Bag-of-features models

Many slides adapted from Fei-Fei Li, Rob Fergus, and Antonio Torralba

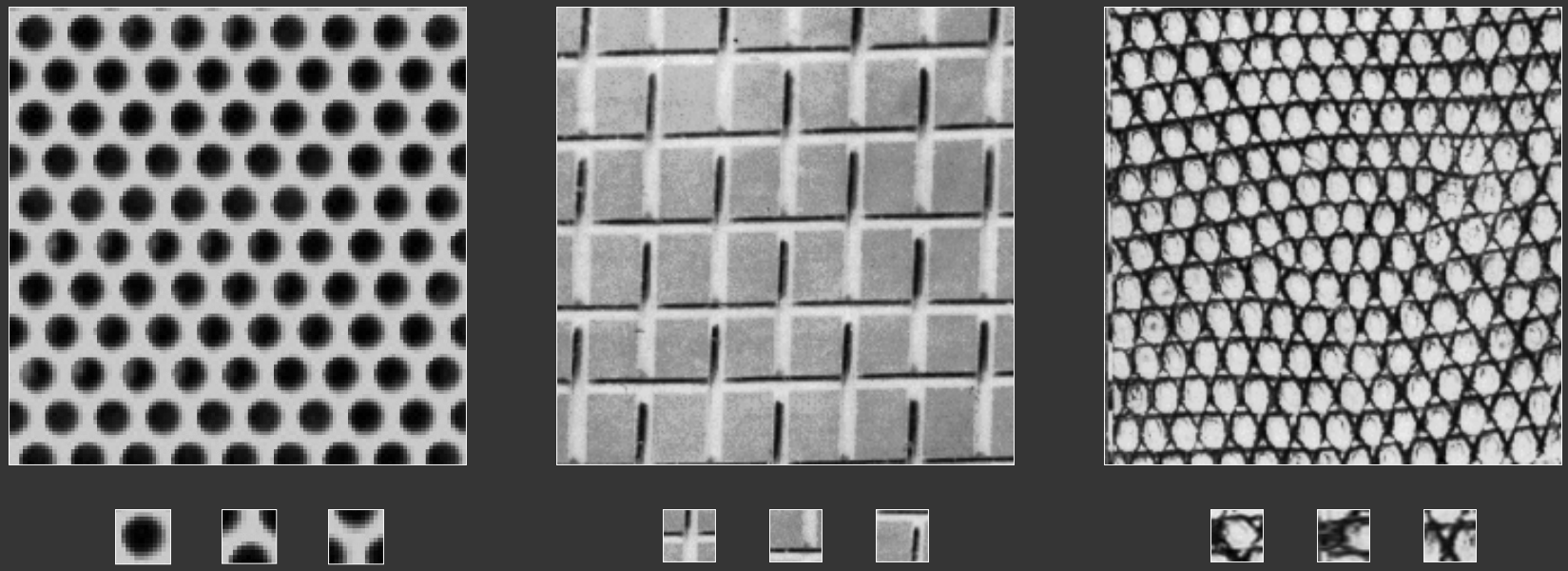
Origins

1. Texton frequency analysis for texture recognition

2. Bag-of-words models in document analysis

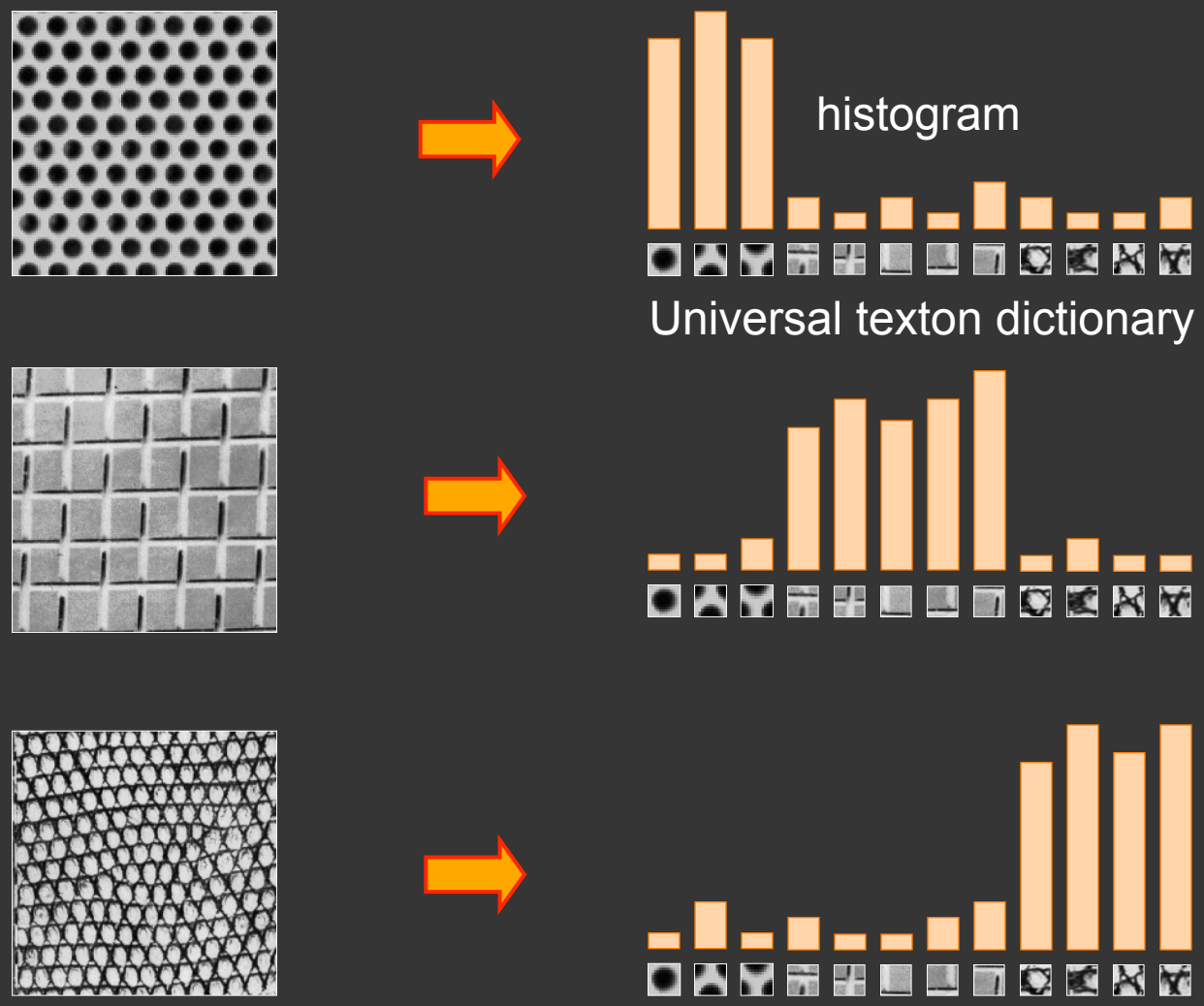
Origin 1: Texture recognition

- Texture is characterized by the repetition of basic elements or *textons*
- Recognition based on identity of the textons, not their spatial arrangement (although that is very important too!)



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

Origin 1: Texture recognition



