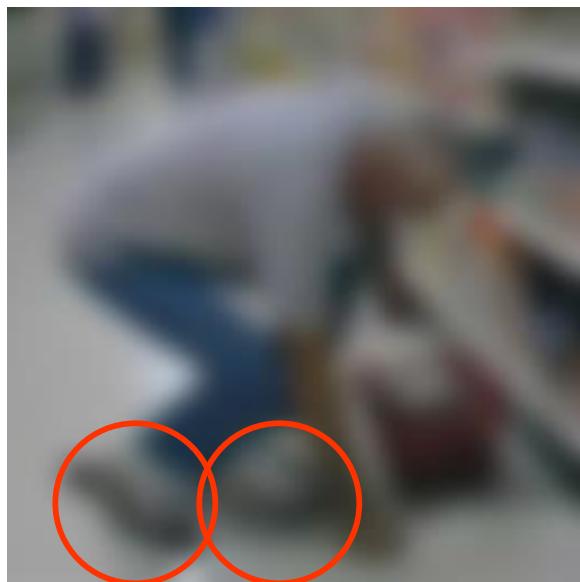
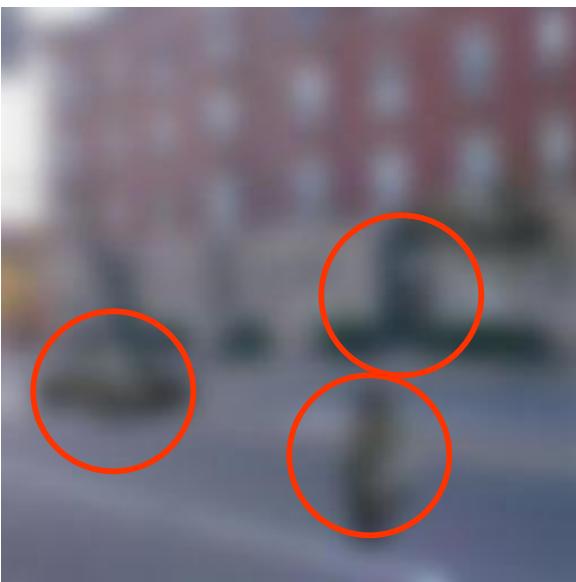
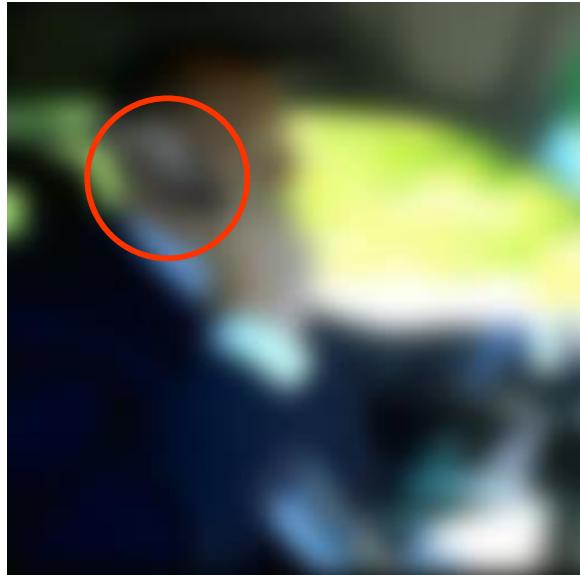
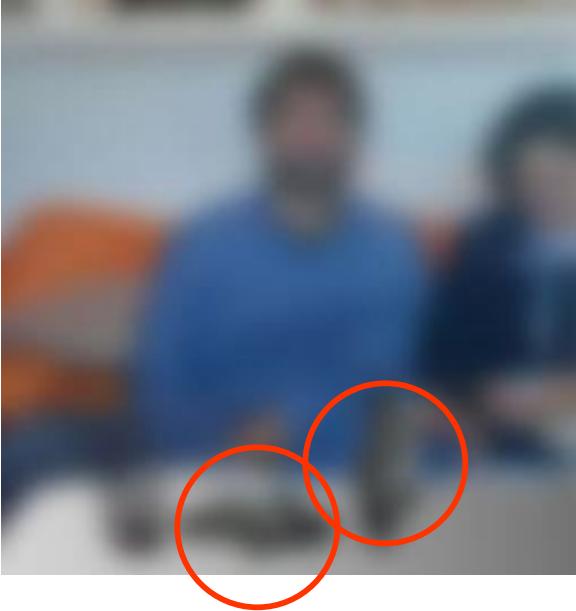
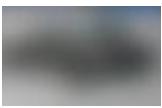


