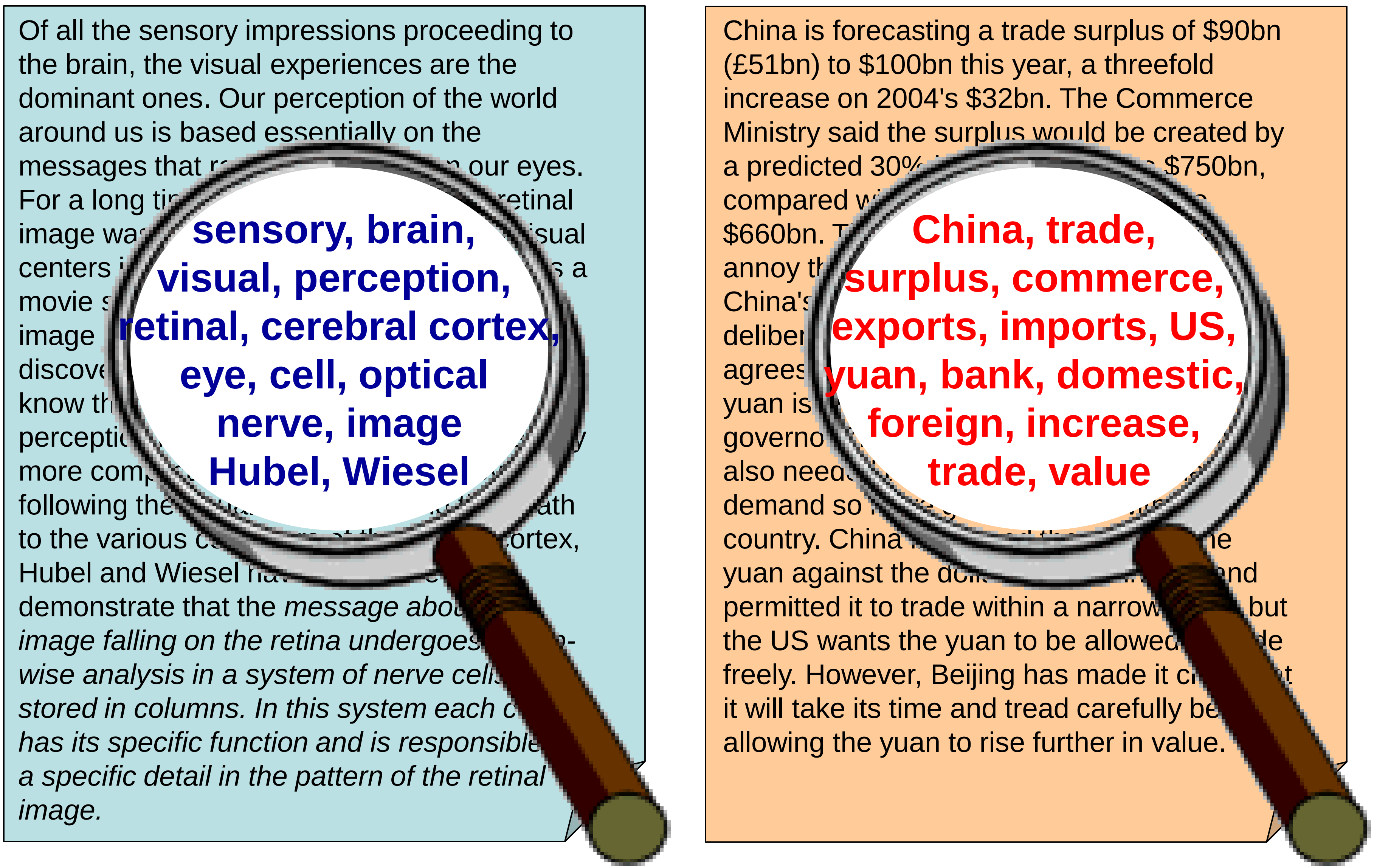
Object Categorization

Bag-of-words models

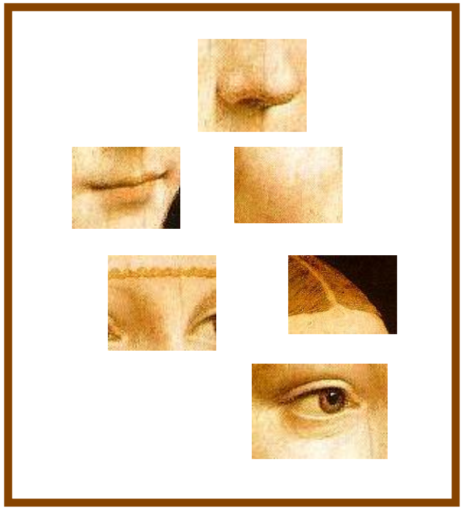
Object → **Bag of ‘words’**

Analogy to documents



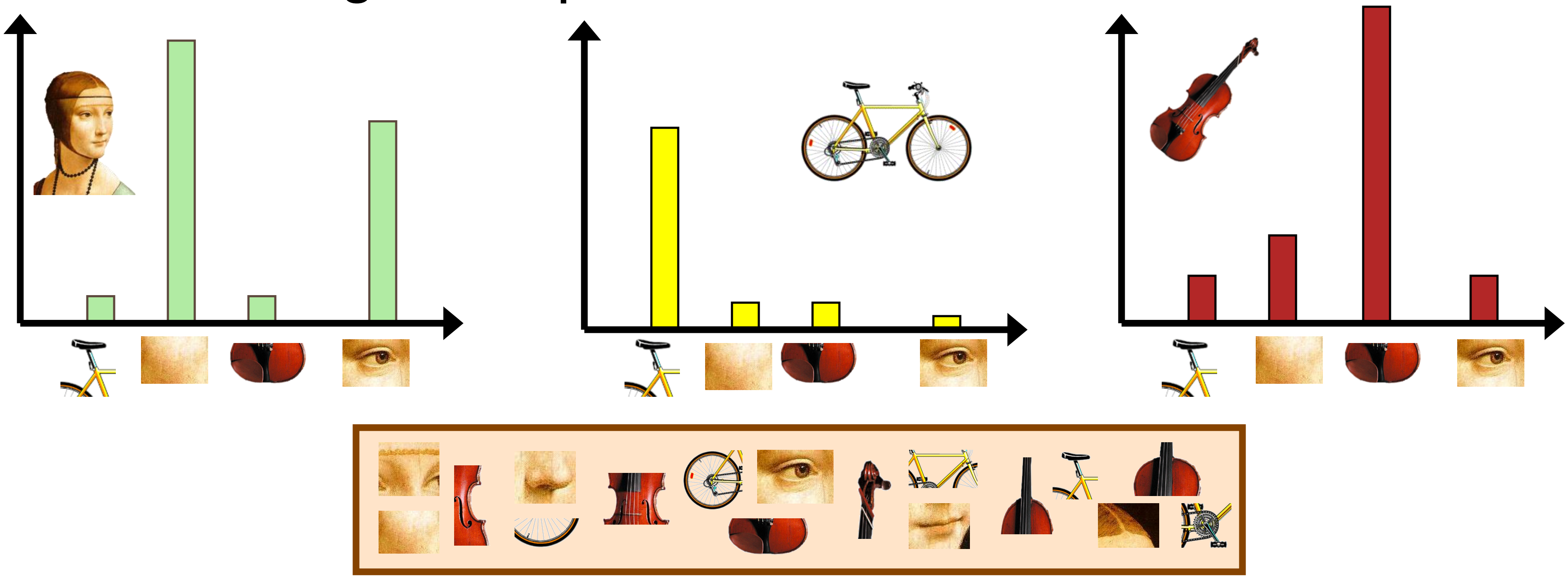
A clarification: definition of “BoW”

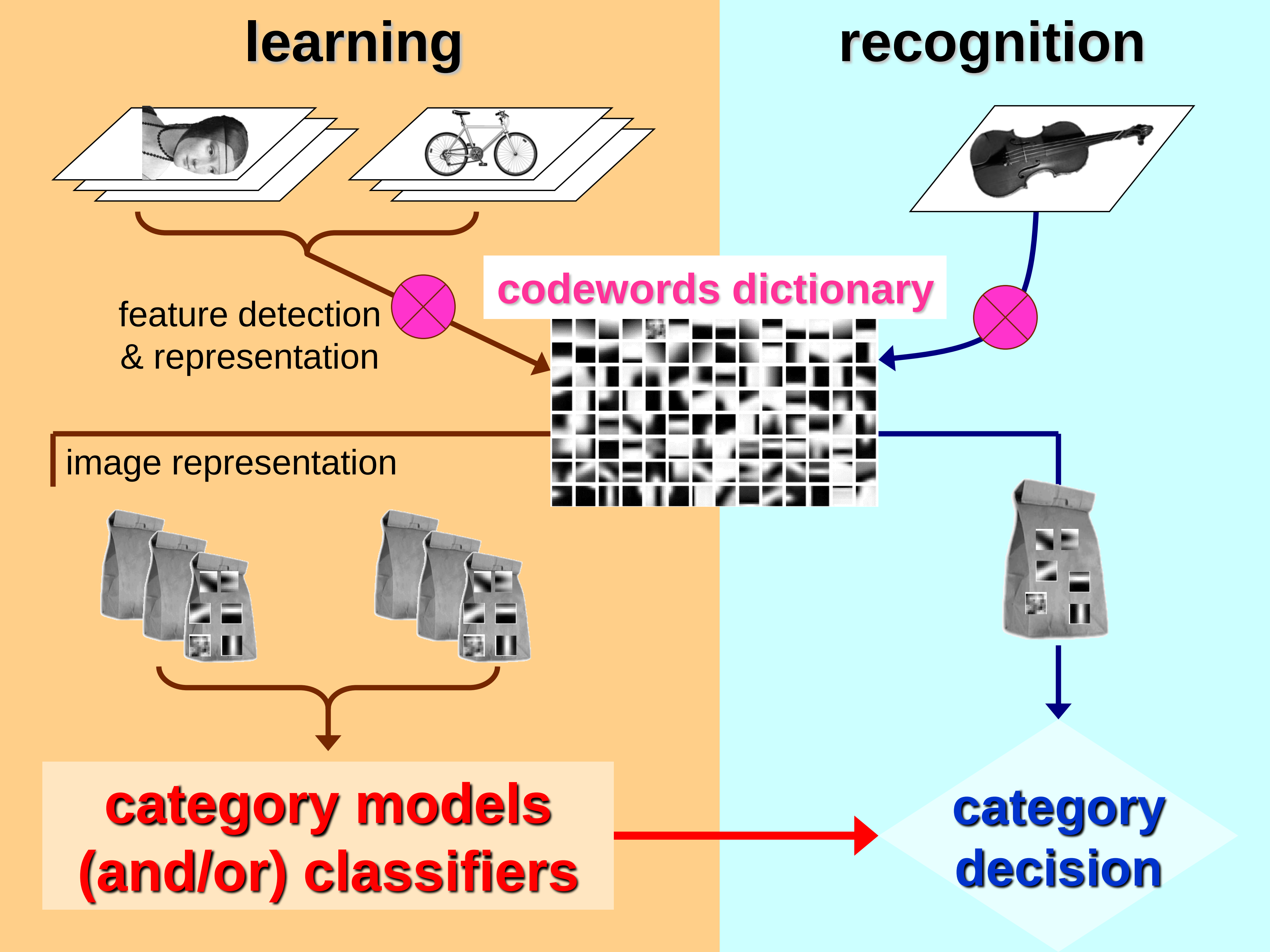
- Looser definition
  - Independent features



