



Object Categorization

Bag-of-words models

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 reach our eyes.

For a long time, the retinal image was considered as a movie screen. It is now known that the image is processed in a more complex way. Following the discovery of the pathway to the various centers of the cortex, Hubel and Wiesel have demonstrated that the *message about the image falling on the retina undergoes a point-by-point analysis in a system of nerve cells stored in columns. In this system each cell has its specific function and is responsible for a specific detail in the pattern of the retinal image.*

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

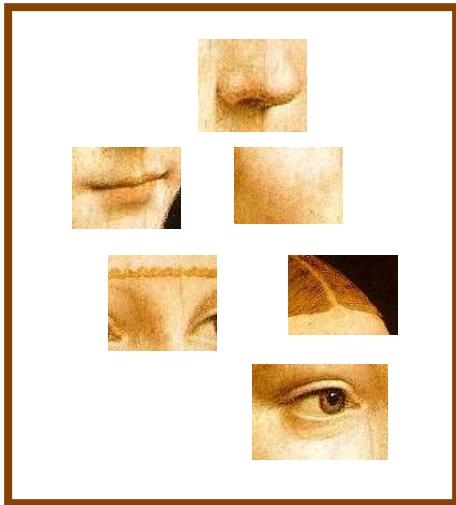
China is forecasting a trade surplus of \$90bn (£51bn) to \$100bn this year, a threefold increase on 2004's \$32bn. The Commerce Ministry said the surplus would be created by a predicted 30% increase in exports to \$750bn, compared with \$560bn in 2004.

The increase will annoy the US, which has long complained of China's deliberate export subsidies. China's government agrees that the yuan is undervalued and also needs to increase demand so that it can stimulate the country. China has agreed to let the yuan against the dollar rise and permitted it to trade within a narrow band but the US wants the yuan to be allowed to move freely. However, Beijing has made it clear that it will take its time and tread carefully before allowing the yuan to rise further in value.

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

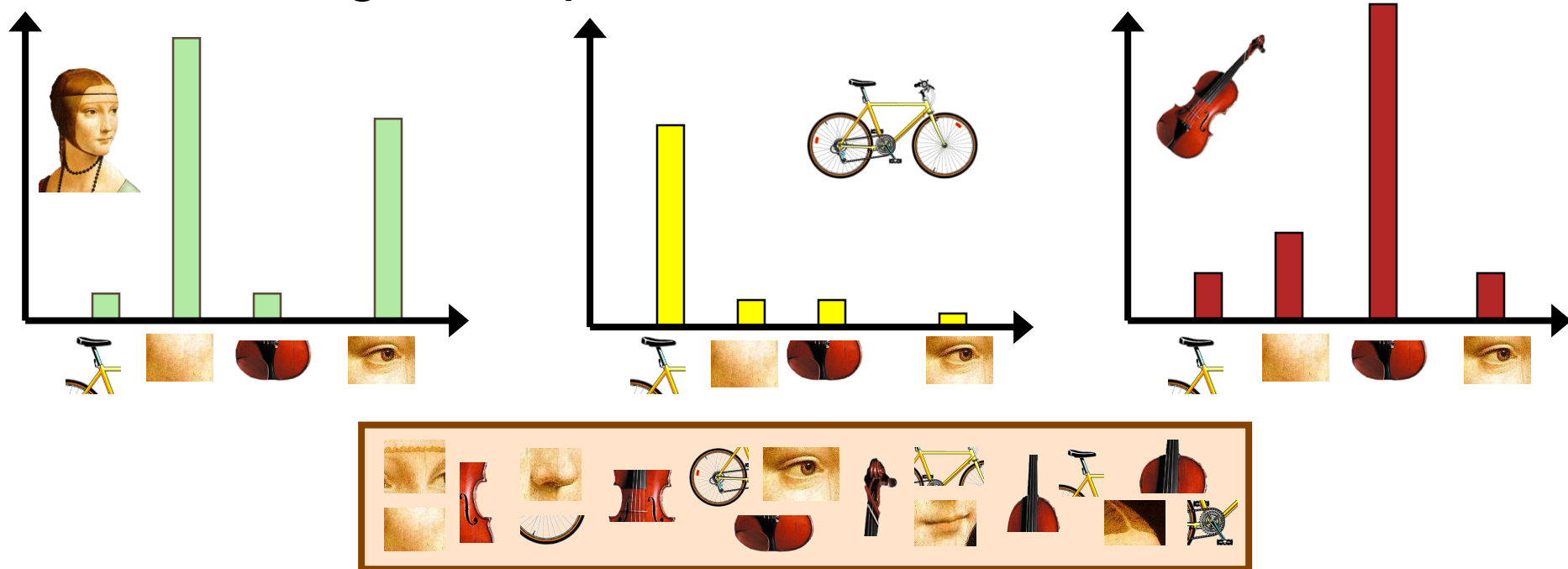
A clarification: definition of “BoW”

- Looser definition
 - Independent features

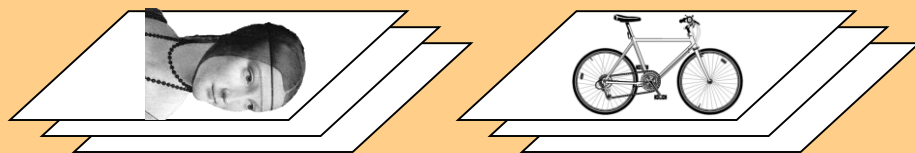


A clarification: definition of “BoW”

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



learning



feature detection
& representation

codewords dictionary

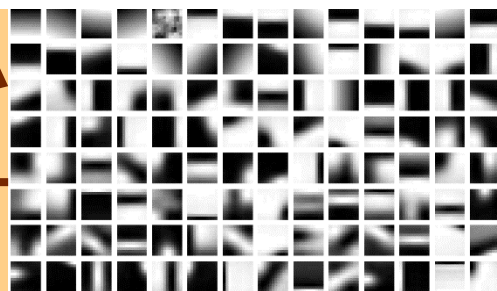
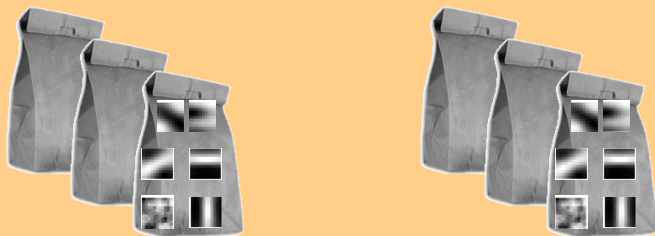


image representation



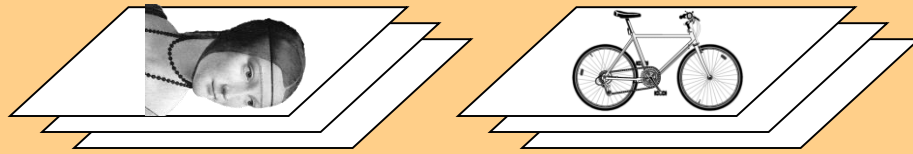
**category models
(and/or) classifiers**

recognition

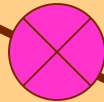


**category
decision**

Representation



1. feature detection
& representation



2. **codewords dictionary**

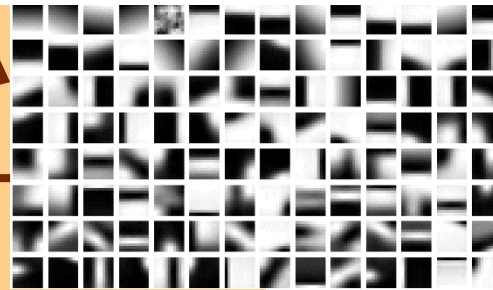
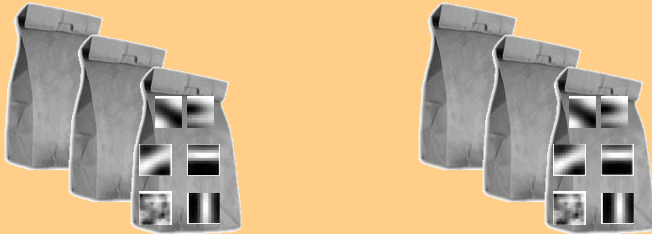
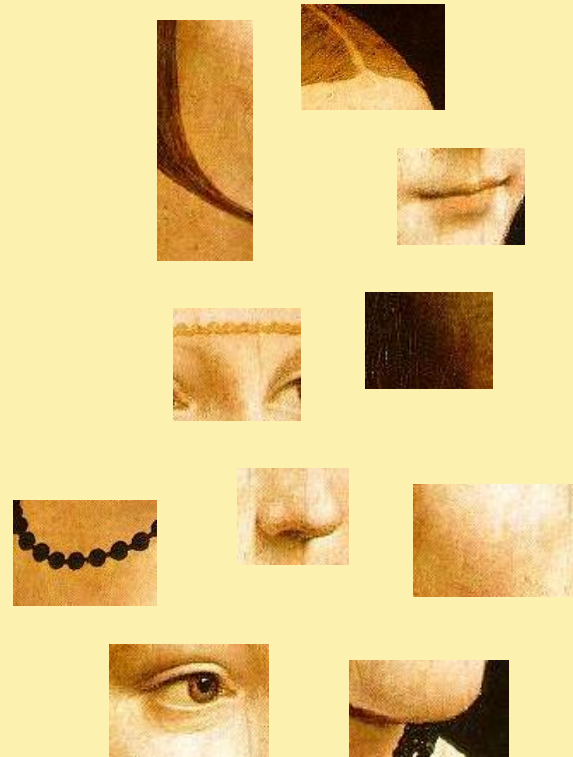


image representation

3.