**~10,000 to 30,000**

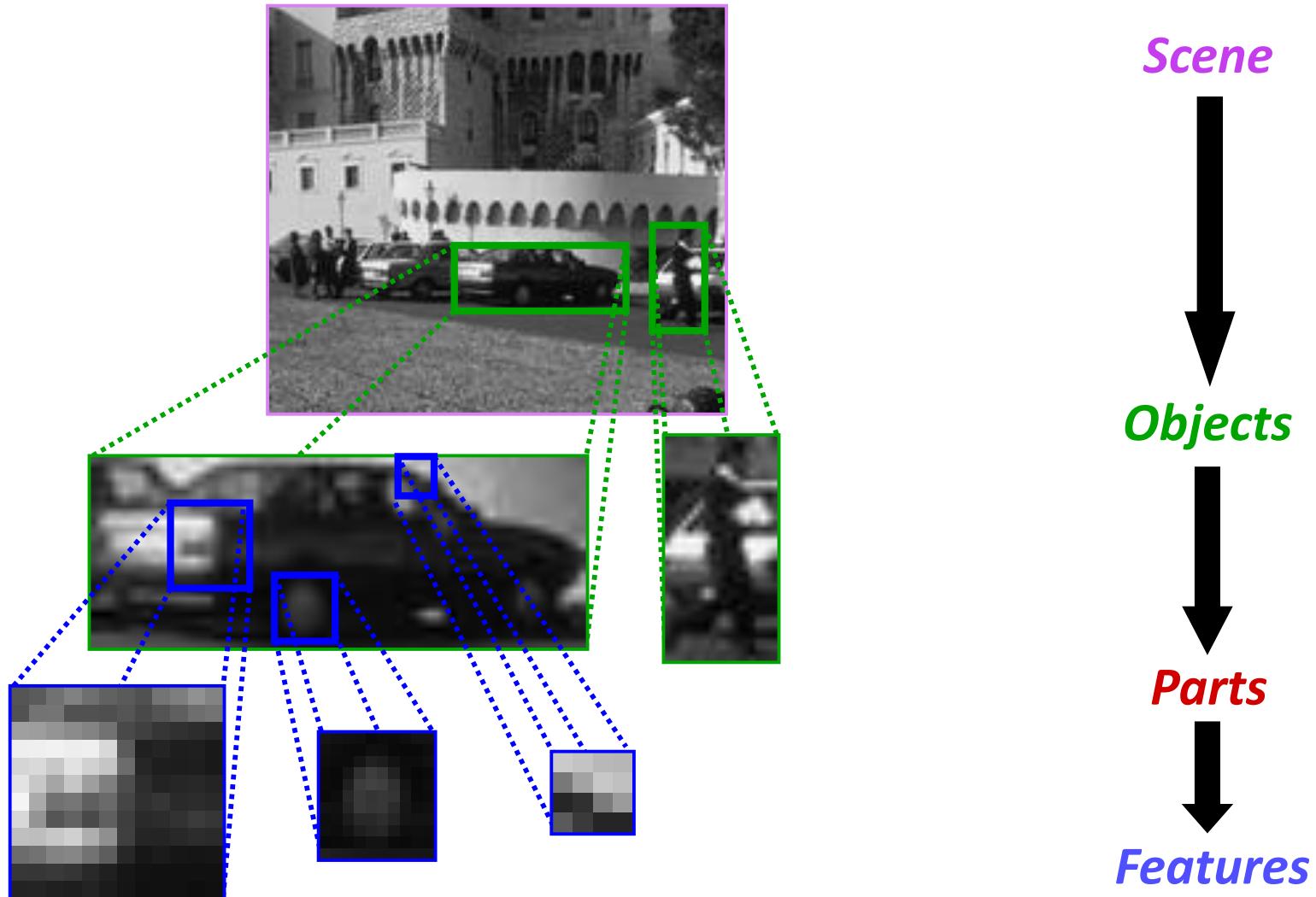


# OBJECTS





# Scenes, Objects, and Parts



# Object categorization: the statistical viewpoint



$$p(\text{zebra} \mid \text{image})$$

vs.

$$p(\text{no zebra} \mid \text{image})$$

- Bayes rule:

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

# Object categorization: the statistical viewpoint

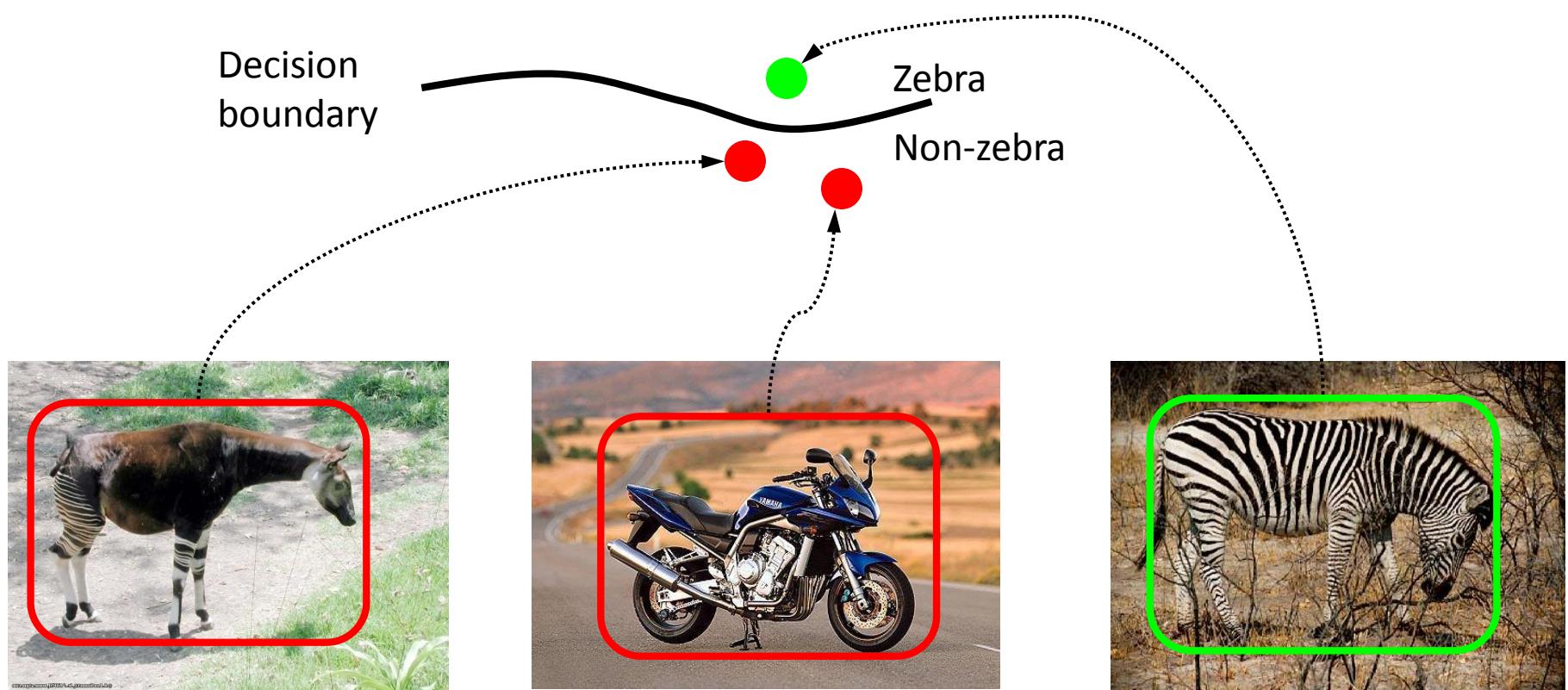
$$\frac{p(\text{zebra} | \text{image})}{p(\text{no zebra} | \text{image})} = \underbrace{\frac{p(\text{image} | \text{zebra})}{p(\text{image} | \text{no zebra})}}_{\text{likelihood ratio}} \cdot \underbrace{\frac{p(\text{zebra})}{p(\text{no zebra})}}_{\text{prior ratio}}$$

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

# Discriminative

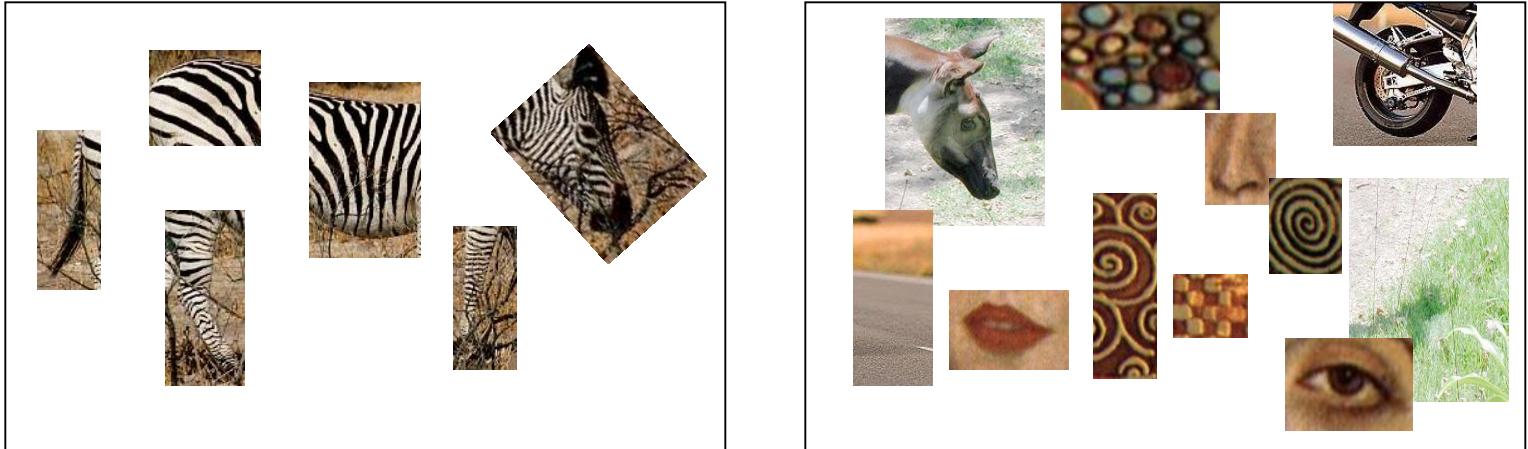
- Direct modeling of

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



# Generative

- Model  $p(\text{image} | \text{zebra})$  and  $p(\text{image} | \text{no zebra})$