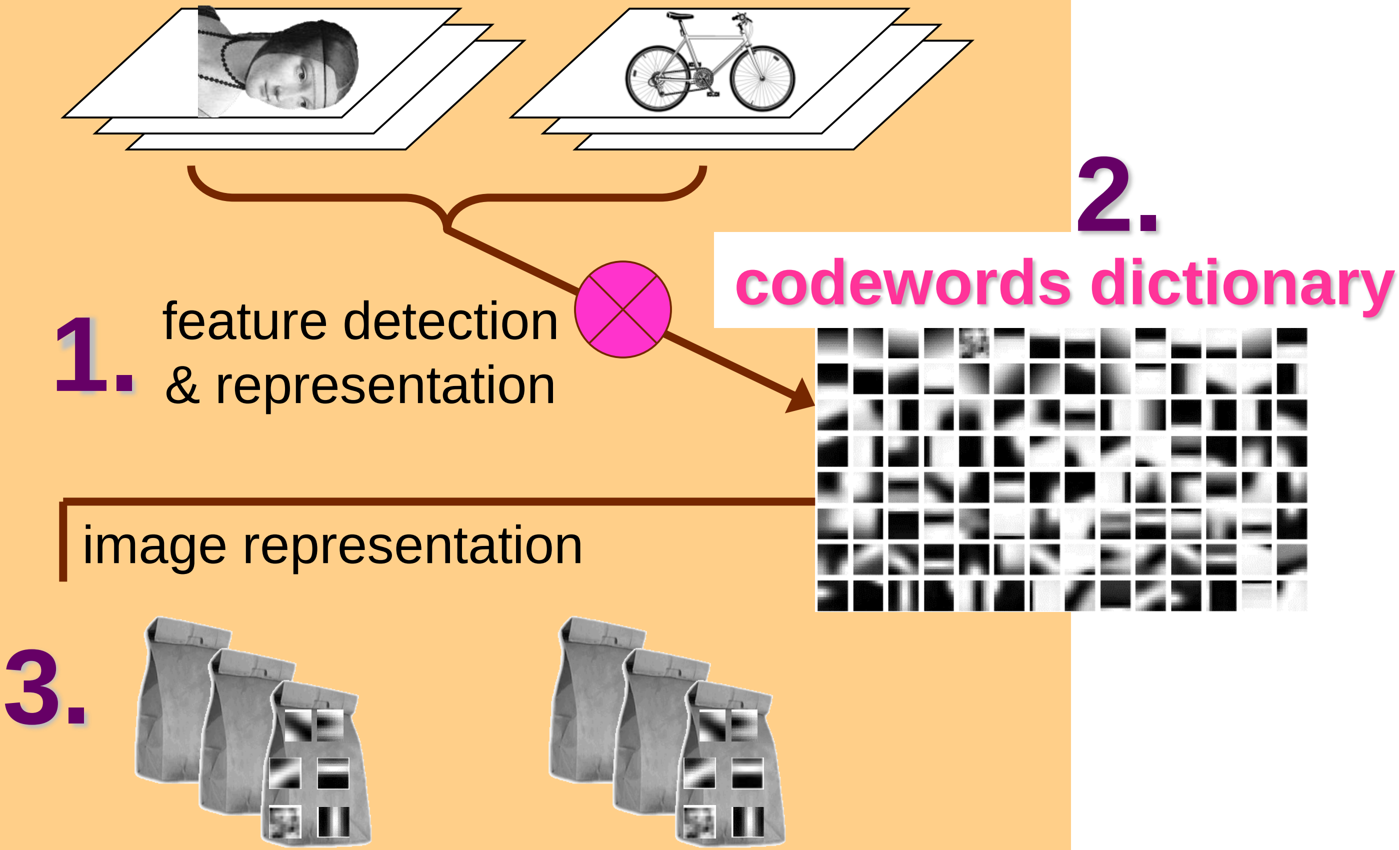
A clarification: definition of “BoW”

- Looser definition
  - Independent features
- Stricter definition
  - Independent features
  - histogram representation

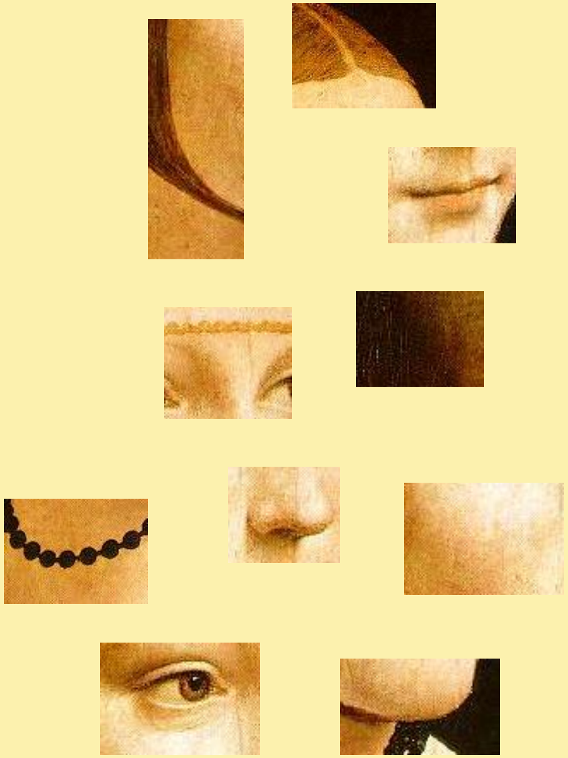



learning
recognition
feature detection
& representation
codewords dictionary
image representation
category models
(and/or) classifiers
category
decision

Representation

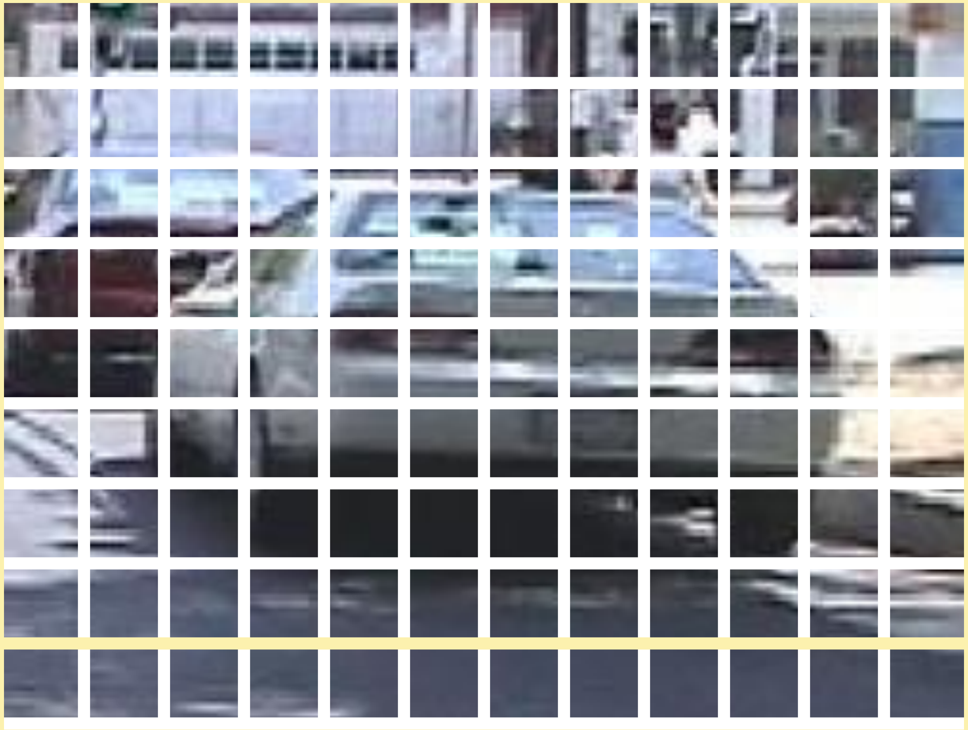


1.Feature detection and representation



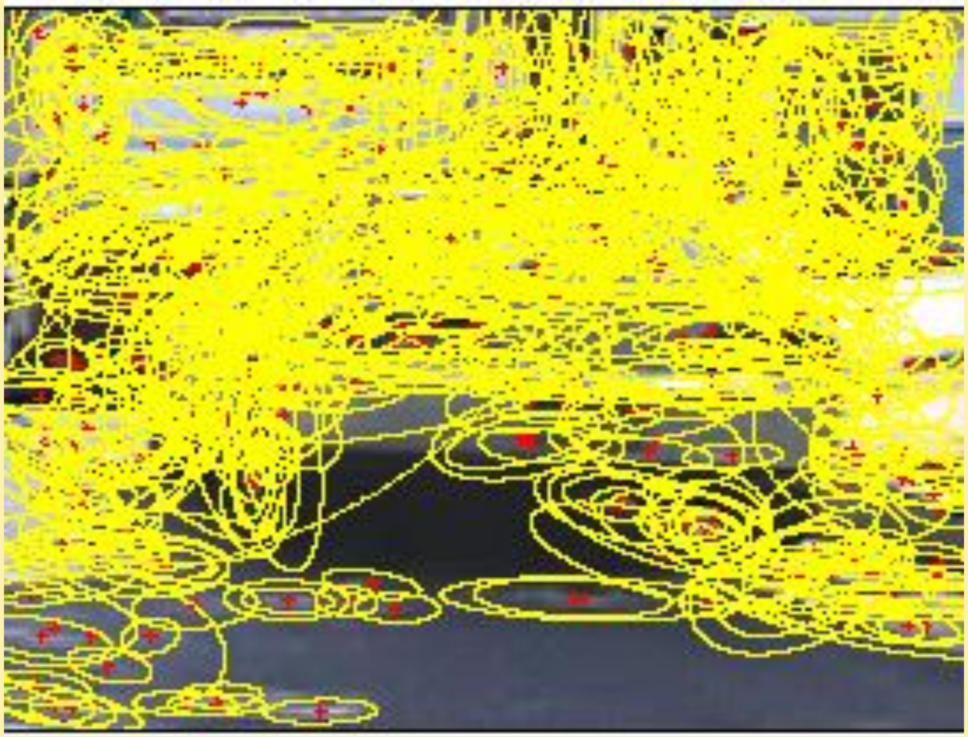
1.Feature detection and representation

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



1.Feature detection and representation

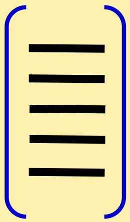
- Regular grid
  - Vogel & Schiele, 2003
  - Fei-Fei & Perona, 2005
- Interest point detector
  - Csurka, et al. 2004
  - Fei-Fei & Perona, 2005
  - Sivic, et al. 2005



1.Feature detection and representation

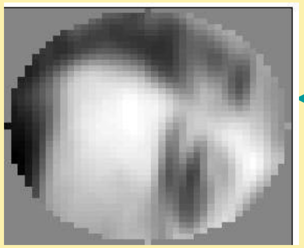
- Regular grid
  - Vogel & Schiele, 2003
  - Fei-Fei & Perona, 2005
- **Interest point detector**
  - Csurka, Bray, Dance & Fan, 2004
  - Fei-Fei & Perona, 2005
  - Sivic, Russell, Efros, Freeman & Zisserman, 2005
- Other methods
  - Random sampling (Vidal-Naquet & Ullman, 2002)
  - Segmentation based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

1.Feature detection and representation

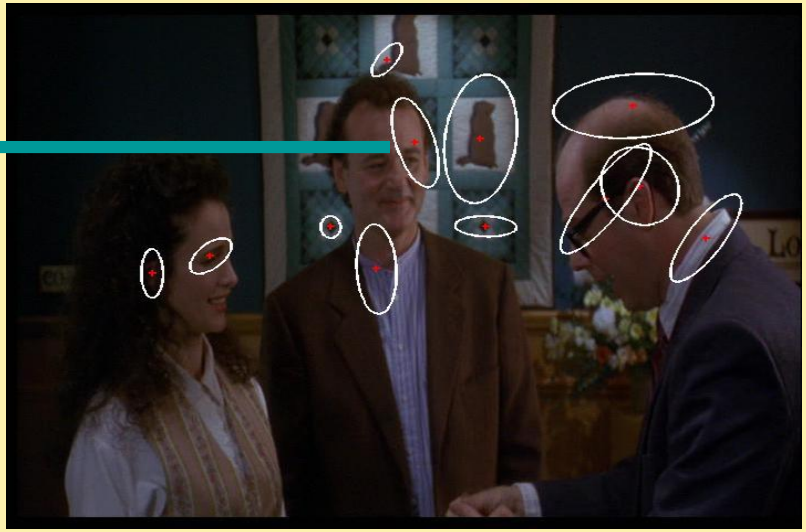


Compute SIFT descriptor

[Lowe'99]



Normalize patch



Detect patches

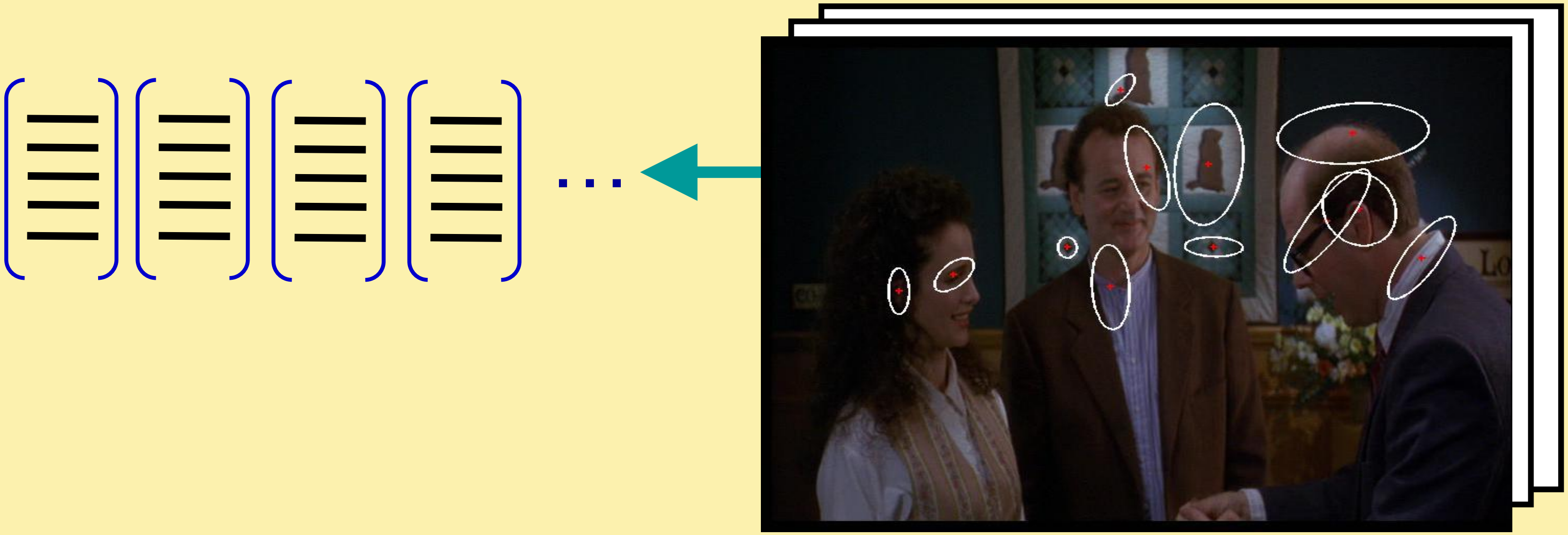
[Mikojaczyk and Schmid '02]