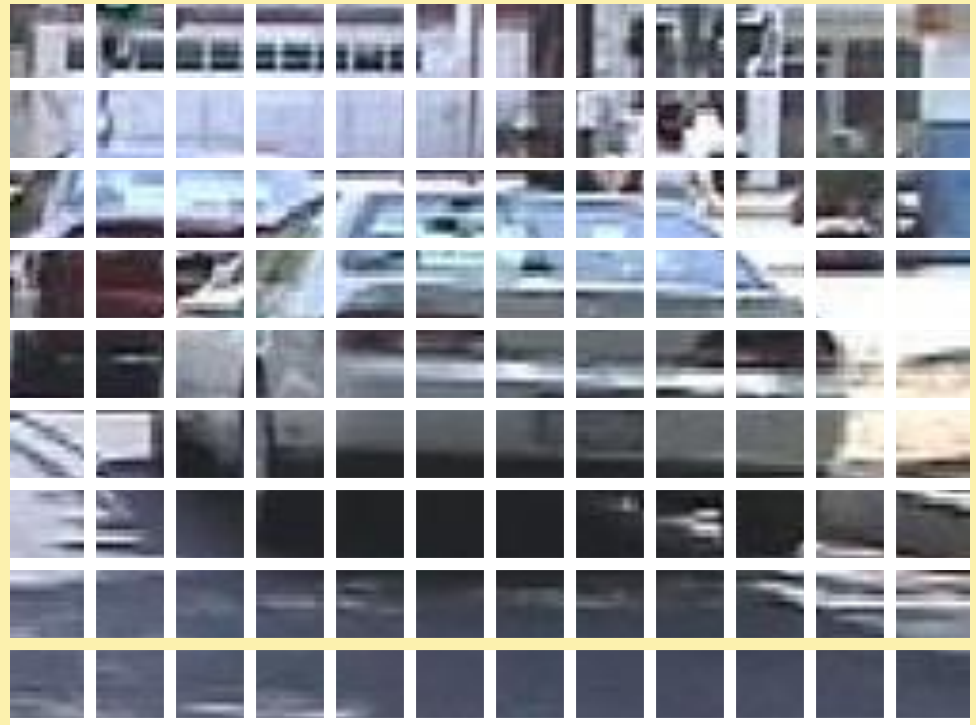


1.Feature detection and representation



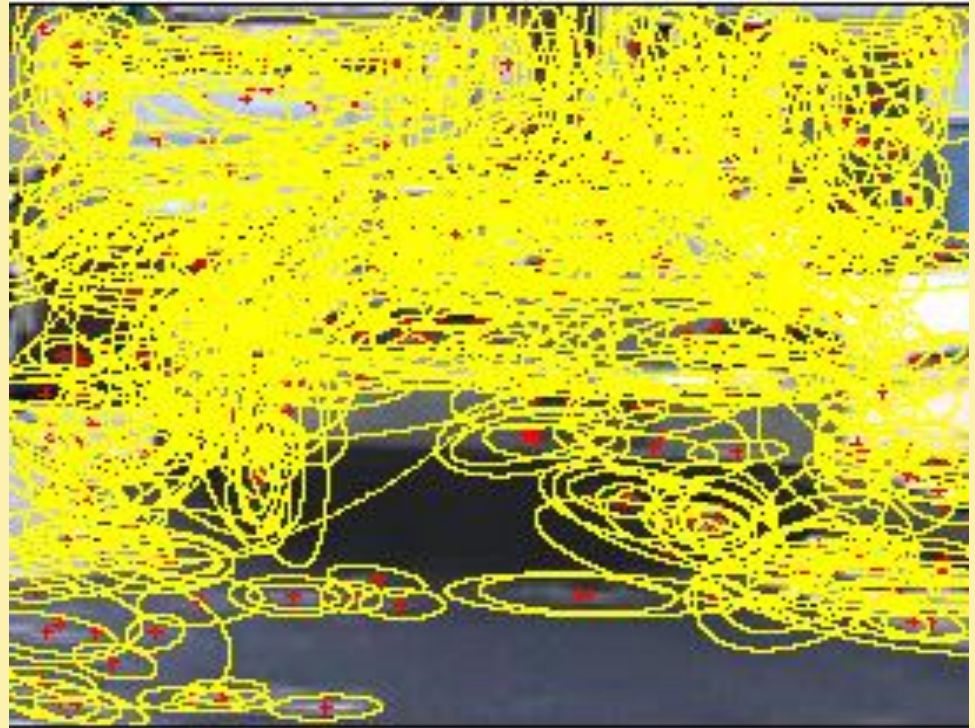
1.Feature detection and representation

- 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



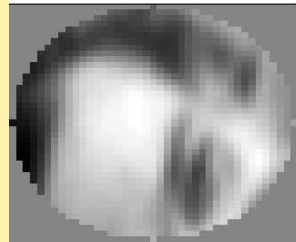
1.Feature detection and representation

- 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)

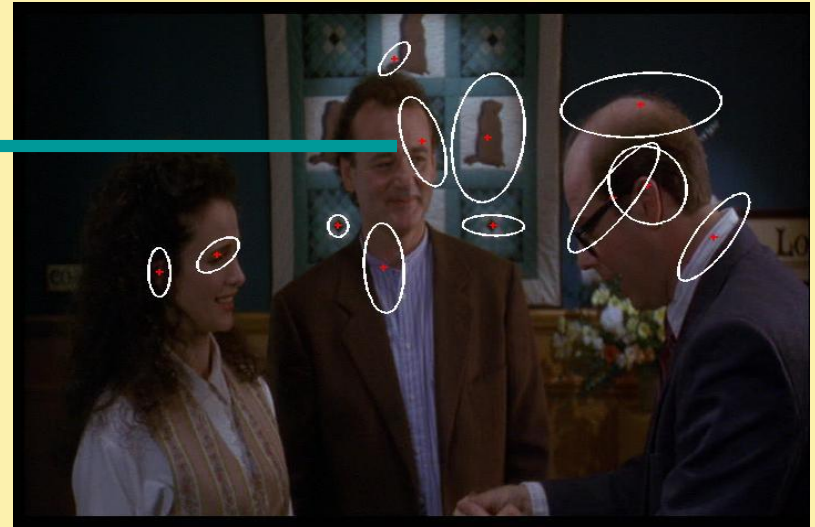
1. Feature detection and representation



**Compute
SIFT
descriptor**
[Lowe'99]



**Normalize
patch**



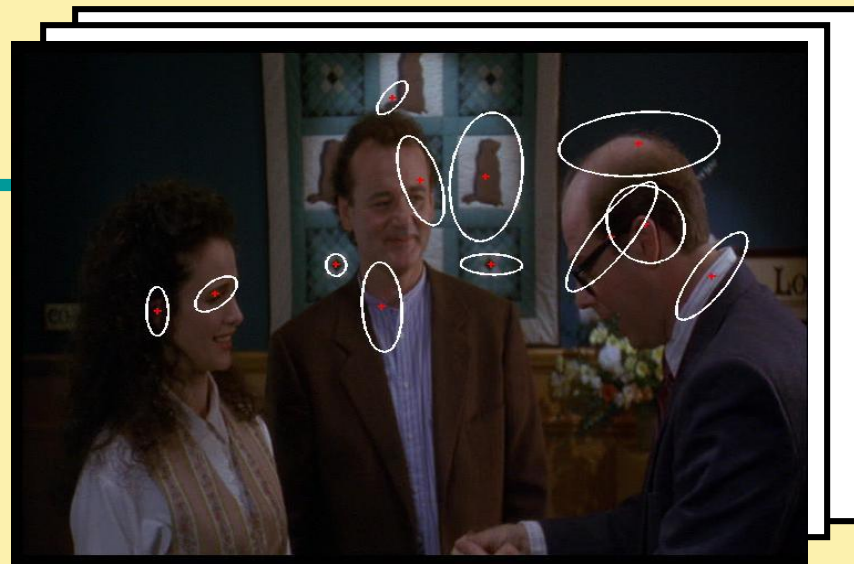
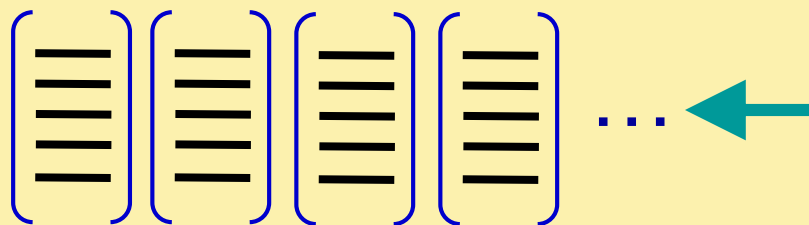
Detect patches

[Mikojaczyk and Schmid '02]