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

Origin 2: Bag-of-words models

- Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)

Origin 2: Bag-of-words models

- Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)

US Presidential Speeches Tag Cloud
http://chir.ag/phernalia/preztags/

Origin 2: Bag-of-words models

- Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)

US Presidential Speeches Tag Cloud
`http://chir.ag/phernalia/preztags/`

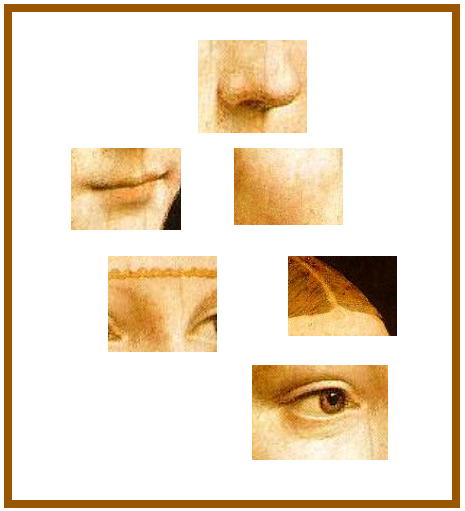
Origin 2: Bag-of-words models

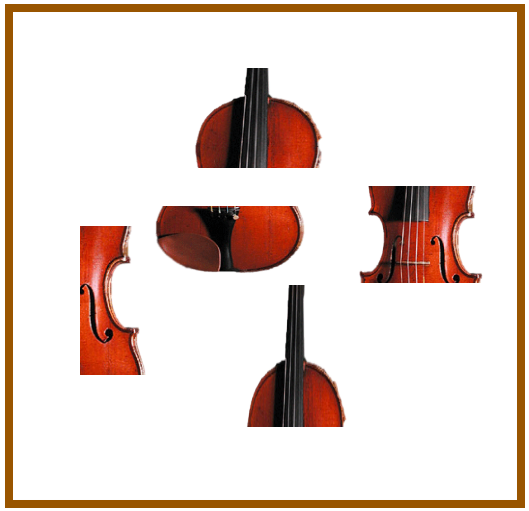
- Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)