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

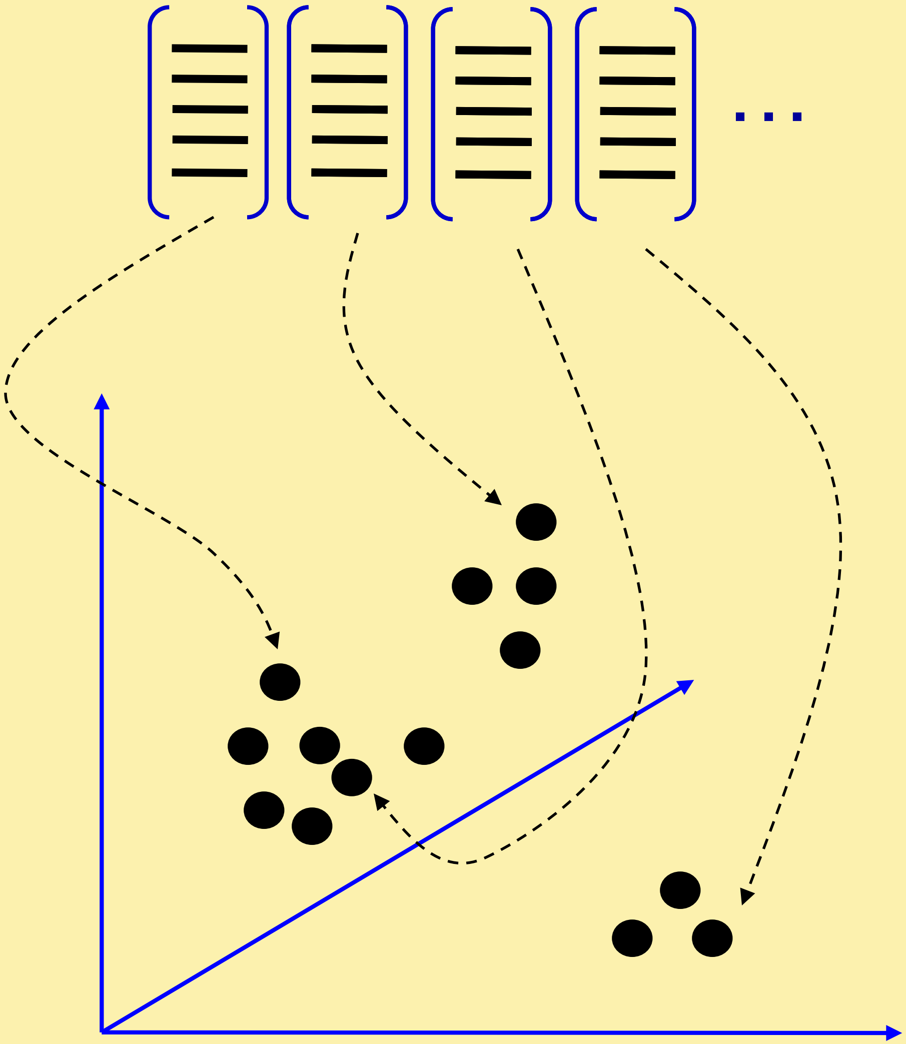
[Sivic & Zisserman, '03]

