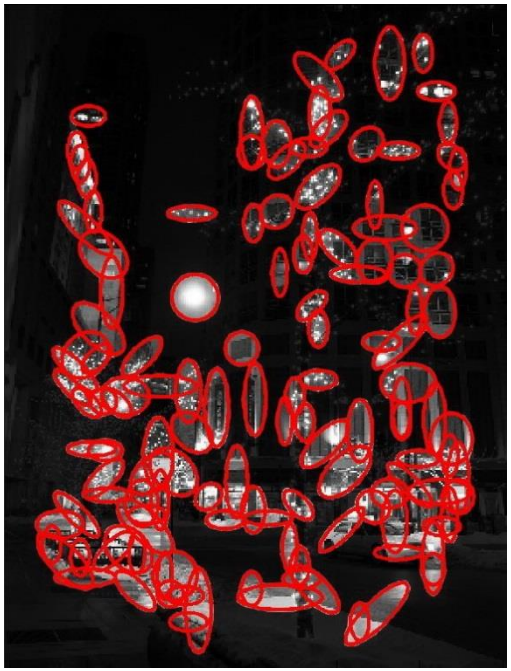
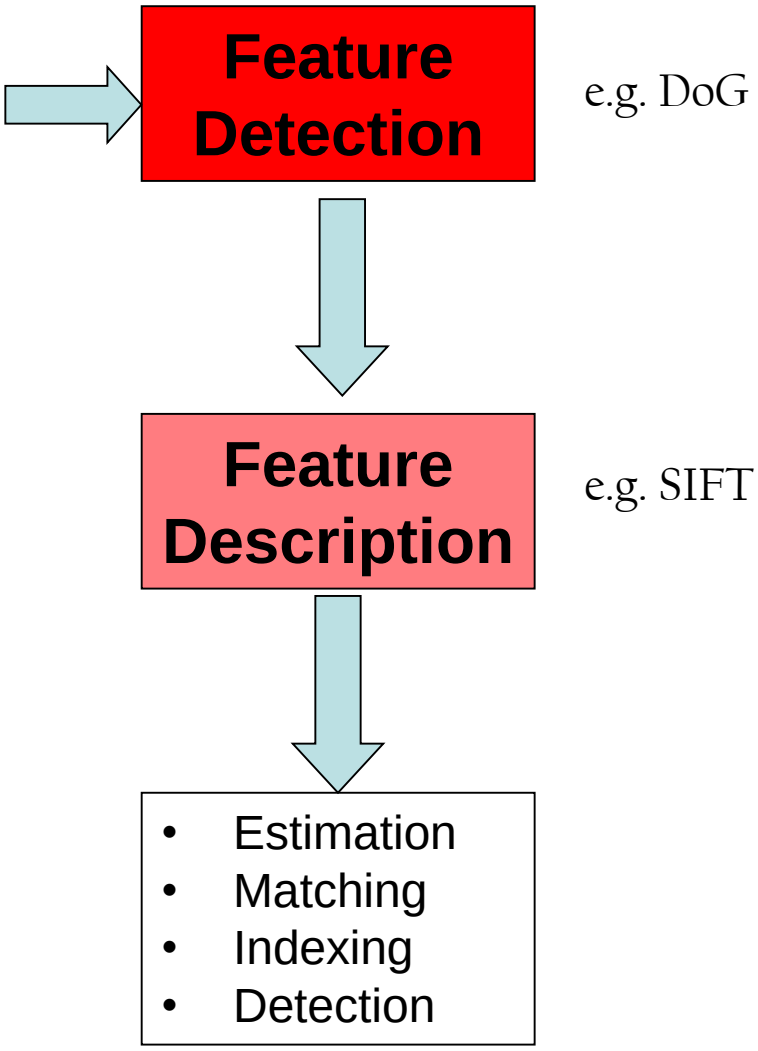
Lecture 11
Visual recognition

- Descriptors (wrapping up)
- An introduction to recognition
- Image classification

The big picture…

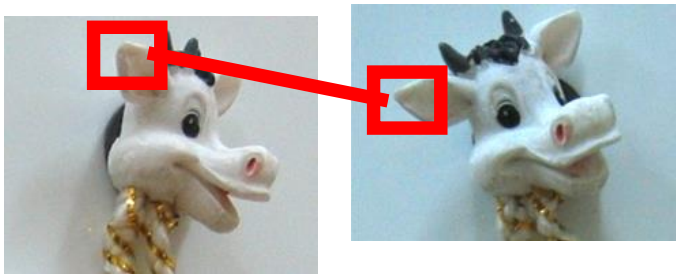


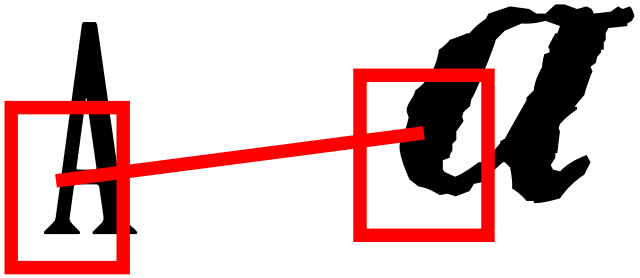


Properties

Depending on the application a descriptor must incorporate information that is:

- Invariant w.r.t:
  - Illumination
  - Pose
  - Scale
  - Intraclass variability



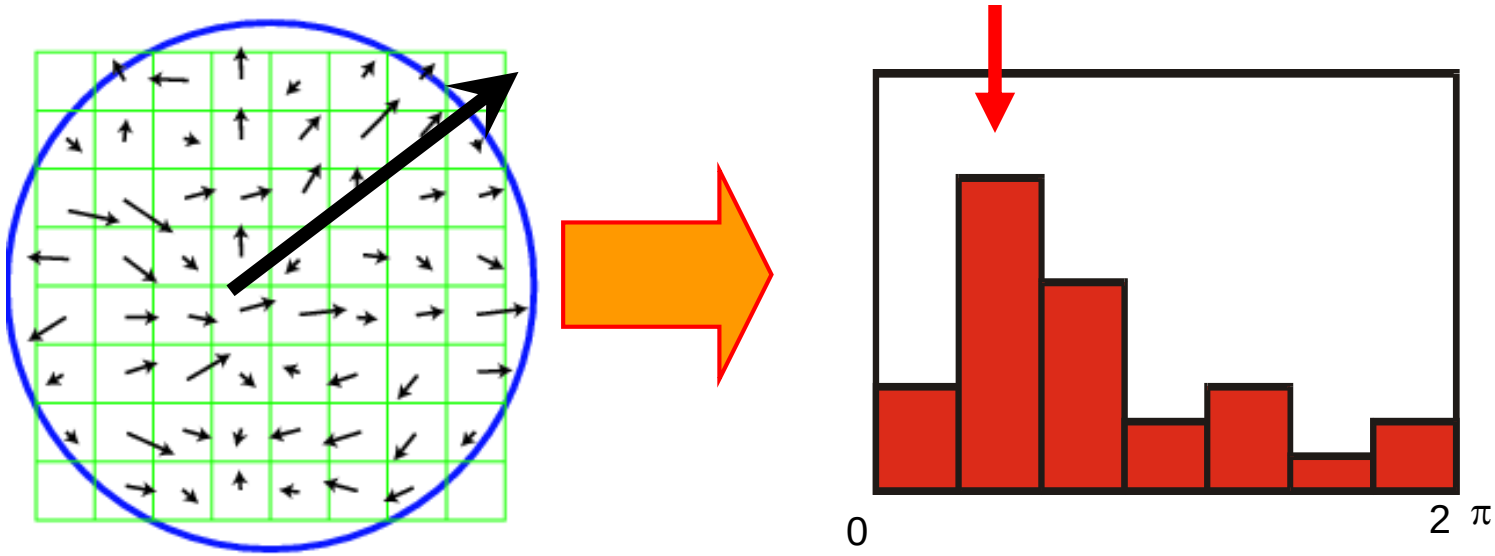


- Highly distinctive (allows a single feature to find its correct match with good probability in a large database of features)

Descriptor	Illumination	Pose	Intra-class variab.
PATCH	Good	Poor	Poor
FILTERS	Good	Medium	Medium
SIFT	Good	Good	Medium

Rotational invariance

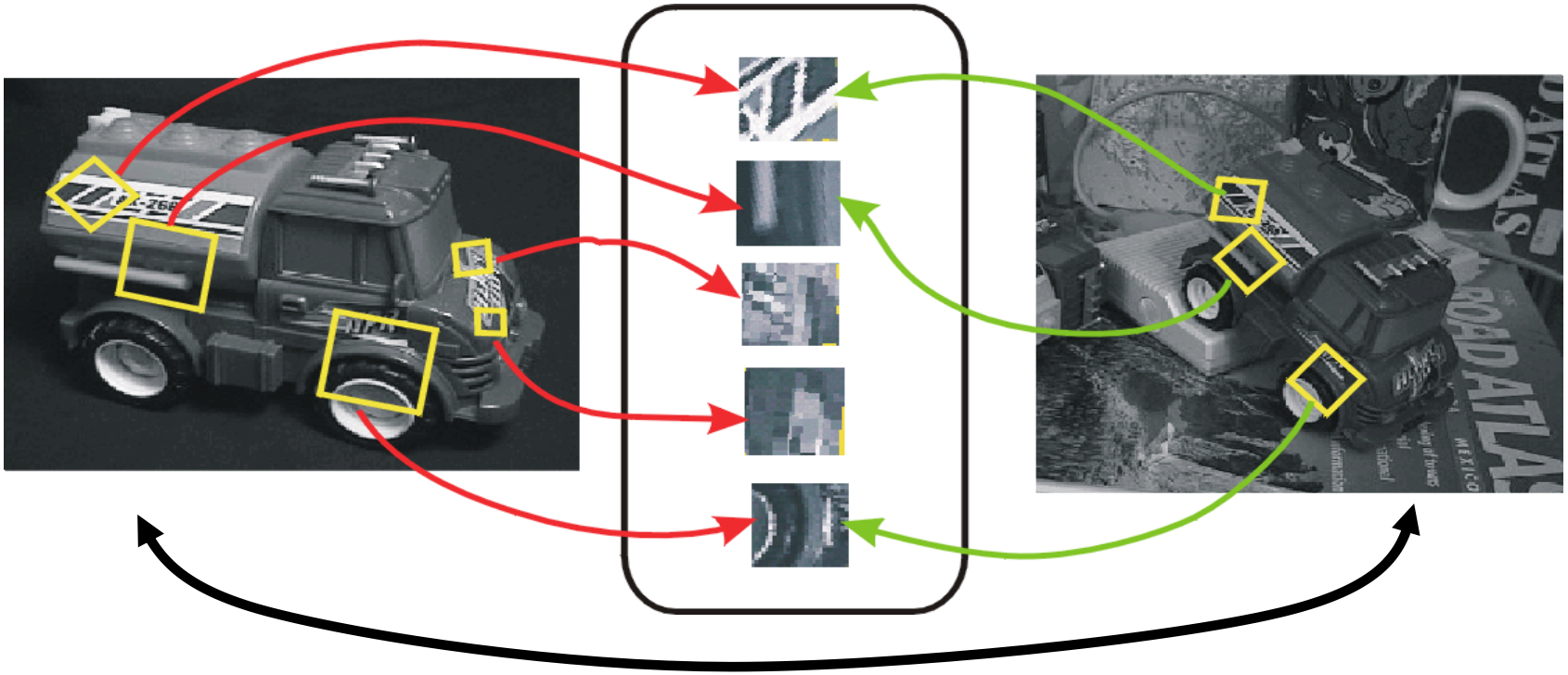
- Find dominant orientation by building a orientation histogram
- Rotate all orientations by the dominant orientation



This makes the SIFT descriptor rotational invariant

Pose normalization

- Keypoints are transformed in order to be invariant to translation, rotation, scale, and other geometrical parameters [Lowe 2000]

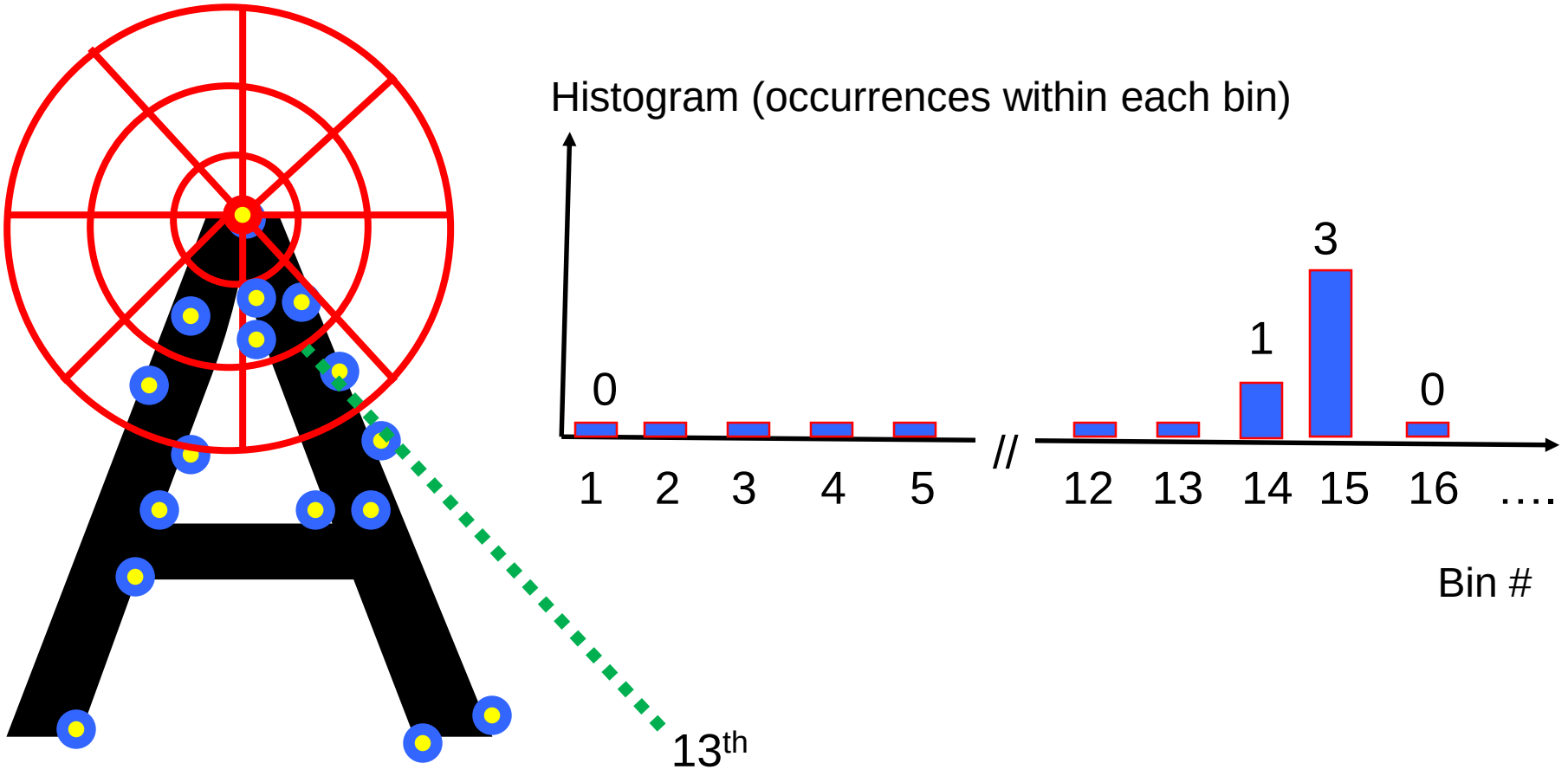


Courtesy of D. Lowe

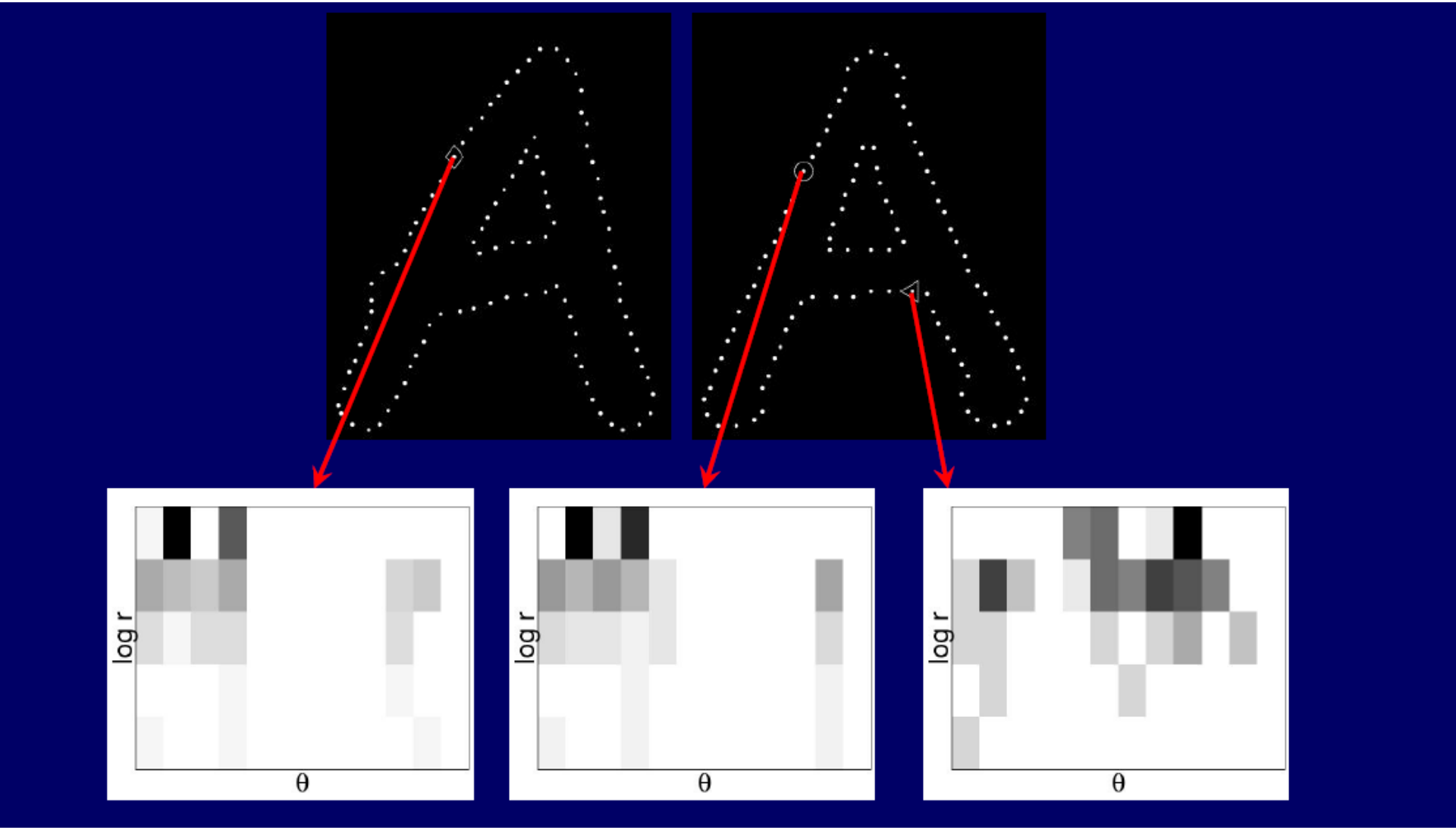
Change of scale, pose, illumination…

Shape context descriptor

Belongie et al. 2002



Shape context descriptor



Belongie et al. 02

Other detectors/descriptors

- **HOG: Histogram of oriented gradients**

  Dalal & Triggs, 2005

- **SURF: Speeded Up Robust Features**

  Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool, "SURF: Speeded Up Robust Features", Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346--359, 2008

- **FAST (corner detector)**

  Rosten. Machine Learning for High-speed Corner Detection, 2006.