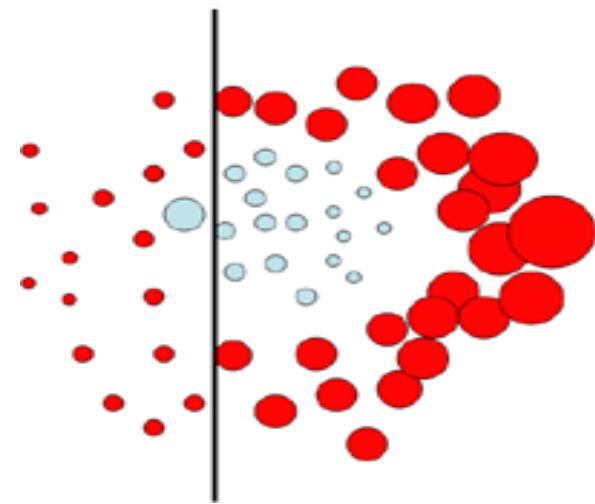
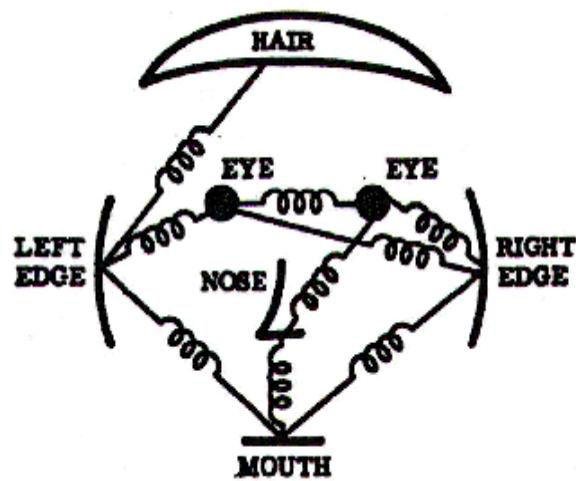
$p(\text{image}   \text{zebra})$	$p(\text{image}   \text{no zebra})$
Low	Middle
High	Middle → Low

# 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

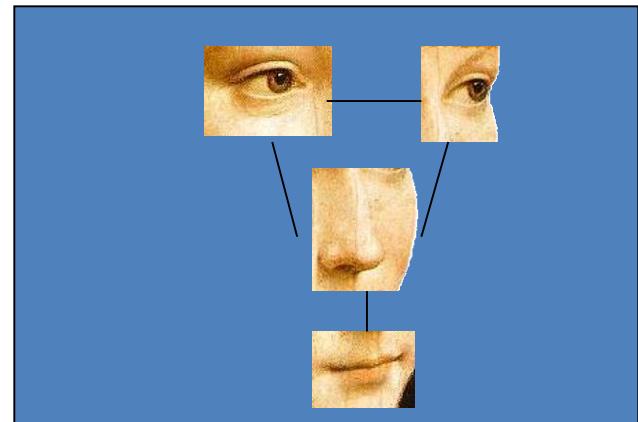
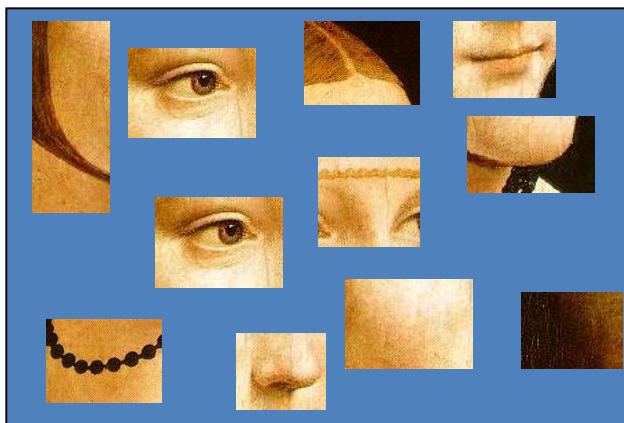
# Representation

- Generative /  
discriminative / hybrid



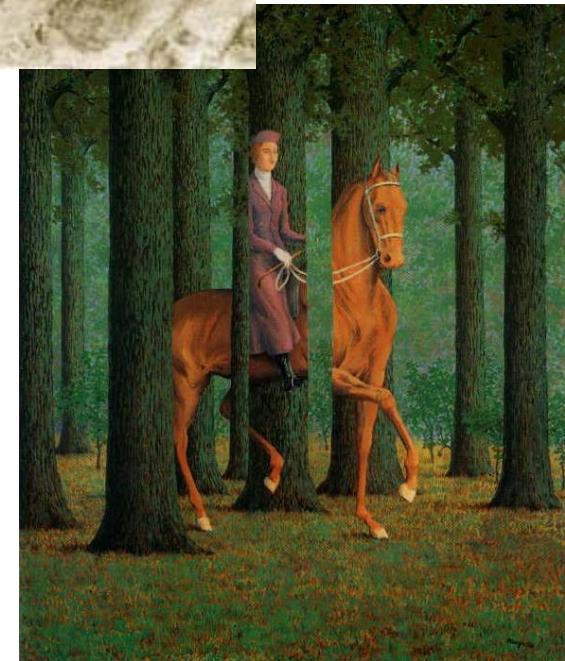
# Representation

- Generative /  
discriminative / hybrid
- Appearance only or  
location and appearance



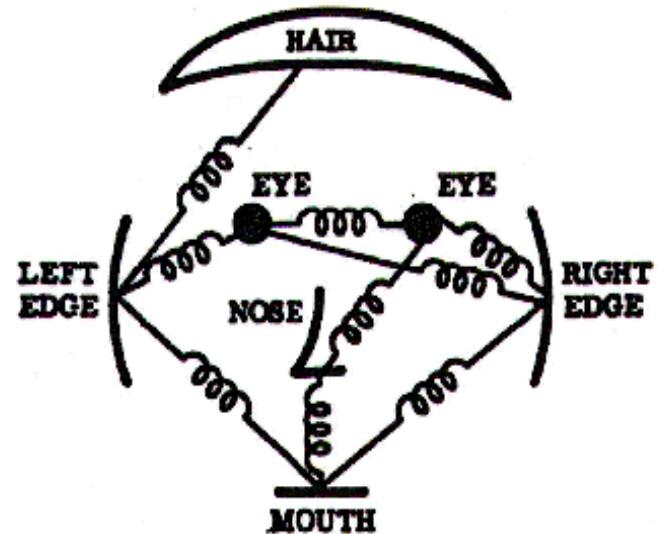
# Representation

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



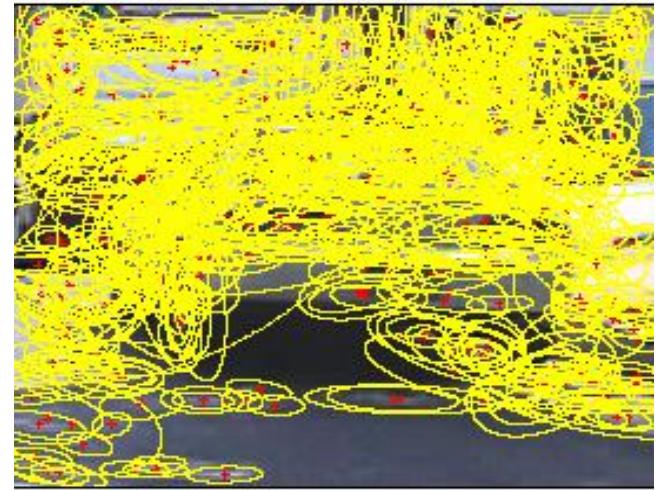
# Representation

- Generative /  
discriminative / hybrid
- Appearance only or  
location and appearance
- invariances
- Part-based or global  
w/sub-window