Slide credit: Josef Sivic

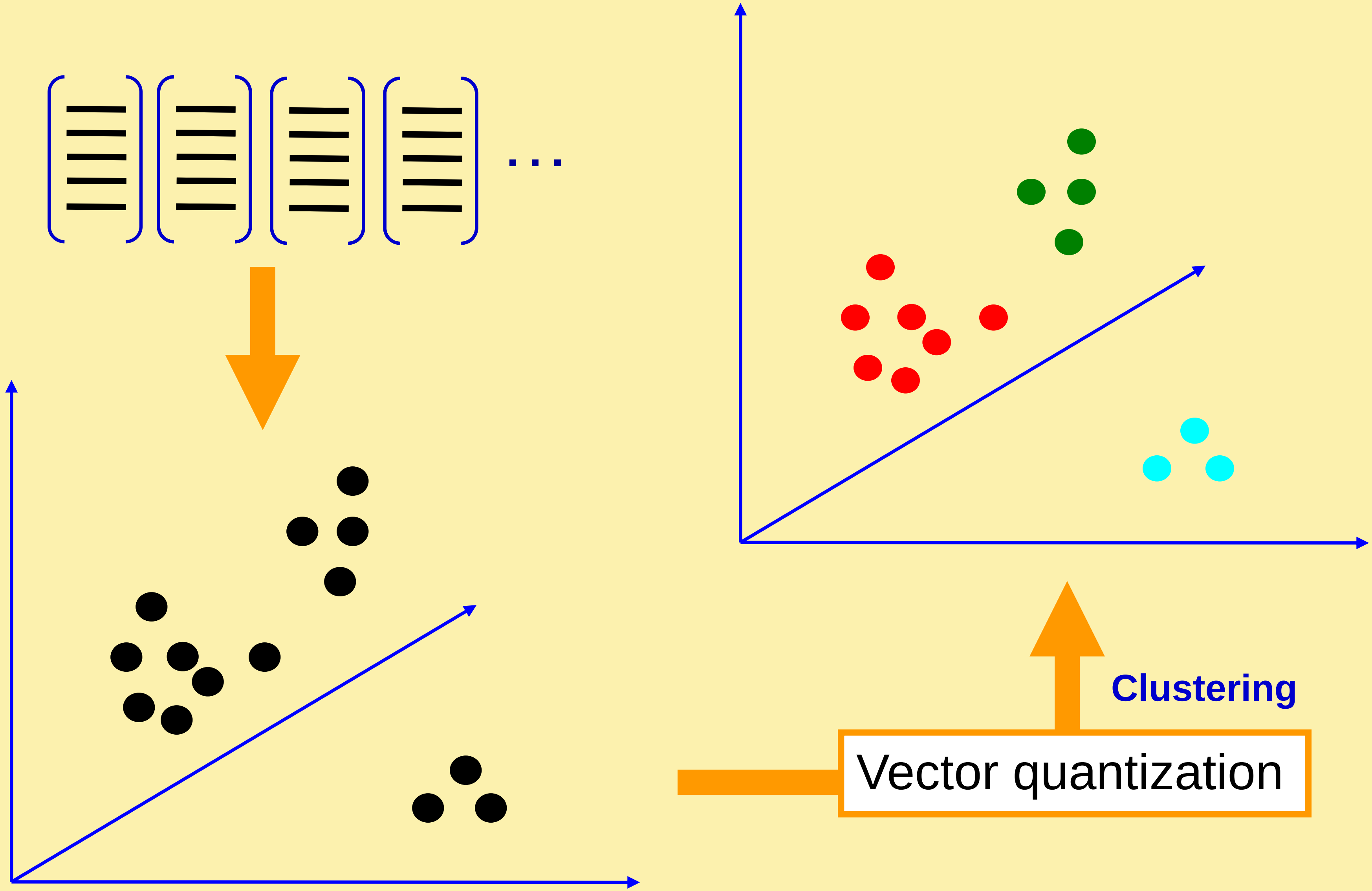
1.Feature detection and representation



2. Codewords dictionary formation

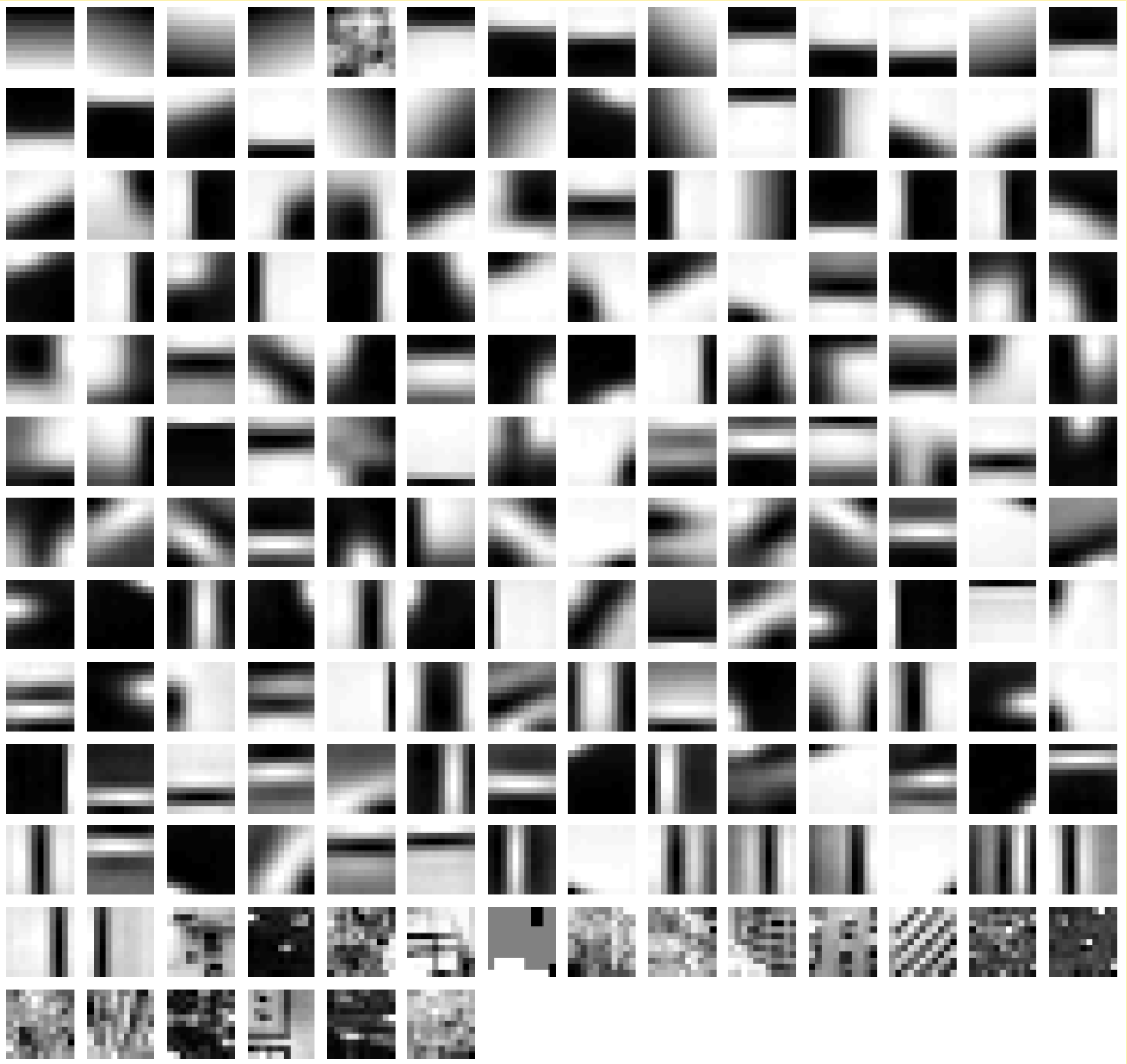


2. Codewords dictionary formation



Slide credit: Josef Sivic

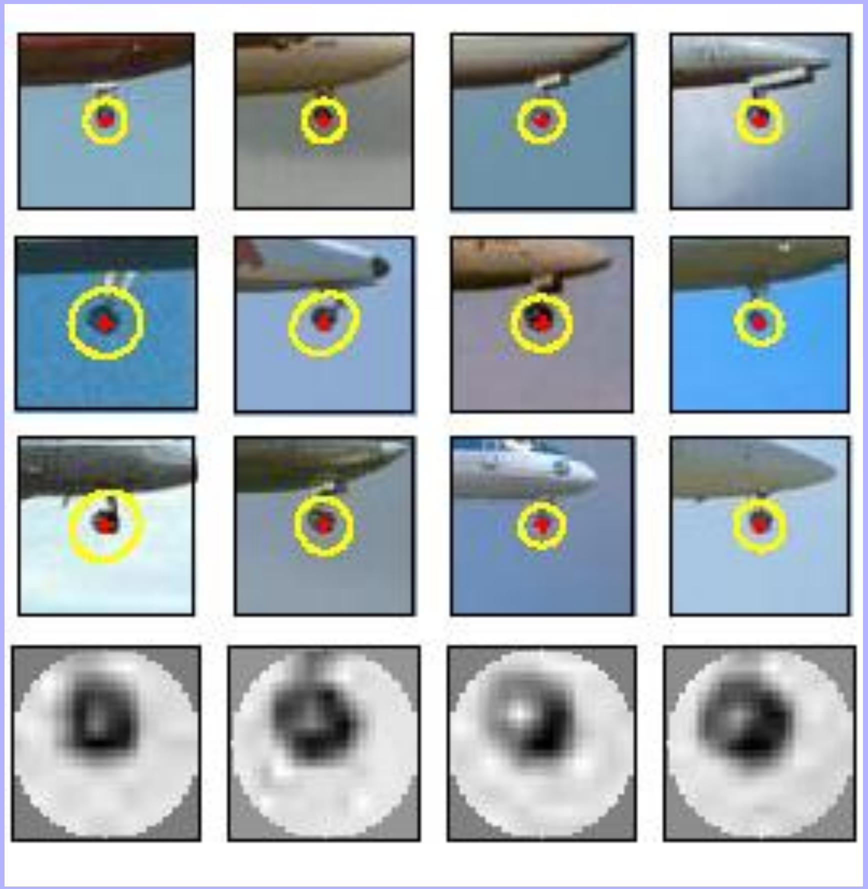
2. Codewords dictionary formation

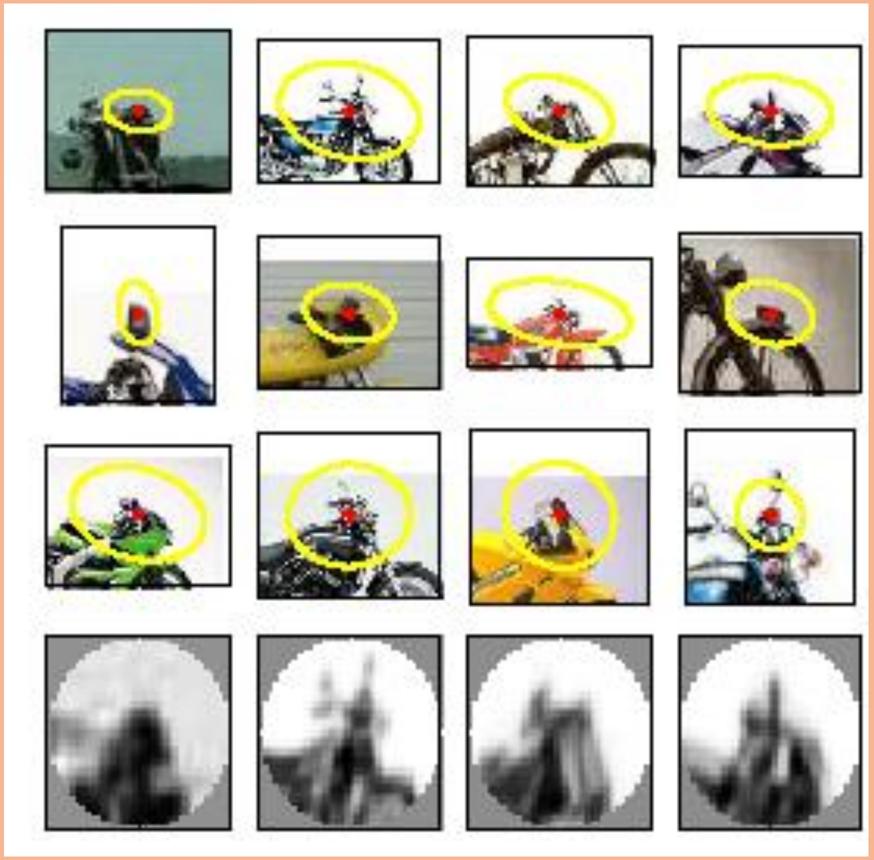


Regular grid

Fei-Fei et al. 2005

Image patch examples of codewords

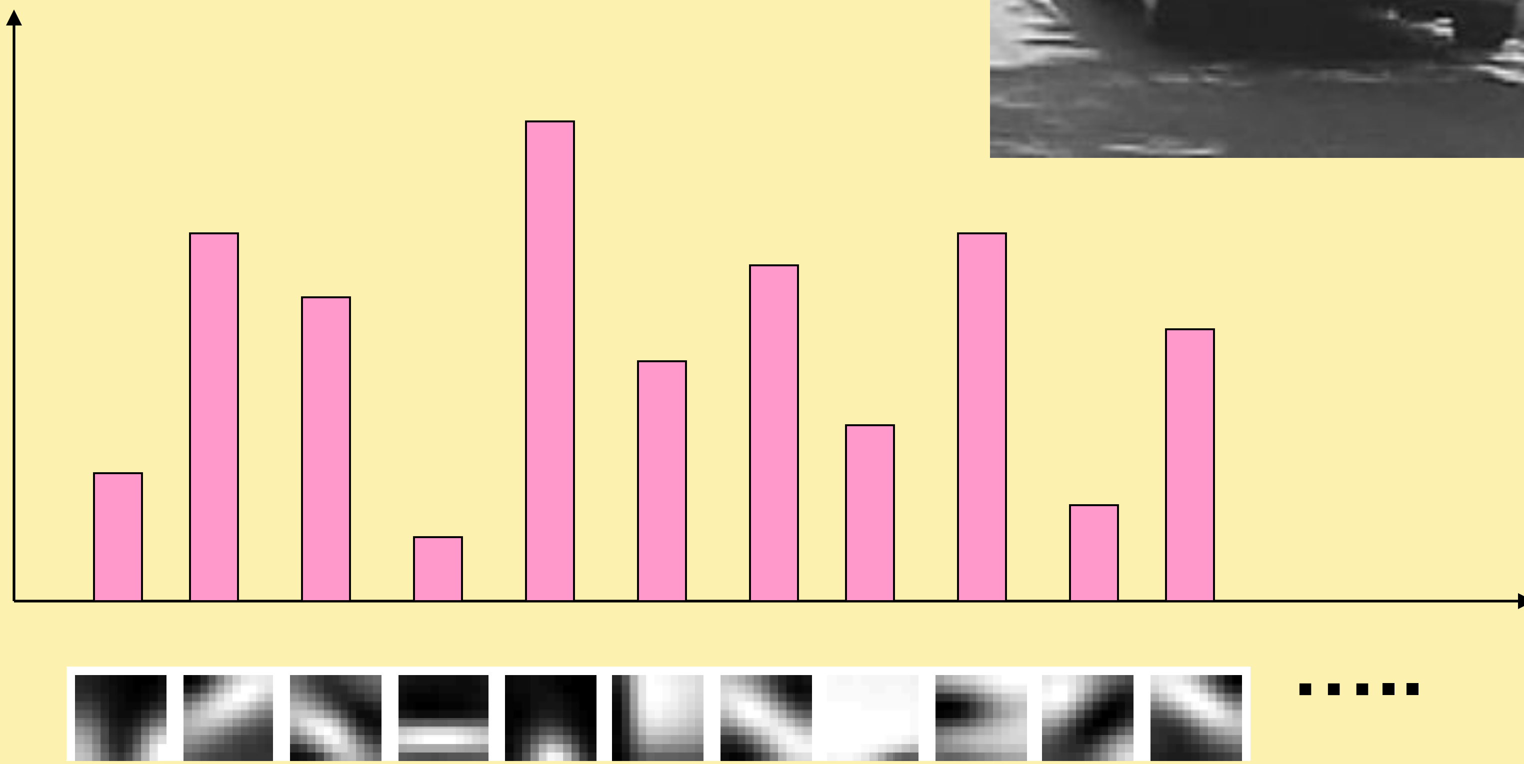




Sivic et al. 2005

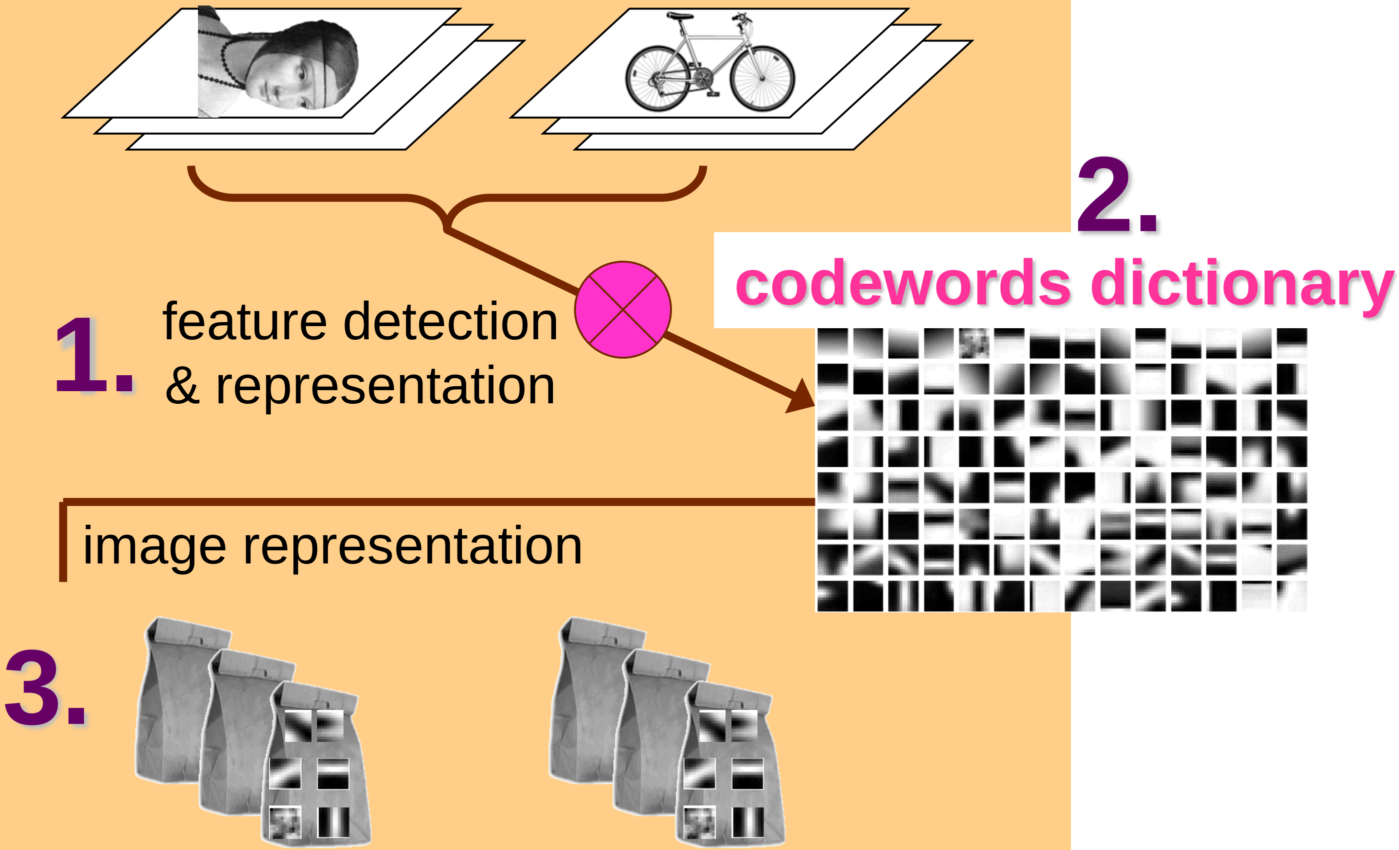
3. Image representation

frequency

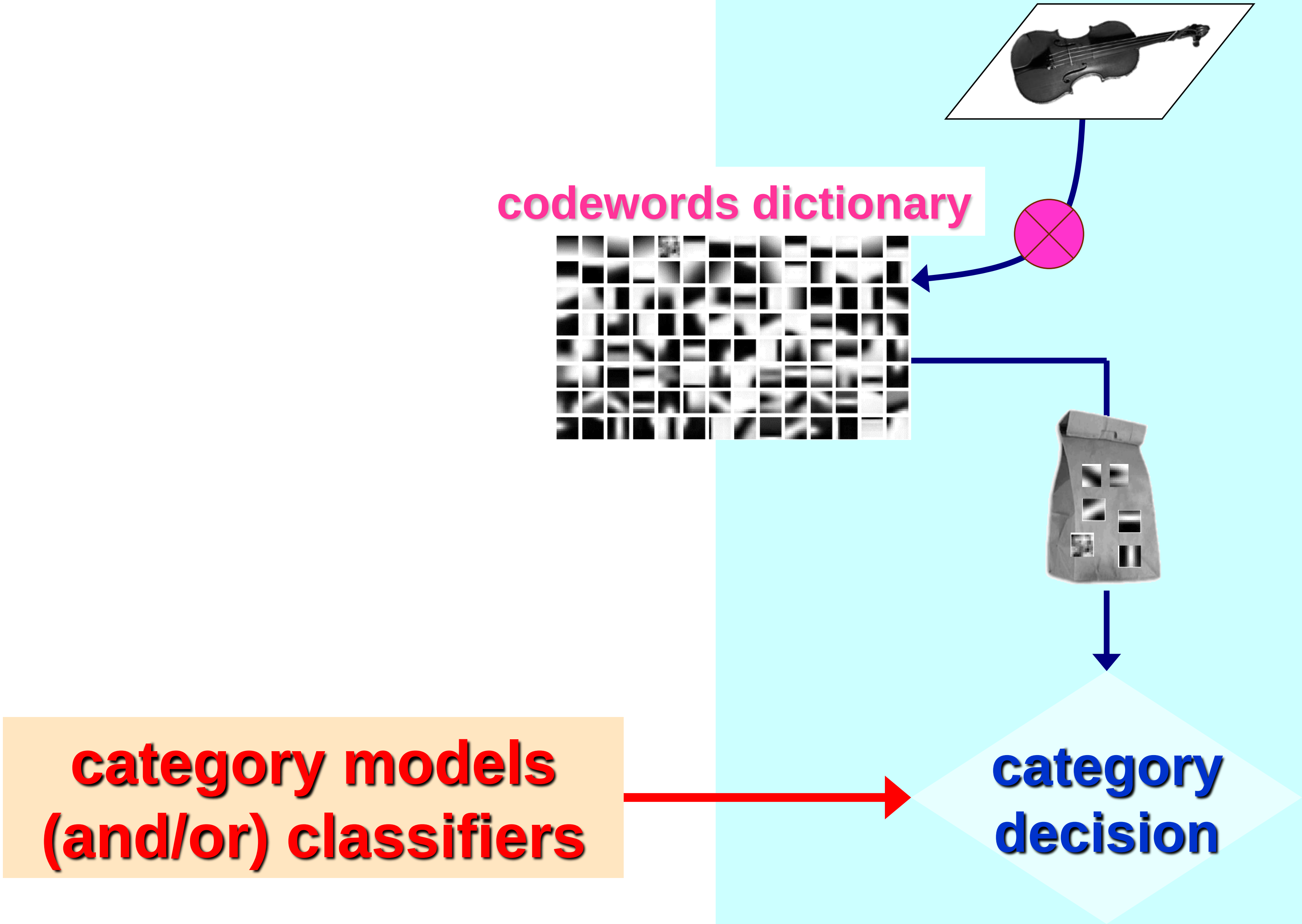


codewords

Representation



Learning and Recognition

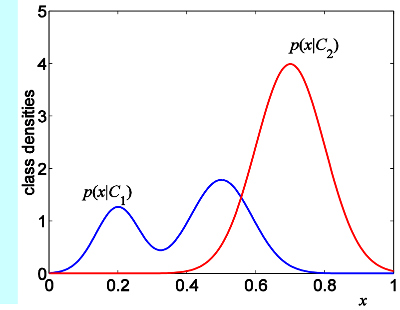


BoW-based Object Categorization

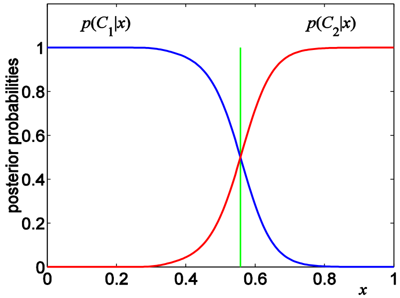
- Basic steps
  - Feature extraction and representation
  - Building codebook (codewords dictionary) from training samples with clustering
  - Represent an image with histogram of codebook (i.e. Bag-of-words of an image)
  - Classify an unknown image with its BoW.

Learning and Recognition

1. Generative method:
   - graphical models



2. Discriminative method:
   - SVM



category models (and/or) classifiers

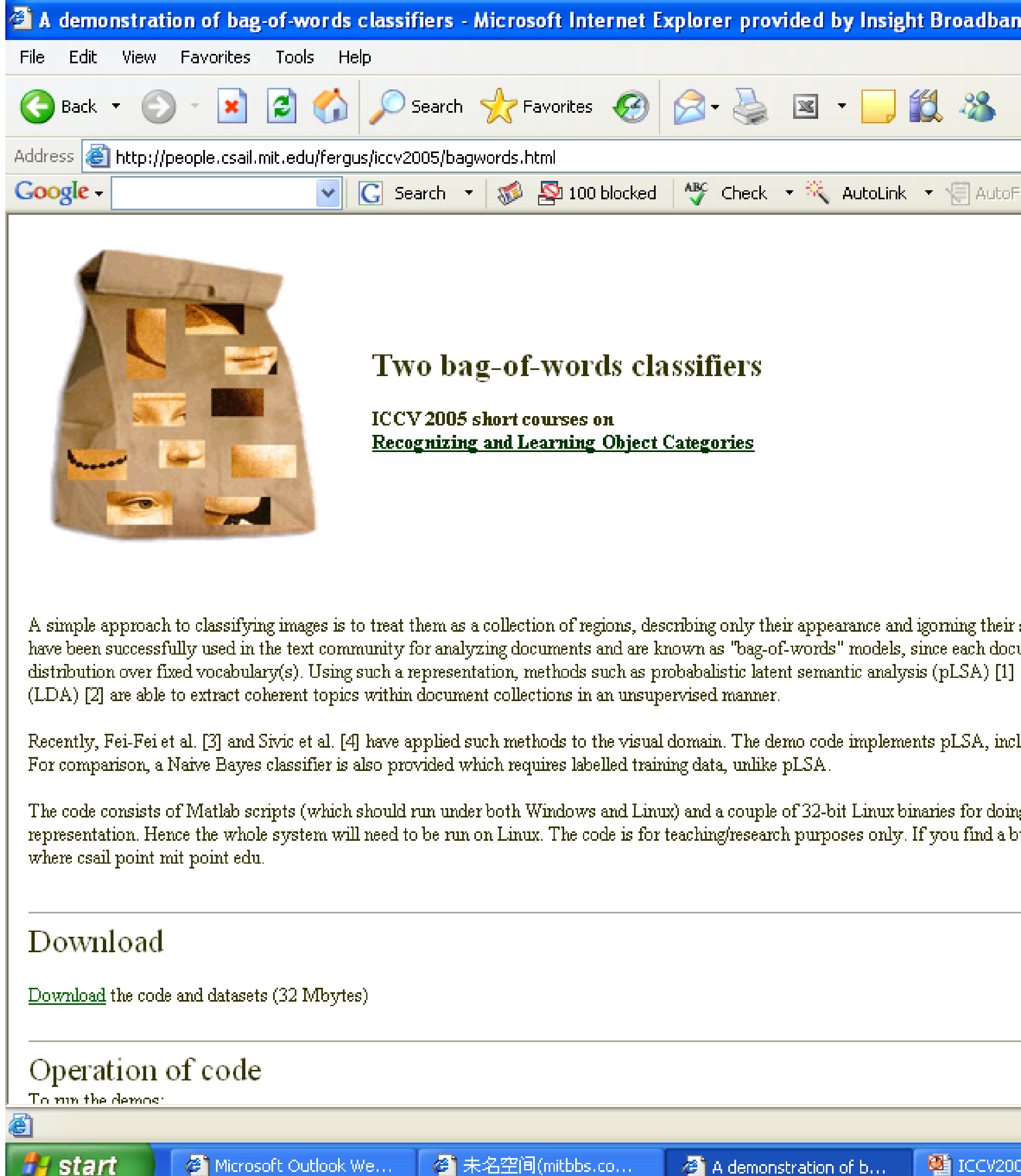