- **ORB: an efficient alternative to SIFT or SURF**

  Ethan Rublee, Vincent Rabaud, Kurt Konolige, Gary R. Bradski: ORB: An efficient alternative to SIFT or SURF. ICCV 2011

- **Fast Retina Key- point (FREAK)**

  A. Alahi, R. Ortiz, and P. Vandergheynst. FREAK: Fast Retina Keypoint. In IEEE Conference on Computer Vision and Pattern Recognition, 2012. CVPR 2012 Open Source Award Winner.

Lecture 12
Visual recognition

- Descriptors (wrapping up)
- An introduction to recognition
- Image classification


ONE
WAY
macy's

Classification:

Does this image contain a building? [yes/no]

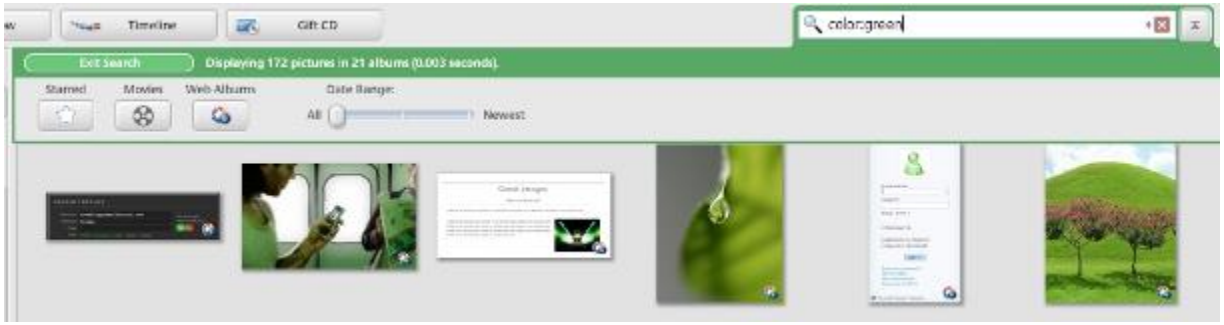
Classification:

Is this an beach?

Image Search

flickr webshots Ask Images Cydral picsearch Google Image Search altavista

Organizing photo collections

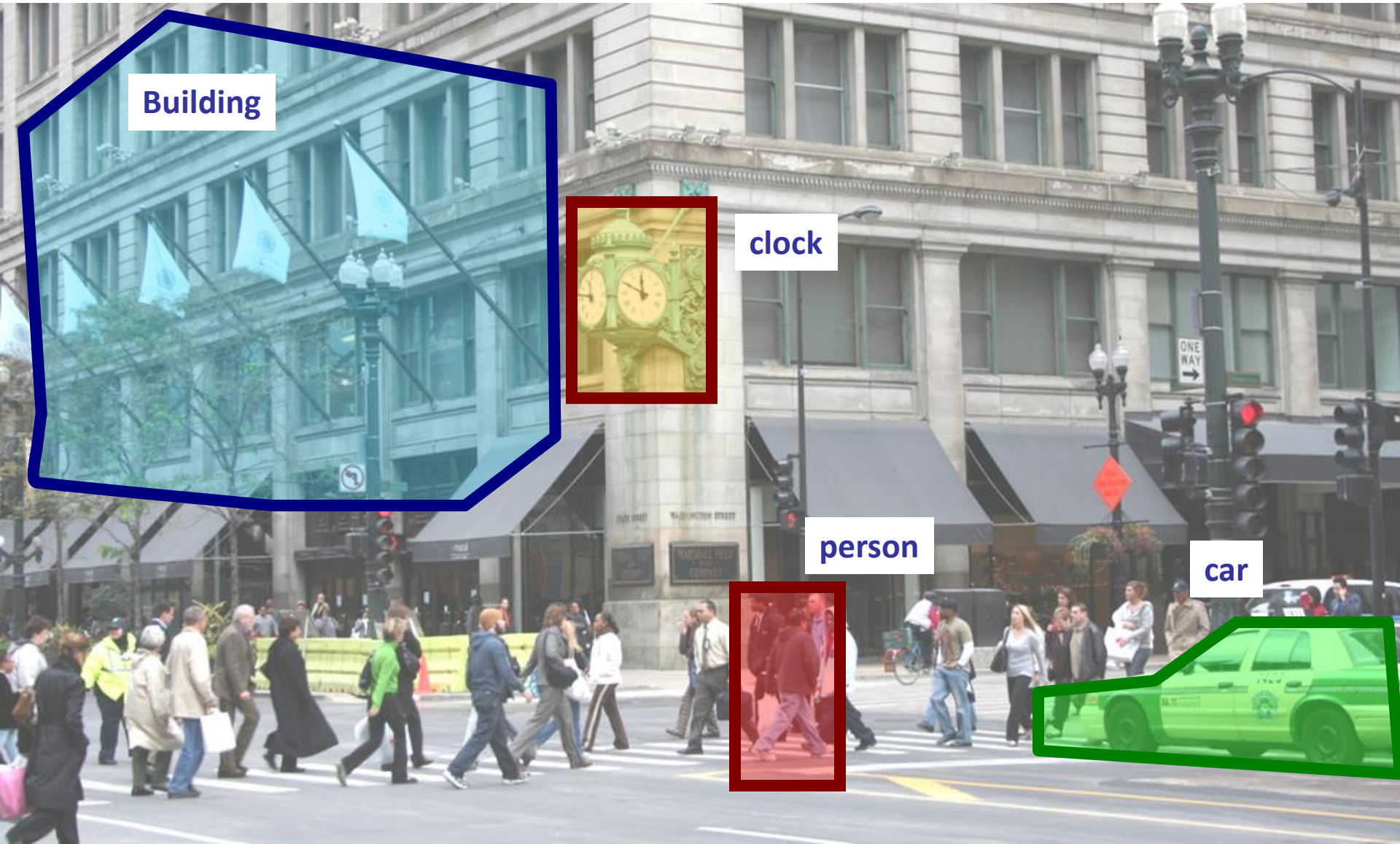


Detection:

Does this image contain a car? [where?]

Detection:

Which object does this image contain? [where?]



Detection:

Accurate localization (segmentation)

Object detection is useful…

Computational photography

Assistive technologies

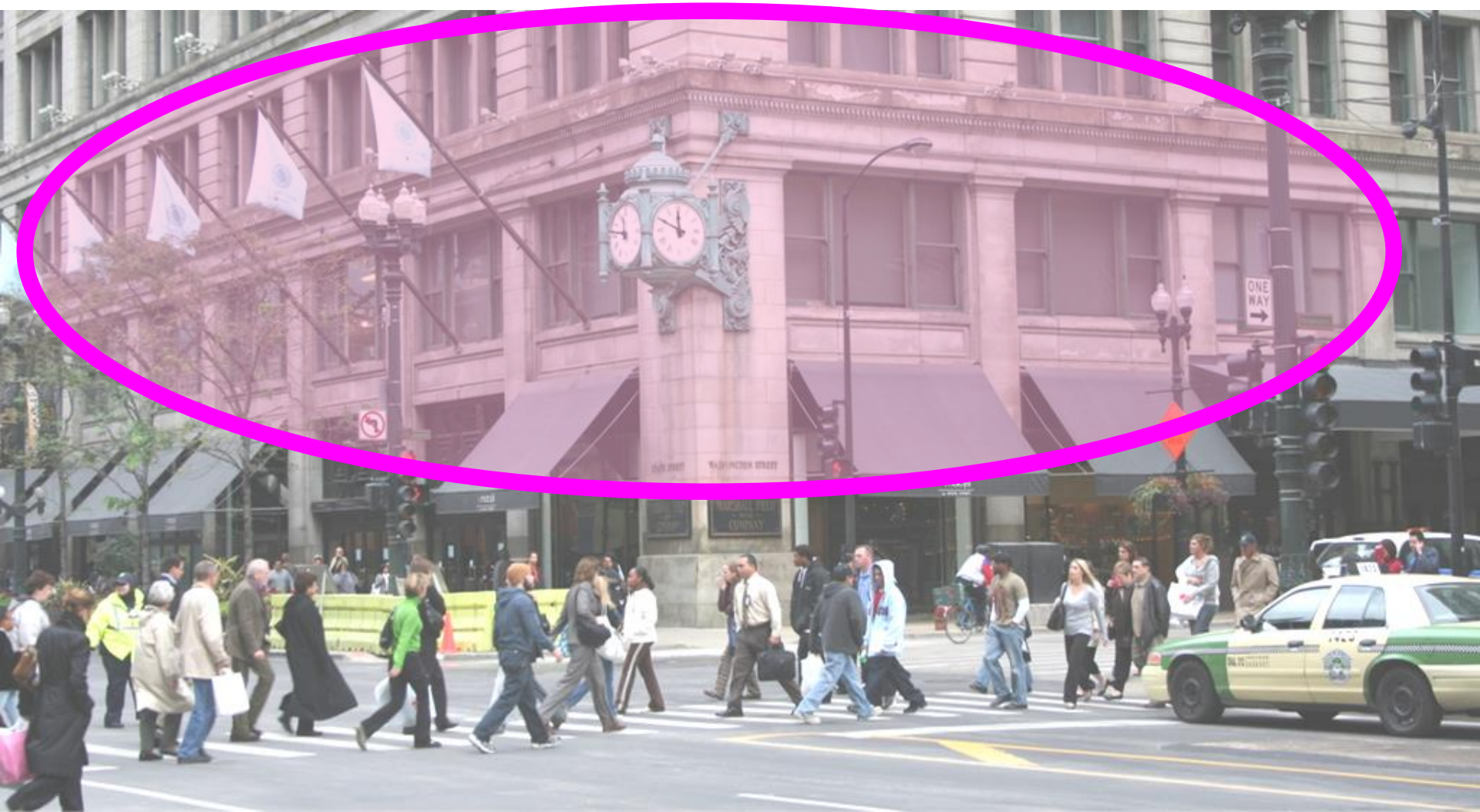
Surveillance

Security

Assistive driving

Categorization vs Single instance recognition

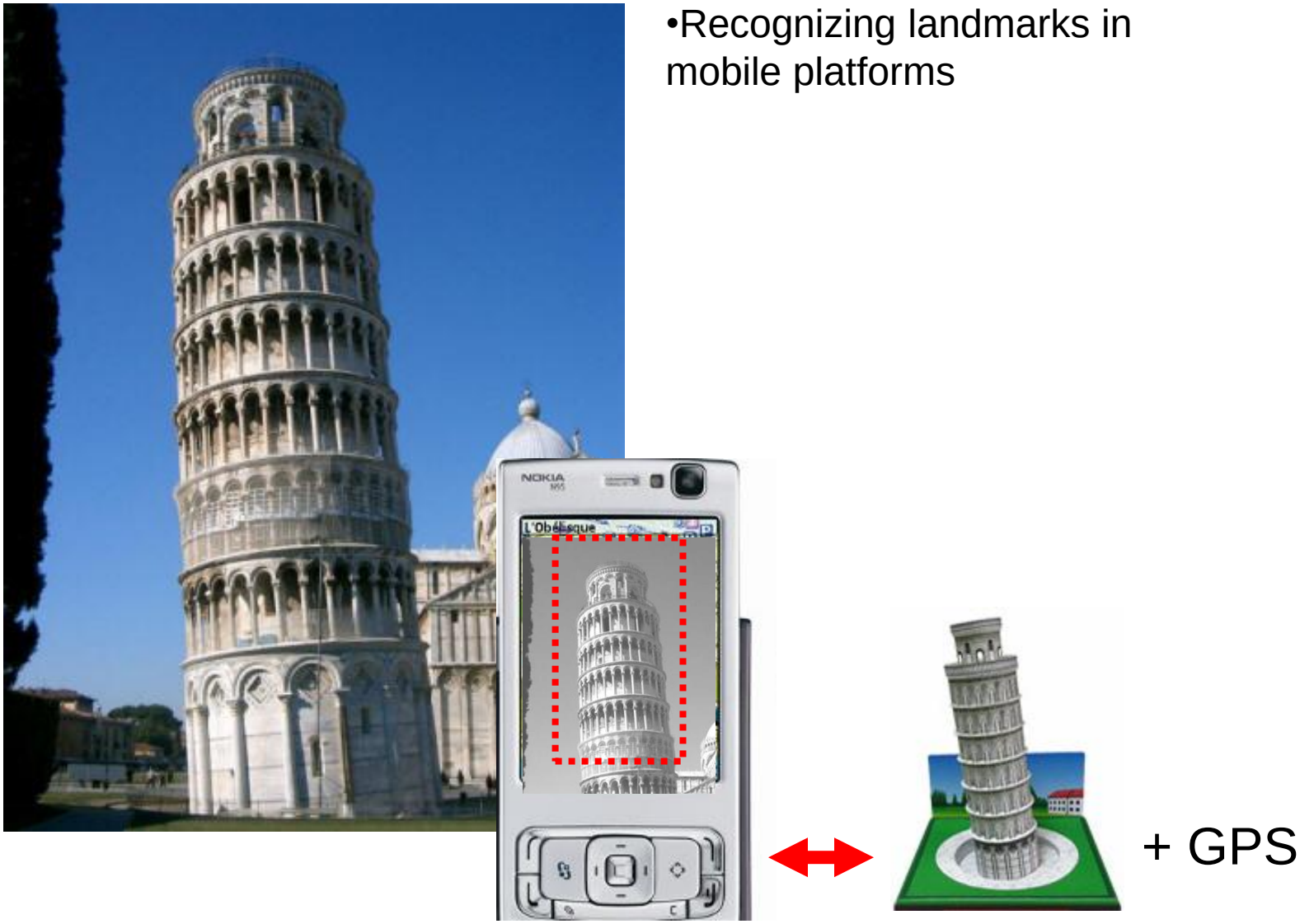
Which building is this? *Marshall Field* building in Chicago



Categorization vs Single instance recognition

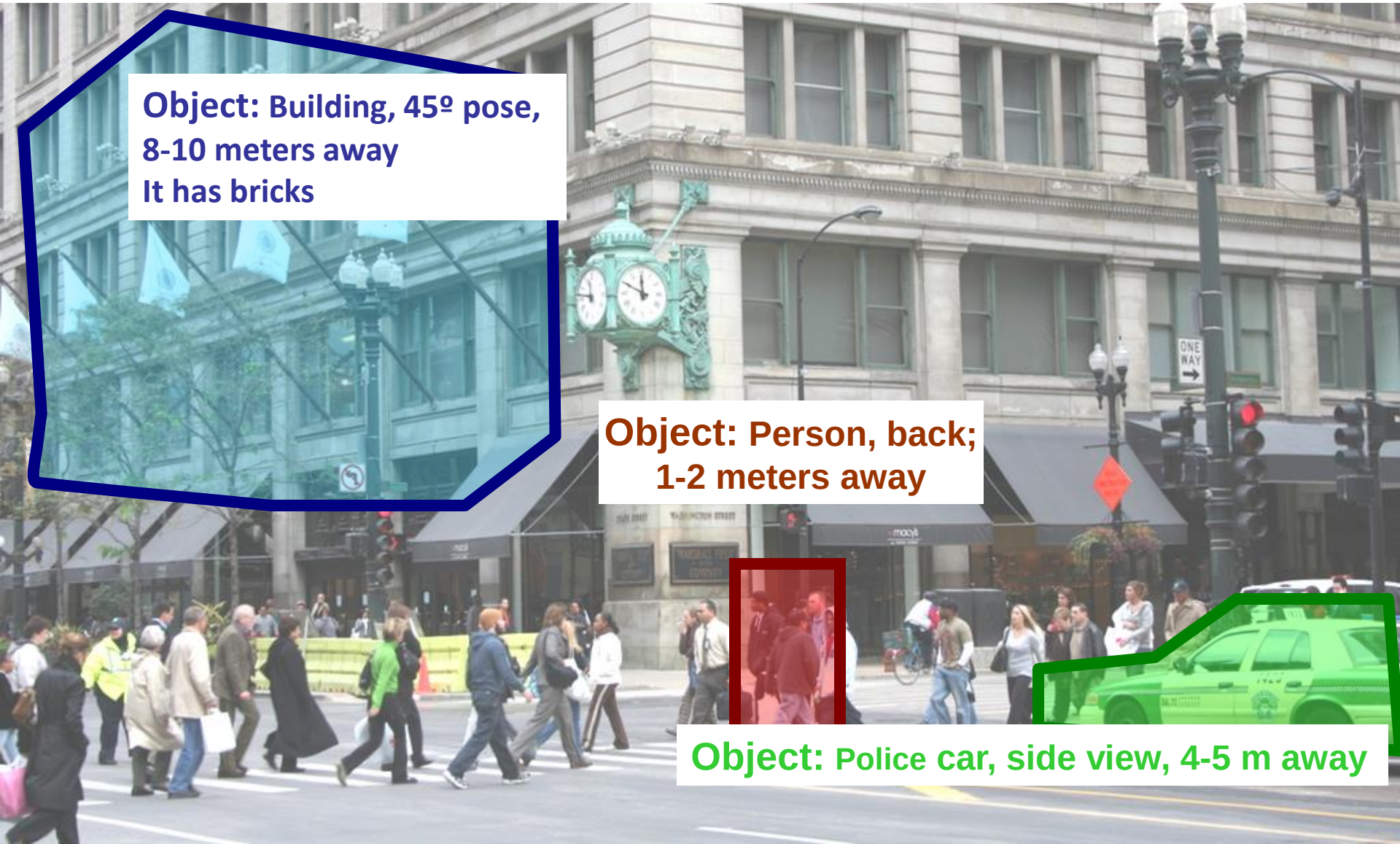
Where is the crunchy nut?

Applications of computer vision



•Recognizing landmarks in mobile platforms

Detection: Estimating object semantic & geometric attributes



Activity or Event recognition

What are these people doing?

Visual Recognition

- Design algorithms that are capable to
  - Classify images or videos
  - Detect and localize objects
  - Estimate semantic and geometrical attributes
  - Classify human activities and events

Why is this challenging?

How many object categories are there?