US Presidential Speeches Tag Cloud
`http://chir.ag/phernalia/preztags/`

Bag of features: overview

1. Extract features from a training set of images





Bag of features: overview

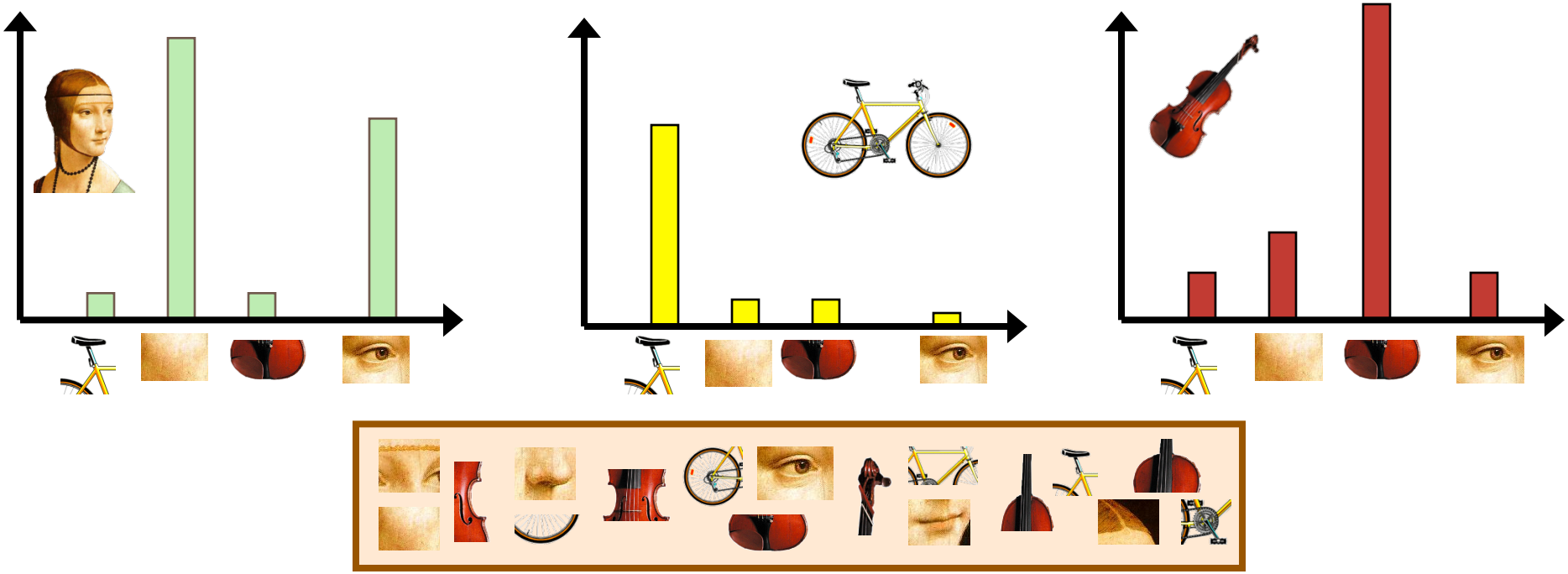
1. Extract features from a training set
2. Learn “visual vocabulary” from these features