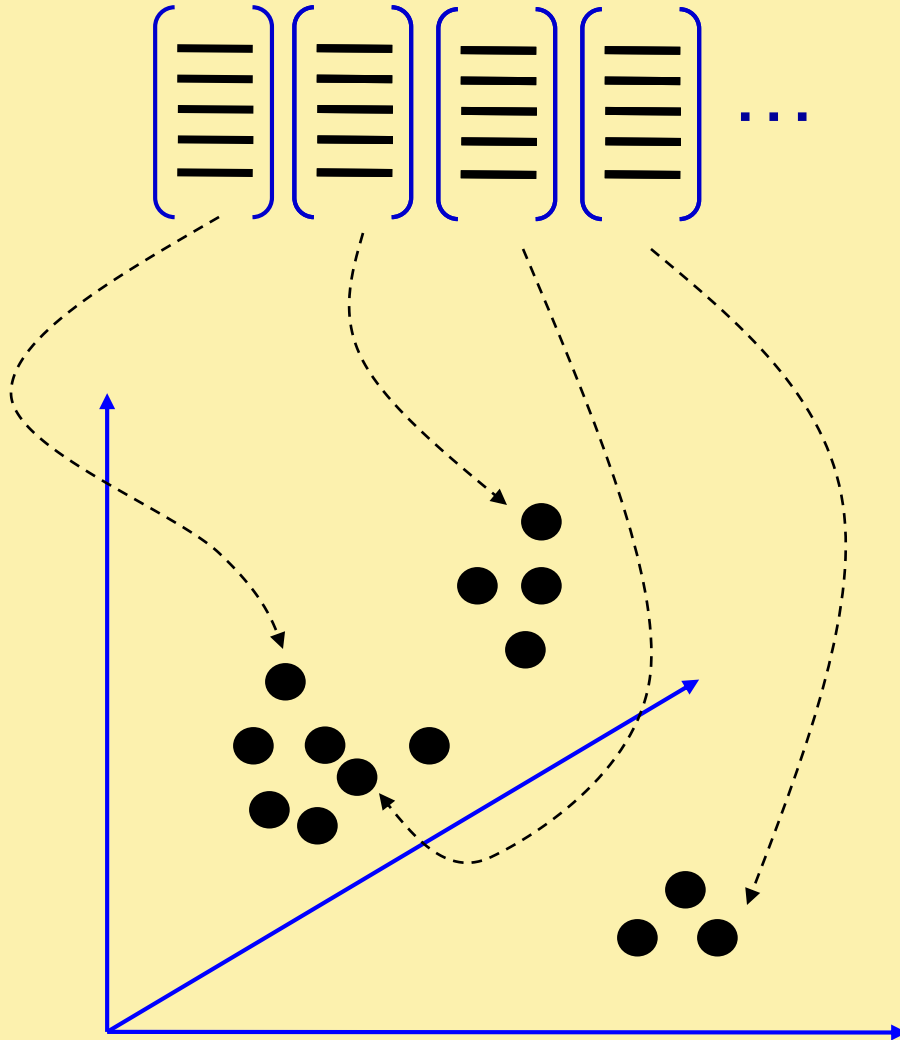
[Mata, Chum, Urban & Pajdla, '02]

[Sivic & Zisserman, '03]