Challenges: viewpoint variation

Michelangelo 1475-1564

slide credit: Fei-Fei, Fergus & Torralba

Challenges: illumination

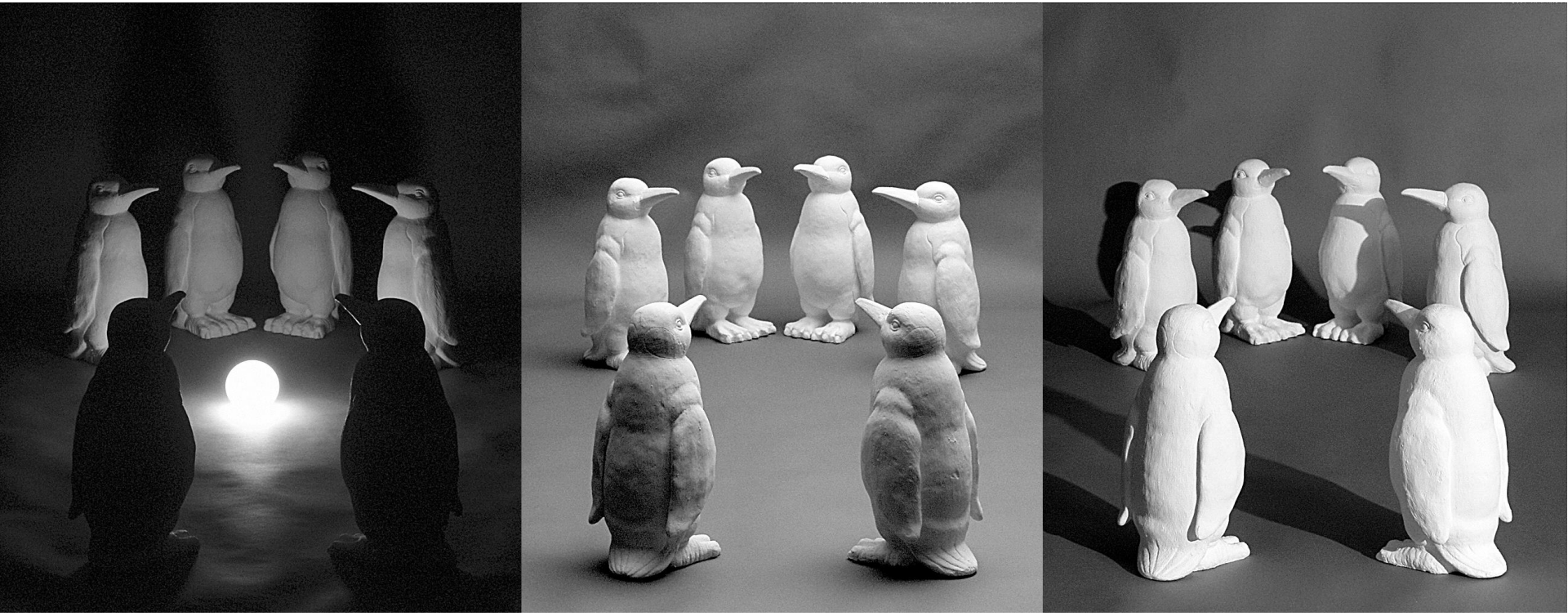


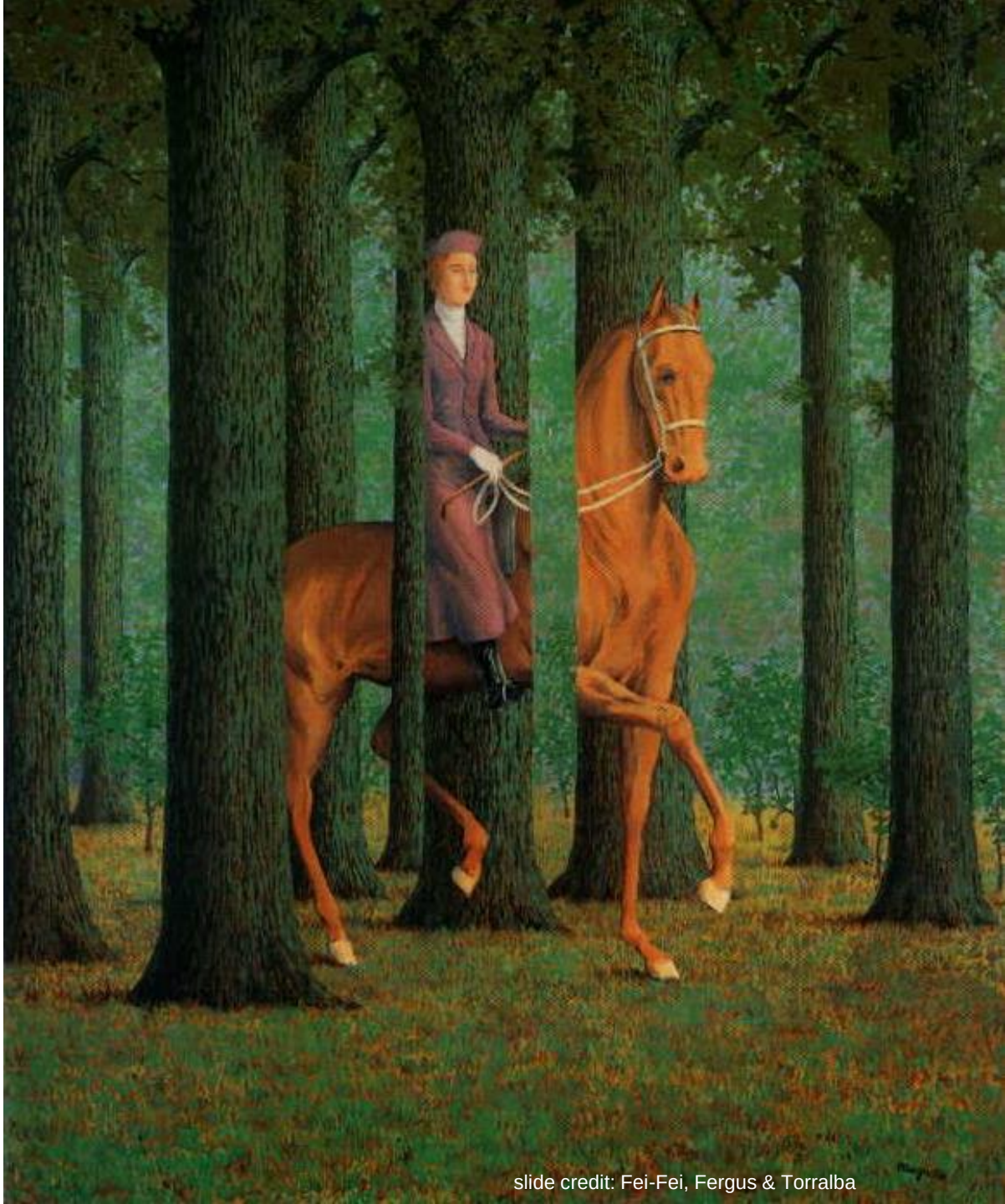
image credit: J. Koenderink

Challenges: scale

slide credit: Fei-Fei, Fergus & Torralba

Challenges: deformation

Challenges: occlusion



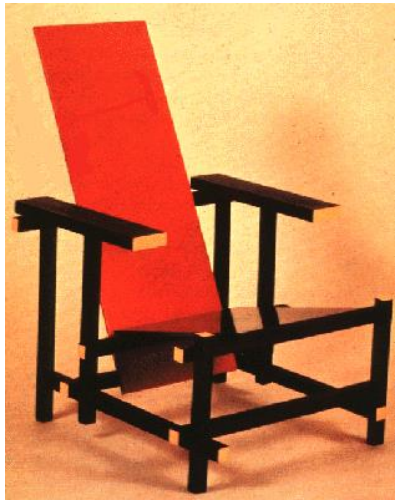
Magritte, 1957

slide credit: Fei-Fei, Fergus & Torralba

Challenges: background clutter

Kilmeny Niland. 1995

Challenges: intra-class variation



Some early works on object categorization

- Turk and Pentland, 1991
- Belhumeur, Hespanha, & Kriegman, 1997
- Schneiderman & Kanade 2004
- Viola and Jones, 2000

- Amit and Geman, 1999
- LeCun et al. 1998
- Belongie and Malik, 2002

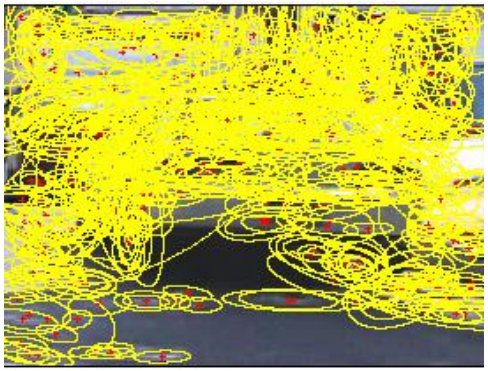
- Schneiderman & Kanade, 2004
- Argawal and Roth, 2002
- Poggio et al. 1993

Basic properties

- Representation
  - How to represent an object category; which classification scheme?
- Learning
  - How to learn the classifier, given training data
- Recognition
  - How the classifier is to be used on novel data

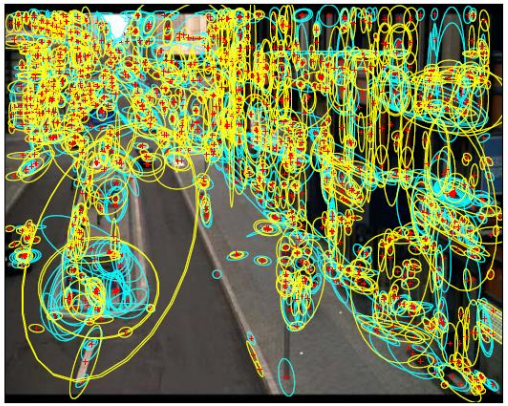
Representation

- Building blocks: Sampling strategies



Interest operators

Dense, uniformly



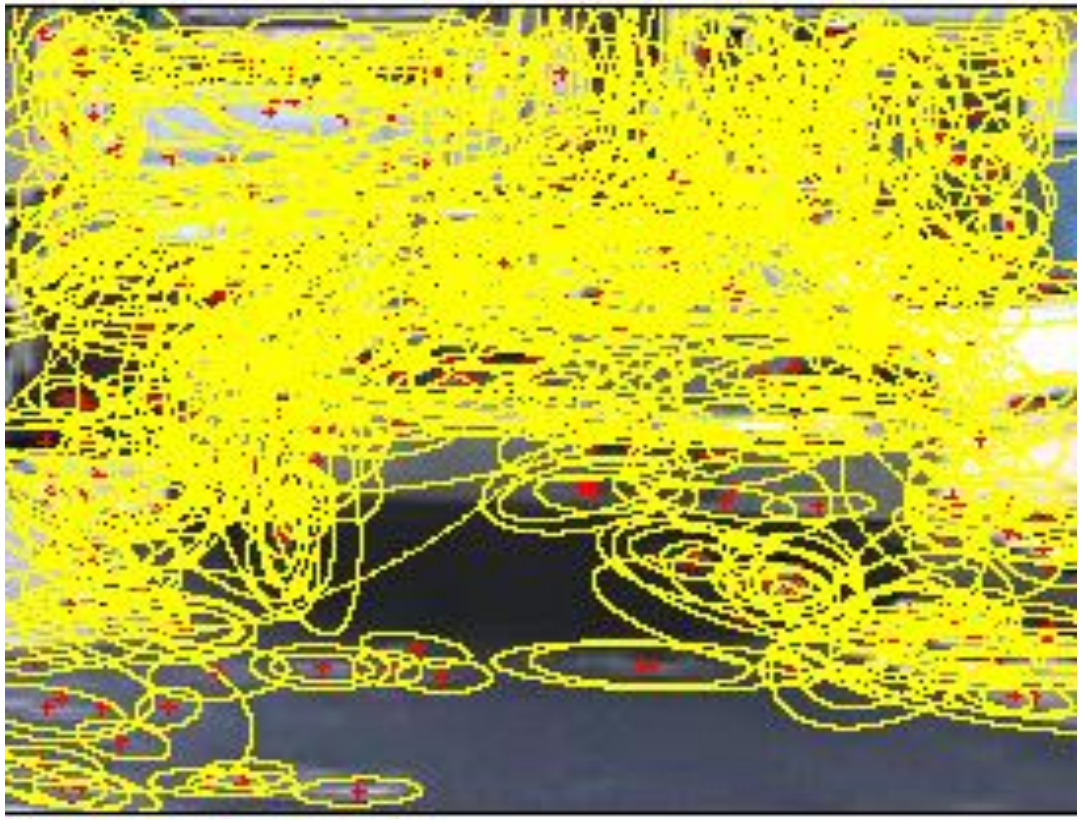
Multiple interest operators

Randomly

Image credits: F-F. Li, E. Nowak, J. Sivic

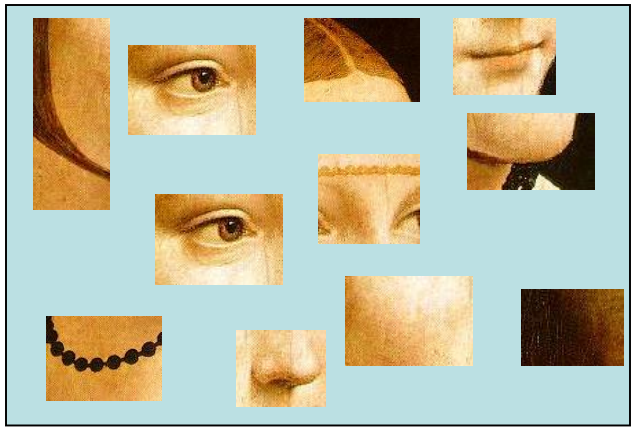
Representation

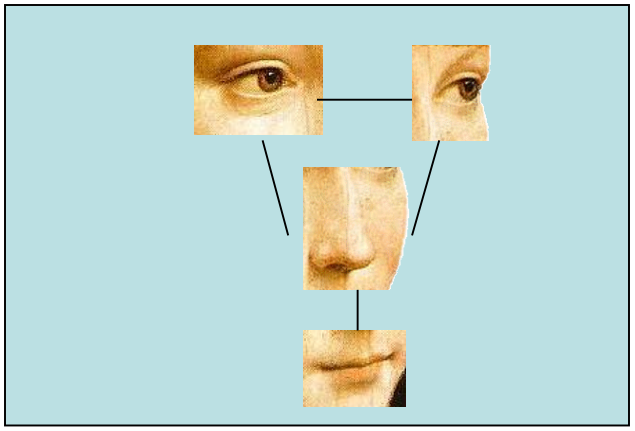
- Building blocks: Choice of descriptors
  [SIFT, HOG, codewords….]

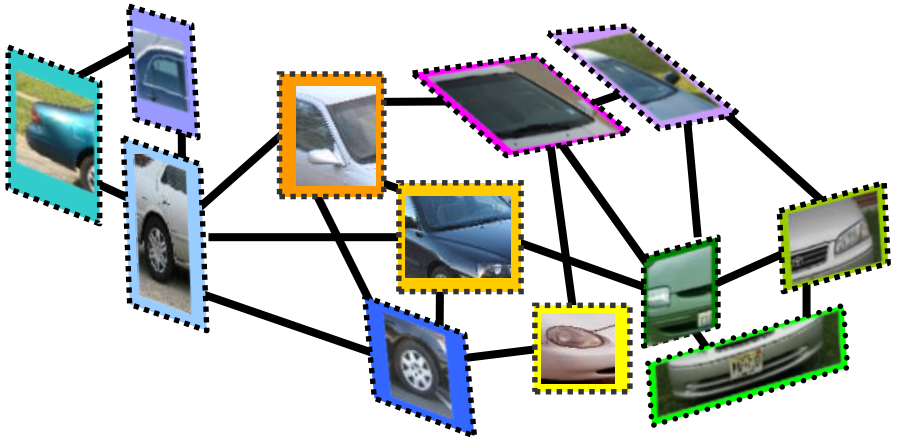


Representation

– Appearance only or location and appearance

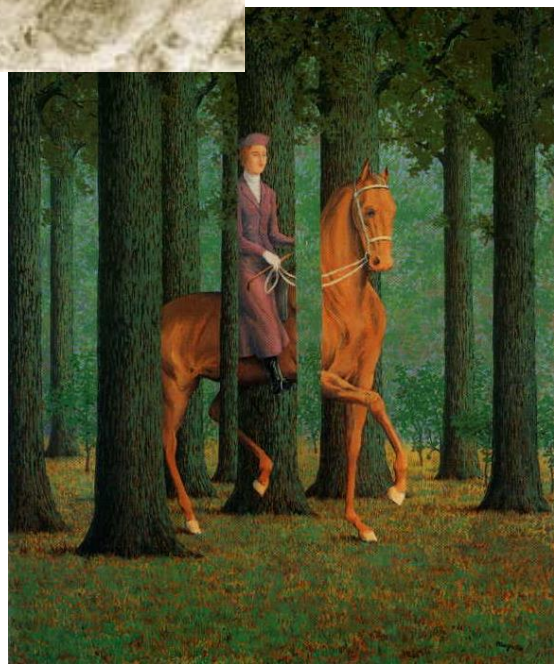






Representation

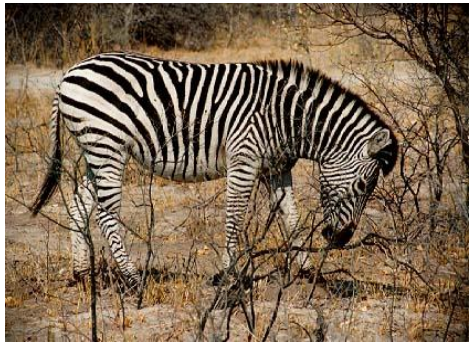
- Invariances
  - View point
  - Illumination
  - Occlusion
  - Scale
  - Deformation
  - Clutter
  - etc.



Representation

- To handle intra-class variability, it is convenient to describe an object categories using probabilistic models
- Object models: Generative vs Discriminative vs hybrid

Object categorization: the statistical viewpoint



$$p(zebra \mid image)$$

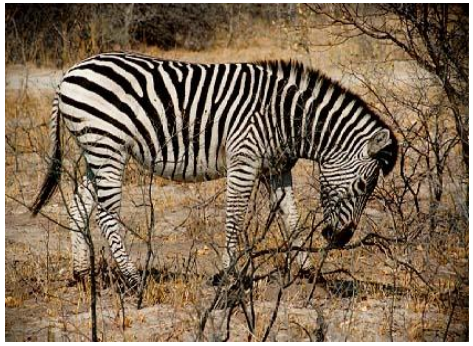
vs.

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

- Bayes rule: $P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$

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

Object categorization: the statistical viewpoint



$$p(zebra \mid image)$$

vs.

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

- Bayes rule: $P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$

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

Object categorization: the statistical viewpoint

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

- Bayes rule:

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

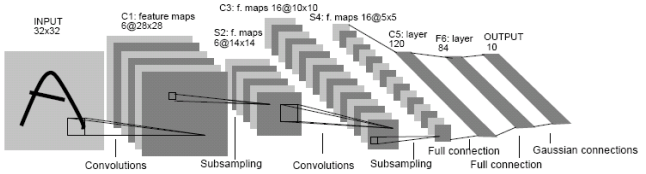
Discriminative models

Nearest neighbor

10^6 examples

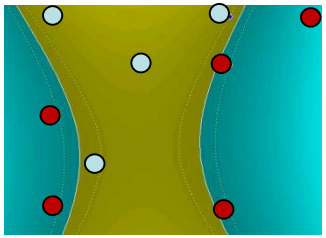
Shakhnarovich, Viola, Darrell 2003
Berg, Berg, Malik 2005...

Neural networks



LeCun, Bottou, Bengio, Haffner 1998
Rowley, Baluja, Kanade 1998
…

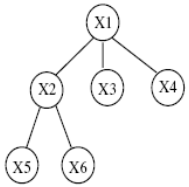
Support Vector Machines



Guyon, Vapnik, Heisele, Serre, Poggio…

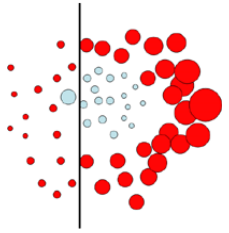
Latent SVM

Structural SVM



Felzenszwalb 00
Ramanan 03…

Boosting



Viola, Jones 2001,
Torralba et al. 2004,
Opelt et al. 2006,…

Courtesy of Vittorio Ferrari
Slide credit: Kristen Grauman

Slide adapted from Antonio Torralba

Generative models

- Naïve Bayes classifier
  - Csurka Bray, Dance & Fan, 2004
- Hierarchical Bayesian topic models (e.g. pLSA and LDA)
  - Object categorization: Sivic et al. 2005, Sudderth et al. 2005
  - Natural scene categorization: Fei-Fei et al. 2005
- 2D Part based models
  - Constellation models: Weber et al 2000; Fergus et al 200
  - Star models: ISM (Leibe et al 05)
- 3D part based models:
  - multi-aspects: Sun, et al, 2009

Basic properties

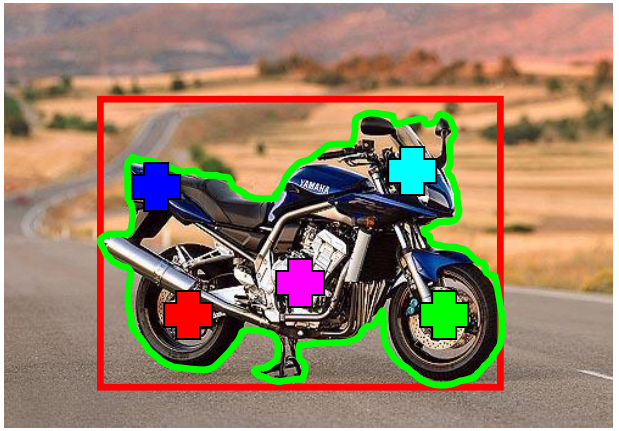
- Representation
  - How to represent an object category; which classification scheme?
- Learning
  - How to learn the classifier, given training data
- Recognition
  - How the classifier is to be used on novel data

Learning

- Learning parameters: What are you maximizing? Likelihood (Gen.) or performances on train/validation set (Disc.)