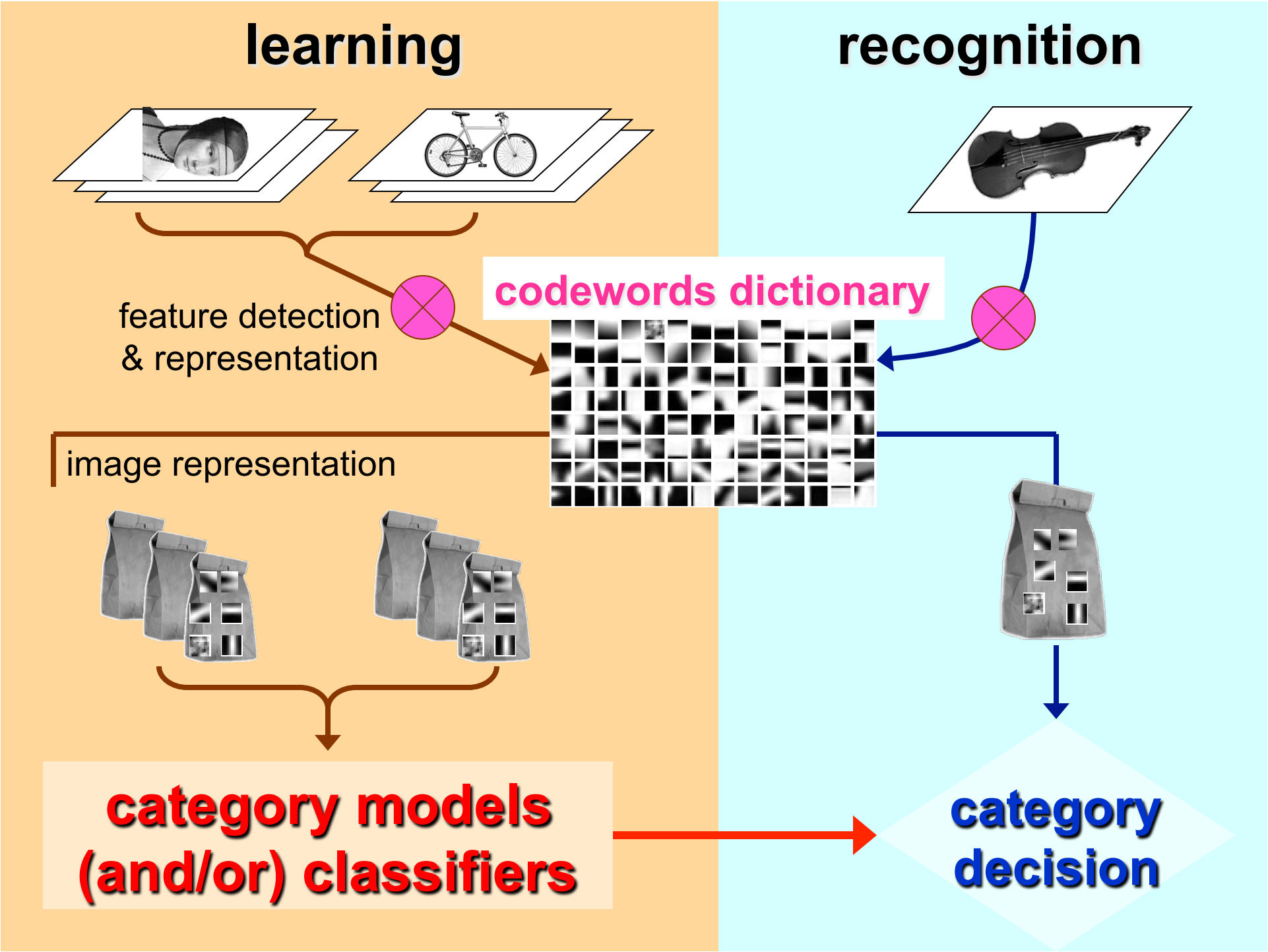
Visual words

This is a clustering process.
Recall our lectures on k-means, EM, and *-shift methods.