2 generative models

1. Naïve Bayes classifier
   – Csurka Bray, Dance & Fan, 2004

2. Hierarchical Bayesian text models (pLSA and LDA)
   – Background: Hoffman 2001, Blei, Ng & Jordan, 2004
   – Object categorization: Sivic et al. 2005, Sudderth et al. 2005
   – Natural scene categorization: Fei-Fei et al. 2005

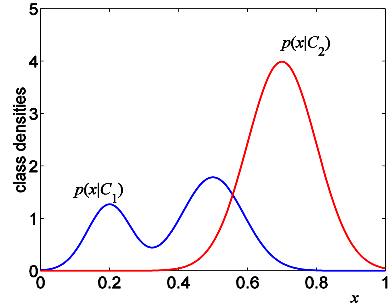
Demo

- Course website

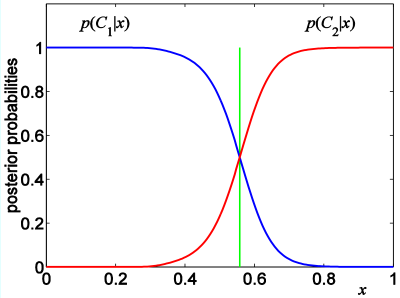


Learning and Recognition

1. Generative method:
   - graphical models

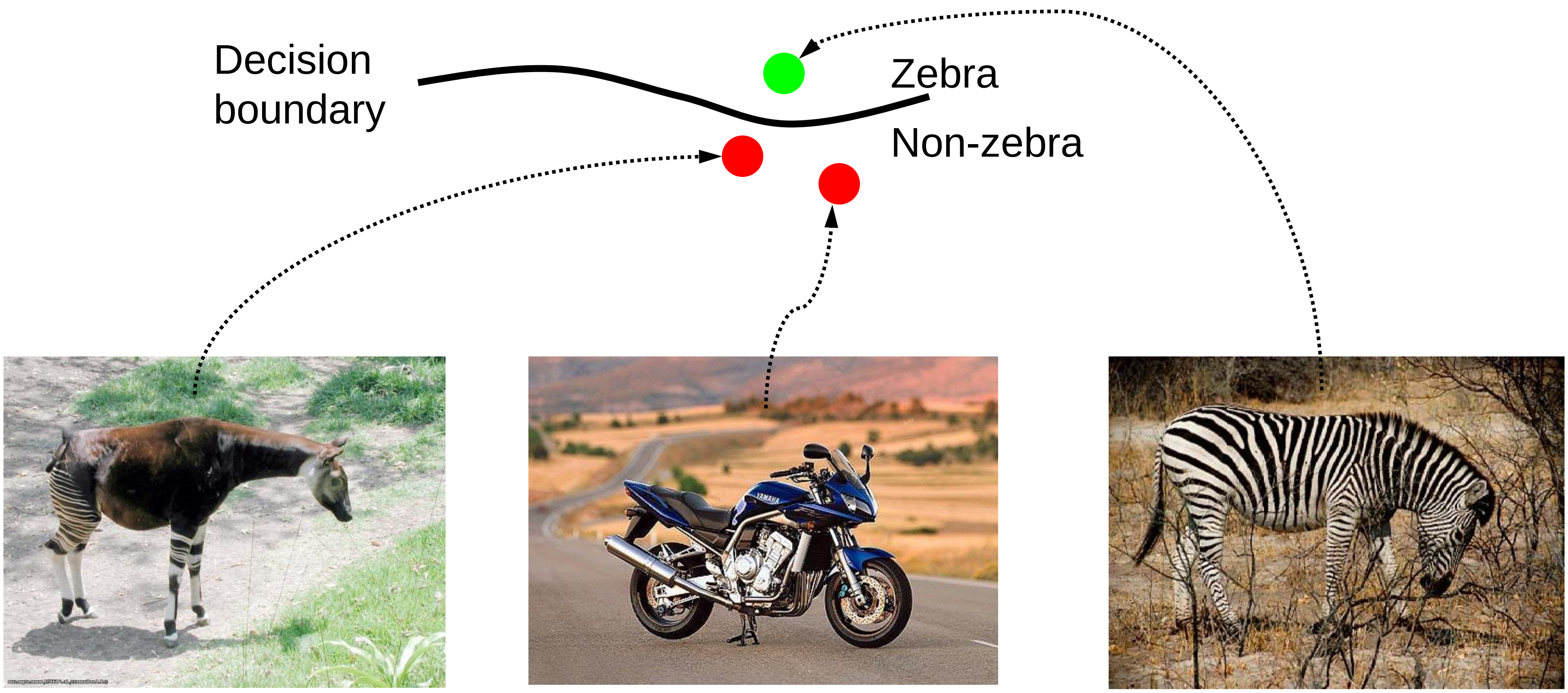


2. Discriminative method:
   - SVM



category models (and/or) classifiers

Discriminative methods based on ‘bag of words’ representation



Discriminative methods based on ‘bag of words’ representation

- Grauman & Darrell, 2005, 2006:
  - SVM w/ Pyramid Match kernels
- Others
  - Csurka, Bray, Dance & Fan, 2004
  - Serre & Poggio, 2005

Object recognition results

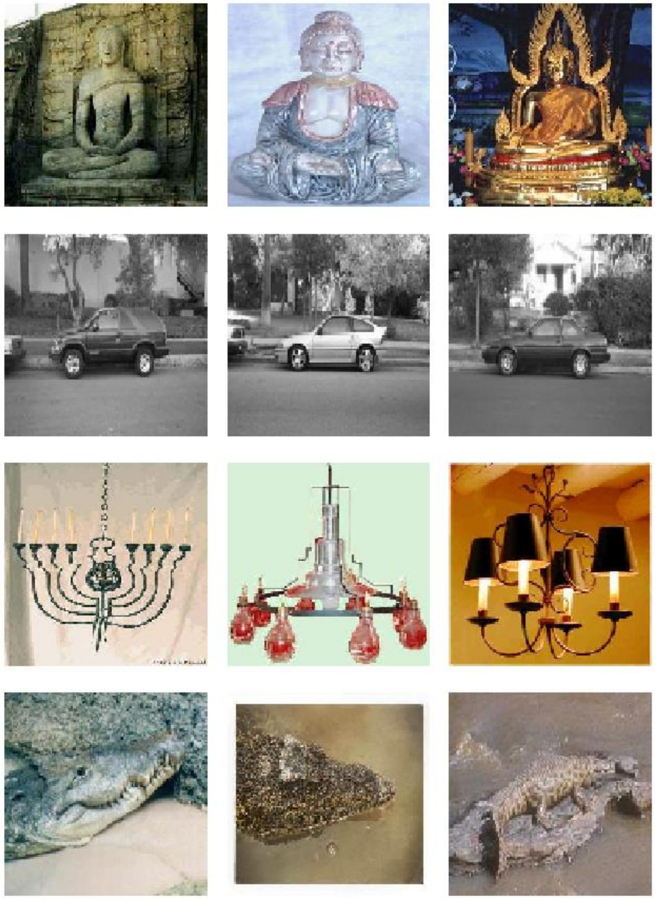
- ETH-80 database
  8 object classes
  (Eichhorn and Chapelle 2004)
- Features:
  - Harris detector
  - PCA-SIFT descriptor, *d*=10

Kernel	Complexity	Recognition rate
Match *[Wallraven et al.]*	$O(dm^2)$	84%
Bhattacharyya affinity *[Kondor & Jebara]*	$O(dm^3)$	85%
Pyramid match	$O(dmL)$	84%