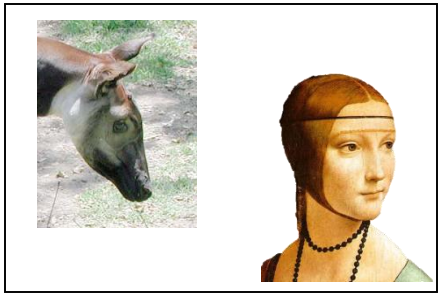
Learning

- Learning parameters: What are you maximizing? Likelihood (Gen.) or performances on train/validation set (Disc.)
- Level of supervision
  - Manual segmentation; bounding box; image labels; noisy labels
- Batch/incremental
- Priors



Learning

- Learning parameters: What are you maximizing? Likelihood (Gen.) or performances on train/validation set (Disc.)
- Level of supervision
  - Manual segmentation; bounding box; image labels; noisy labels
- Batch/incremental
- Priors
- Training images:
  - Issue of overfitting
  - Negative images for discriminative methods



Basic properties

- Representation
  - How to represent an object category; which classification scheme?
- Learning
  - How to learn the classifier, given training data
- Recognition
  - How the classifier is to be used on novel data

Recognition

– Recognition task: classification, detection, etc..

Recognition

– Recognition task

– Search strategy: Sliding Windows Viola, Jones 2001,

- Simple
- Computational complexity (x,y, S, θ, N of classes)

  - BSW by Lampert et al 08

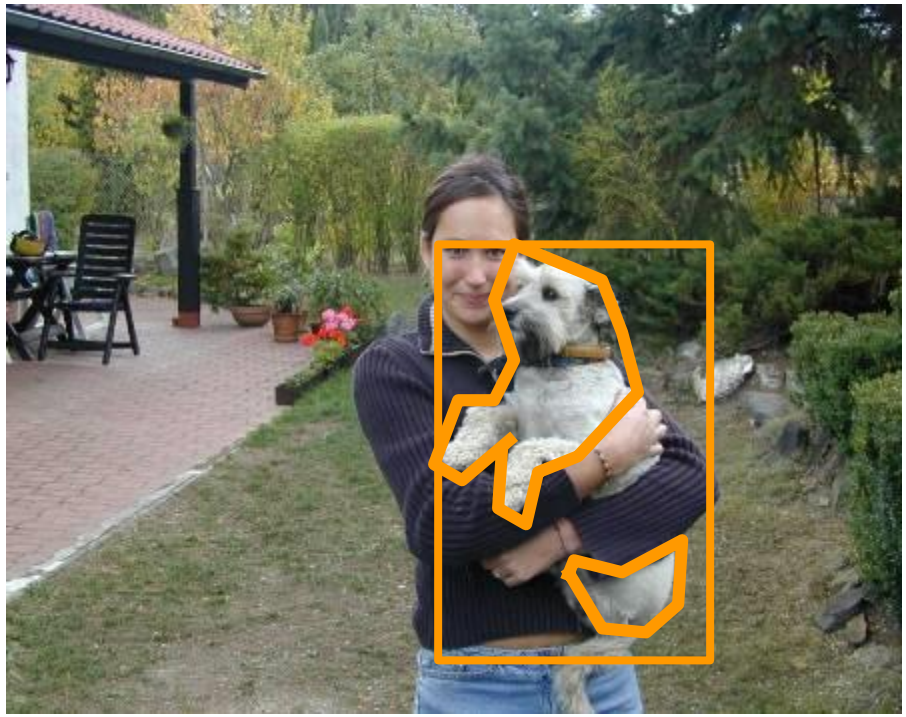
  - Also, Alexe, et al 10

Recognition

– Recognition task

– Search strategy: Sliding Windows Viola, Jones 2001,

- Simple
- Computational complexity (x,y, S, θ, N of classes)
  - BSW by Lampert et al 08
  - Also, Alexe, et al 10
- Localization
  - Objects are not boxes



Recognition

– Recognition task

– Search strategy: Sliding Windows Viola, Jones 2001,

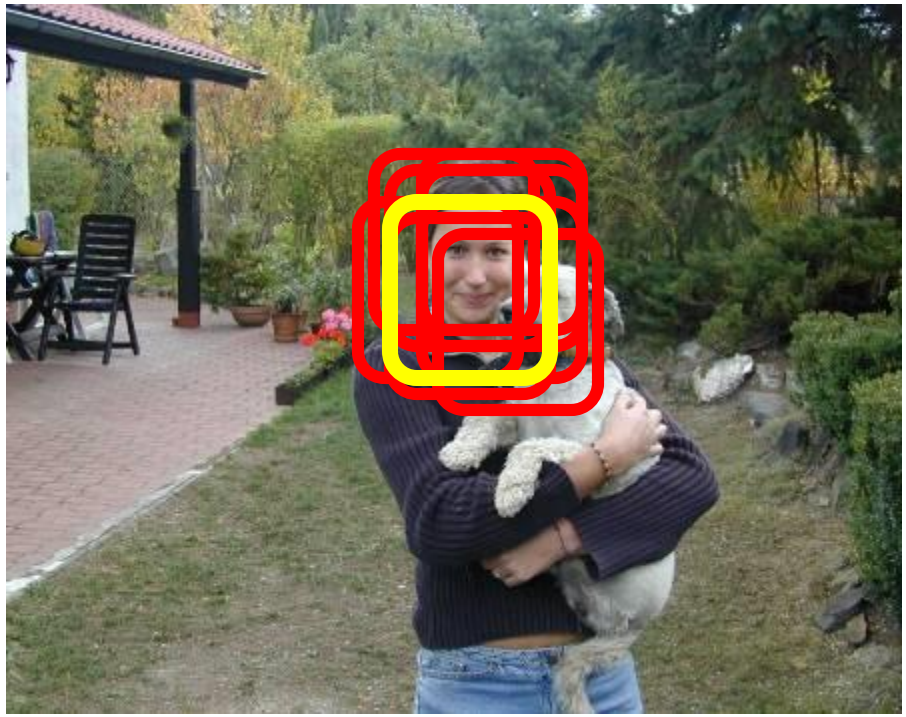
- Simple
- Computational complexity (x,y, S, θ, N of classes)

  - BSW by Lampert et al 08

  - Also, Alexe, et al 10

- Localization
  - Objects are not boxes
  - Prone to false positive

    Non max suppression:
    Canny ’86
    ….
    Desai et al , 2009

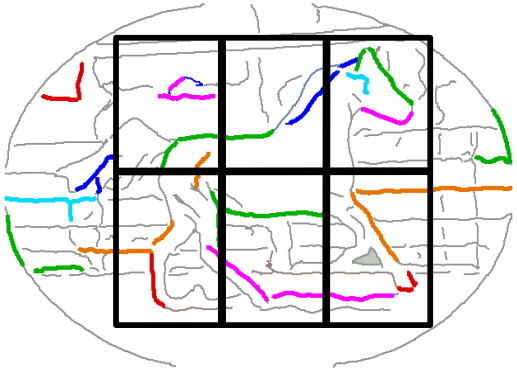


Successful methods using sliding windows

- Subdivide scanning window
- In each cell compute histogram of gradients orientation.

Code available: http://pascal.inrialpes.fr/soft/olt/

[Dalal & Triggs, CVPR 2005]



- Subdivide scanning window
- In each cell compute histogram of codewords of adjacent segments

Code available: http://www.vision.ee.ethz.ch/~calvin

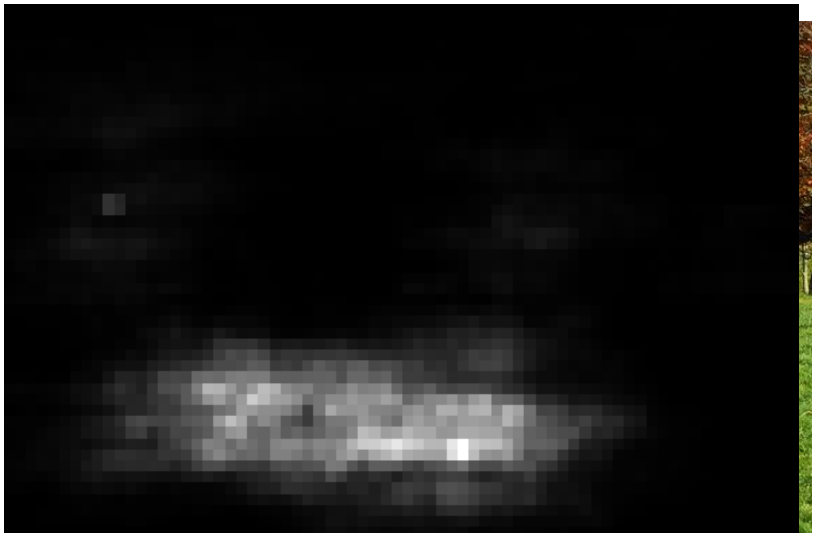
[Ferrari & al, PAMI 2008]

Recognition

– Recognition task

– Search strategy : Probabilistic “heat maps”

- Fergus et al 03
- Leibe et al 04



Recognition

- Recognition task
- Search strategy :
  - Hypothesis generation + verification

Recognition

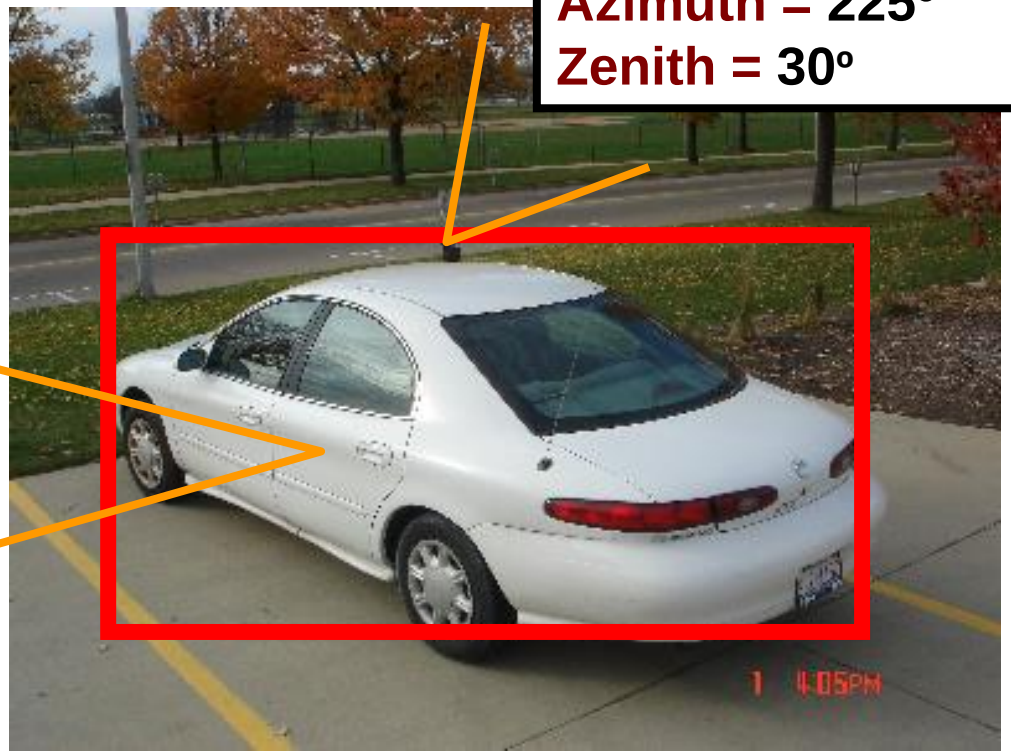
– Recognition task

– Search strategy

– Attributes

•Savarese, 2007
•Sun et al 2009
• Liebelt et al., ’08, 10
•Farhadi et al 09

Category: car
Azimuth = 225º
Zenith = 30º



- It has metal
- it is glossy
- has wheels

•Farhadi et al 09
• Lampert et al 09
• Wang & Forsyth 09

Recognition

– Recognition task

– Search strategy

– Attributes

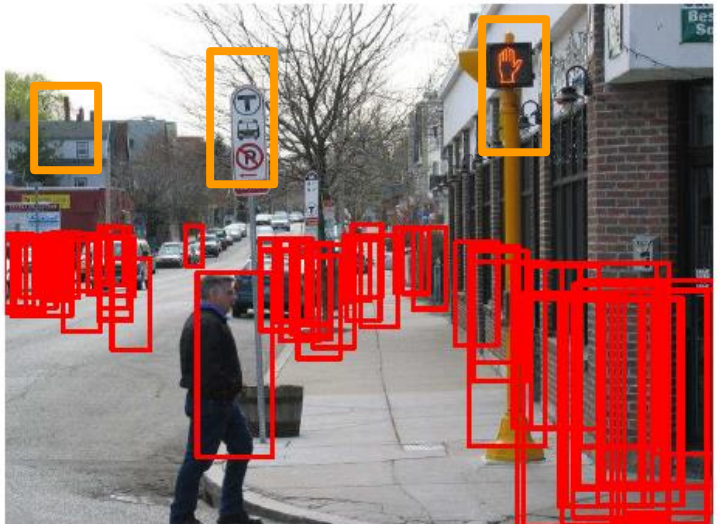
– Context

Semantic:

- Torralba et al 03
- Rabinovich et al 07
- Gupta & Davis 08
- Heitz & Koller 08
- L-J Li et al 08
- Bang & Fei-Fei 10

Geometric

- Hoiem, et al 06
- Gould et al 09
- Bao, Sun, Savarese 10



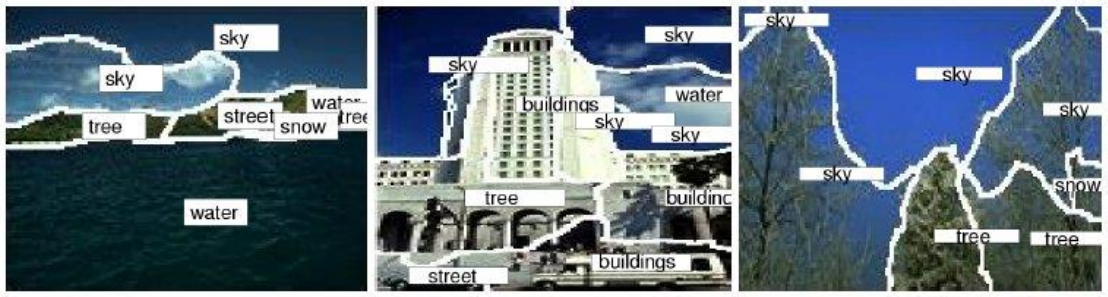
Segmentation

- Bottom up segmentation

Malik et al. 01
Maire et al. 08

Felzenszwalb and Huttenlocher, 2004

- Semantic segmentation



Duygulu et al. 02

Agenda on recognition

- Image classification
  - Bag of words representations
- Object detection
  - 2D object detection
  - 3D object detection
- Scene understanding
- Activity understanding

Lecture 12
Visual recognition

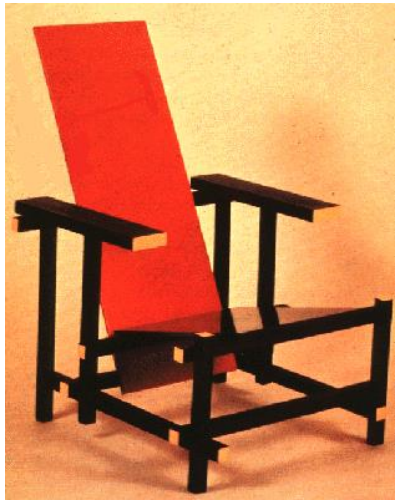
- Descriptors (wrapping up)
- An introduction to recognition
- Image classification
  - Bag of words models

Challenges:

Variability due to:

- View point
- Illumination
- Occlusions
- Etc..

Challenges: intra-class variation



Basic properties

- Representation
  - How to represent an object category; which classification scheme?
- Learning
  - How to learn the classifier, given training data
- Recognition
  - How the classifier is to be used on novel data

Part 1: Bag-of-words models

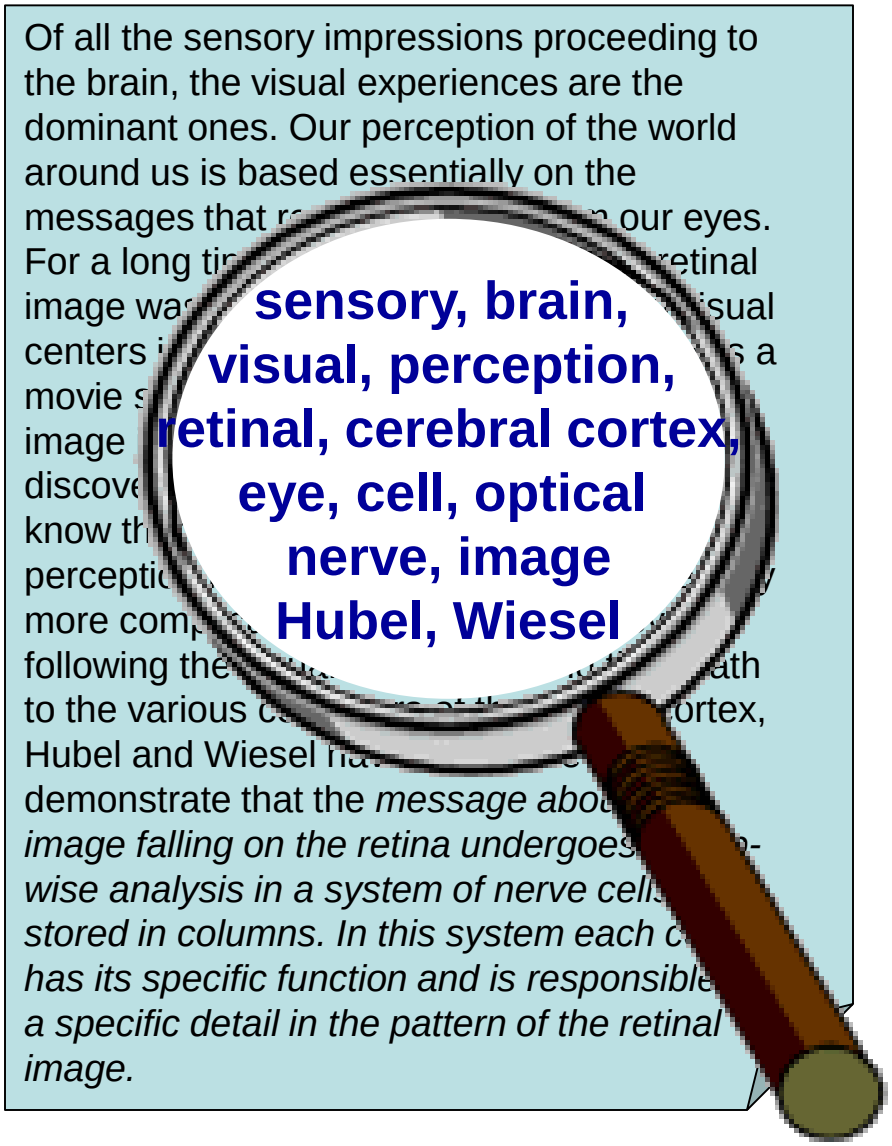
This segment is based on the tutorial *"**Recognizing and Learning Object Categories: Year 2007**",* by Prof A. Torralba, R. Fergus and F. Li

Related works

- Early “bag of words” models: mostly texture recognition
  - Cula & Dana, 2001; Leung & Malik 2001; Mori, Belongie & Malik, 2001; Schmid 2001; Varma & Zisserman, 2002, 2003; Lazebnik, Schmid & Ponce, 2003;
- Hierarchical Bayesian models for documents (pLSA, LDA, etc.)
  - Hoffman 1999; Blei, Ng & Jordan, 2004; Teh, Jordan, Beal & Blei, 2004
- Object categorization
  - Csurka, Bray, Dance & Fan, 2004; Sivic, Russell, Efros, Freeman & Zisserman, 2005; Sudderth, Torralba, Freeman & Willsky, 2005;
- Natural scene categorization
  - Vogel & Schiele, 2004; Fei-Fei & Perona, 2005; Bosch, Zisserman & Munoz, 2006