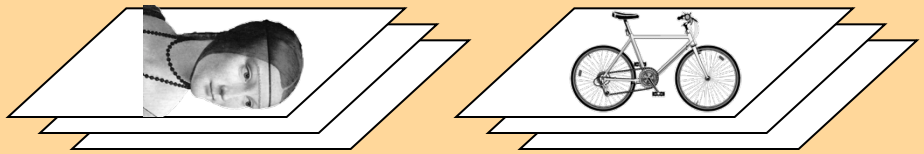
Bag of features: overview

1. Extract features
2. Learn “visual vocabulary” (visual words)
3. Represent images by frequencies of these visual words




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

Representation



1. feature detection & representation

2. codewords dictionary

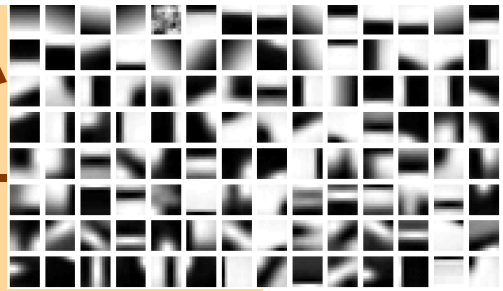
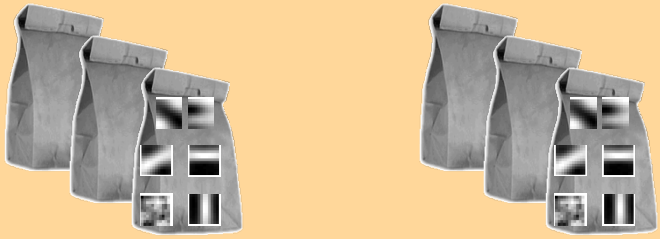
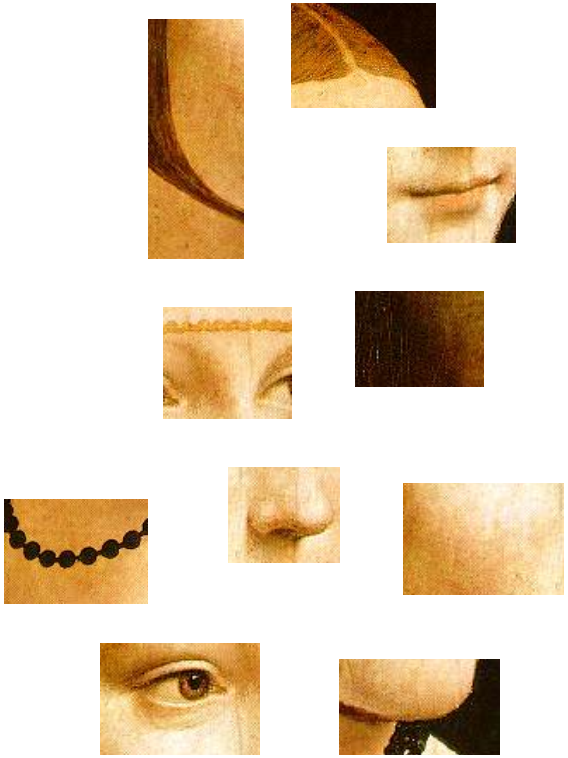


image representation

3.



1.Feature detection and representation



Feature detection

Sliding Window (Dense)

- Leung et al, 1999
- Viola et al, 1999
- Renninger et al 2002

Feature detection

Sliding Window (Dense)

- Leung et al, 1999
- Viola et al, 1999
- Renninger et al 2002

Regular grid

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

Feature detection

Sliding Window (Dense)

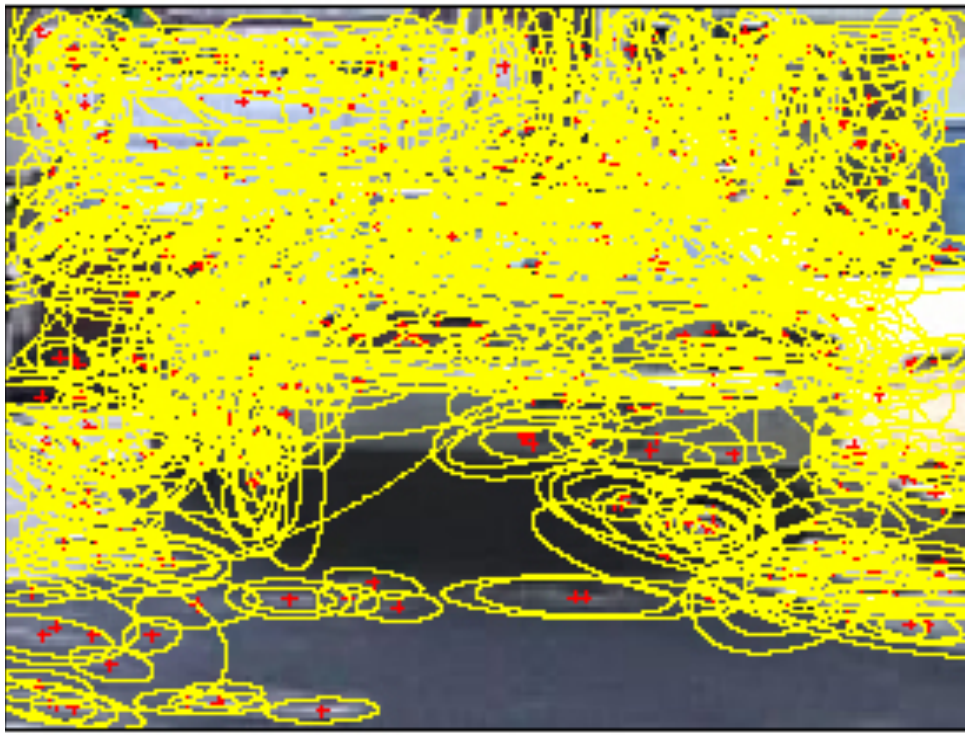
- Leung et al, 1999
- Viola et al, 1999
- Renninger et al 2002

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



Feature detection

Sliding Window (Dense)

- Leung et al, 1999
- Viola et al, 1999
- Renninger et al 2002

Regular grid

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

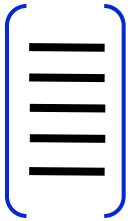
Interest point detector

- Csurka et al. 2004
- Fei-Fei et al. 2005
- Sivic et al. 2005

Other methods

- Random sampling (Nowak, Jurie & Triggs)
- Segmentation based patches
  - Barnard et al. 2003, Russell et al 2006, etc.)