# Representation

- Generative /  
discriminative / hybrid
- Appearance only or  
location and appearance
- invariances
- Parts or global w/sub-  
window
- Use set of features or each  
pixel in image





# Bag-of-words models

Slide credit Li Fei-Fei (Stanford)

**Object**

**Bag of 'words'**



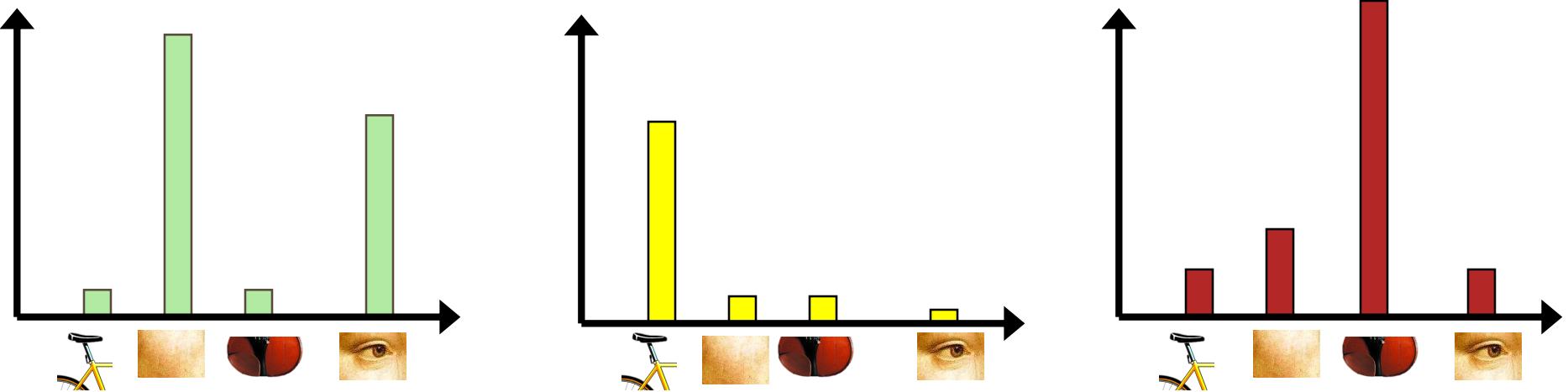
# Analogy to documents

Of all the sensory impressions proceeding to the brain, the visual experiences are the dominant ones. Our perception of the world around us is based essentially upon what we see. Light does not reach the brain from our eyes directly. We have thought that the optic nerve passes through the cerebral cortex before reaching the retina. Upon what basis can we make such a statement? Through the work of Hubel and Wiesel we now know that the visual system is a perceptual system. It is able to analyze more complex features of the image. In analyzing the visual impression, the image is broken up into various cell layers. In the visual cortex, Hubel and Wiesel have been able to show that the message about the image falling on the retina undergoes a step-wise analysis in a systematic way. The image is broken up into nerve cells stored in columns. In this system each column contains a number of nerve cells. Each cell has its specific function and is responsible for detecting a specific detail in the pattern of the retinal image.

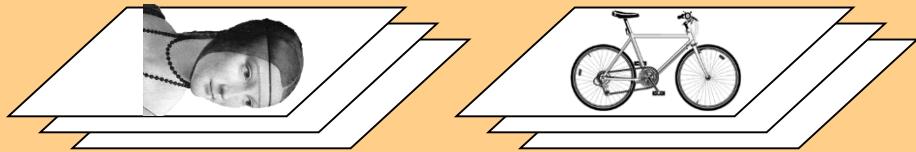
**sensory, brain,  
visual, perception,  
retinal, cerebral cortex,  
eye, cell, optical  
nerve, image  
Hubel, Wiesel**

China is forecasting a trade surplus of \$90bn (£51bn) to \$100bn this year, a threefold increase on 2004's \$32bn. The Commerce Ministry said the surplus would be driven by a predicted 30% jump in exports and a 18% rise in imports. The ministry also predicted a further appreciation of the yuan. China's deliberations over the size of the surplus have one factor in mind: the US' desire for a more to boost its economy. The Chinese government stayed within the range of the central bank's target value of the yuan. The Chinese government allowed the yuan to rise 2% in July and permitted it to fluctuate within a band, but the US wants the yuan to be allowed to trade freely. However, Beijing has made clear that it will take its time and tread carefully, allowing the yuan to rise further in value.

**China, trade,  
surplus, commerce,  
exports, imports, US,  
yuan, bank, domestic,  
foreign, increase,  
trade, value**



# Representation



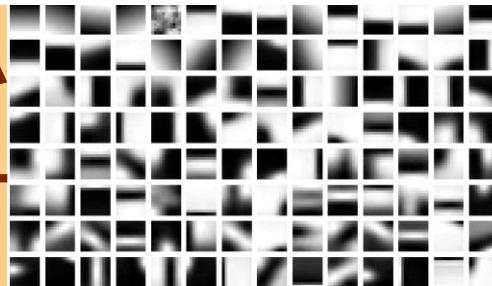
1. feature detection  
& representation



image representation

2.

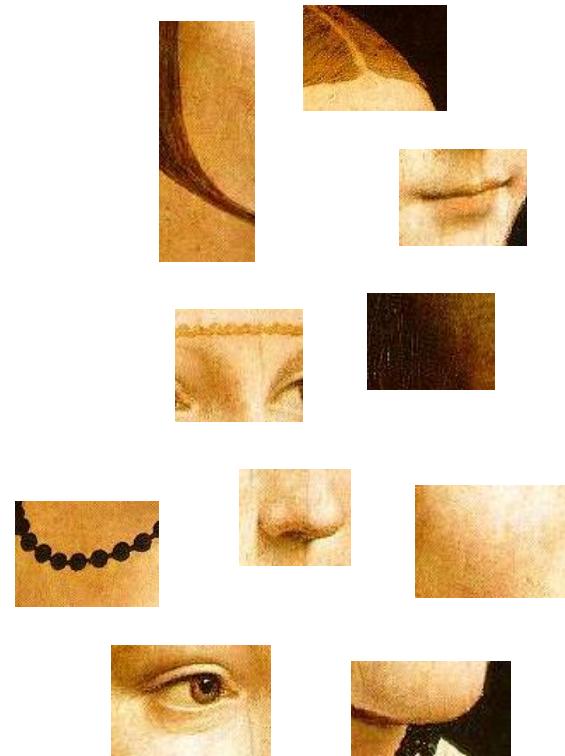
codewords dictionary



3.