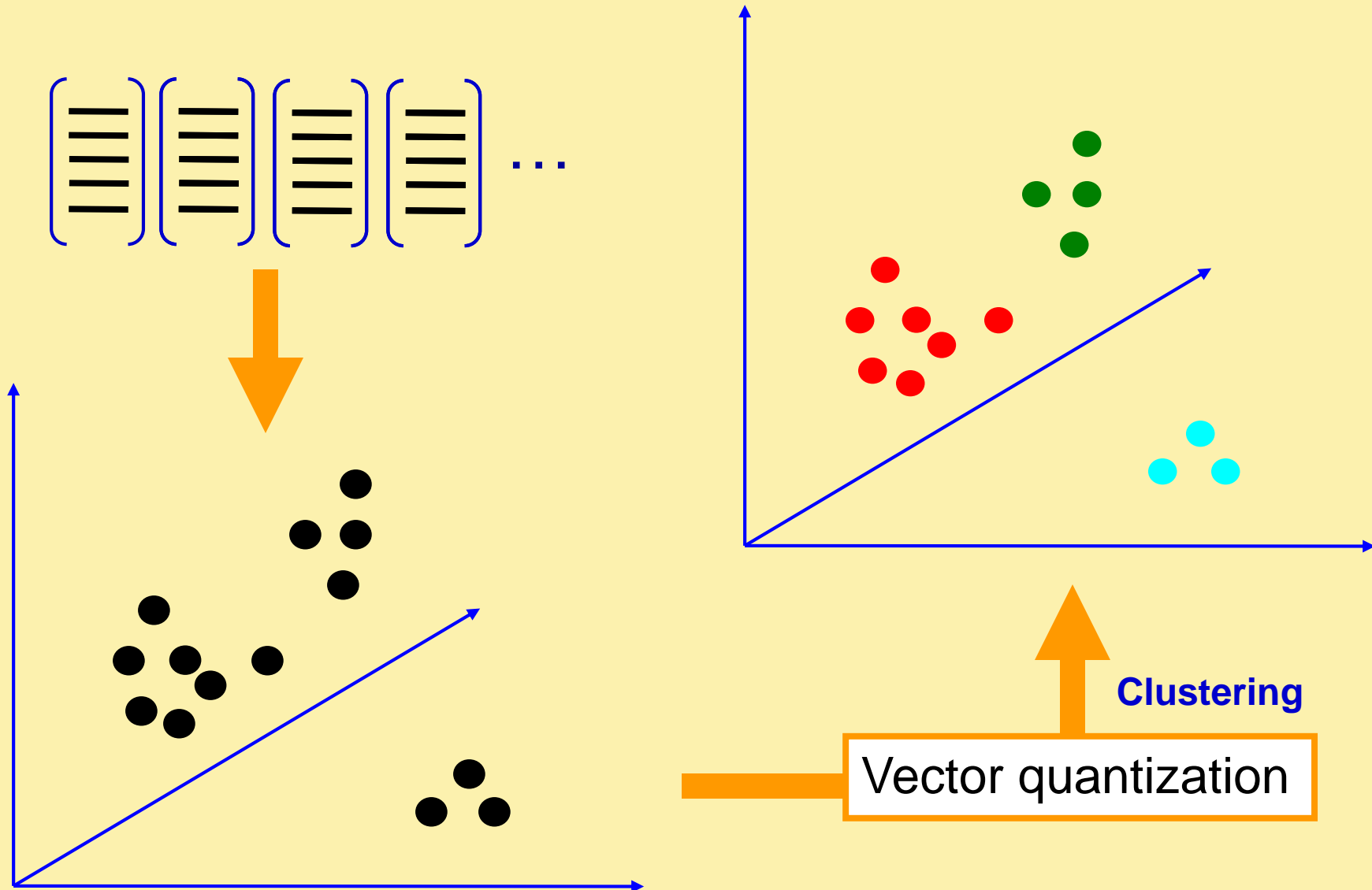
1. Feature detection and representation



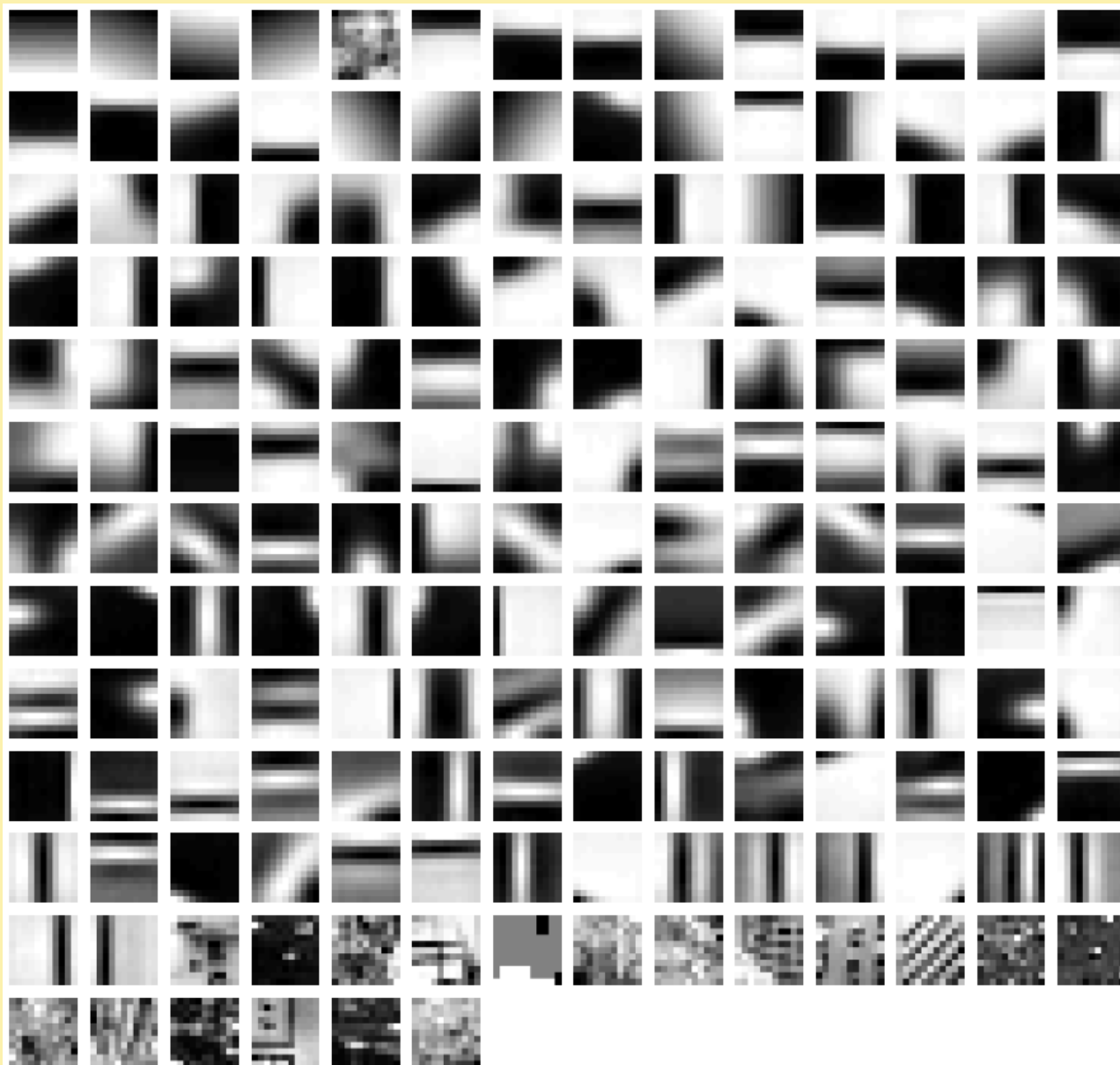
2. Codewords dictionary formation



2. Codewords dictionary formation

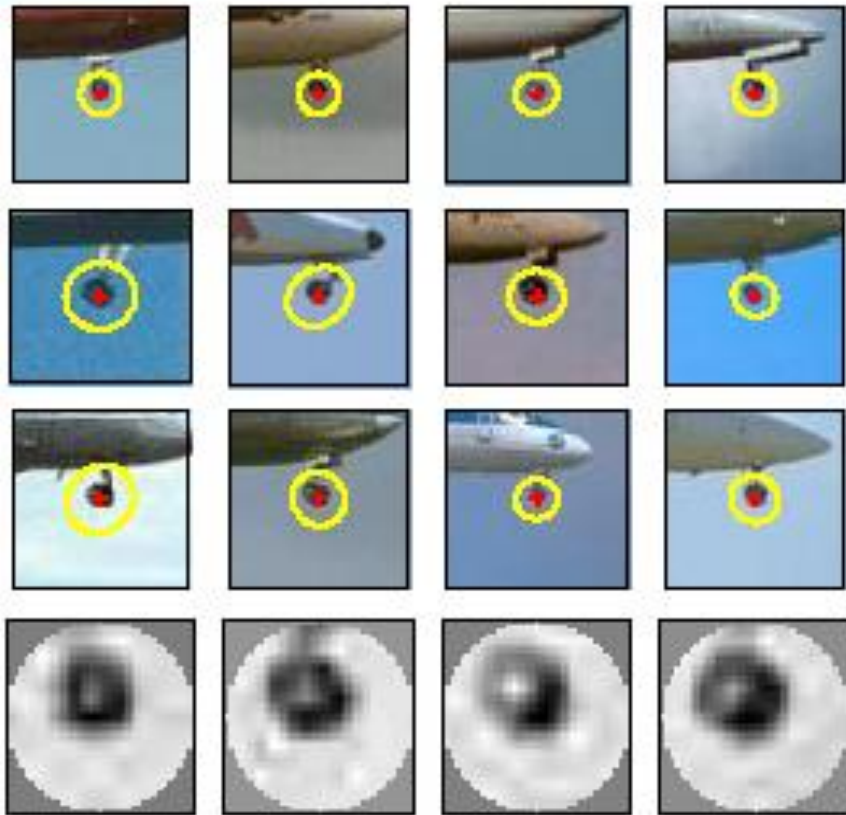


2. Codewords dictionary formation



Regular grid

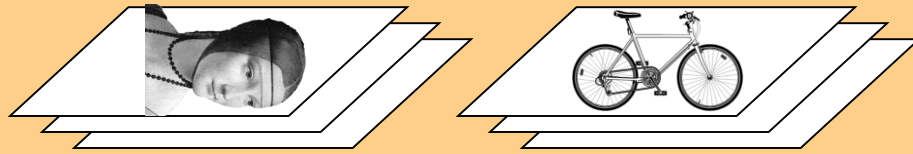
Image patch examples of codewords



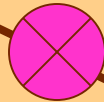
3. Image representation



Representation



1. feature detection
& representation



2. **codewords dictionary**

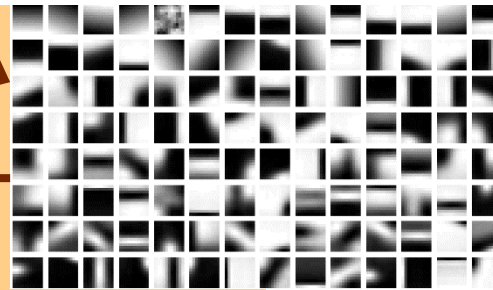
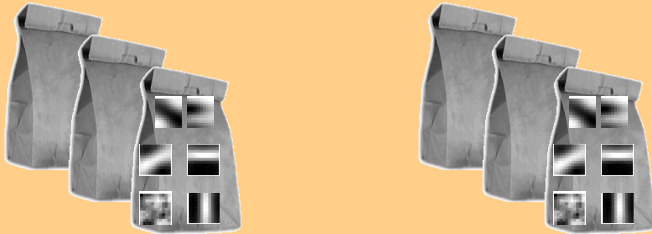
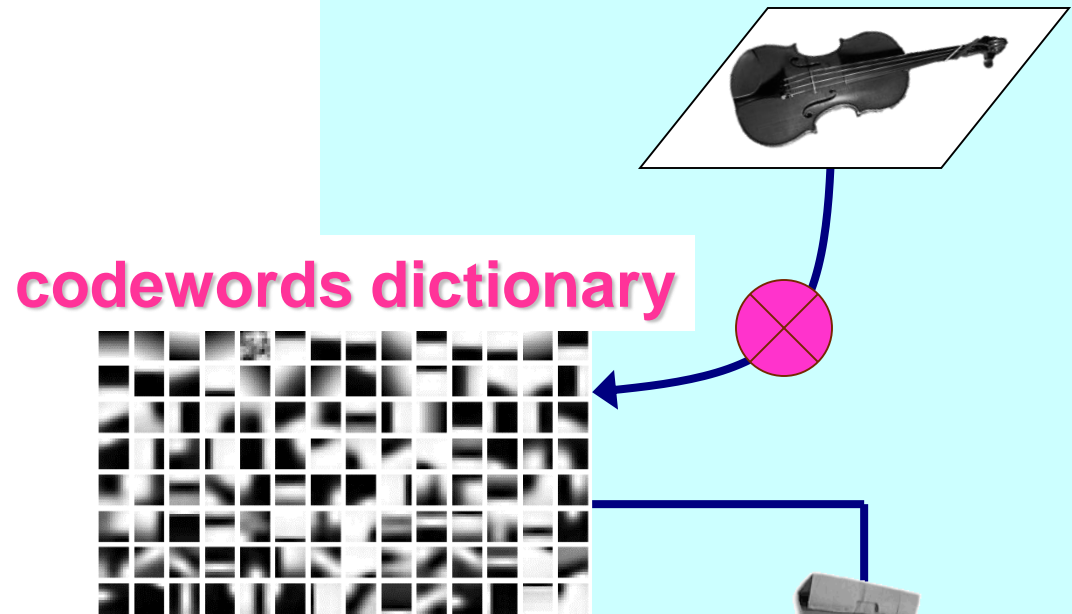


image representation

3.



Learning and Recognition



**category models
(and/or) classifiers**

**category
decision**

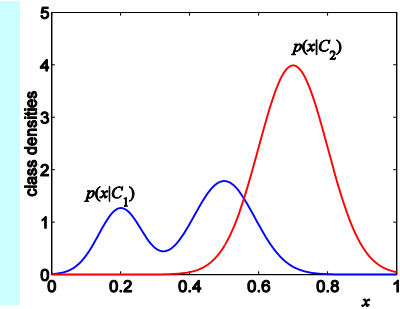
BoW-based Object Categorization



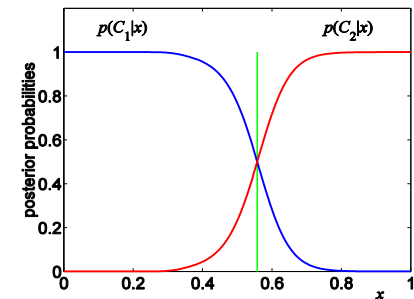
- Basic steps
 - Feature extraction and representation
 - Building codebook (codewords dictionary) from training samples with clustering
 - Represent an image with histogram of codebook (i.e. Bag-of-words of an image)
 - Classify an unknown image with its BoW.