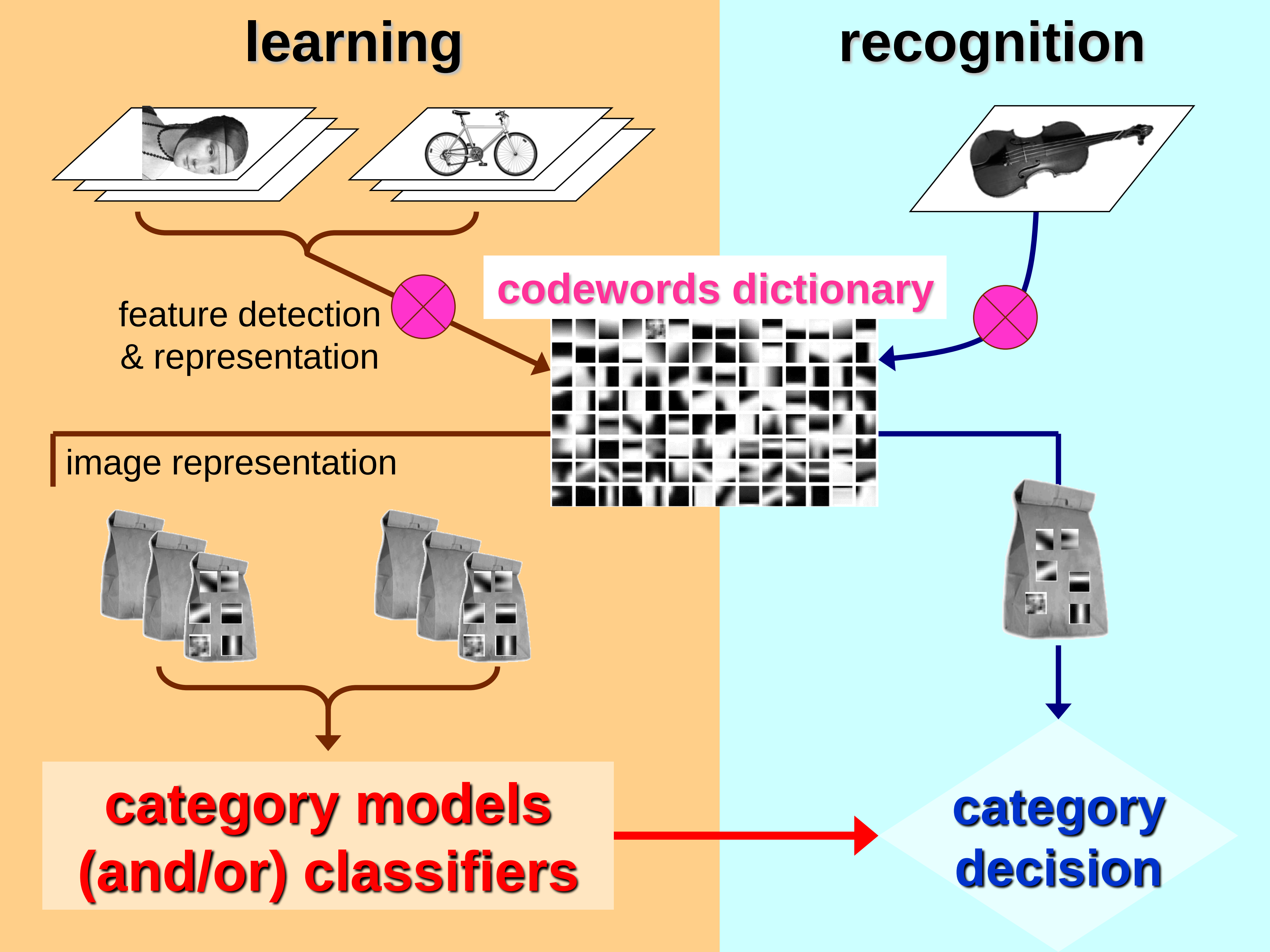
Slide credit: Kristen Grauman

Object recognition results

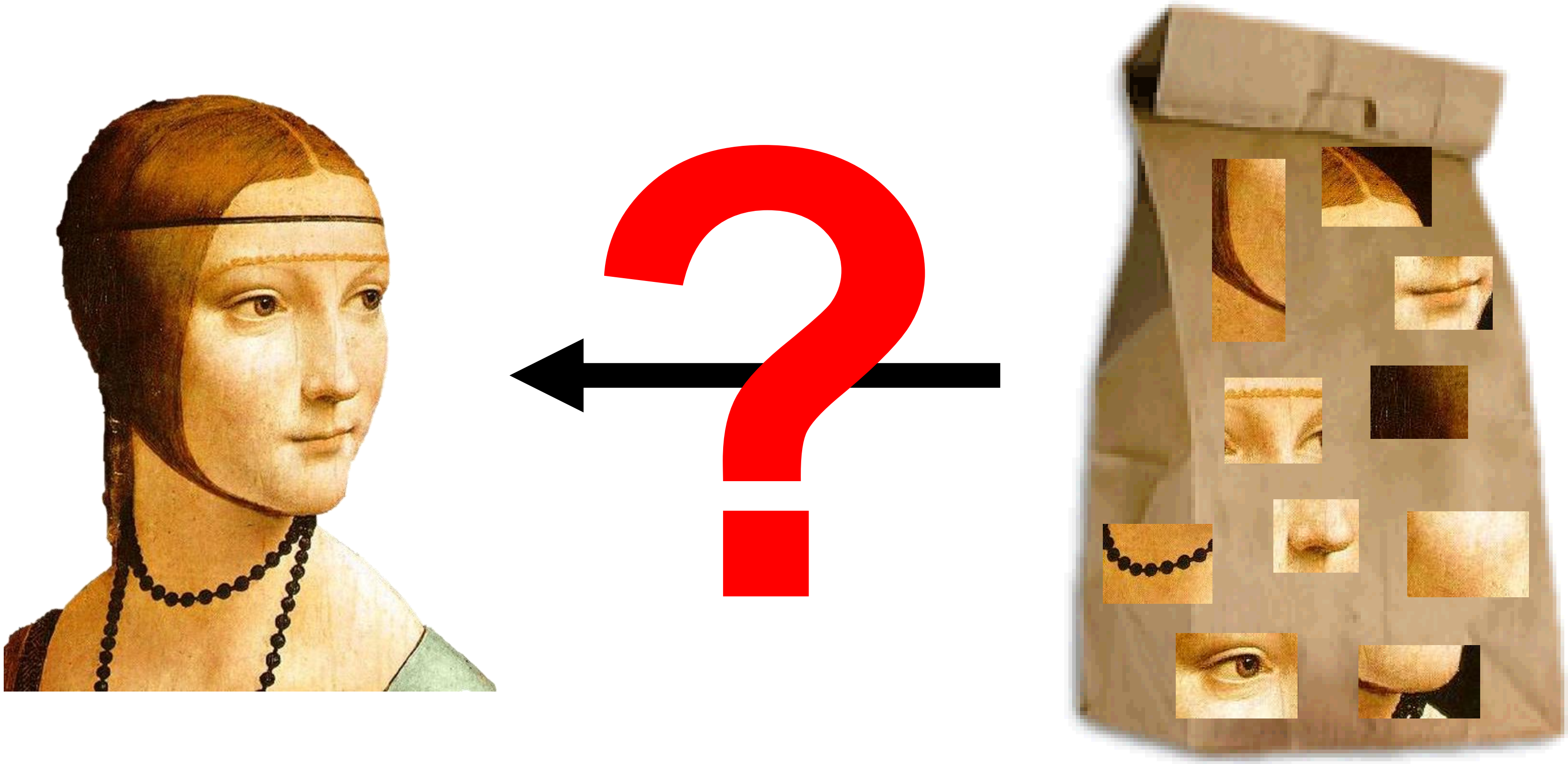
- Caltech objects database
  101 object classes
- Features:
  - SIFT detector
  - PCA-SIFT descriptor, *d*=10
- 30 training images / class
- 43% recognition rate
  (1% chance performance)
- 0.002 seconds per match



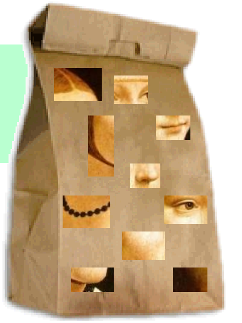
Slide credit: Kristen Grauman


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

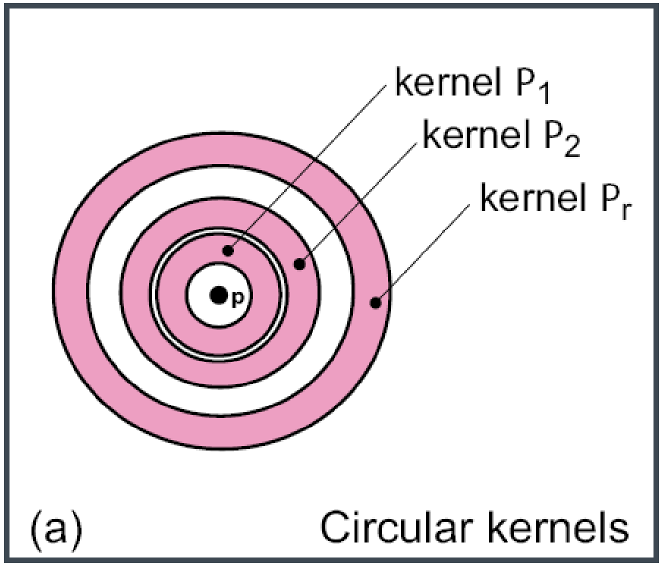
What about spatial info?

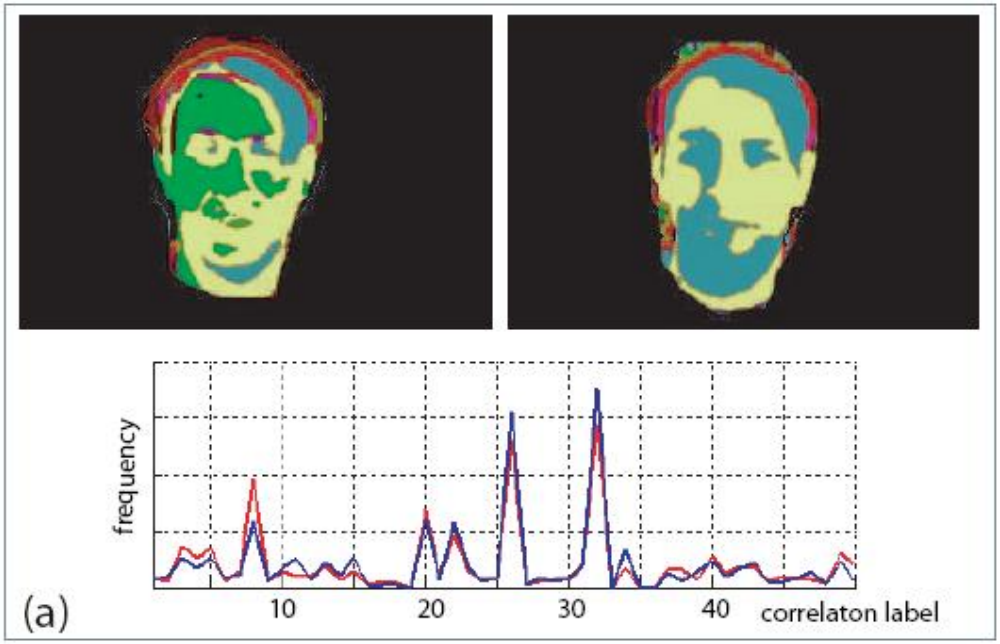


What about spatial info?



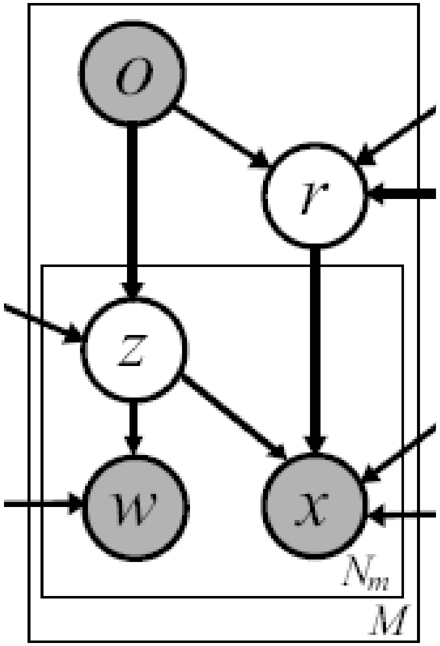
- Feature level
  - Spatial influence through correlogram features: Savarese, Winn and Criminisi, CVPR 2006

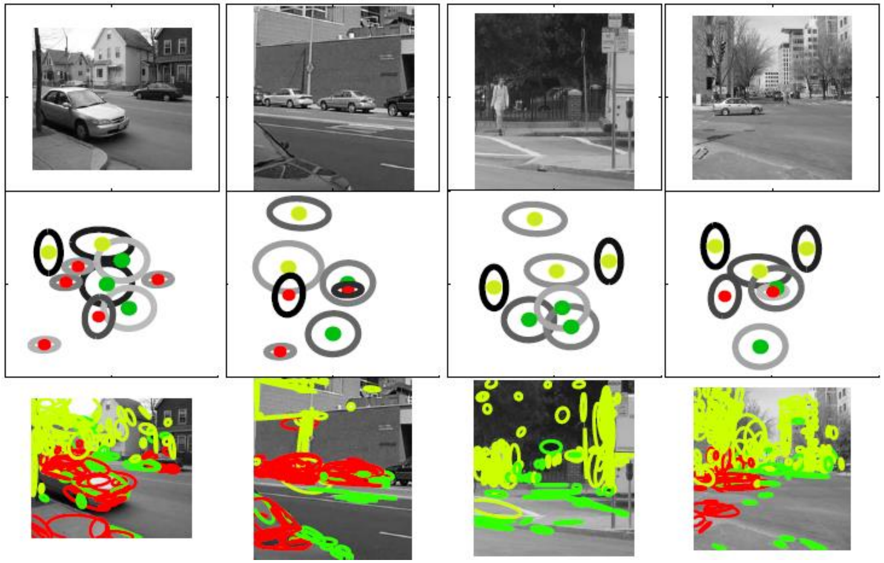




What about spatial info?