Object → Bag of ‘words’

Analogy to documents



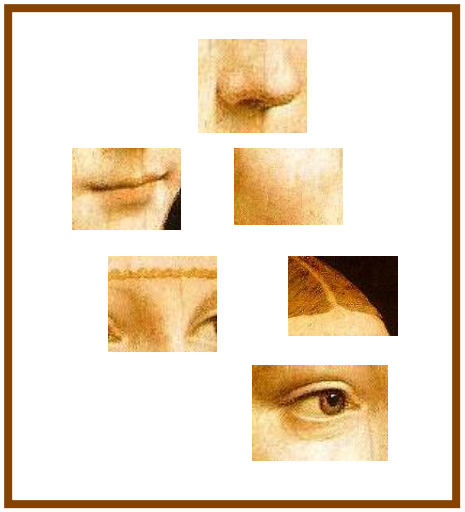
Of all the sensory impressions proceeding to the brain, the visual experiences are the dominant ones. Our perception of the world around us is based essentially on the messages that r… our eyes. For a long ti… retinal image wa… isual centers i… s a movie s… image … discove… know th… perceptio… more com… following the… ath to the various c… ortex, Hubel and Wiesel ha… demonstrate that the *message abou… image falling on the retina undergoes… -wise analysis in a system of nerve cell… stored in columns. In this system each c… has its specific function and is responsible… a specific detail in the pattern of the retinal image.*

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 created by a predicted 30% … $750bn, compared wi… $660bn. T… annoy th… China's … deliber… agrees … yuan is … governo… also need… demand so … country. China … e yuan against the dol… nd permitted it to trade within a narrow … but the US wants the yuan to be allowed … de freely. However, Beijing has made it cl… t it will take its time and tread carefully be… allowing the yuan to rise further in value.

definition of “BoW”

– Independent features

face

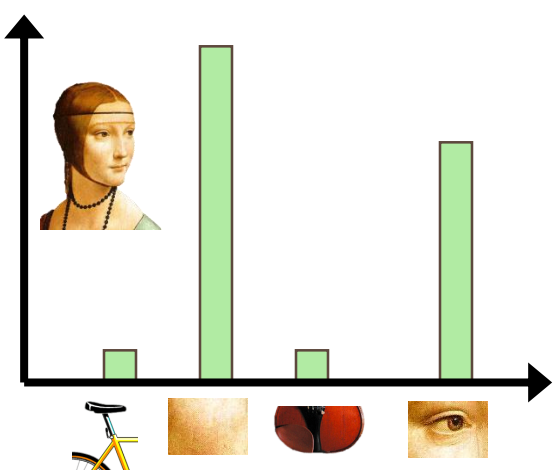


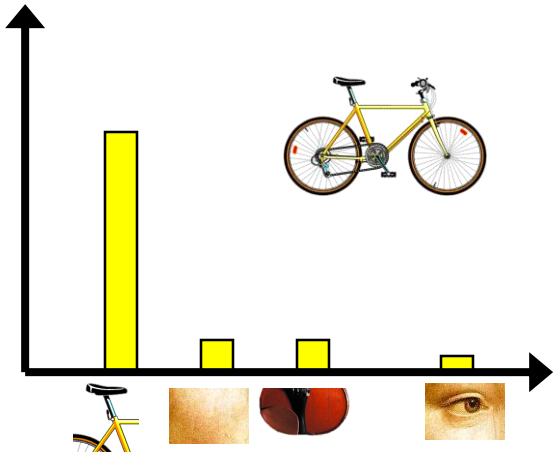
bike

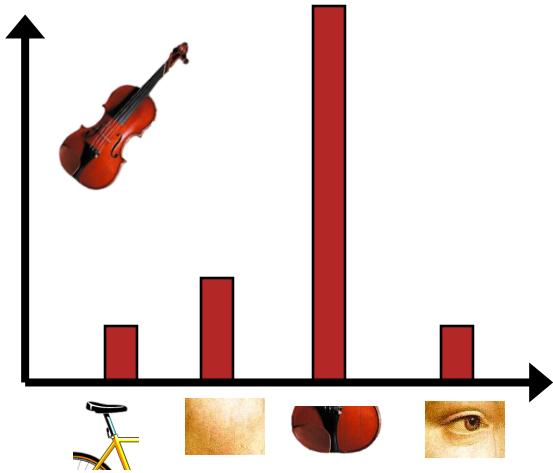
violin

definition of “BoW”

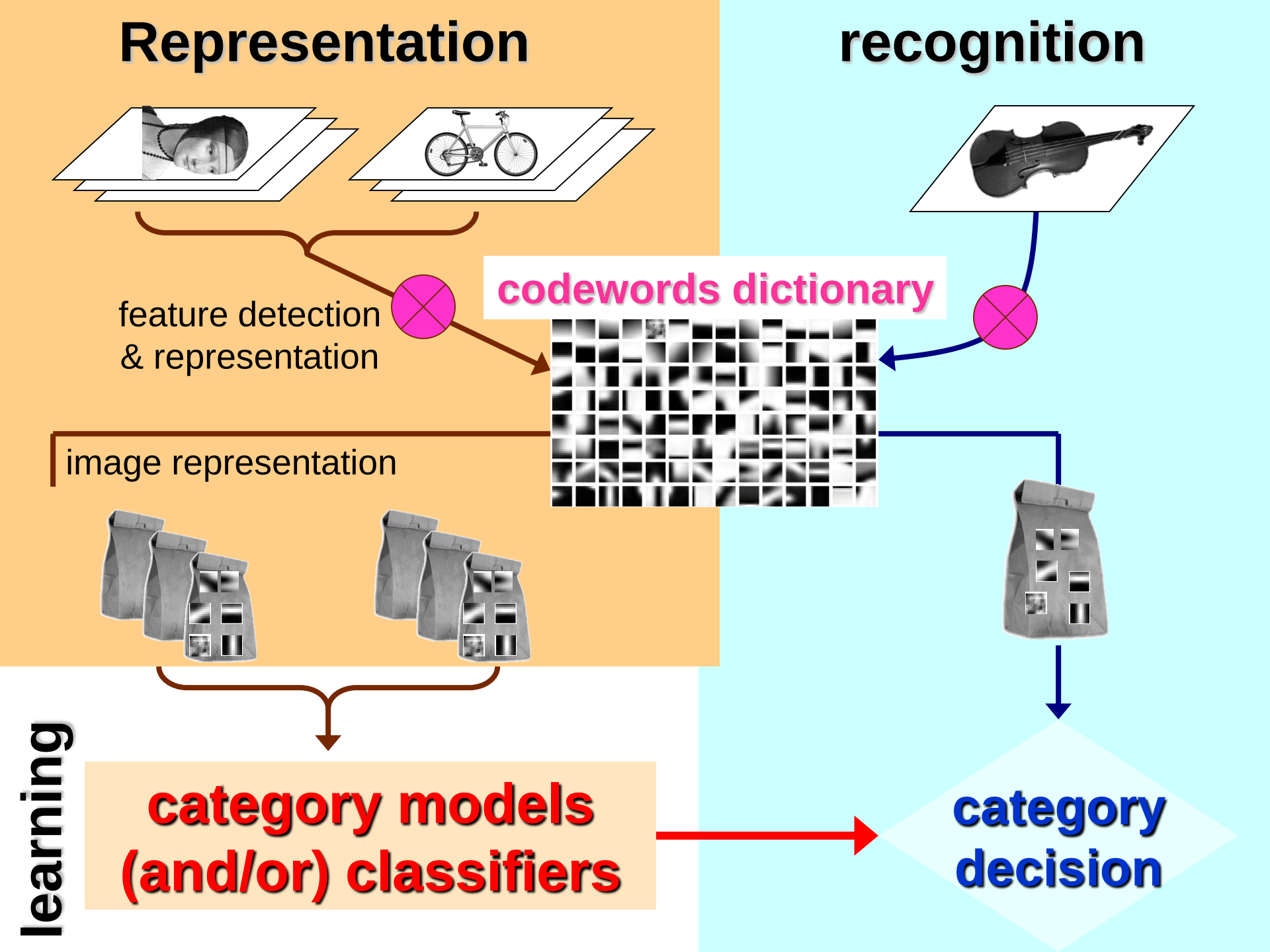
- Independent features
- histogram representation



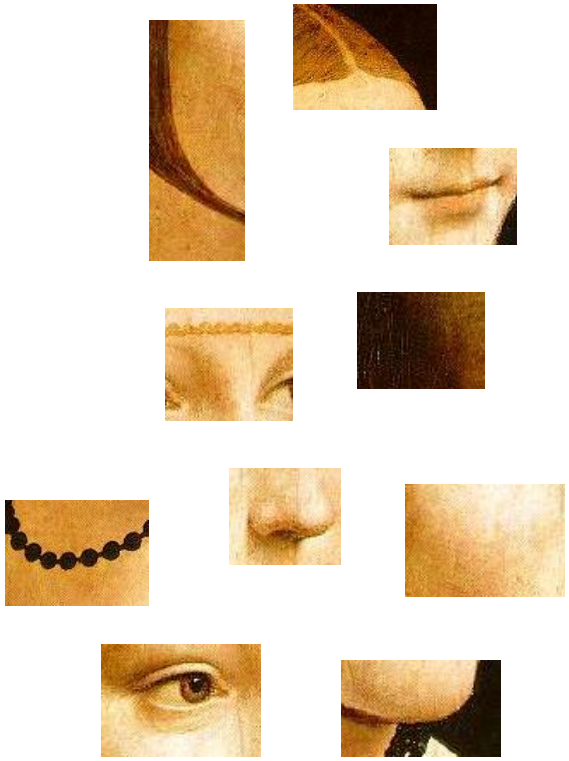




codewords dictionary

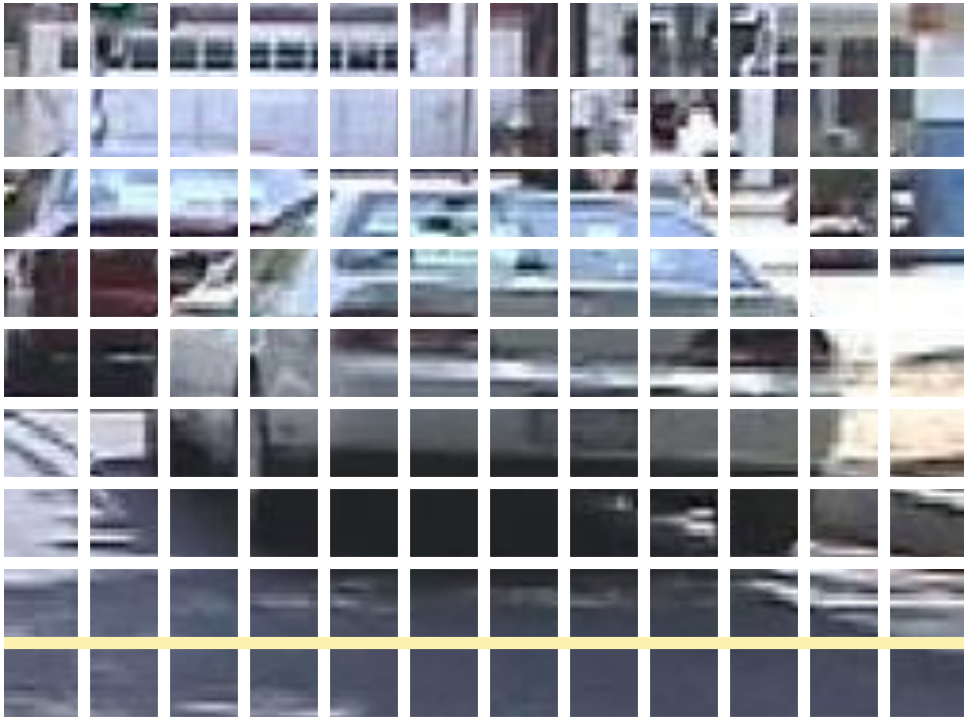

Representation
recognition
feature detection
& representation
codewords dictionary
image representation
learning
category models
(and/or) classifiers
category
decision

1.Feature detection and description



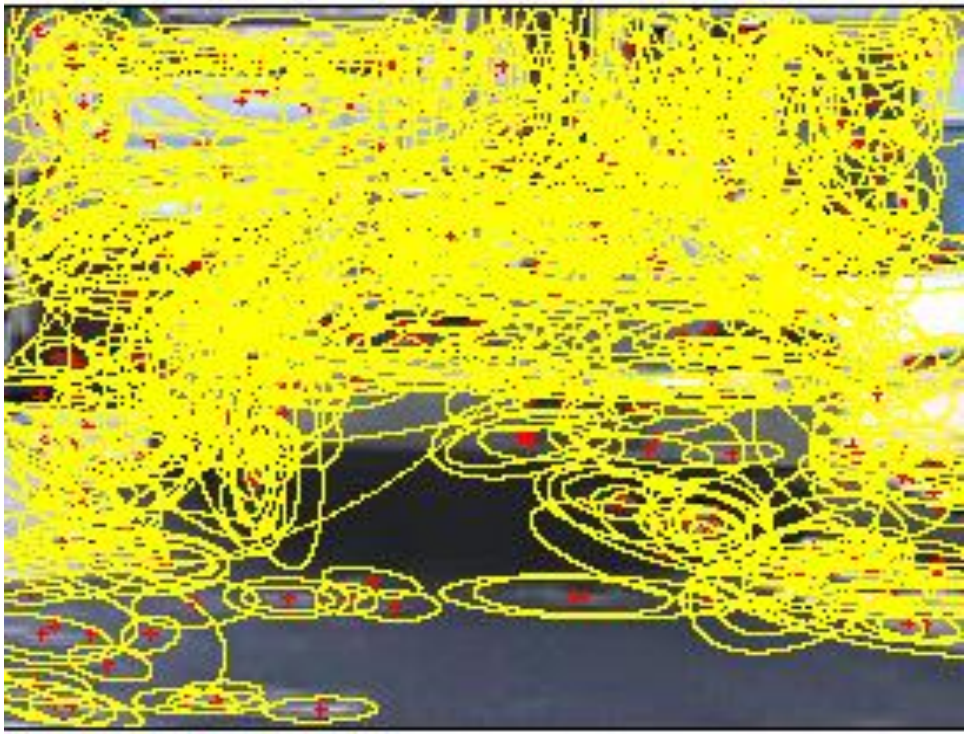
1.Feature detection and description

- Regular grid
  - Vogel & Schiele, 2003
  - Fei-Fei & Perona, 2005



1.Feature detection and description

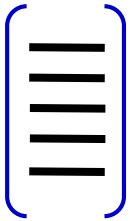
- Regular grid
  - Vogel & Schiele, 2003
  - Fei-Fei & Perona, 2005
- Interest point detector
  - Csurka, et al. 2004
  - Fei-Fei & Perona, 2005
  - Sivic, et al. 2005

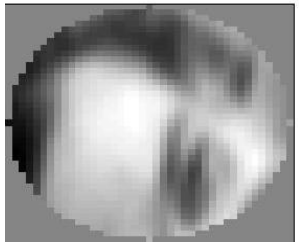


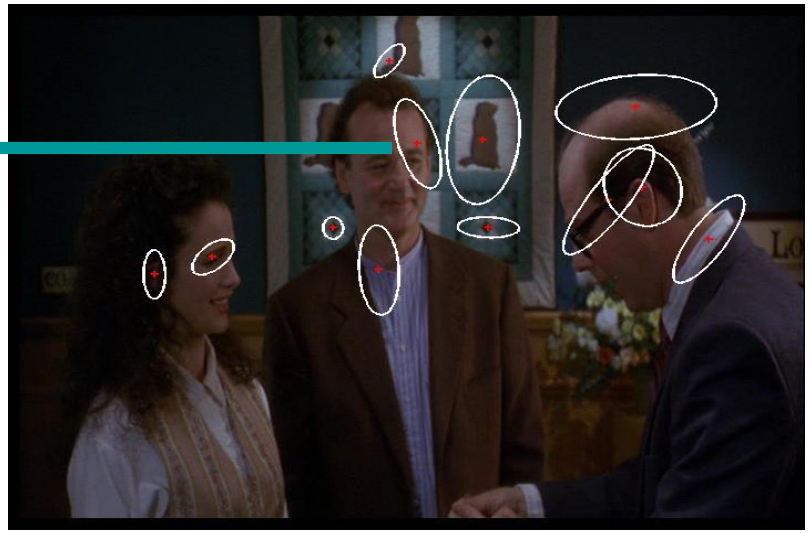
1.Feature detection and description

- Regular grid
  - Vogel & Schiele, 2003
  - Fei-Fei & Perona, 2005
- Interest point detector
  - Csurka, Bray, Dance & Fan, 2004
  - Fei-Fei & Perona, 2005
  - Sivic, Russell, Efros, Freeman & Zisserman, 2005
- Other methods
  - Random sampling (Vidal-Naquet & Ullman, 2002)
  - Segmentation based patches (Barnard, Duygulu, Forsyth, de Freitas, Blei, Jordan, 2003)

1.Feature detection and description







Compute SIFT descriptor

[Lowe'99]

Normalize patch

Detect patches

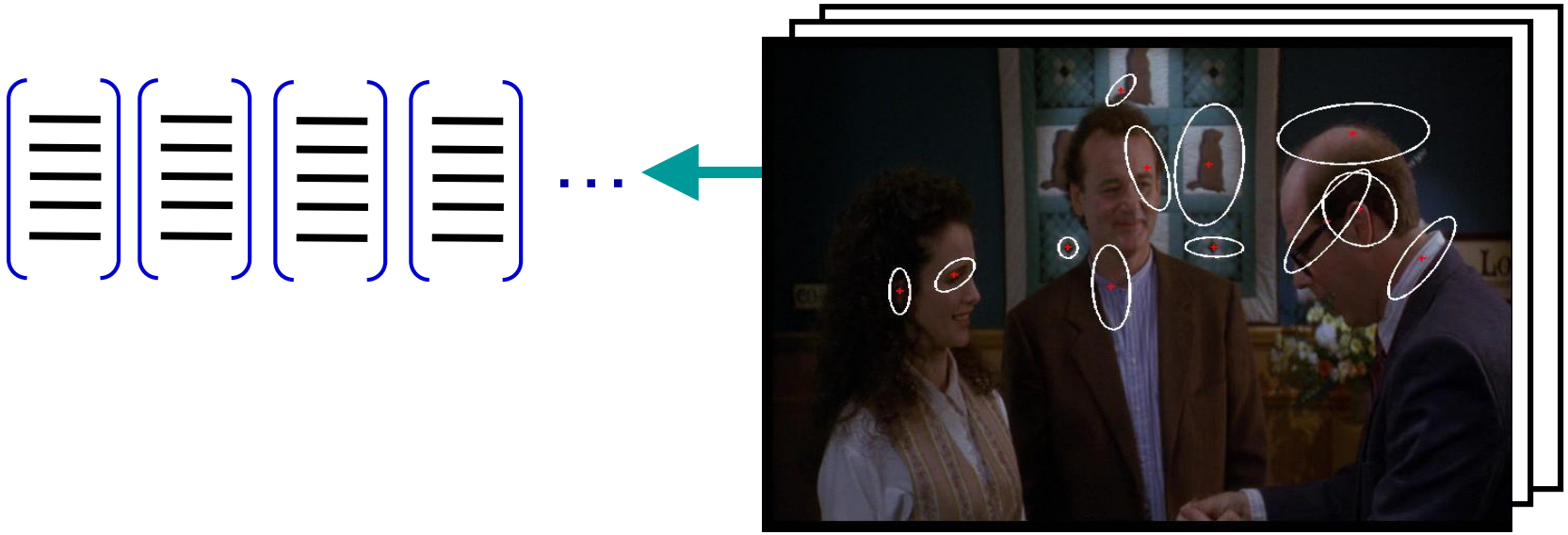
[Mikojaczyk and Schmid '02]