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


A demonstration of bag-of-words classifiers - Microsoft Internet Explorer provided by Insight Broadband

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Refresh Mail Print Send To Favorites

Address <http://people.csail.mit.edu/fergus/iccv2005/bagwords.html>

Google Search 100 blocked Check AutoLink AutoF



Two bag-of-words classifiers

ICCV 2005 short courses on
[Recognizing and Learning Object Categories](#)

A simple approach to classifying images is to treat them as a collection of regions, describing only their appearance and ignoring their location. This approach has been successfully used in the text community for analyzing documents and are known as "bag-of-words" models, since each document is represented by a distribution over fixed vocabulary(s). Using such a representation, methods such as probabilistic latent semantic analysis (pLSA) [1] (LDA) [2] are able to extract coherent topics within document collections in an unsupervised manner.

Recently, Fei-Fei et al. [3] and Sivic et al. [4] have applied such methods to the visual domain. The demo code implements pLSA, including a Naïve Bayes classifier. For comparison, a Naïve Bayes classifier is also provided which requires labelled training data, unlike pLSA.

The code consists of Matlab scripts (which should run under both Windows and Linux) and a couple of 32-bit Linux binaries for doing matrix operations. Hence the whole system will need to be run on Linux. The code is for teaching/research purposes only. If you find a bug, please report it to fergus@csail.mit.edu where csail point mit point edu.

Download

[Download](#) the code and datasets (32 Mbytes)

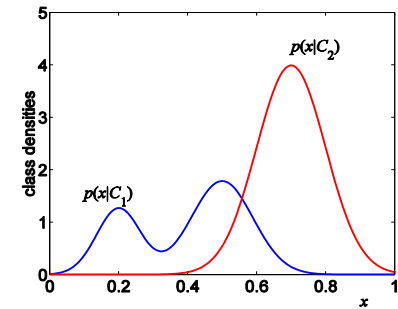
Operation of code

To run the demos:

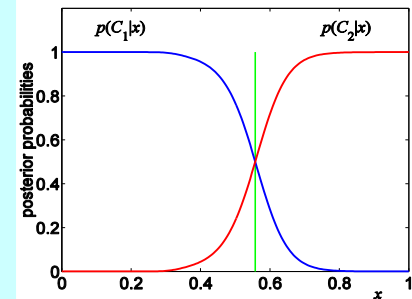
start Microsoft Outlook We... 未名空间(mitbbs.co... A demonstration of b... ICCV2005

Learning and Recognition

1. Generative method:
 - graphical models

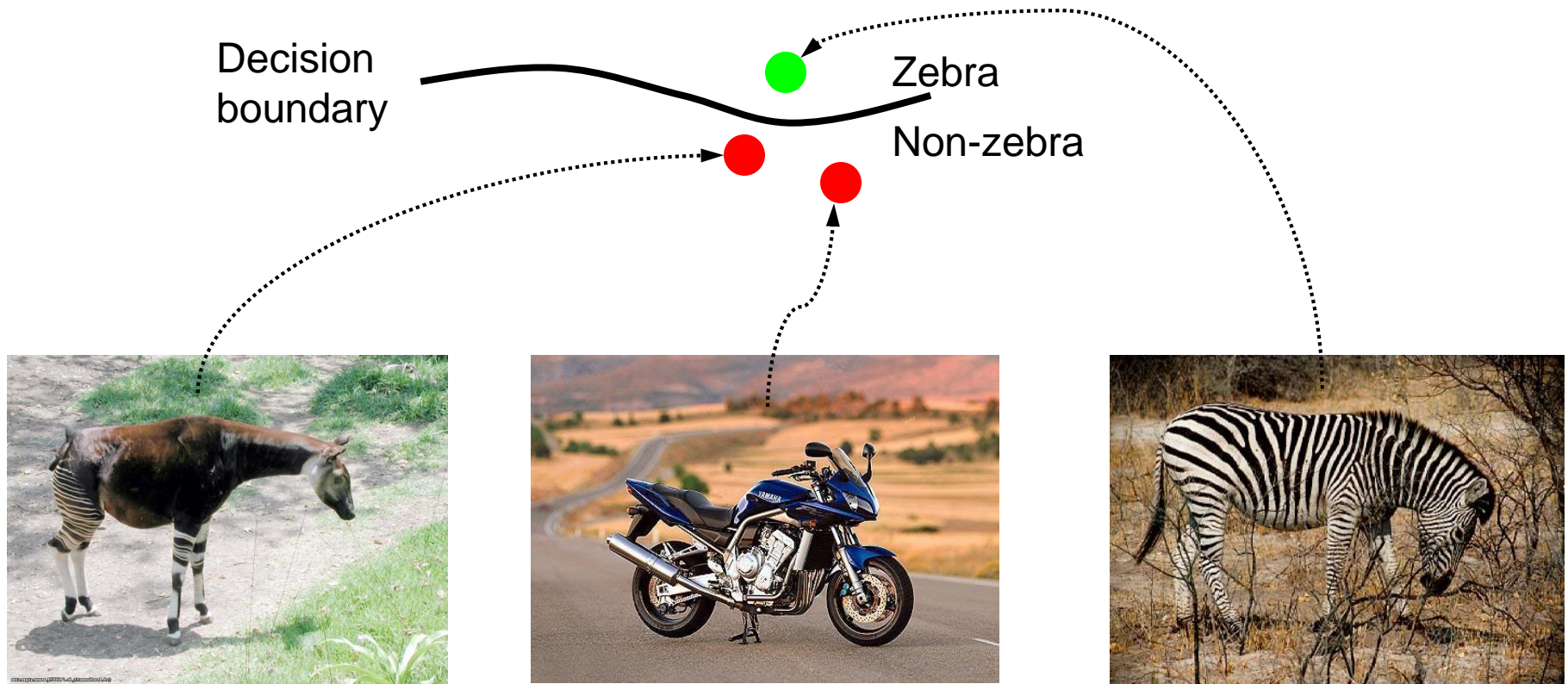


2. Discriminative method:
 - SVM



**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)

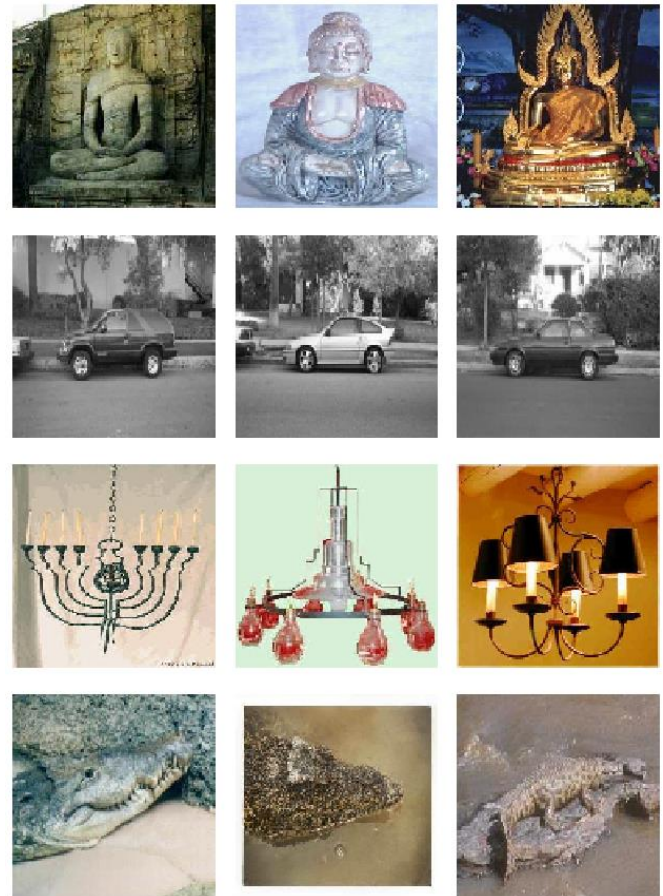
- Features:
 - Harris detector
 - PCA-SIFT descriptor, $d=10$



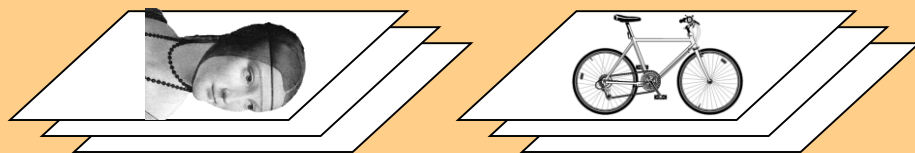
Kernel	Complexity	Recognition rate
Match [Wallraven et al.]	$O(dm^2)$	84%
Bhattacharyya affinity [Kondor & Jebara]	$O(dm^3)$	85%
Pyramid match	$O(dmL)$	84%