Interest Point Descriptors

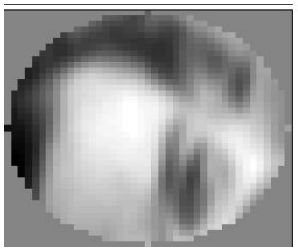


Compute SIFT descriptor

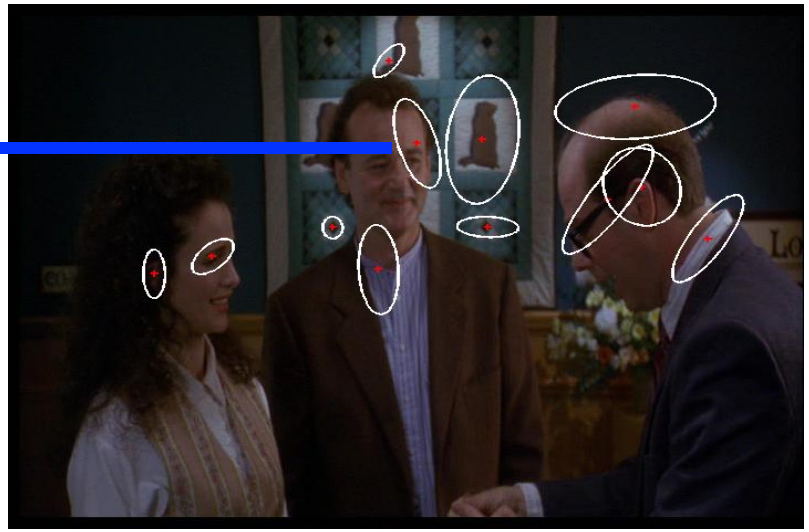
[Lowe'99]

Or HoG

[Dalaal&Triggs]



Normalize patch



Detect patches

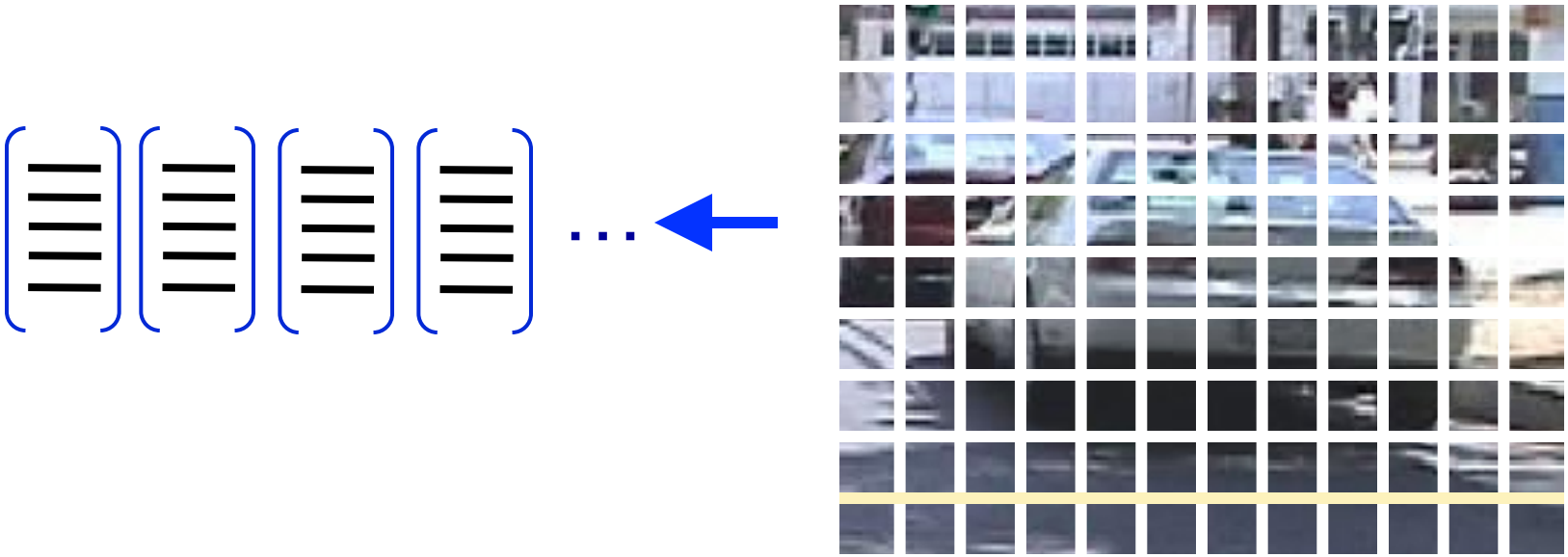
[Mikojaczyk and Schmid '02]