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

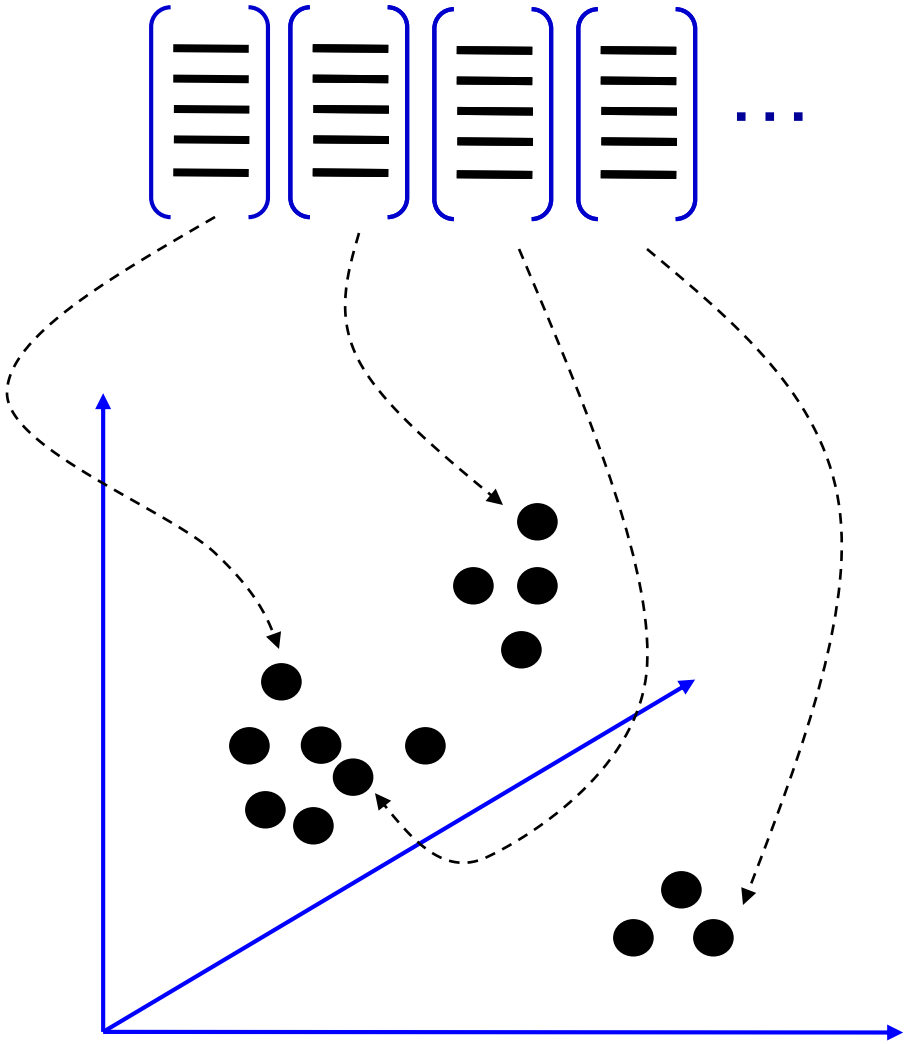
[Sivic & Zisserman, '03]