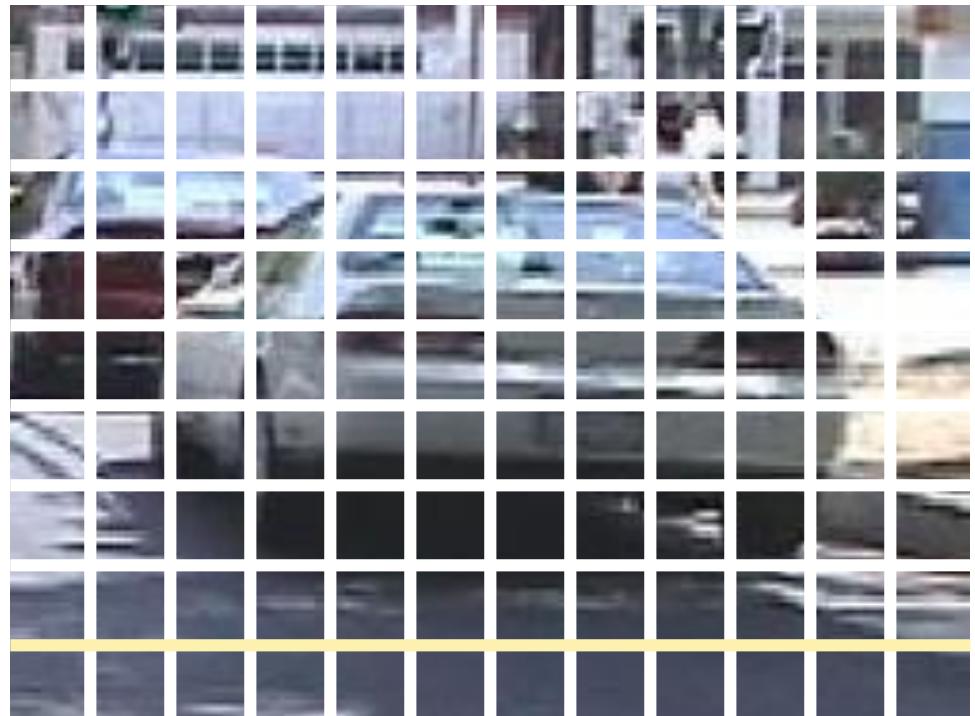


# 1. Feature detection and representation



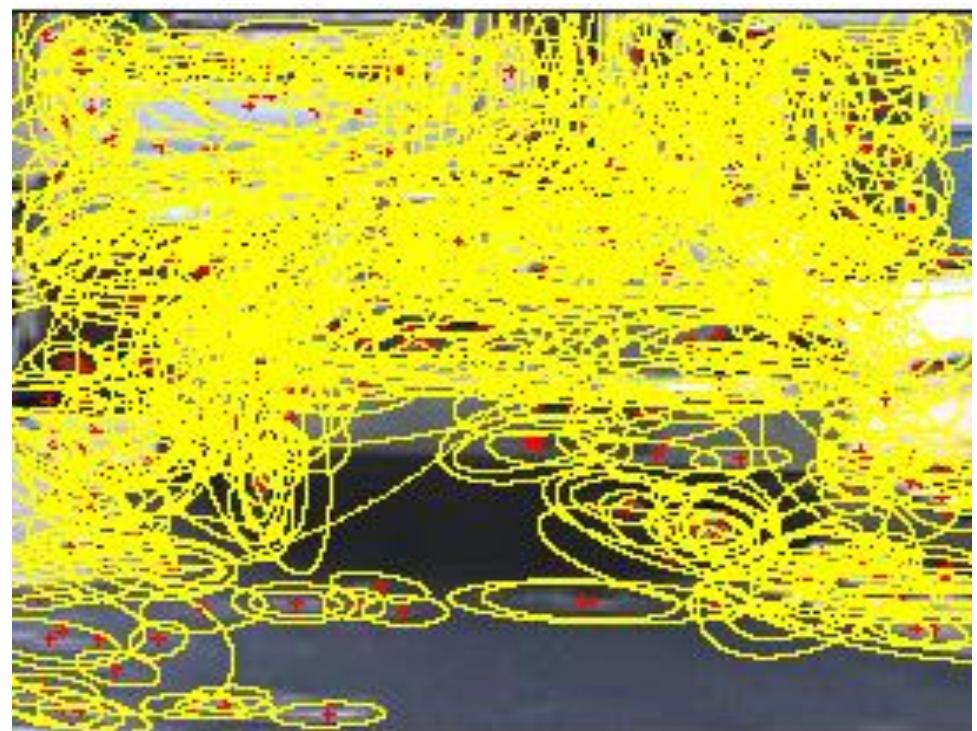
# 1. Feature detection and representation

- Regular grid
  - Vogel et al. 2003
  - Fei-Fei et al. 2005



# 1. Feature detection and representation

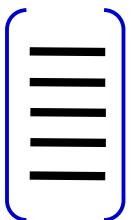
- Regular grid
  - Vogel et al. 2003
  - Fei-Fei et al. 2005
- Interest point detector
  - Csurka et al. 2004
  - Fei-Fei et al. 2005
  - Sivic et al. 2005



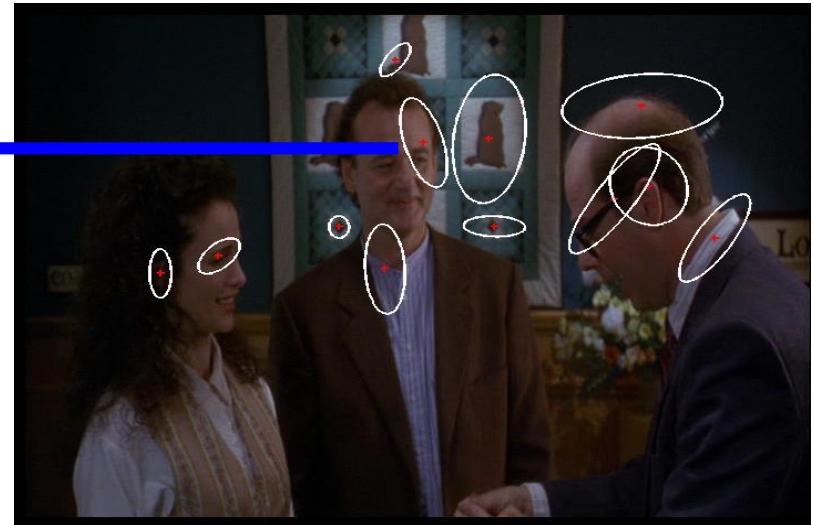
# **1. Feature detection and representation**

- Regular grid
  - Vogel et al. 2003
  - Fei-Fei et al. 2005
- Interest point detector
  - Csurka et al. 2004
  - Fei-Fei et al. 2005
  - Sivic et al. 2005
- Other methods
  - Random sampling (Ullman et al. 2002)
  - Segmentation based patches (Barnard et al. 2003)

# 1. Feature detection and representation

 ←  
Compute SIFT  
descriptor  
[Lowe'99]

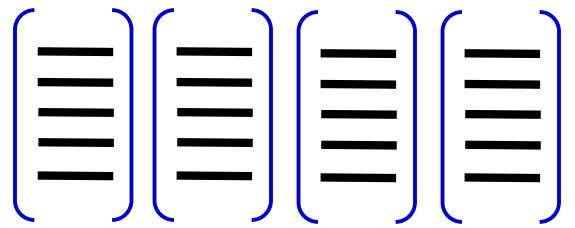
←  
Normalize patch  
[Mikojaczyk and Schmid '02]  
[Matas et al. '02]  
[Sivic et al. '03]