[Matas et al. '02]

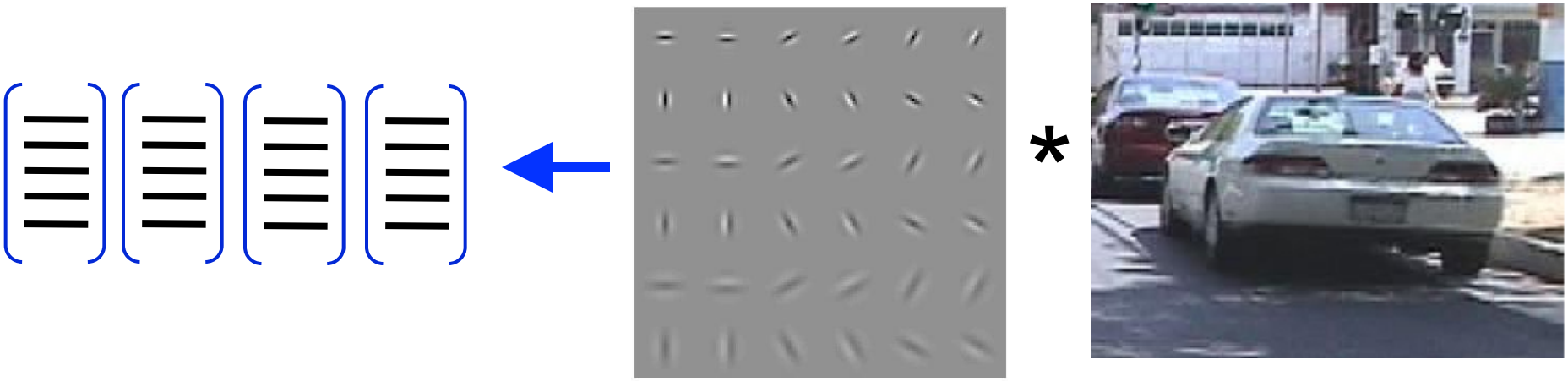
[Sivic et al. '03]