Slide credit: Josef Sivic

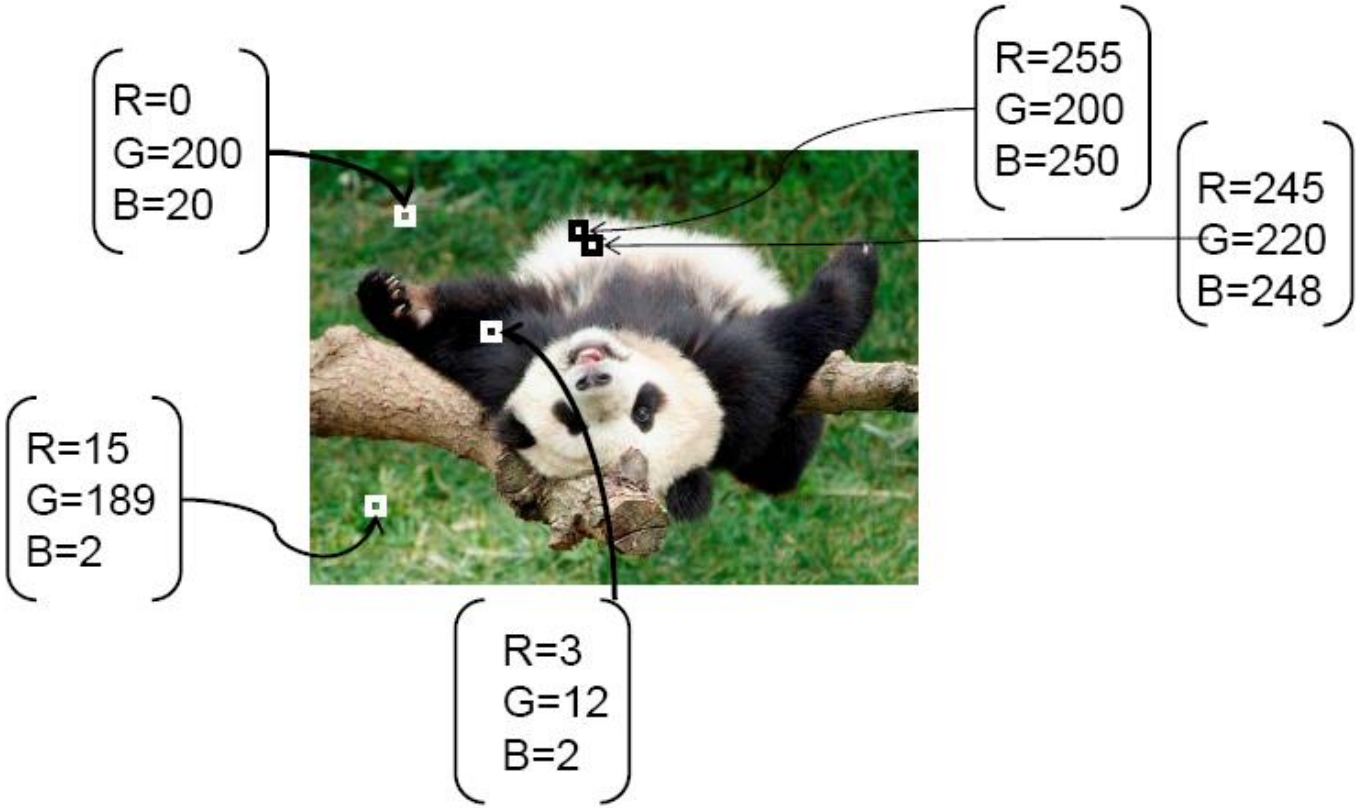
2. Codewords dictionary formation



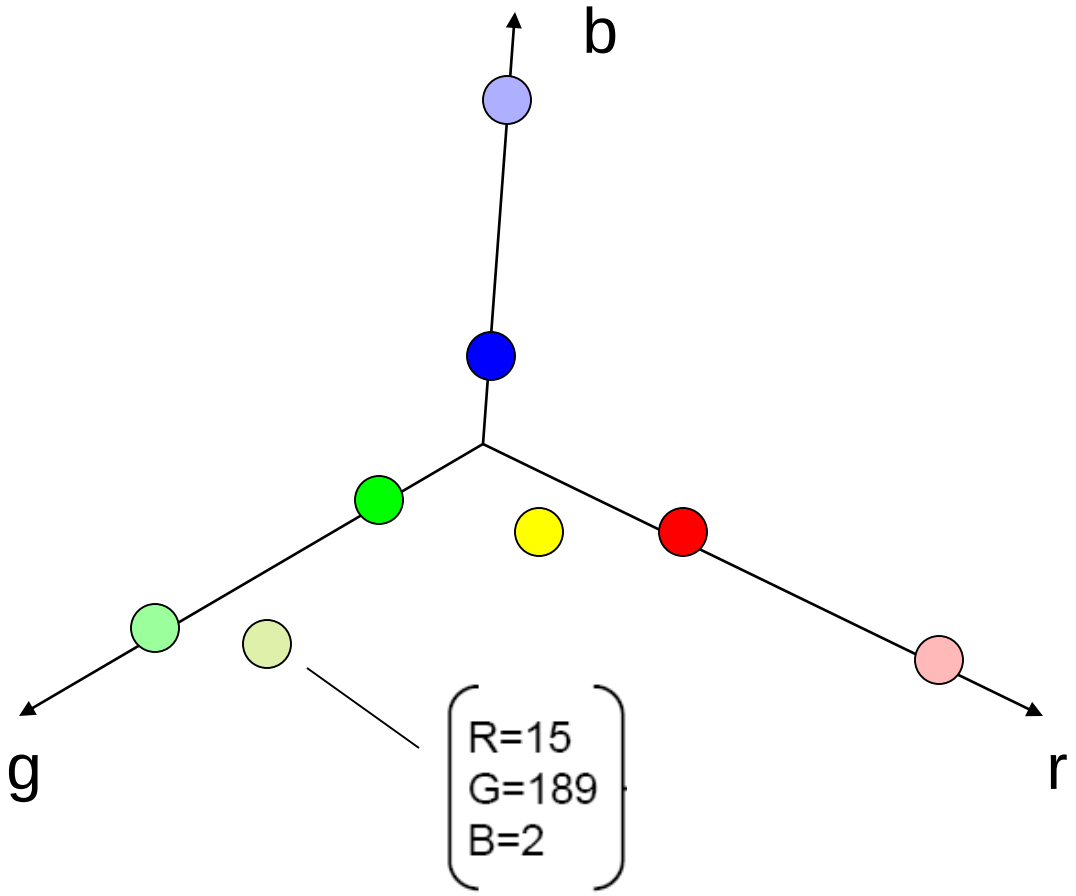
2. Codewords dictionary formation



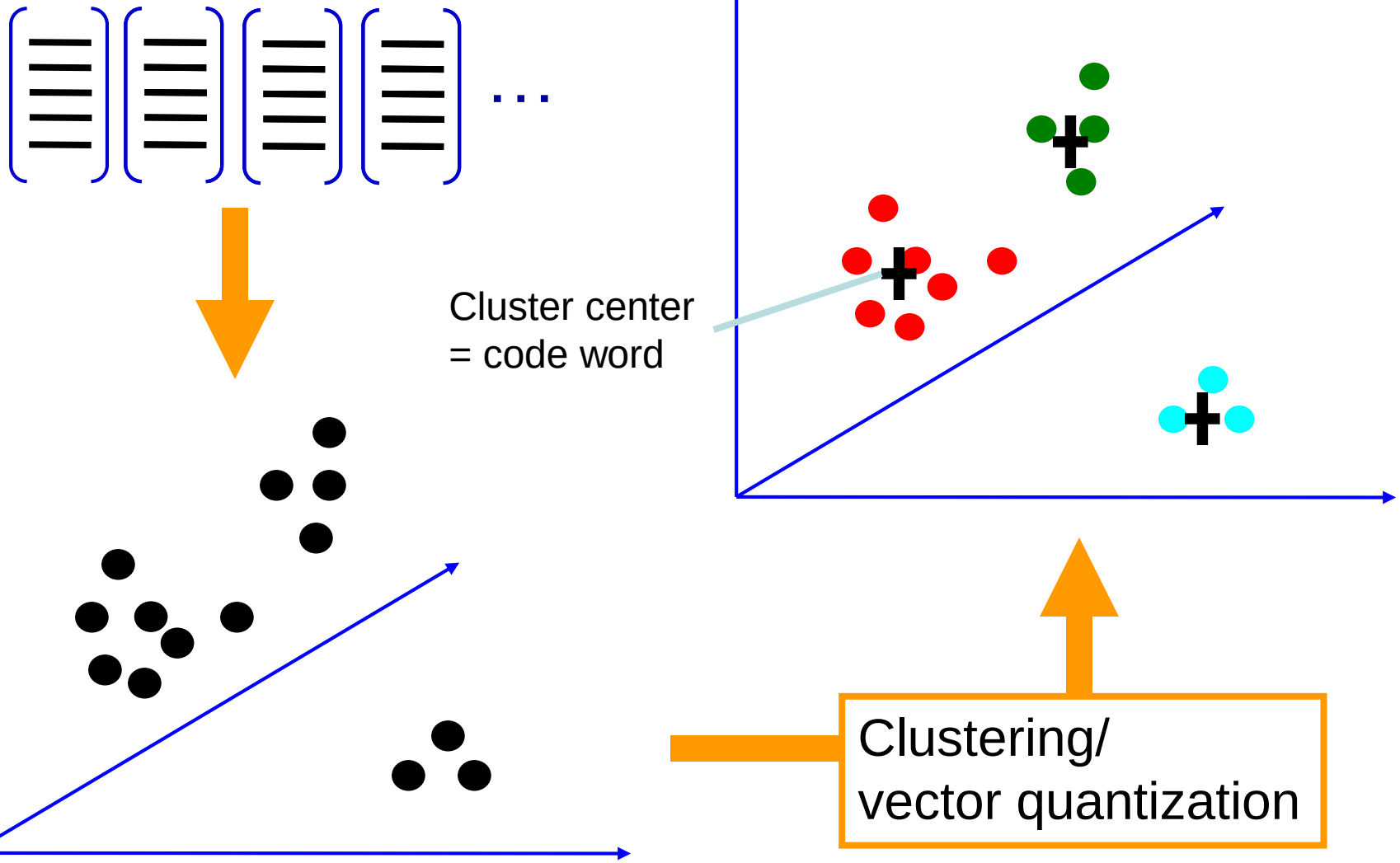
Example: color feature



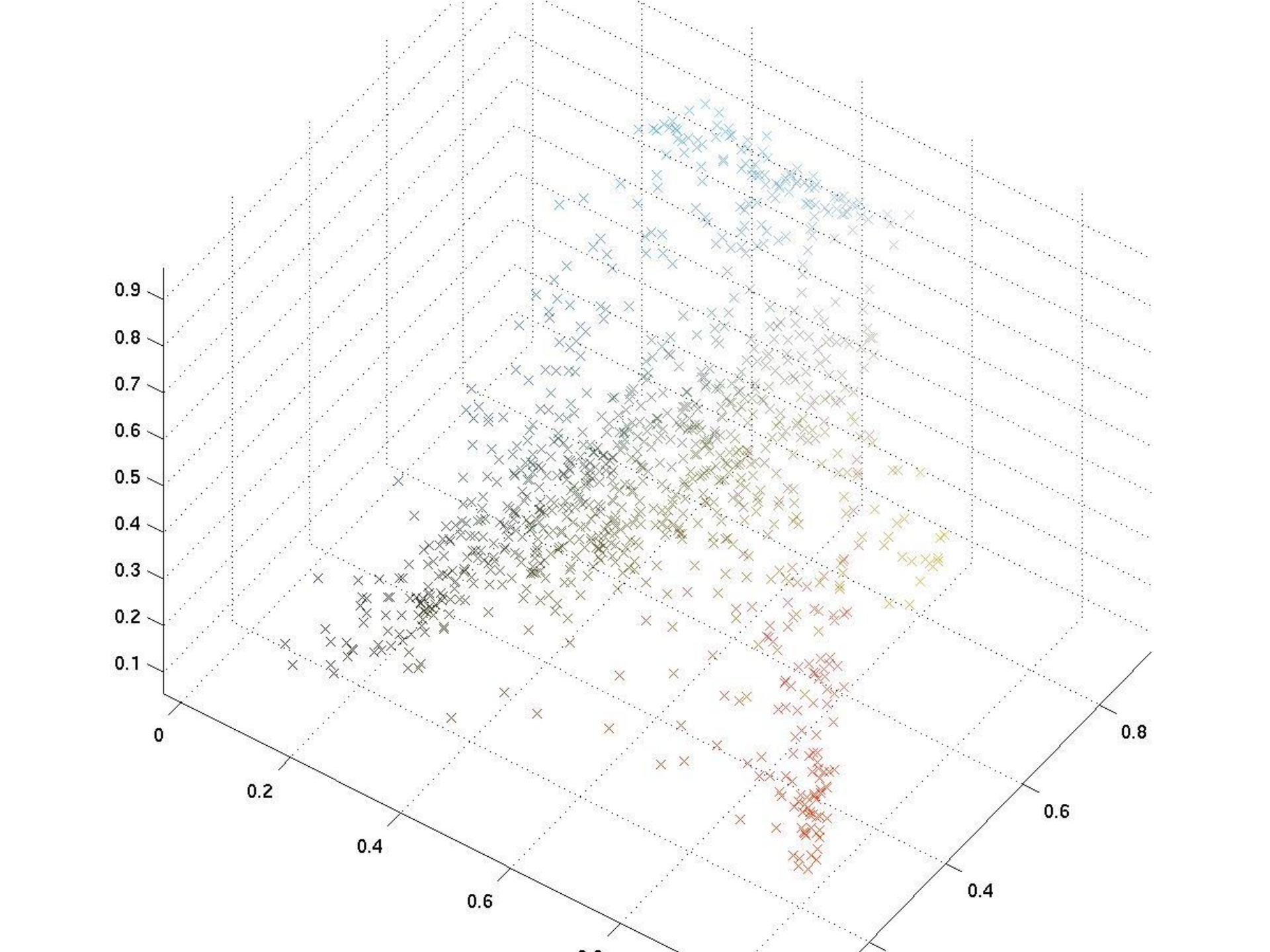
Example: color feature

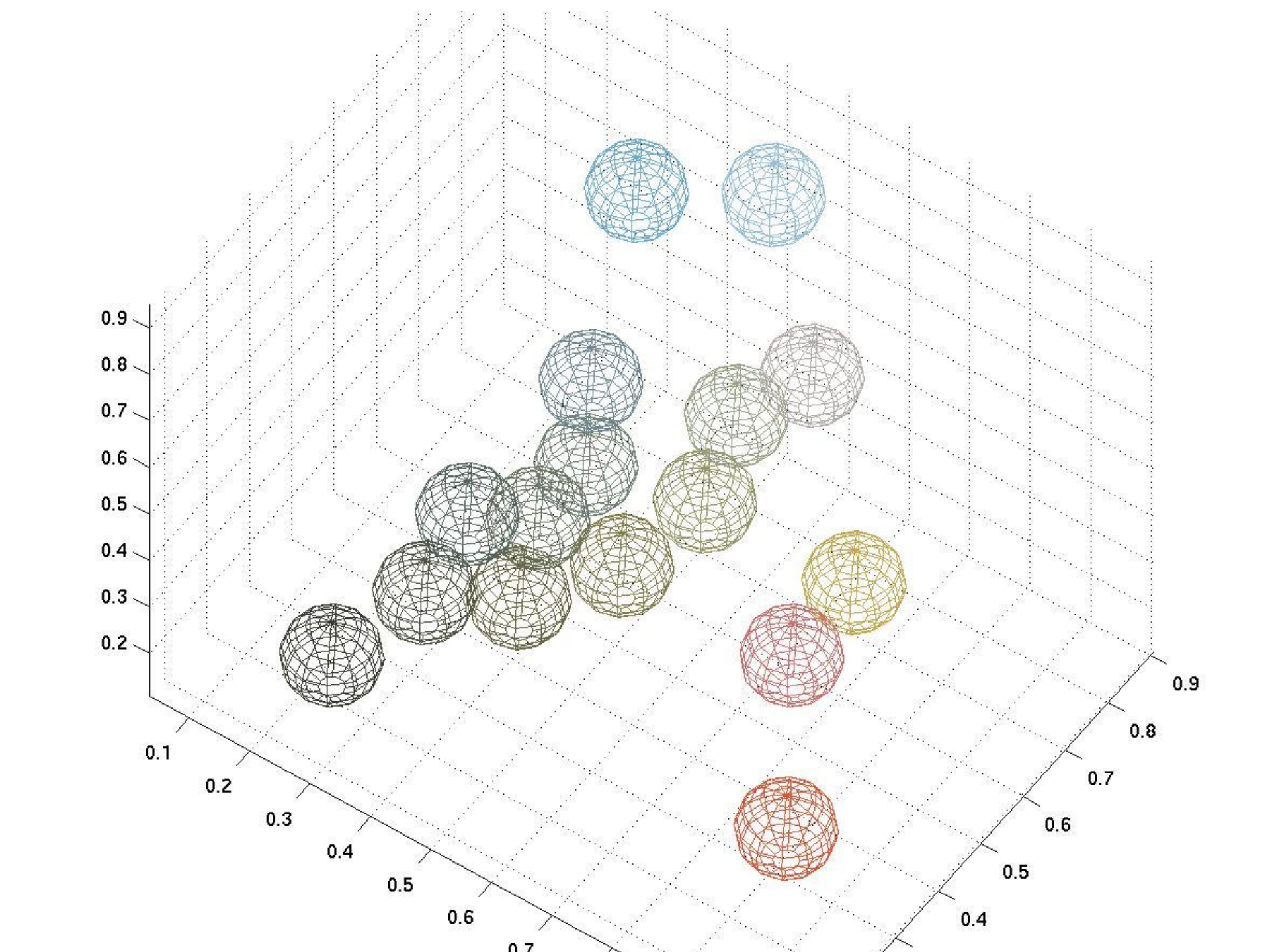


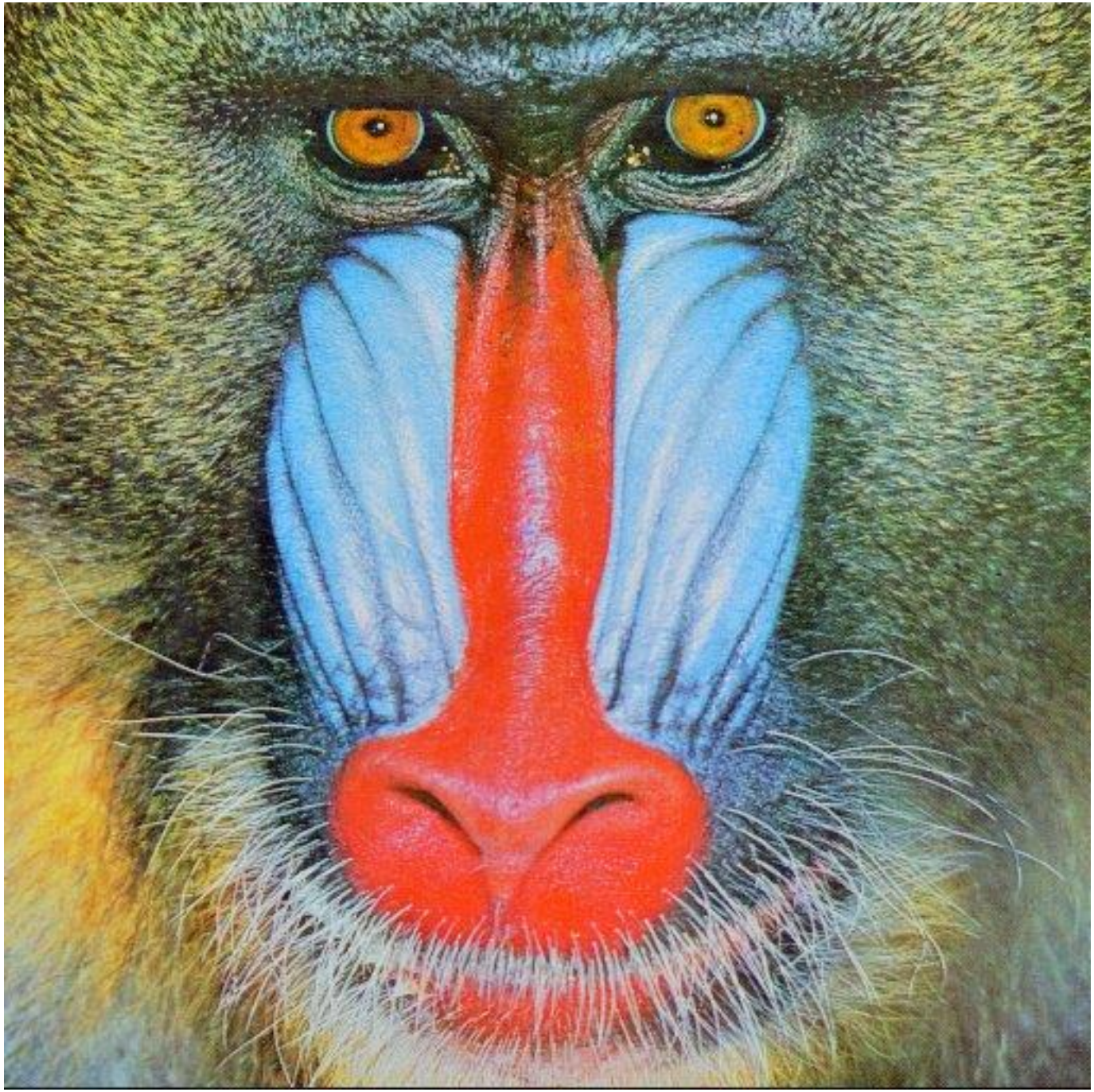
2. Codewords dictionary formation



E.g., Kmeans, see CS131A

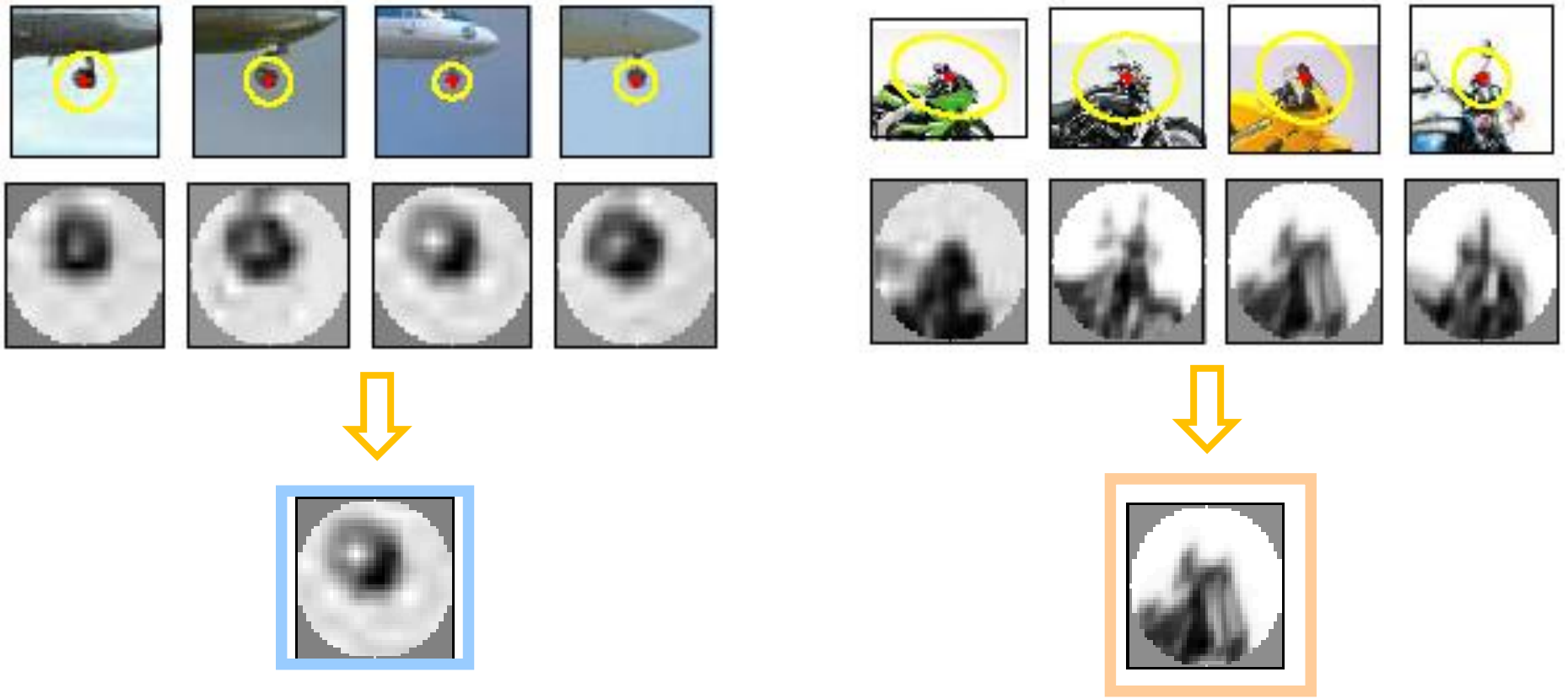

0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.2
0.4
0.6
0.8
0.6
0.4


0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.4
0.5
0.6
0.7
0.8
0.9



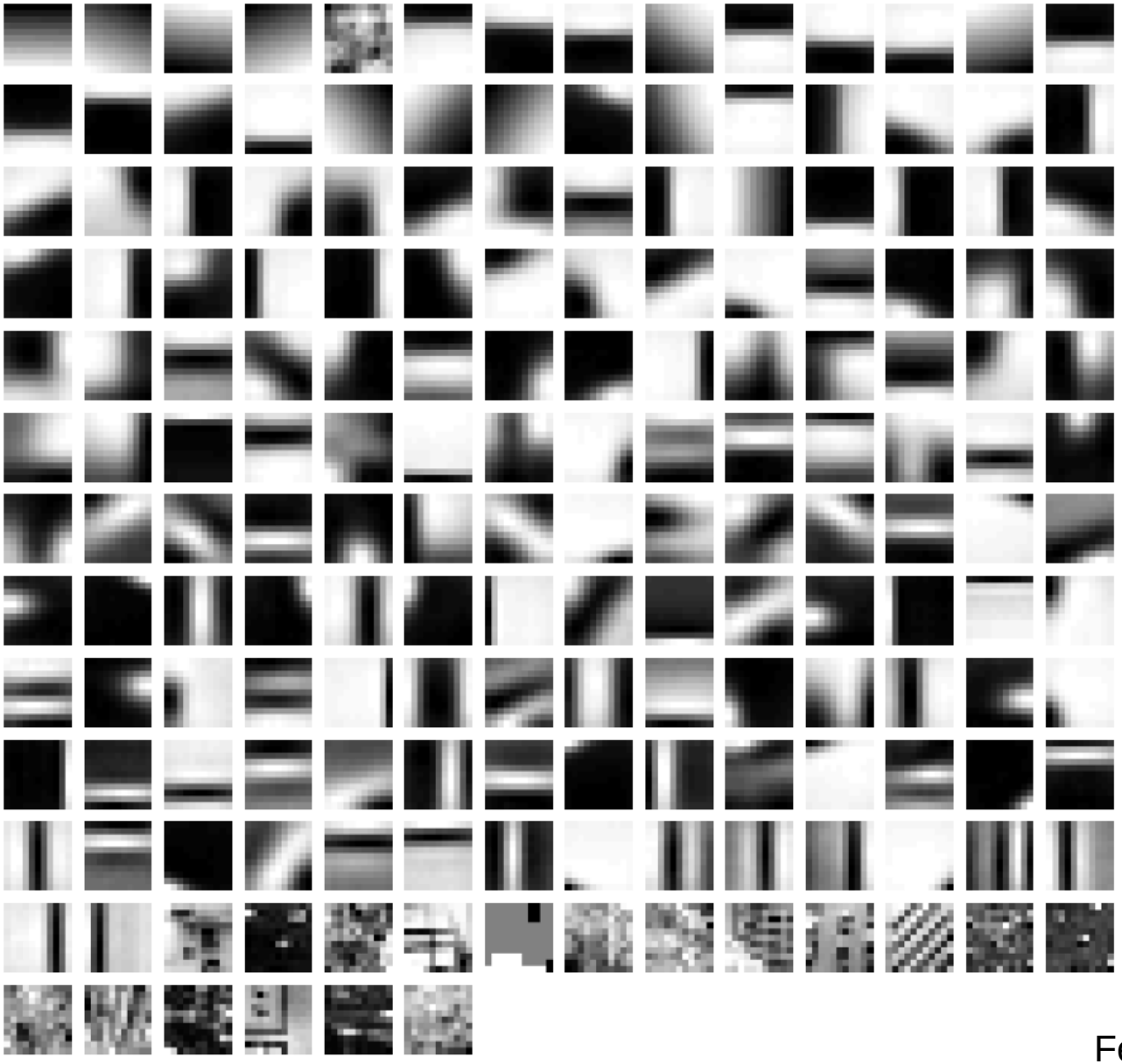
2. Codewords dictionary formation

- **Image patch examples of codewords**



Sivic et al. 2005

2. Codewords dictionary formation



Fei-Fei et al. 2005

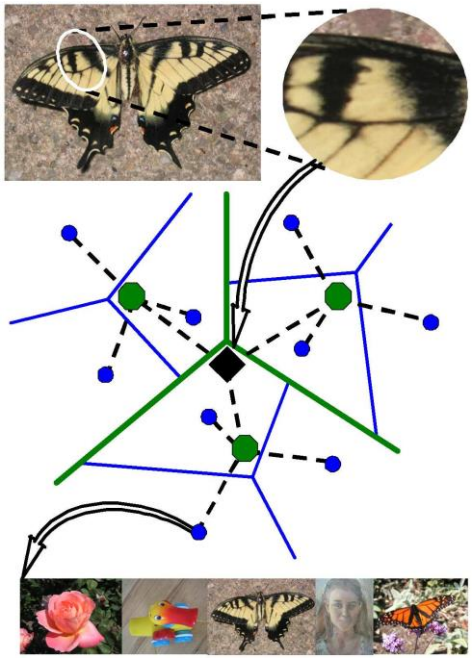
2. Codewords dictionary formation

- **Typically a codeword dictionary is obtained from a training set comprising all the object classes of interests**

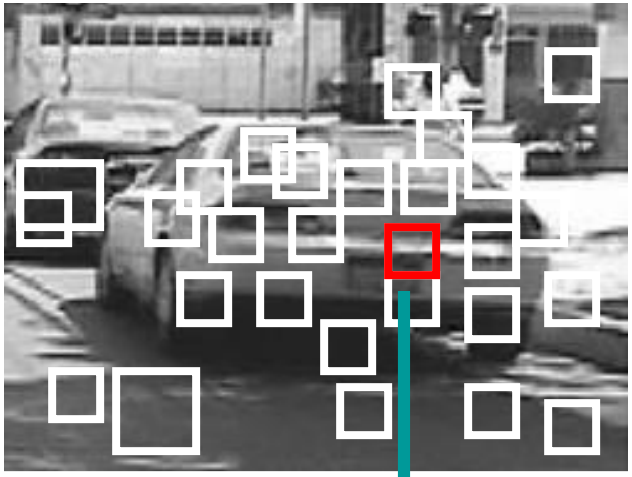
Visual vocabularies: Issues

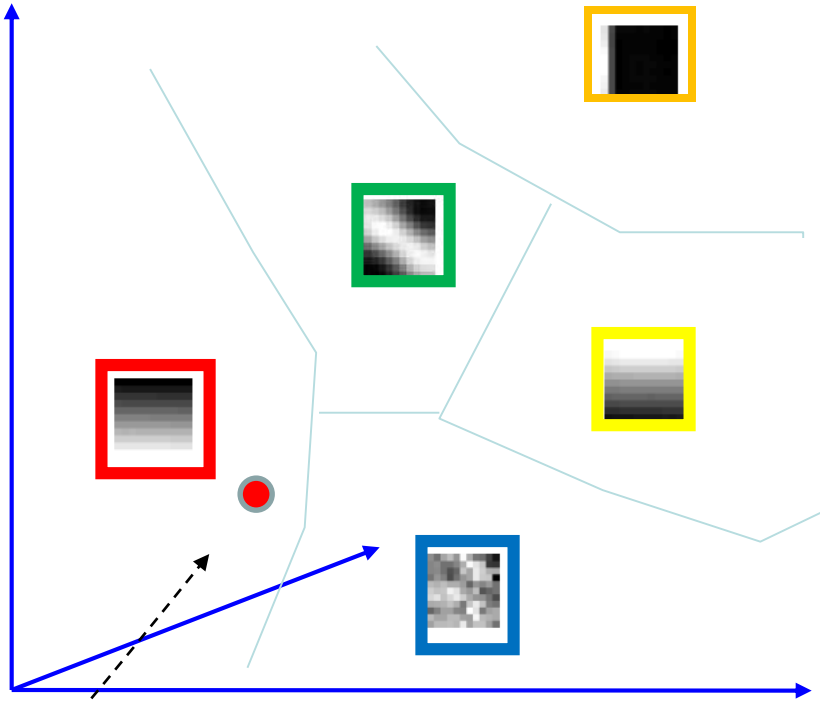
- How to choose vocabulary size?
  - Too small: visual words not representative of all patches
  - Too large: quantization artifacts, overfitting

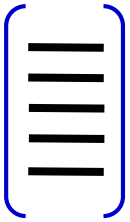
- Computational efficiency
  - Vocabulary trees
    (Nister & Stewenius, 2006)

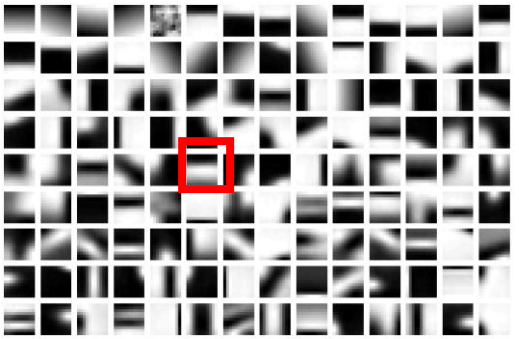


3. Bag of word representation





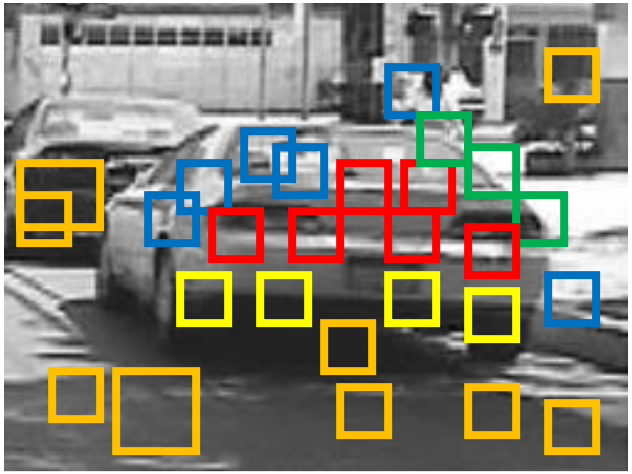


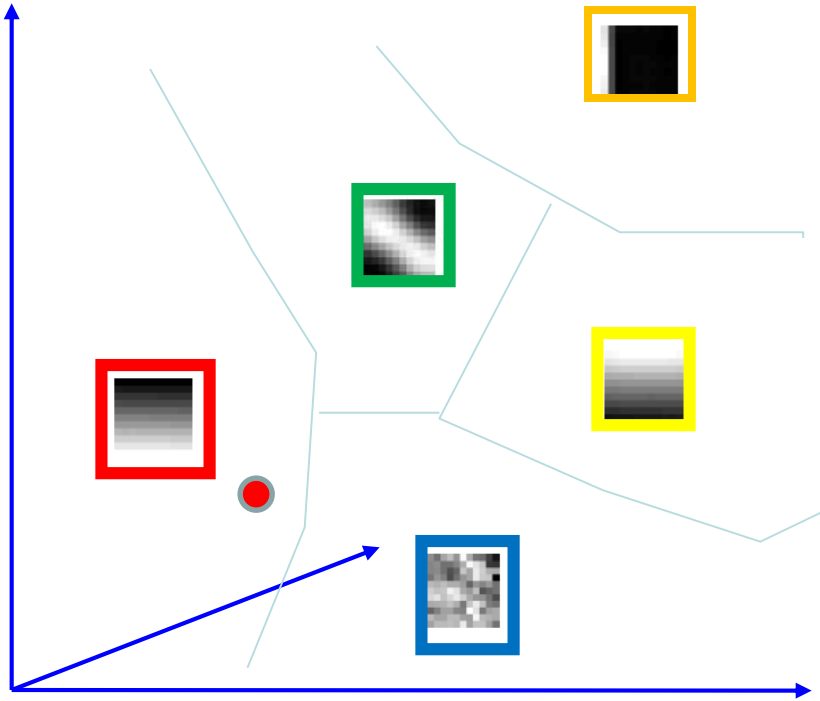


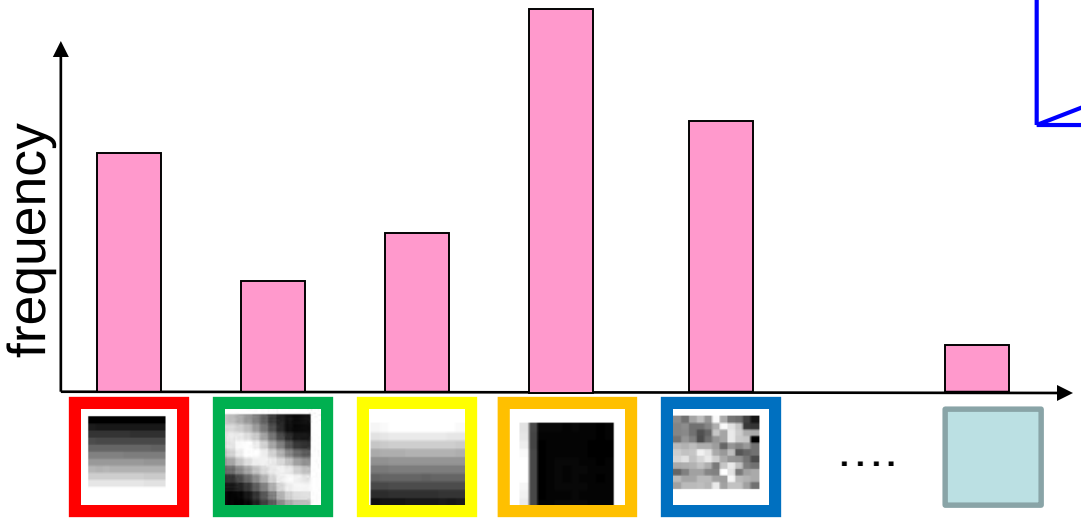
Codewords dictionary

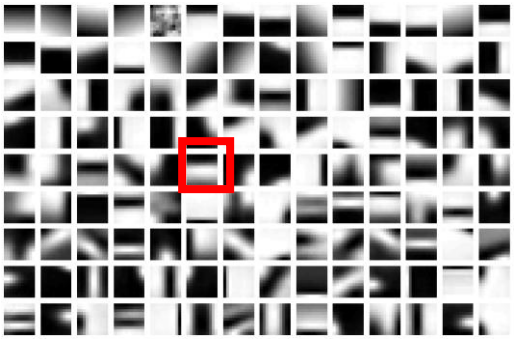
- Nearest neighbors assignment
- K-D tree search strategy