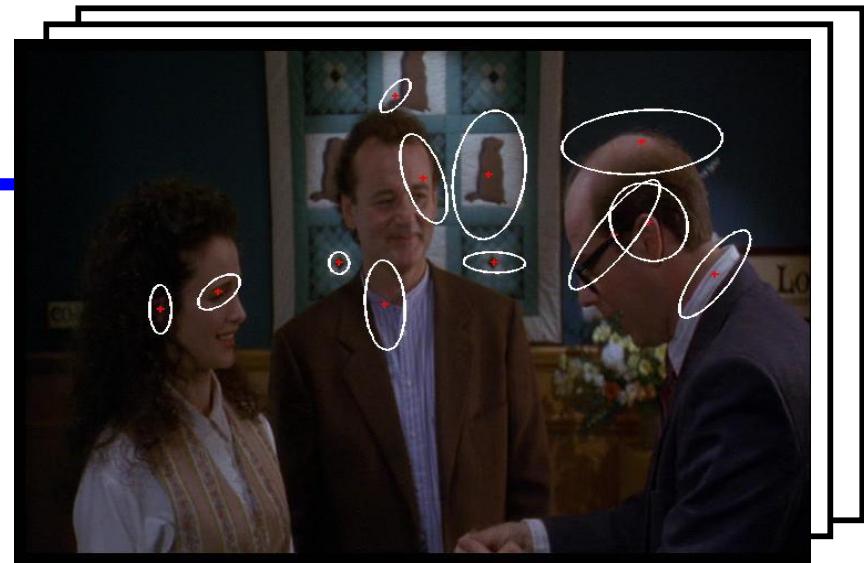
Detect patches

[Mikojaczyk and Schmid '02]  
[Matas et al. '02]  
[Sivic et al. '03]

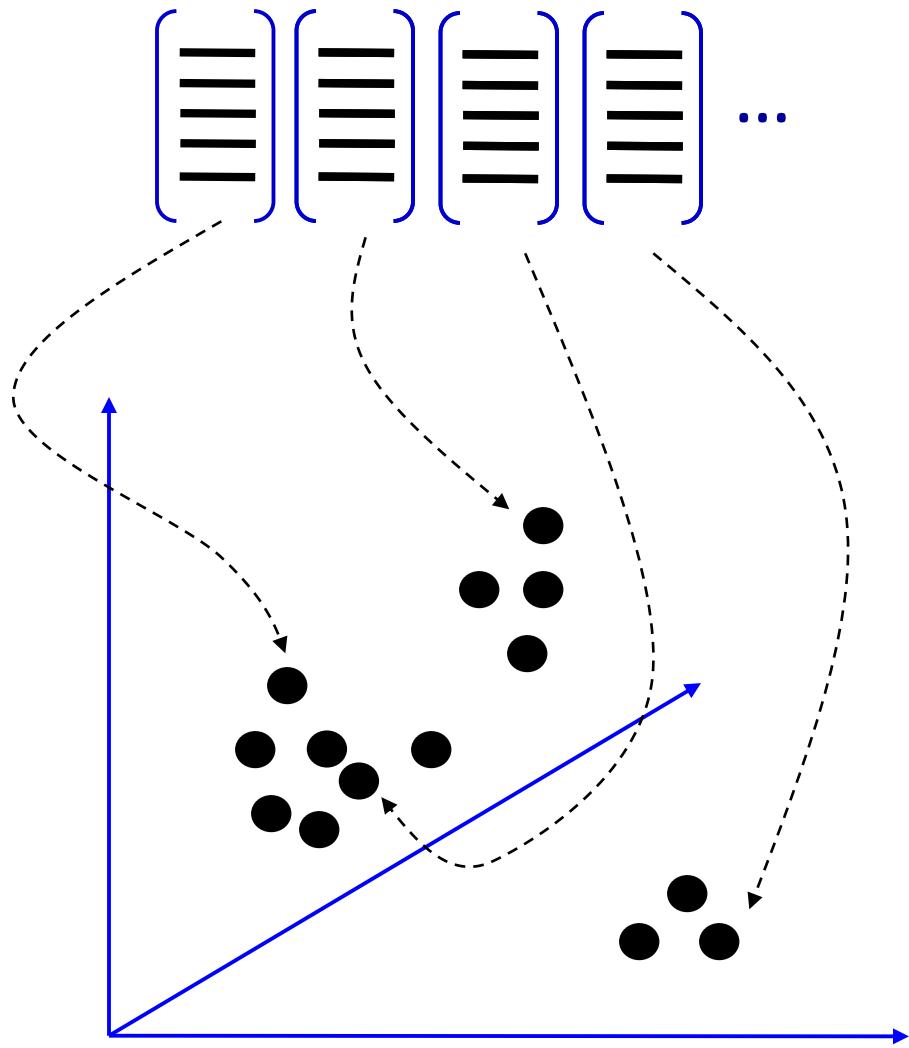
# 1. Feature detection and representation