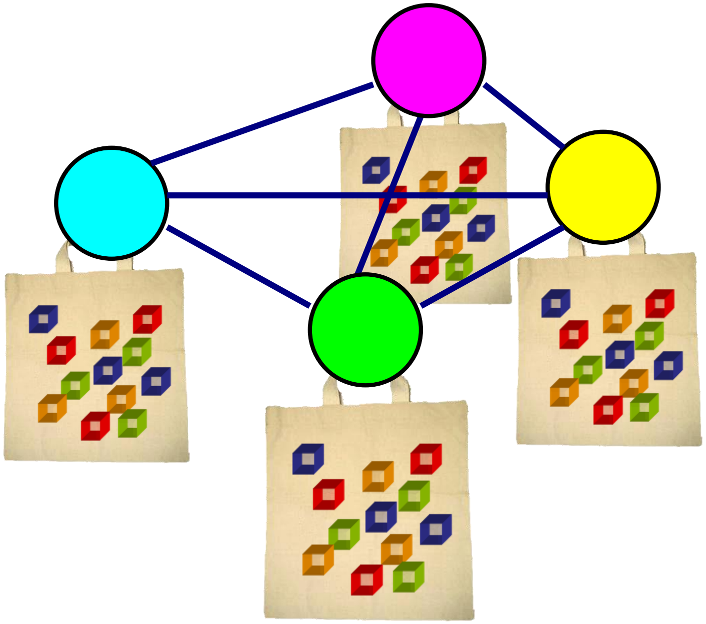
- Feature level
- Generative models
  - Sudderth, Torralba, Freeman & Willsky, 2005, 2006
  - Niebles & Fei-Fei, CVPR 2007

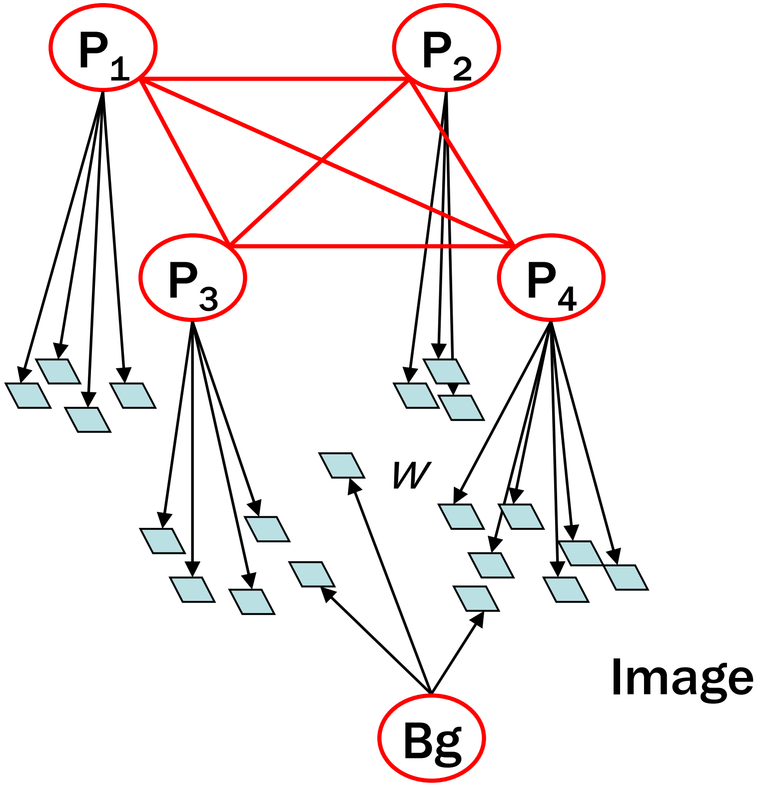




What about spatial info?

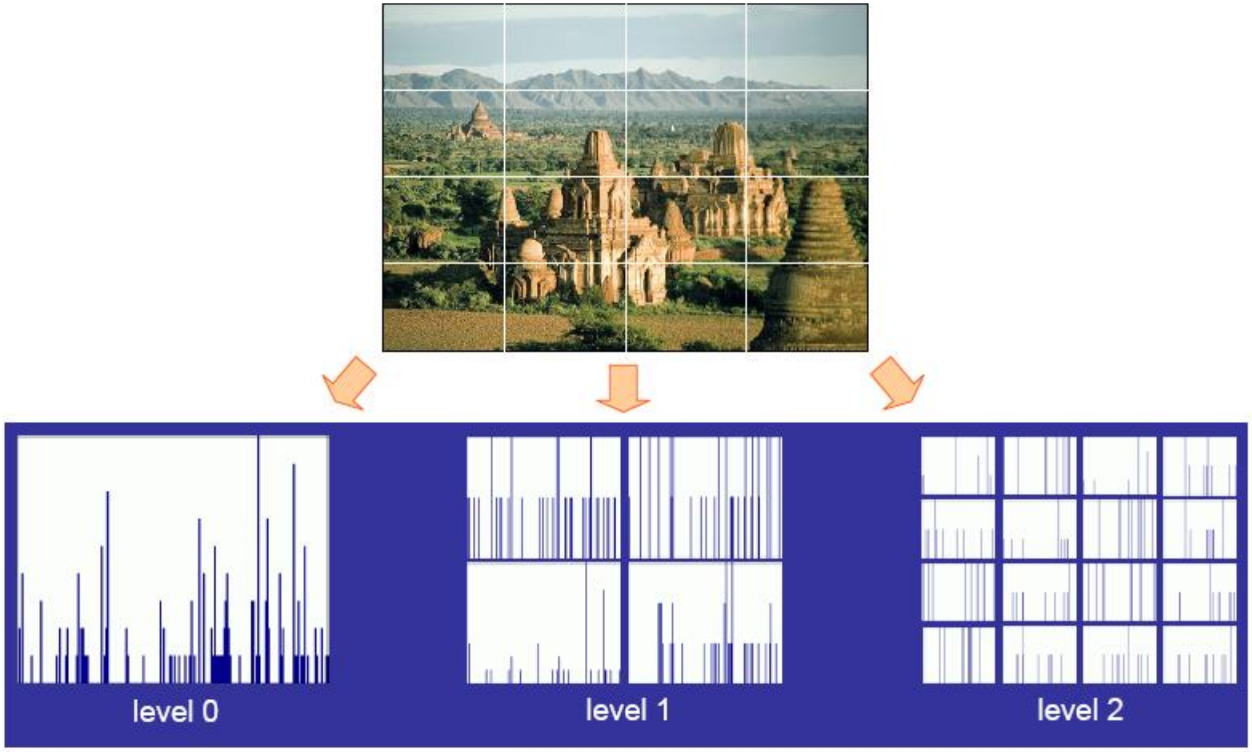
- Feature level
- Generative models
  - Sudderth, Torralba, Freeman & Willsky, 2005, 2006
  - Niebles & Fei-Fei, CVPR 2007





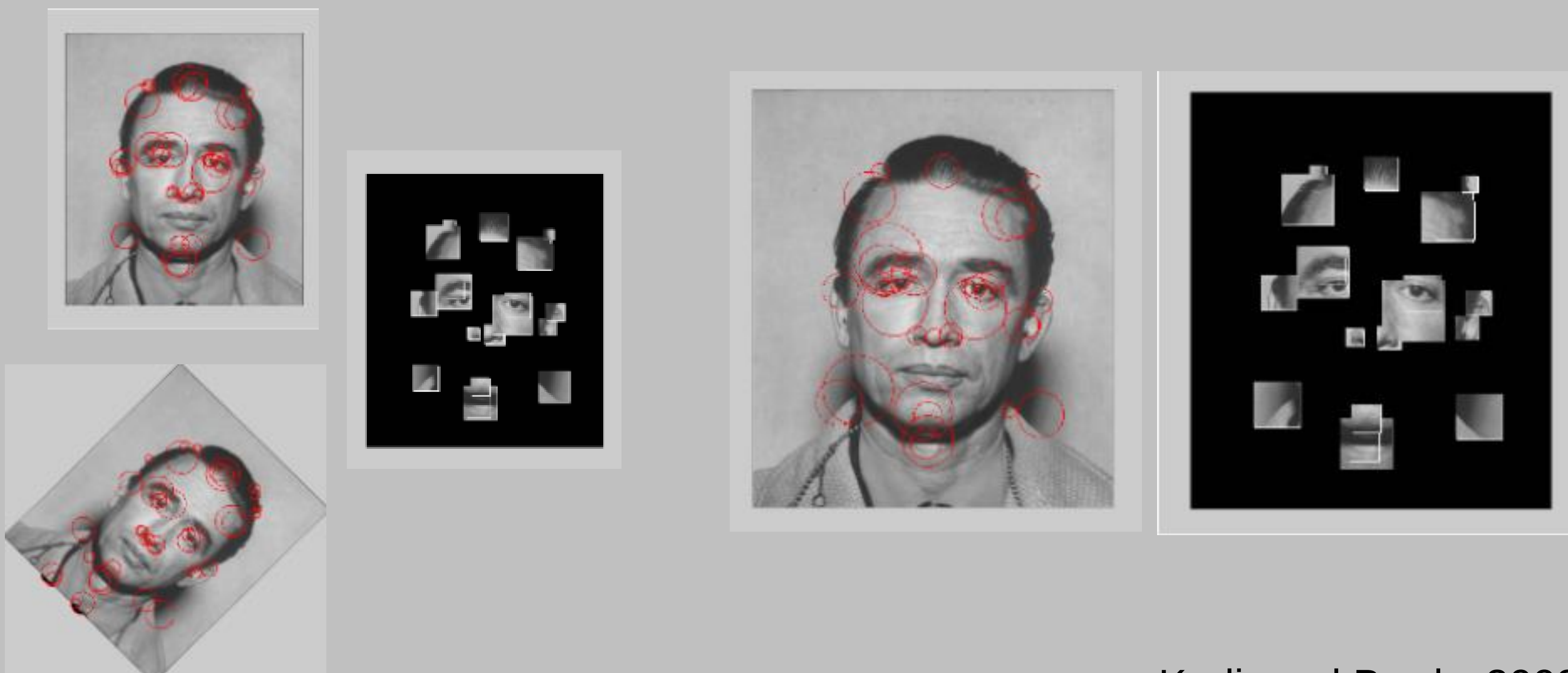
What about spatial info?

- Feature level
- Generative models
- Discriminative methods
  - Lazebnik, Schmid & Ponce, 2006



Invariance issues

- Scale and rotation
  - Implicit
  - Detectors and descriptors

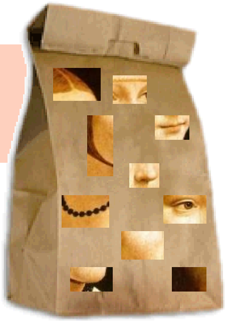


Kadir and Brady. 2003

Invariance issues

- Scale and rotation
- Occlusion
  - Implicit in the models
  - Codeword distribution: small variations
  - (In theory) Theme (z) distribution: different occlusion patterns

Invariance issues



- Scale and rotation
- Occlusion
- Translation
  - Encode (relative) location information
    - Sudderth, Torralba, Freeman & Willsky, 2005, 2006
    - Niebles & Fei-Fei, 2007