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



learning



feature detection
& representation

codewords dictionary

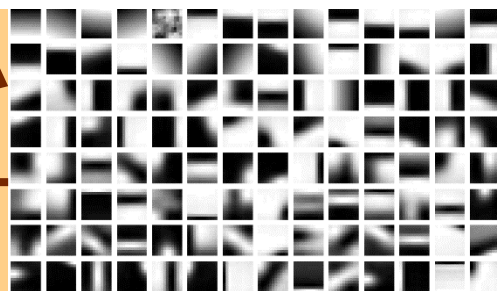
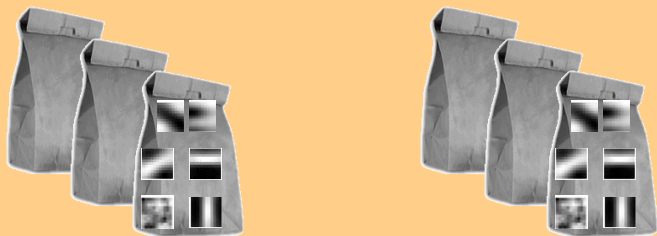
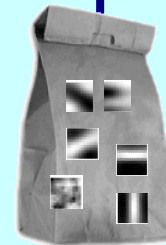
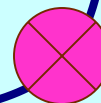


image representation



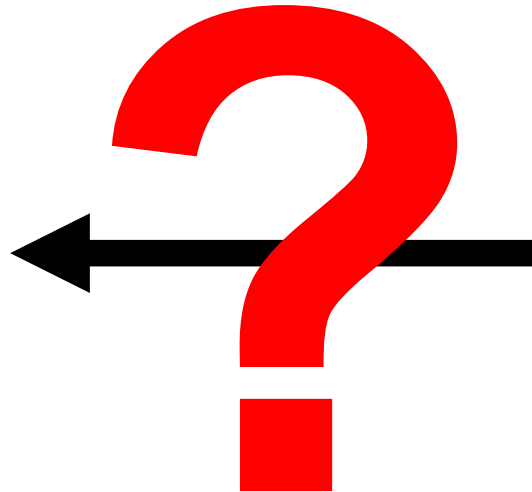
**category models
(and/or) classifiers**

recognition



**category
decision**

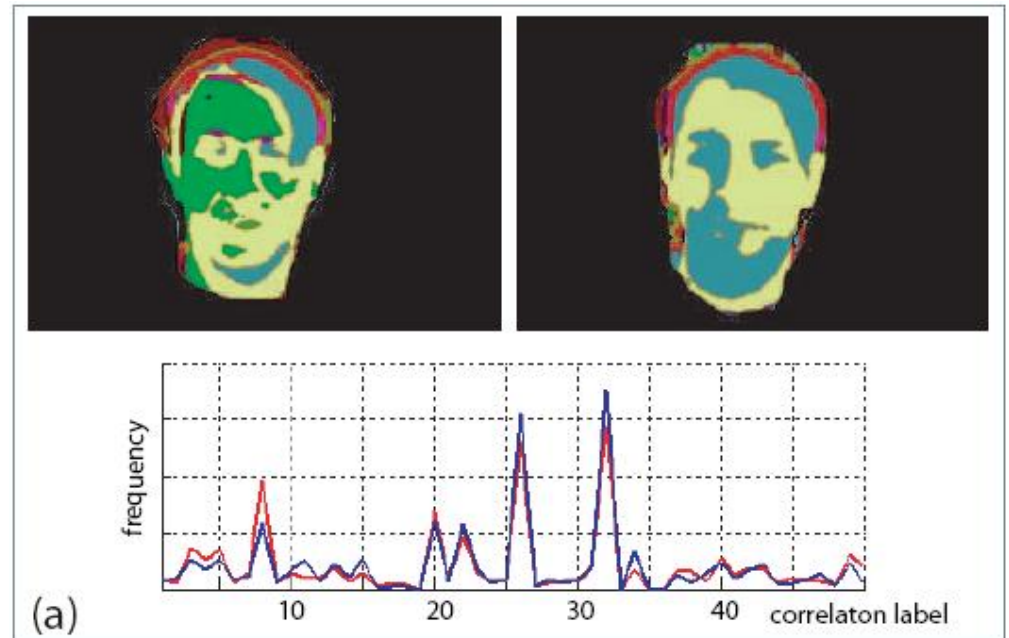
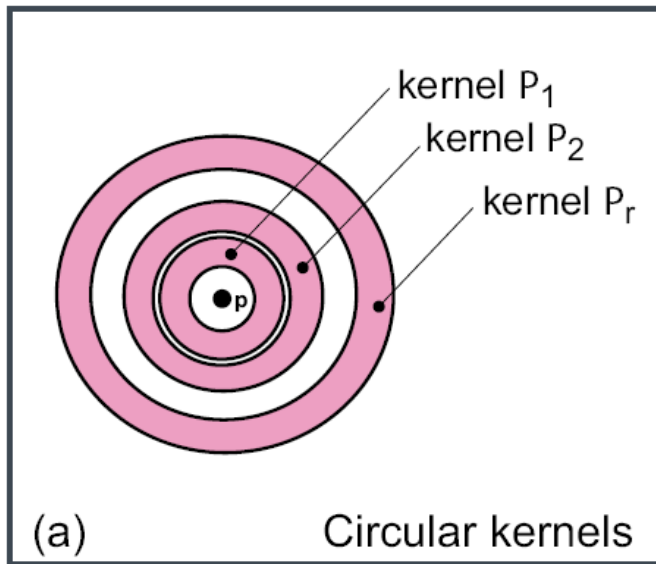
What about spatial info?



What about spatial info?



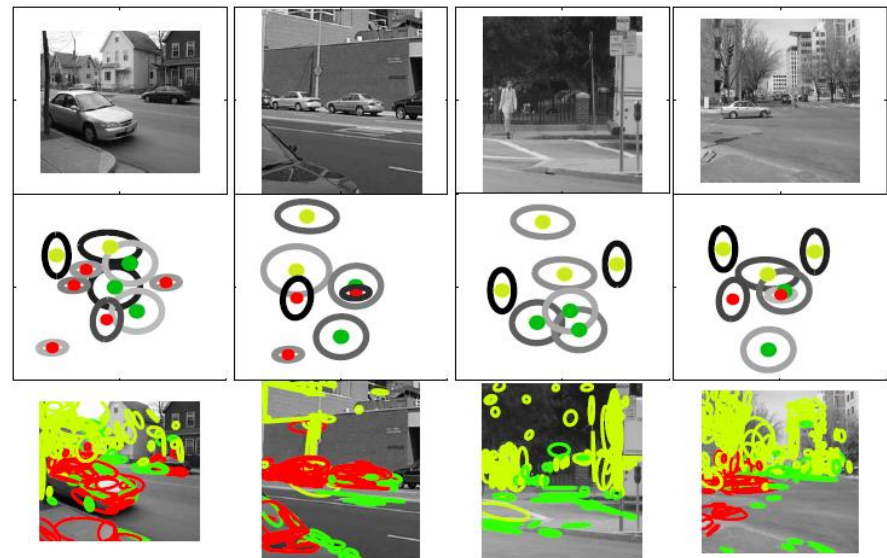
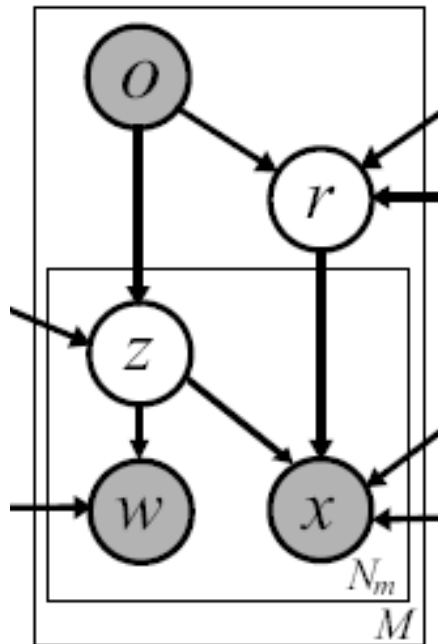
- Feature level
 - Spatial influence through correlogram features: Savarese, Winn and Criminisi, CVPR 2006



What about spatial info?



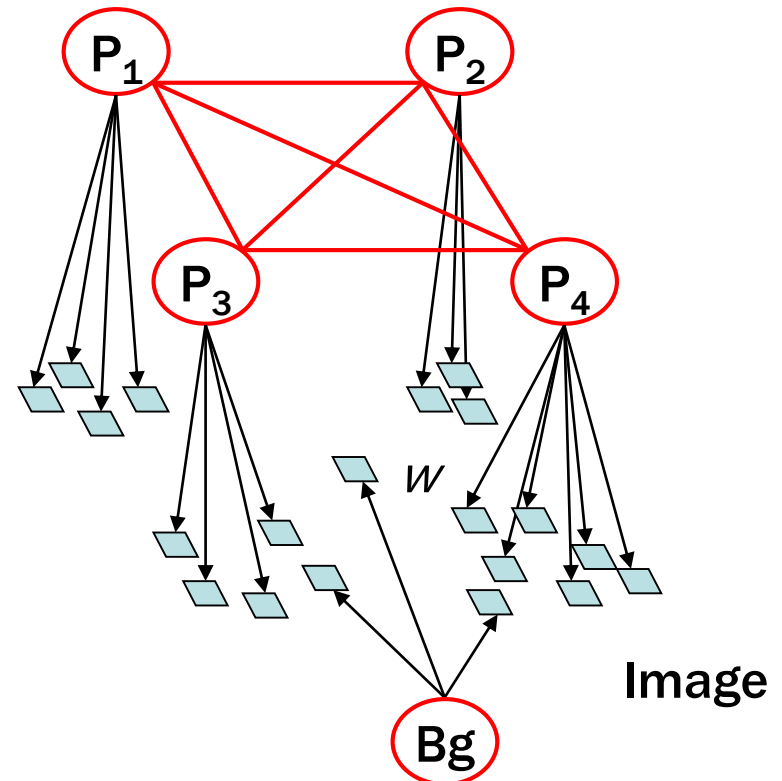
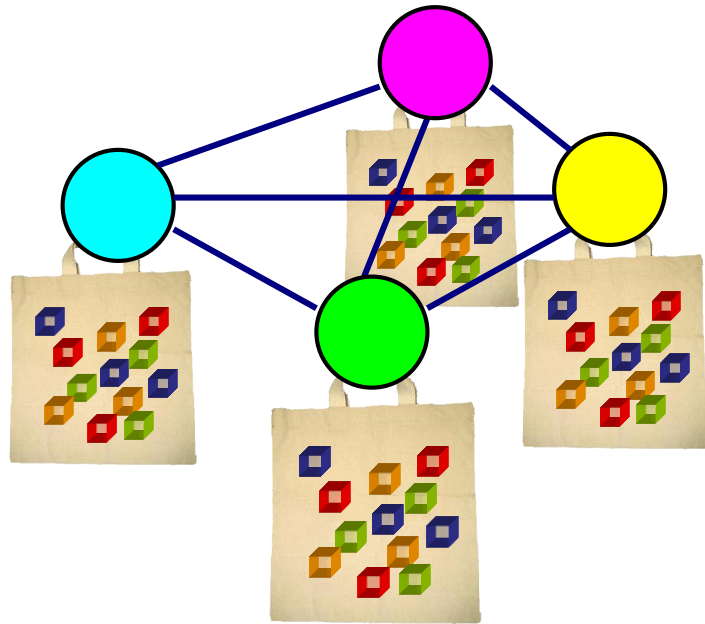
- Feature level
- Generative models
 - Sudderth, Torralba, Freeman & Willsky, 2005, 2006
 - Niebles & Fei-Fei, CVPR 2007