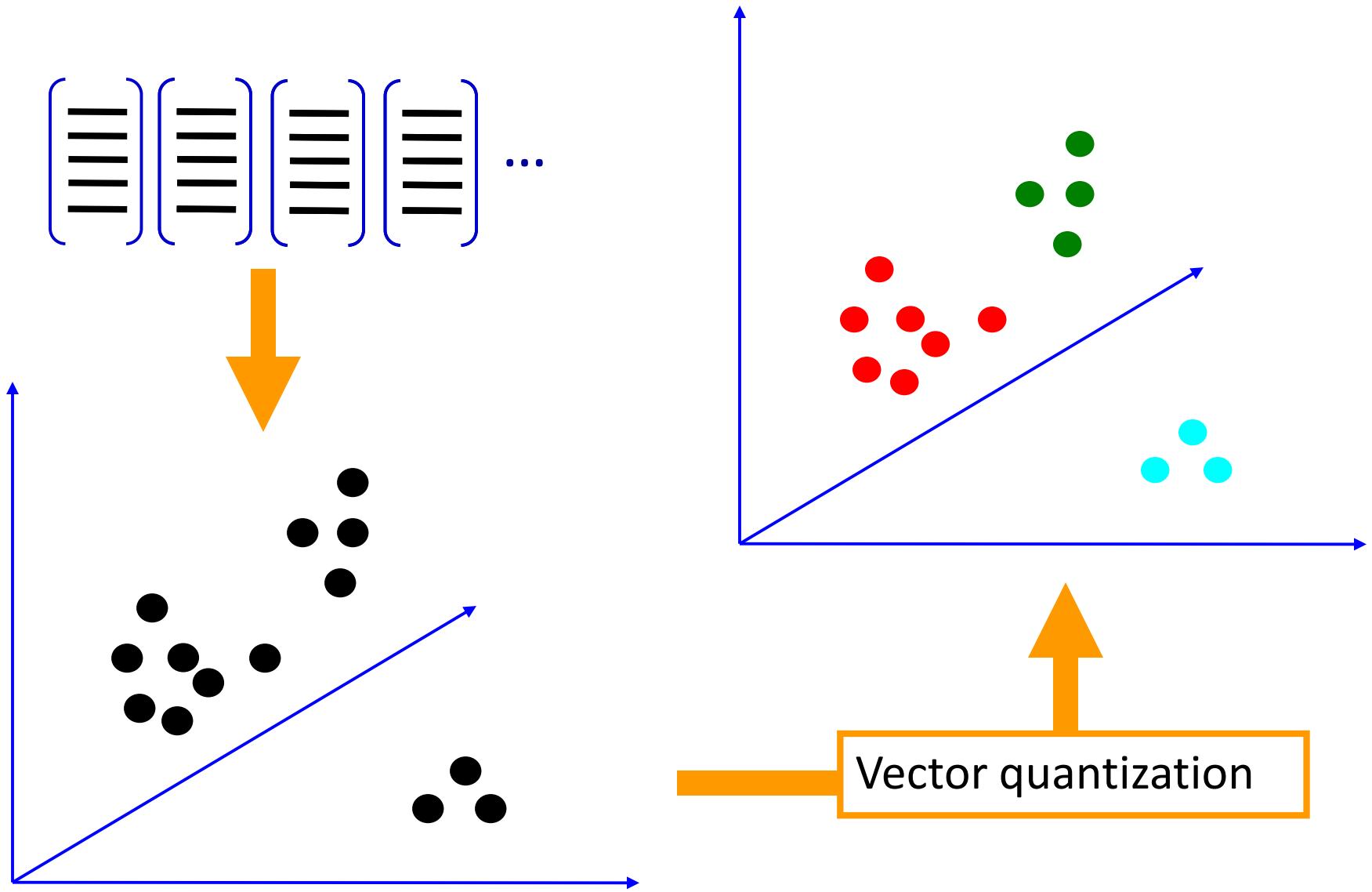
... ←



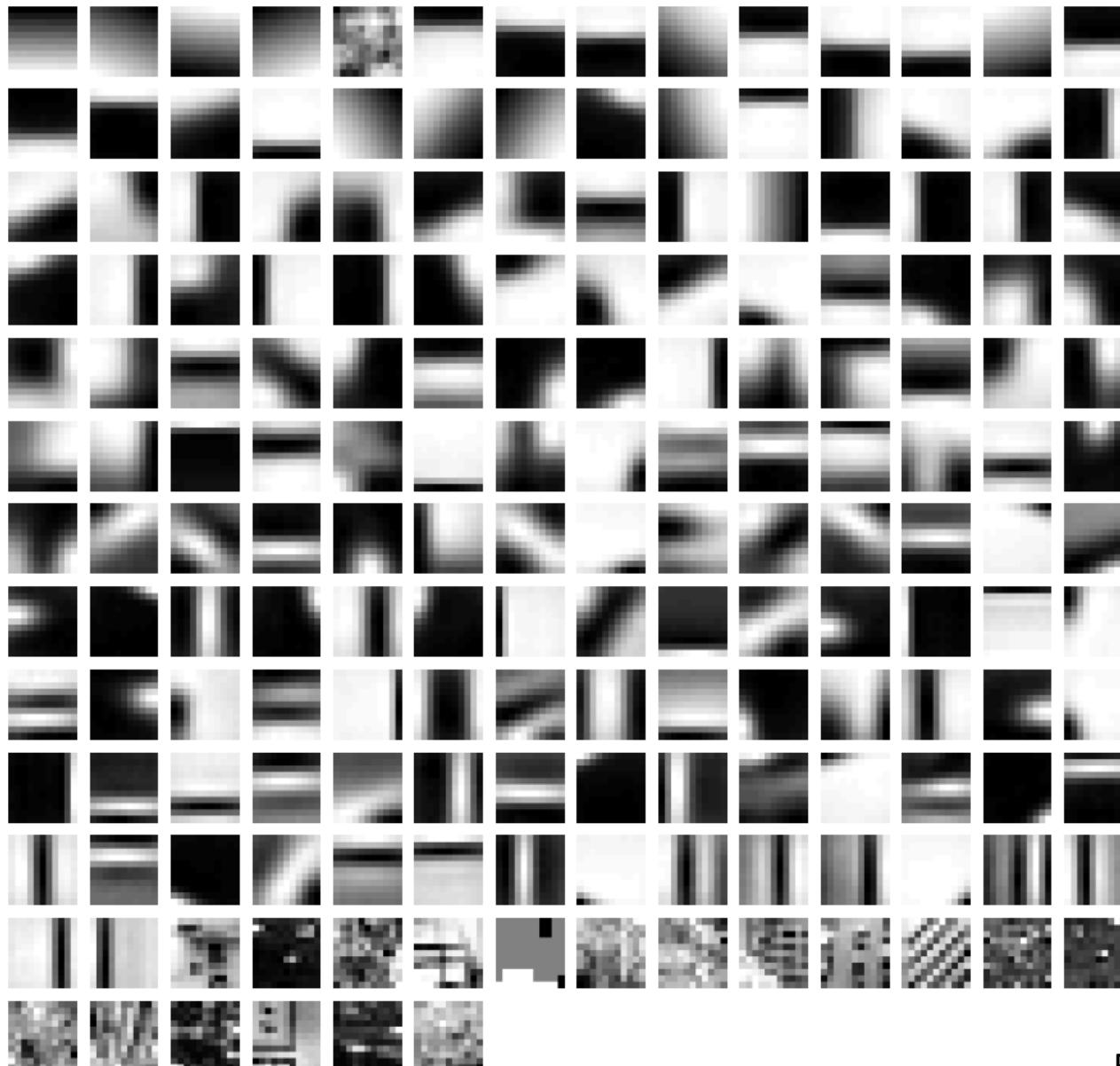
## 2. Codewords dictionary formation



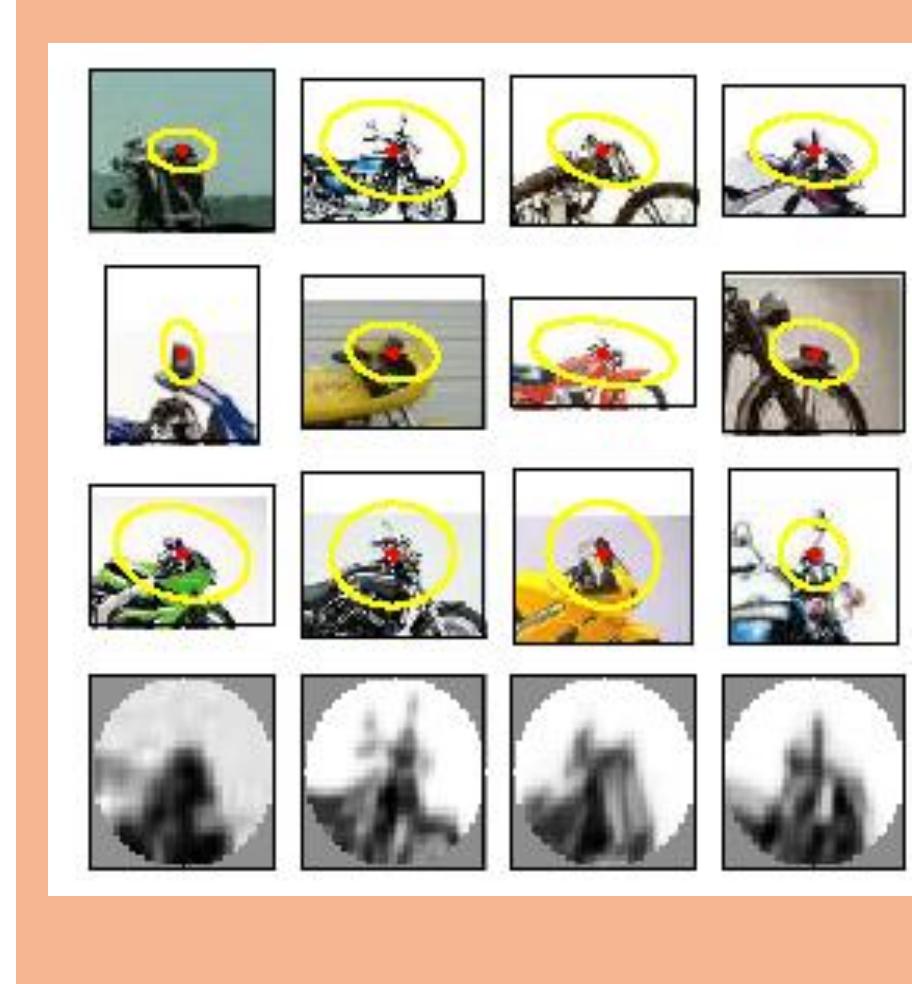
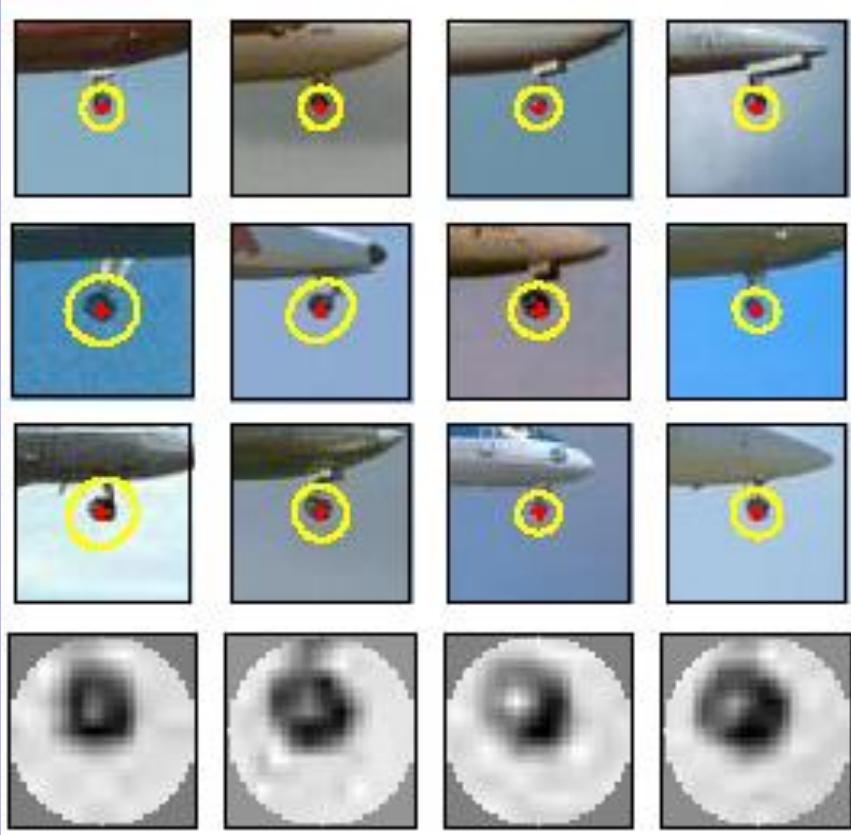