3. Bag of word representation





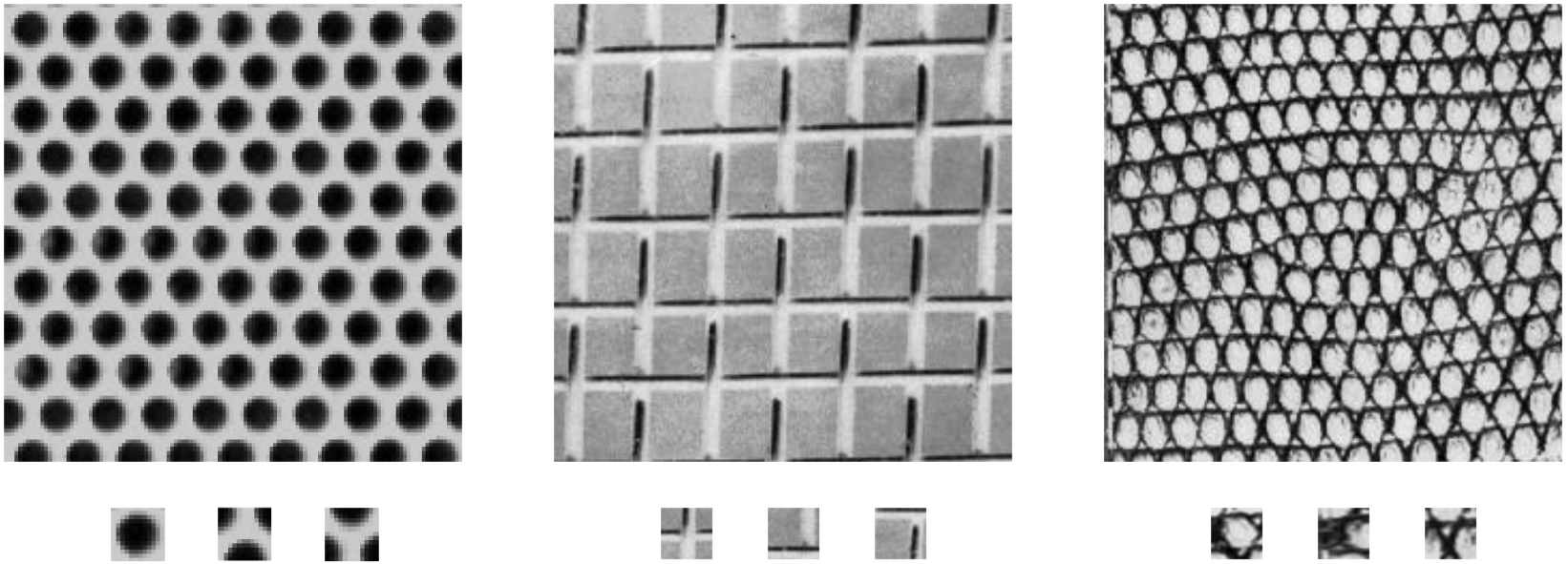




Codewords dictionary

Representing textures

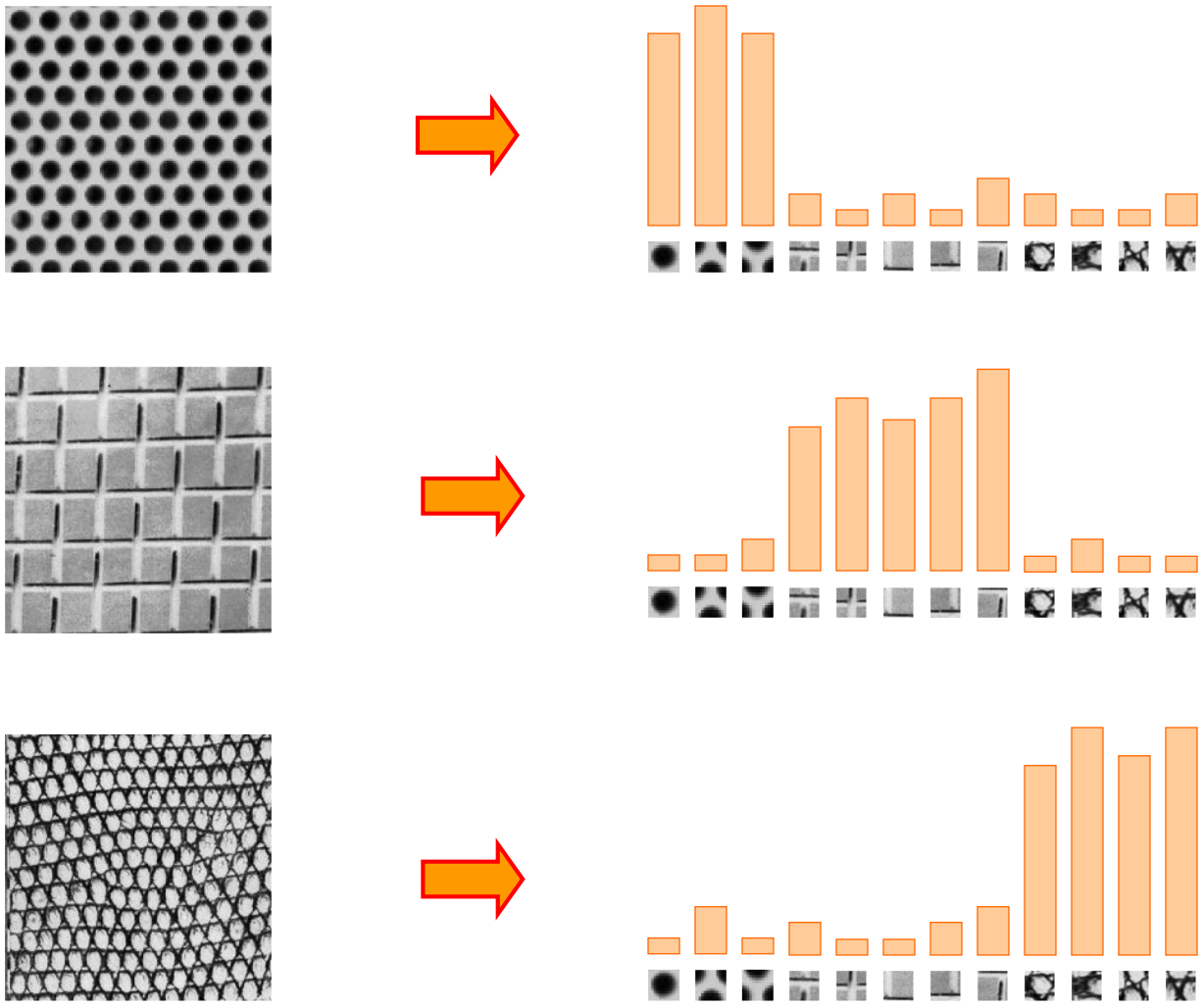
- Texture is characterized by the repetition of basic elements or *textons*
- For stochastic textures, it is the identity of the textons, not their spatial arrangement, that matters



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

Credit slide: S. Lazebnik

Representing textures

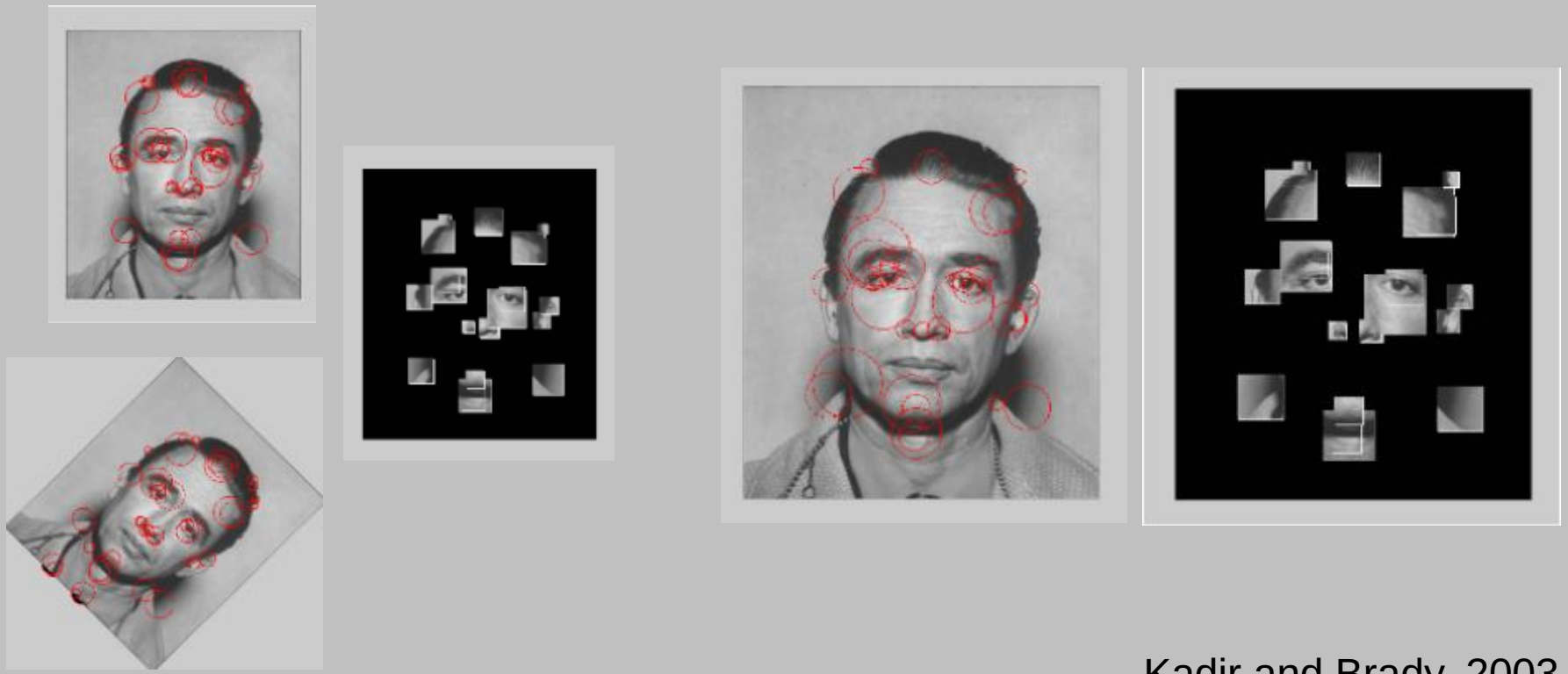


Julesz, 1981; Cula & Dana, 2001; Leung & Malik 2001; Mori, Belongie & Malik, 2001; Schmid 2001; Varma & Zisserman, 2002, 2003; Lazebnik, Schmid & Ponce, 2003

Credit slide: S. Lazebnik

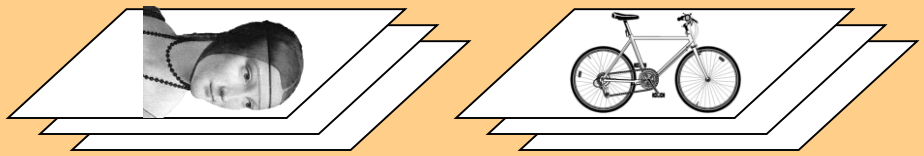
Invariance issues

- Scale? Rotation? View point? Occlusions?
  - Implicit;
  - depends on detectors and descriptors



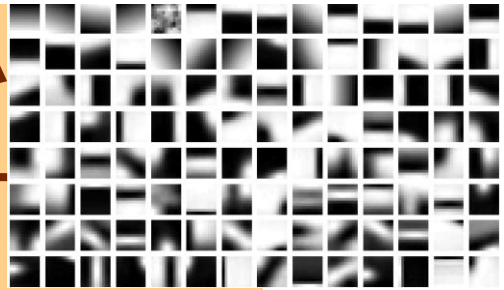
Kadir and Brady. 2003

Representation



2.

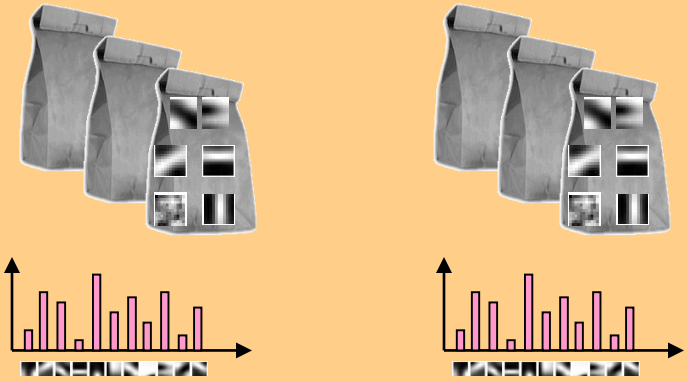
codewords dictionary



1. feature detection & representation

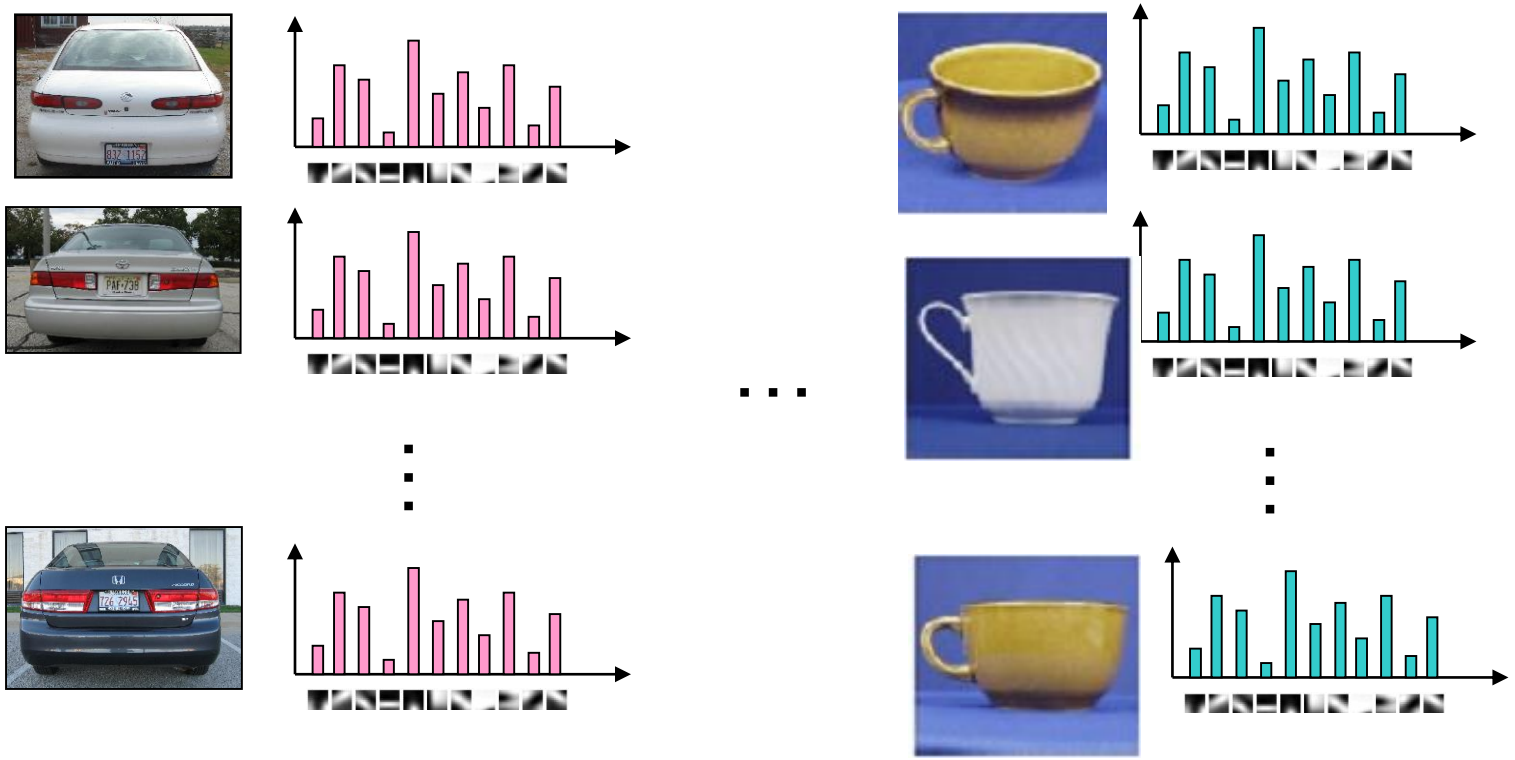
image representation

3.



category models

Category models

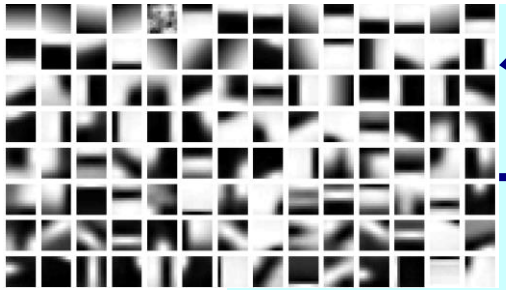


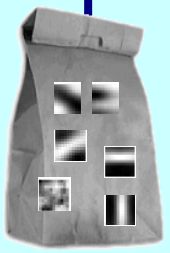
Class 1

Class N

Recognition

codewords dictionary





category models (and/or) classifiers

category decision

Next Lecture

- Bag of words models – part 2