What about spatial info?



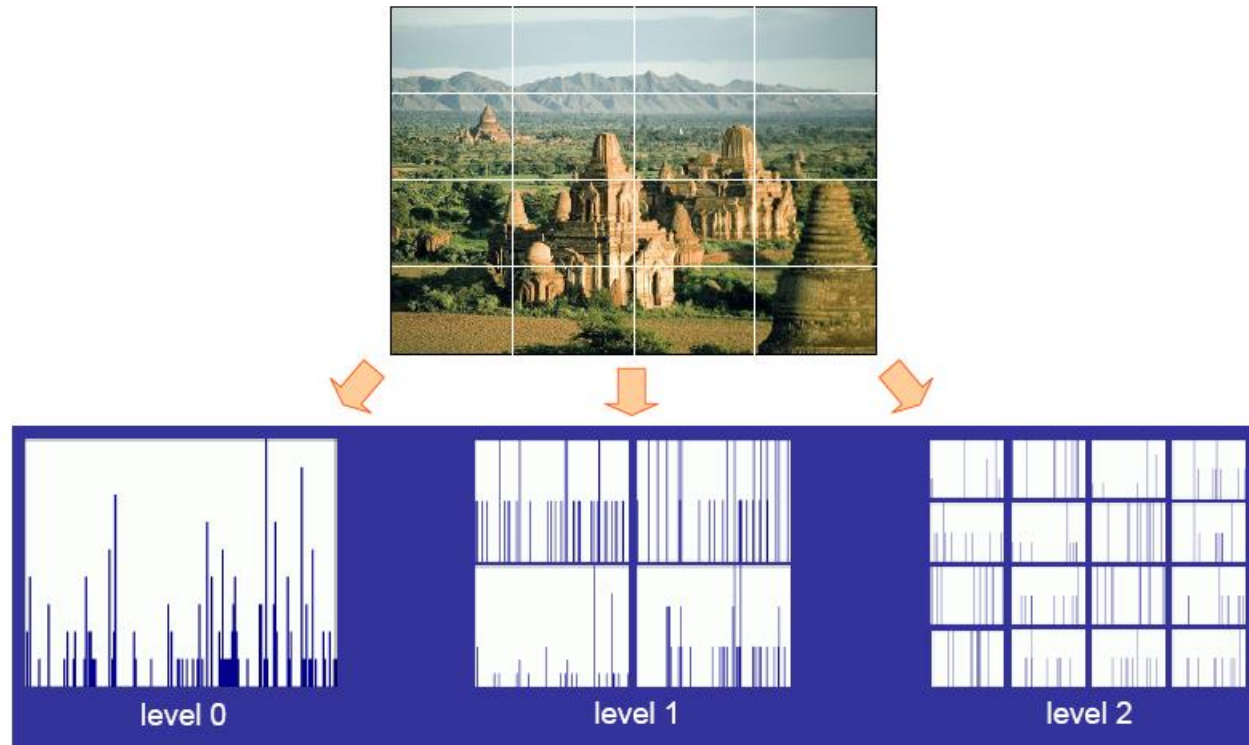
- Feature level
- Generative models
 - Sudderth, Torralba, Freeman & Willsky, 2005, 2006
 - Niebles & Fei-Fei, CVPR 2007



What about spatial info?

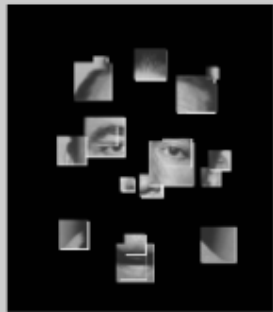


- Feature level
- Generative models
- Discriminative methods
 - Lazebnik, Schmid & Ponce, 2006



Invariance issues

- Scale and rotation
 - Implicit
 - Detectors and descriptors



Invariance issues

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



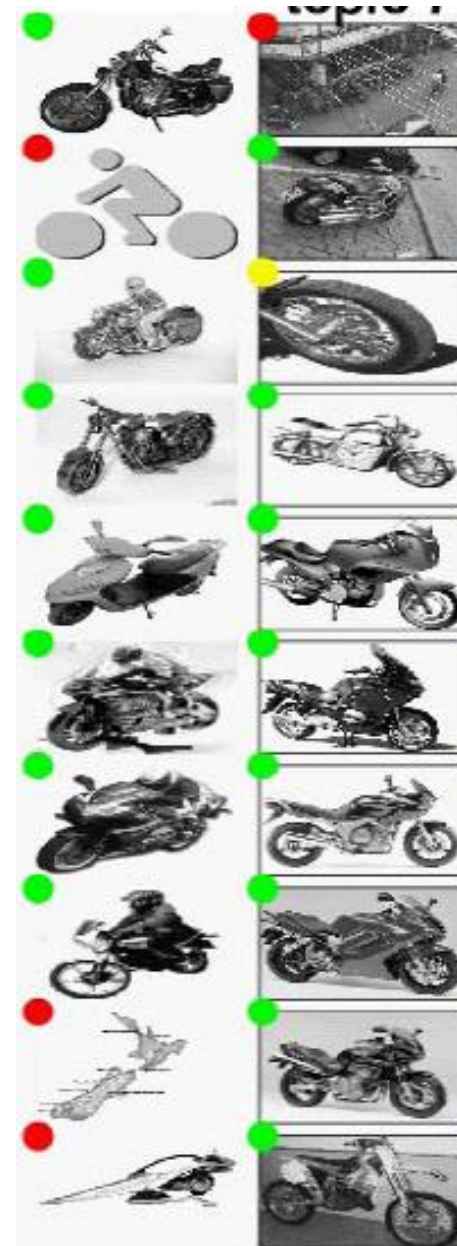
Invariance issues

- 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





Model properties

- Intuitive
 - Analogy to documents

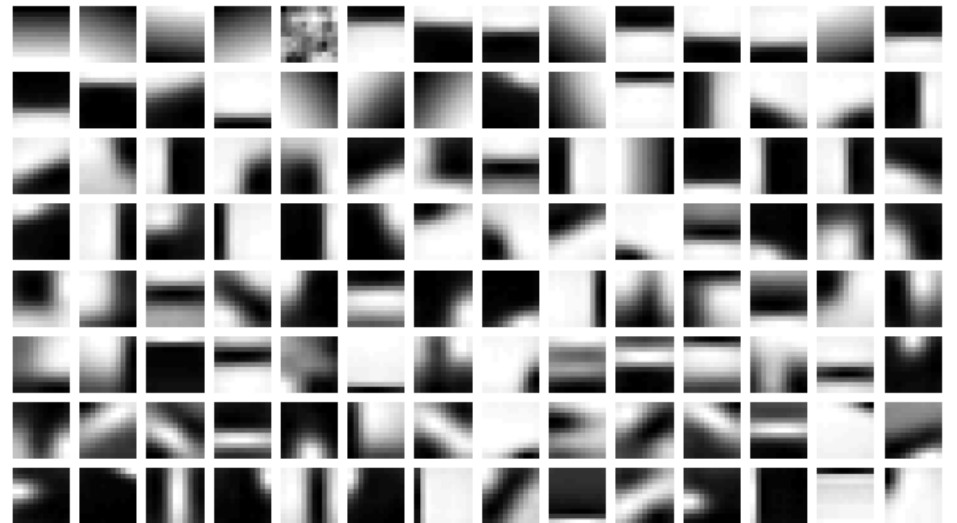
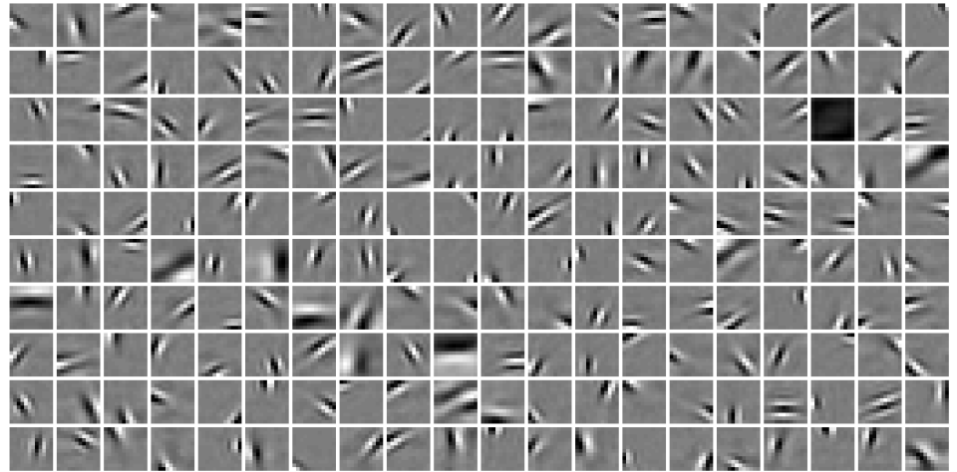
Of all the sensory impressions proceeding to the brain, the visual experiences are the dominant ones. Our perception of the world around us is based essentially on the messages that reach our eyes. For a long time, the retinal image was considered as a movie screen. The visual centers in the brain were discovered. We now know that perception is a more complex process. Following the discovery of the path to the various centers of the cortex, Hubel and Wiesel have demonstrated that the *message about the image falling on the retina undergoes a point-by-point analysis in a system of nerve cells stored in columns. In this system each cell has its specific function and is responsible for a specific detail in the pattern of the retinal image.*