## 2. Codewords dictionary formation



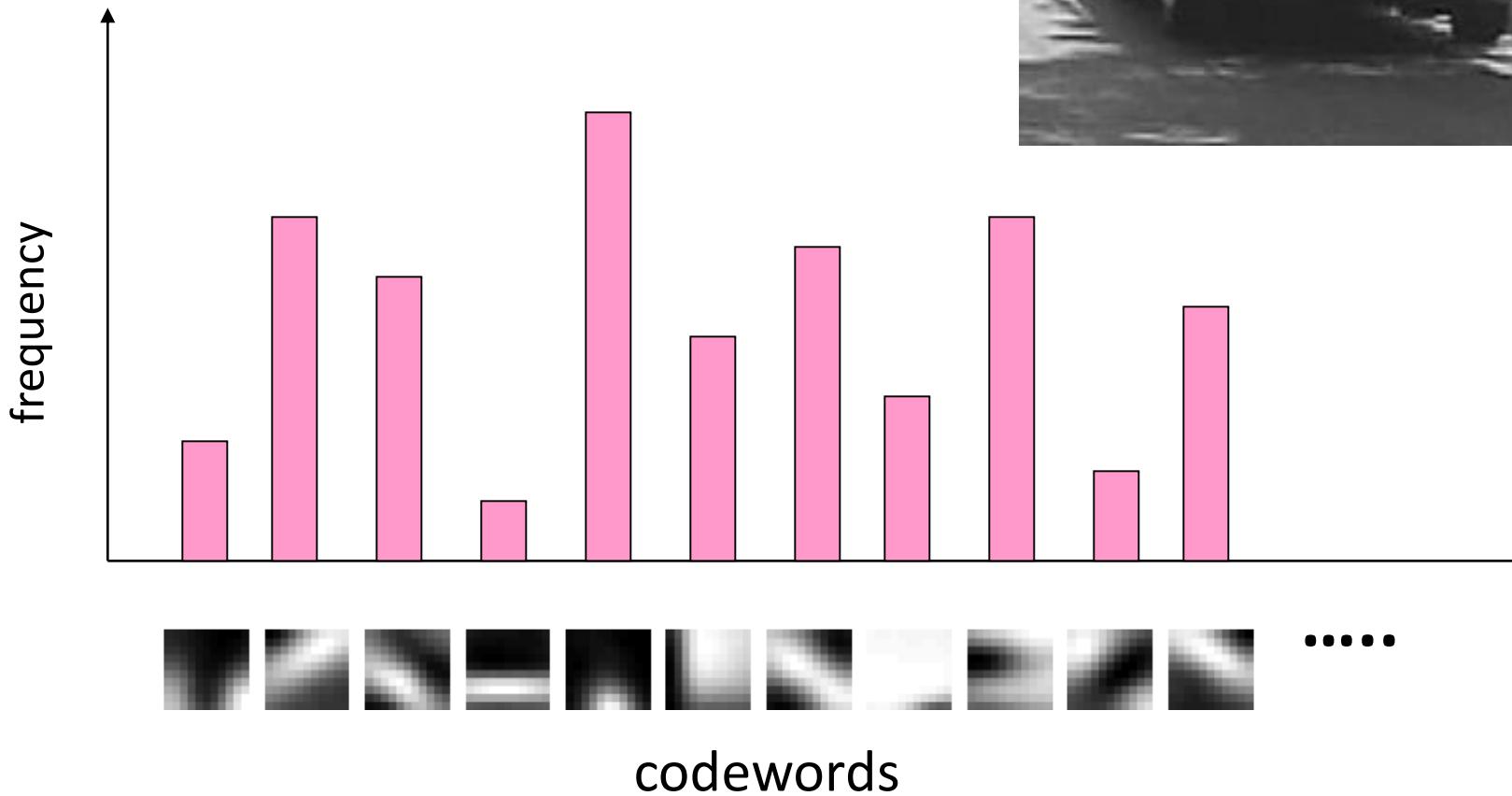
## 2. Codewords dictionary formation



# Image patch examples of codewords



### 3. Image representation



# Next Class

- 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