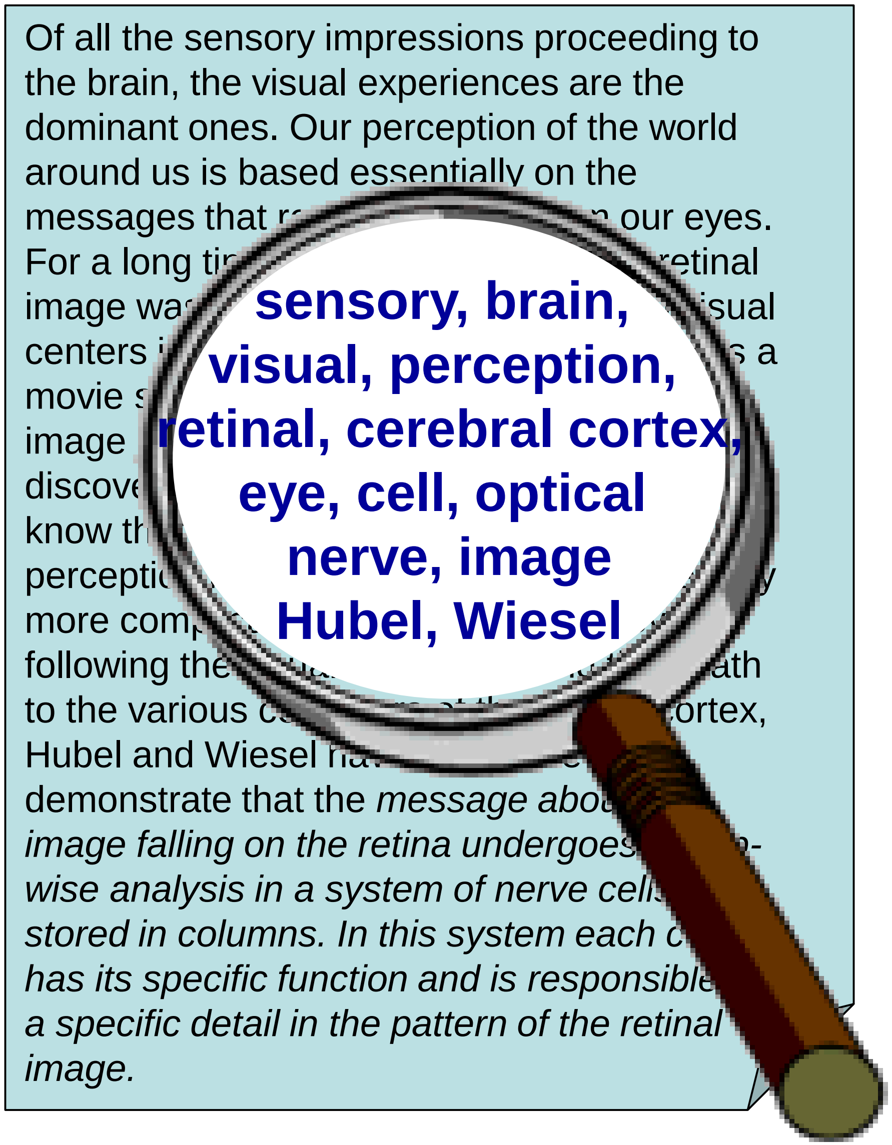
Invariance issues

- Scale and rotation
- Occlusion
- Translation
- View point (in theory)
  - Codewords: detector and descriptor
  - Theme distributions: different view points

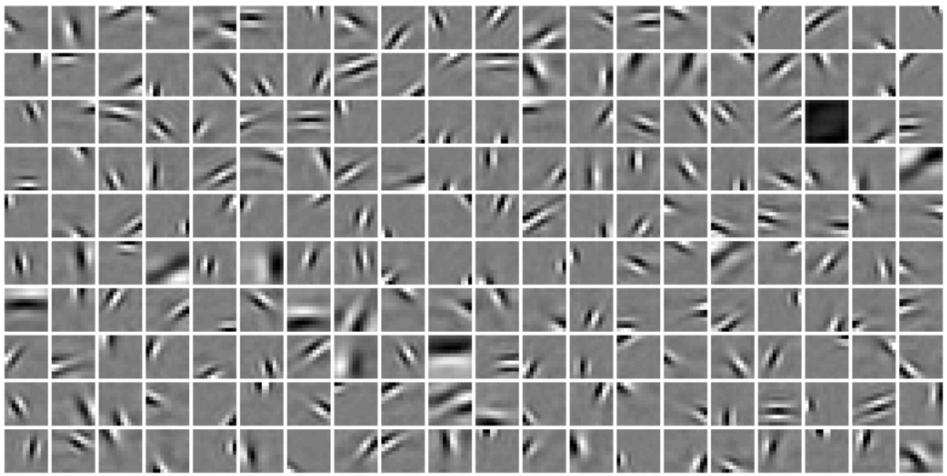
Fergus, Fei-Fei, Perona & Zisserman, 2005

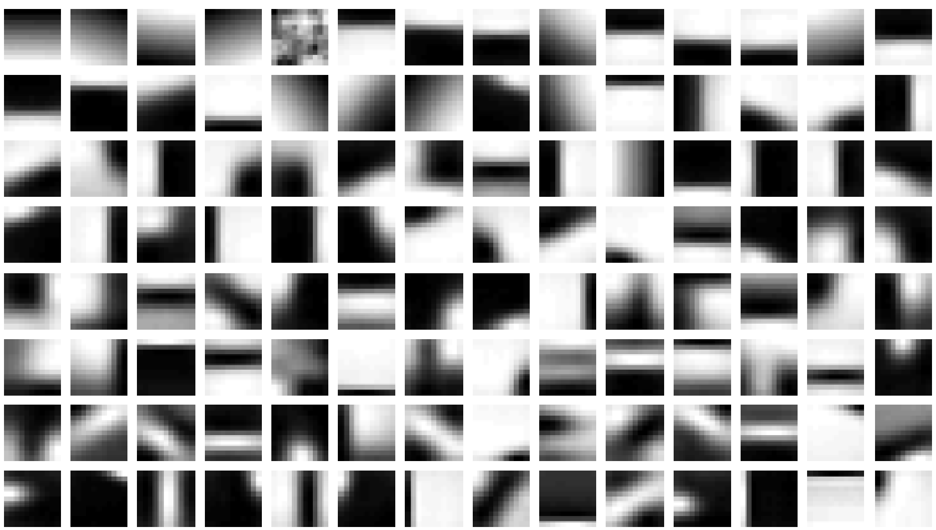
Model properties

- Intuitive
  – Analogy to documents



Model properties

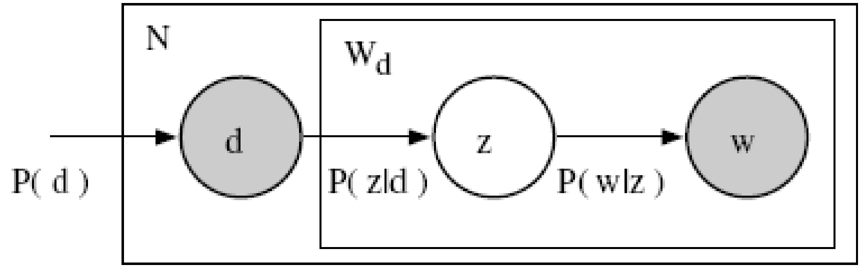




- Intuitive
  - Analogy to documents
  - Analogy to human vision

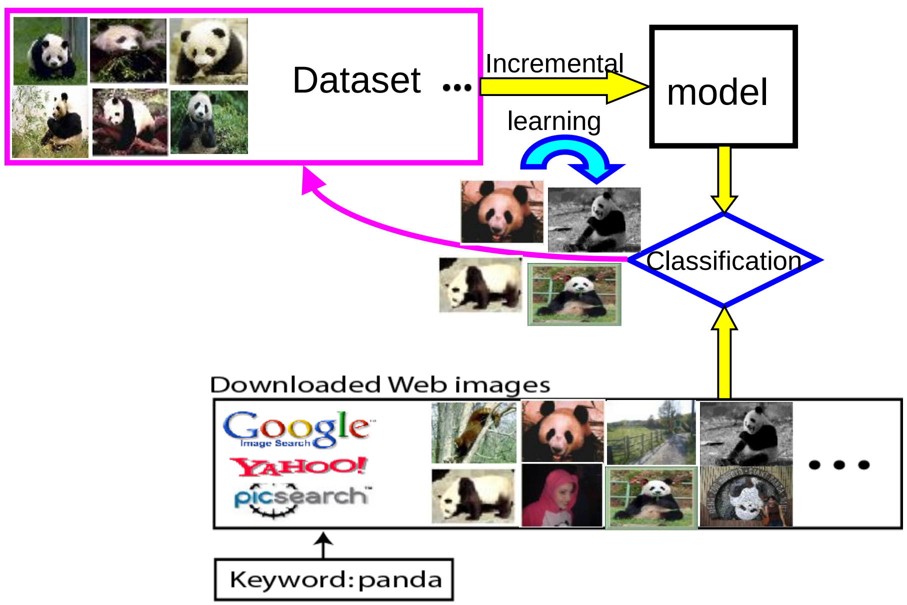
Olshausen and Field, 2004, Fei-Fei and Perona, 2005

Model properties



Sivic, Russell, Efros, Freeman, Zisserman, 2005

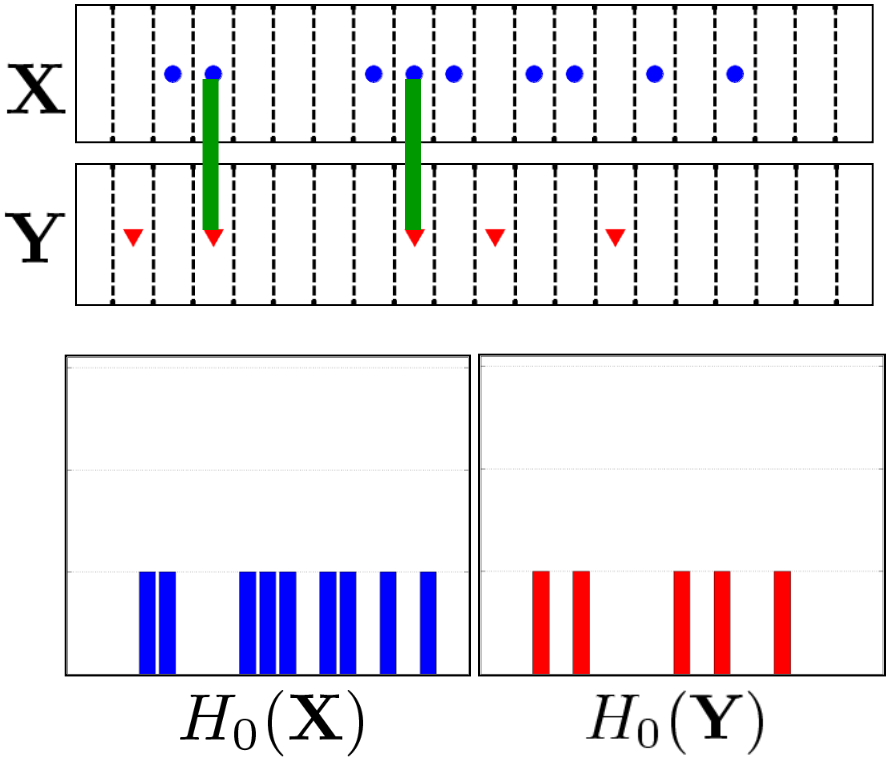
- Intuitive
- generative models
  - Convenient for weakly- or un-supervised, incremental training
  - Prior information
  - Flexibility (e.g. HDP)



Li, Wang & Fei-Fei, CVPR 2007

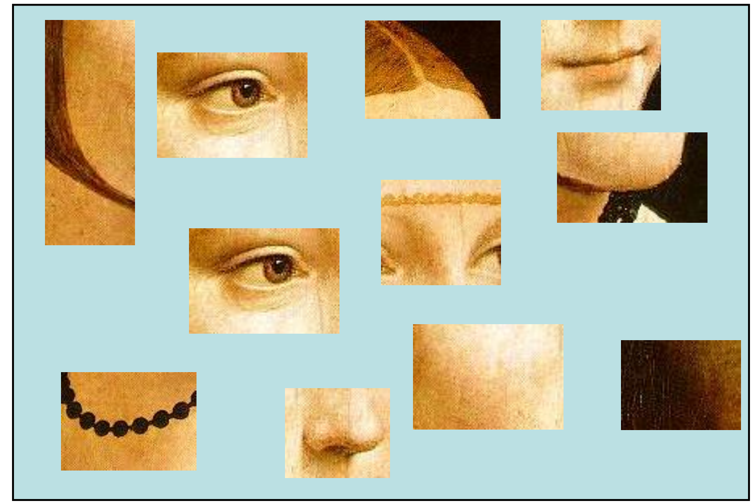
Model properties

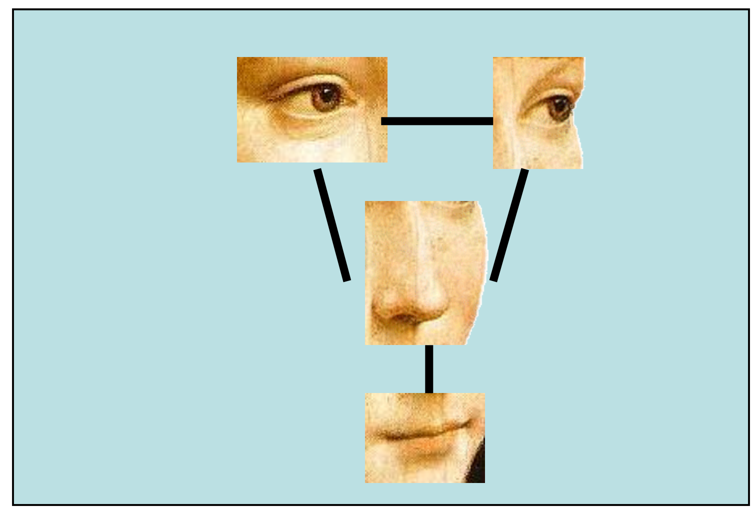
- Intuitive
- generative models
- Discriminative method
  - Computationally efficient



Grauman et al. CVPR 2005

Model properties





- Intuitive
- generative models
- Discriminative method
- Learning and recognition relatively fast
  - Compare to other methods

Weakness of the model

- No rigorous geometric information of the object components
- It's intuitive to most of us that objects are made of parts – no such information
- Not extensively tested yet for
  - View point invariance
  - Scale invariance
- Segmentation and localization unclear