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

Model properties

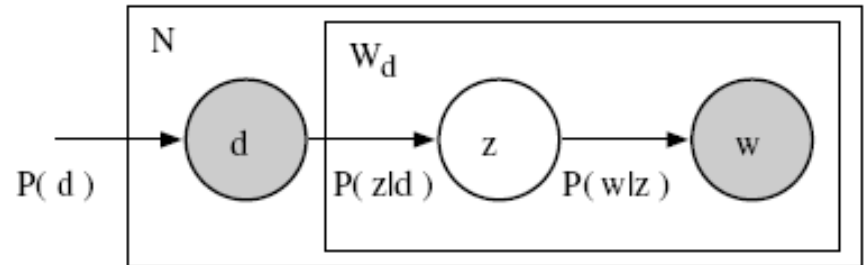


- Intuitive
 - Analogy to documents
 - Analogy to human vision



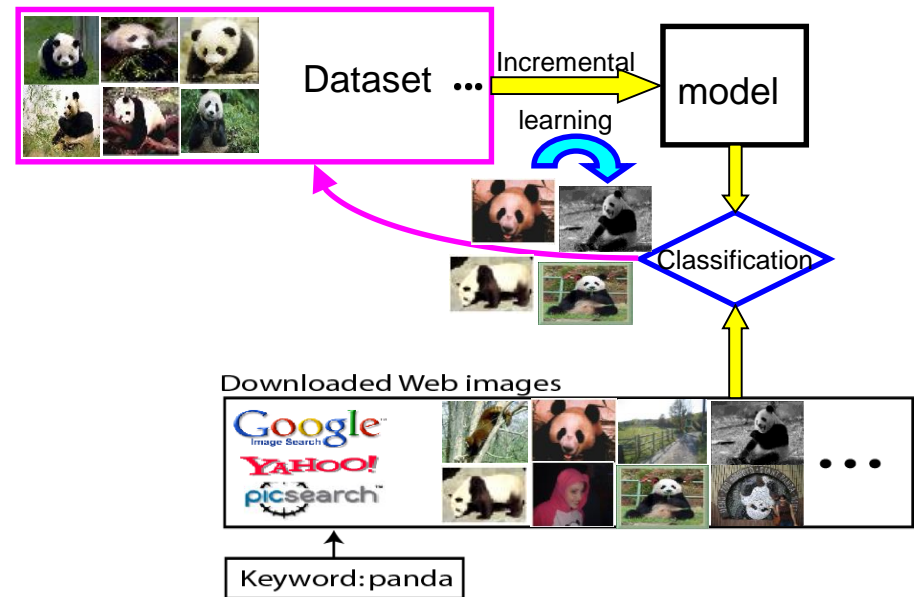


Model properties



Sivic, Russell, Efros, Freeman, Zisserman, 2005

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

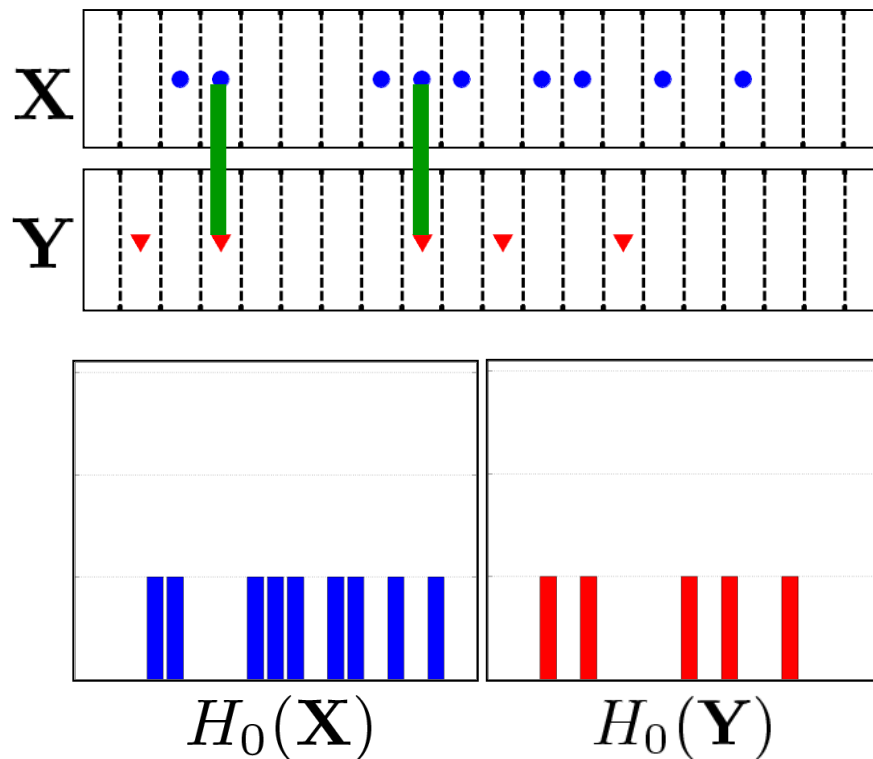


Li, Wang & Fei-Fei, CVPR 2007

Model properties



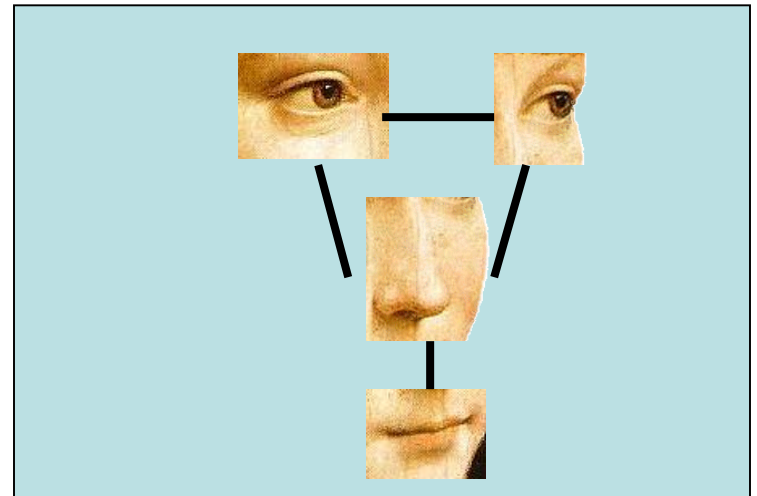
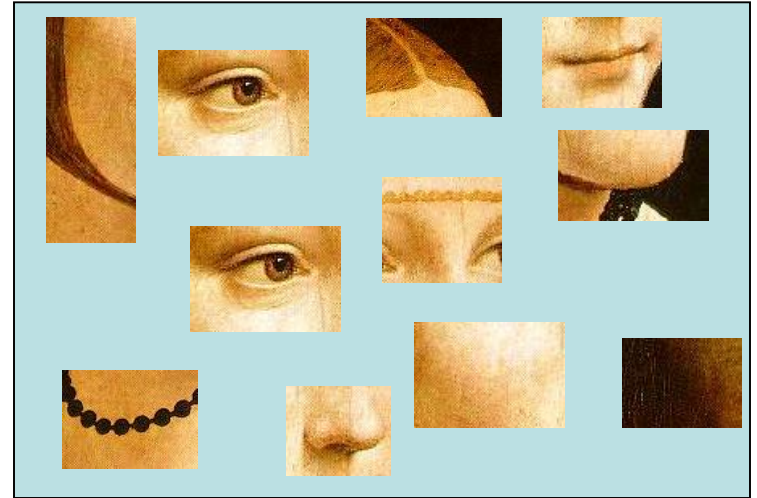
- Intuitive
- generative models
- Discriminative method
 - Computationally efficient



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