Slide credit: Josef Sivic

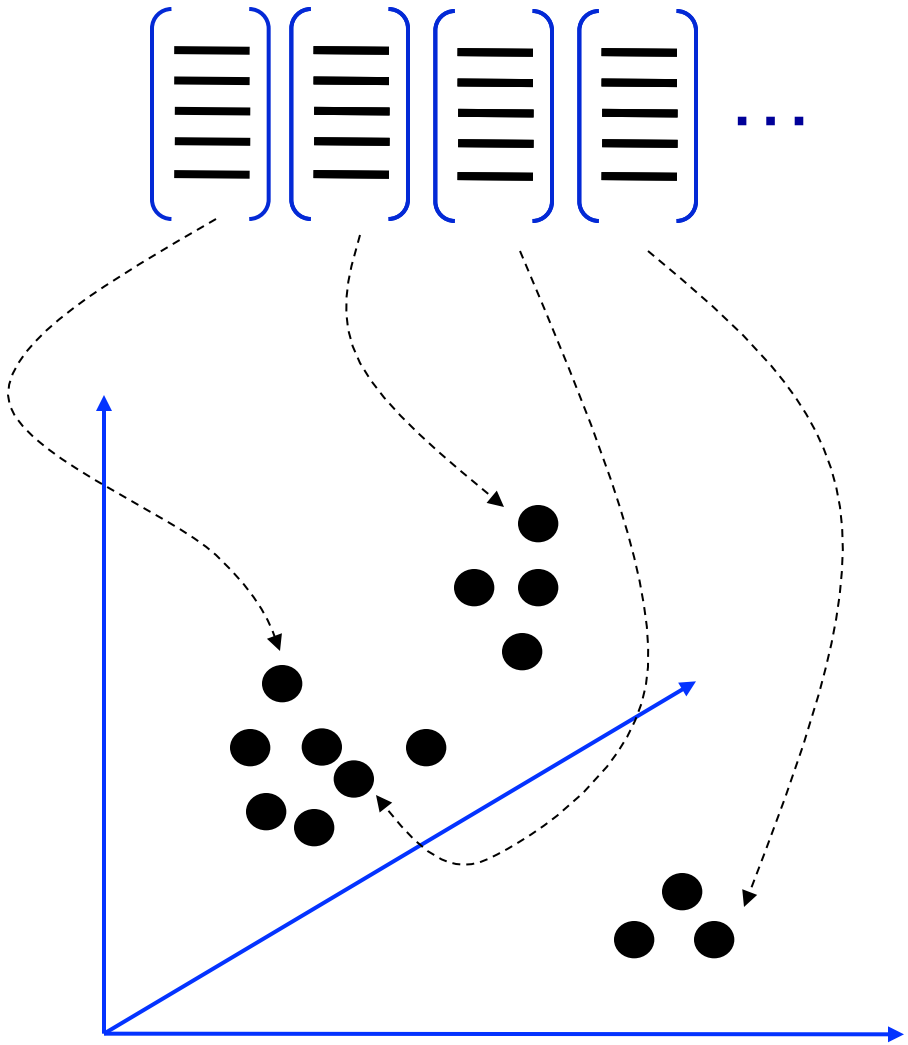
Patch-based Descriptors



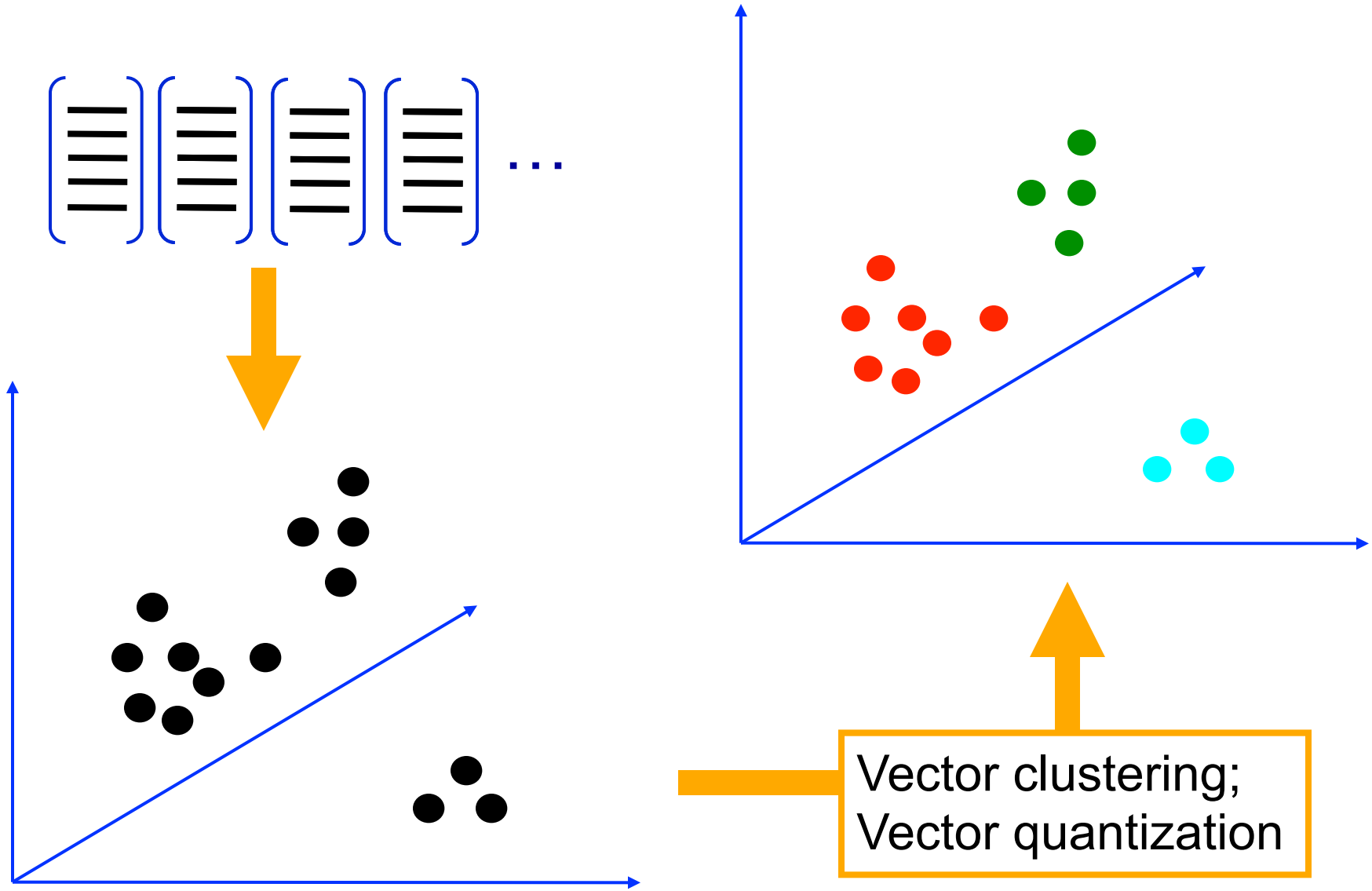
Filterbank Response Descriptors



2. Codewords dictionary formation

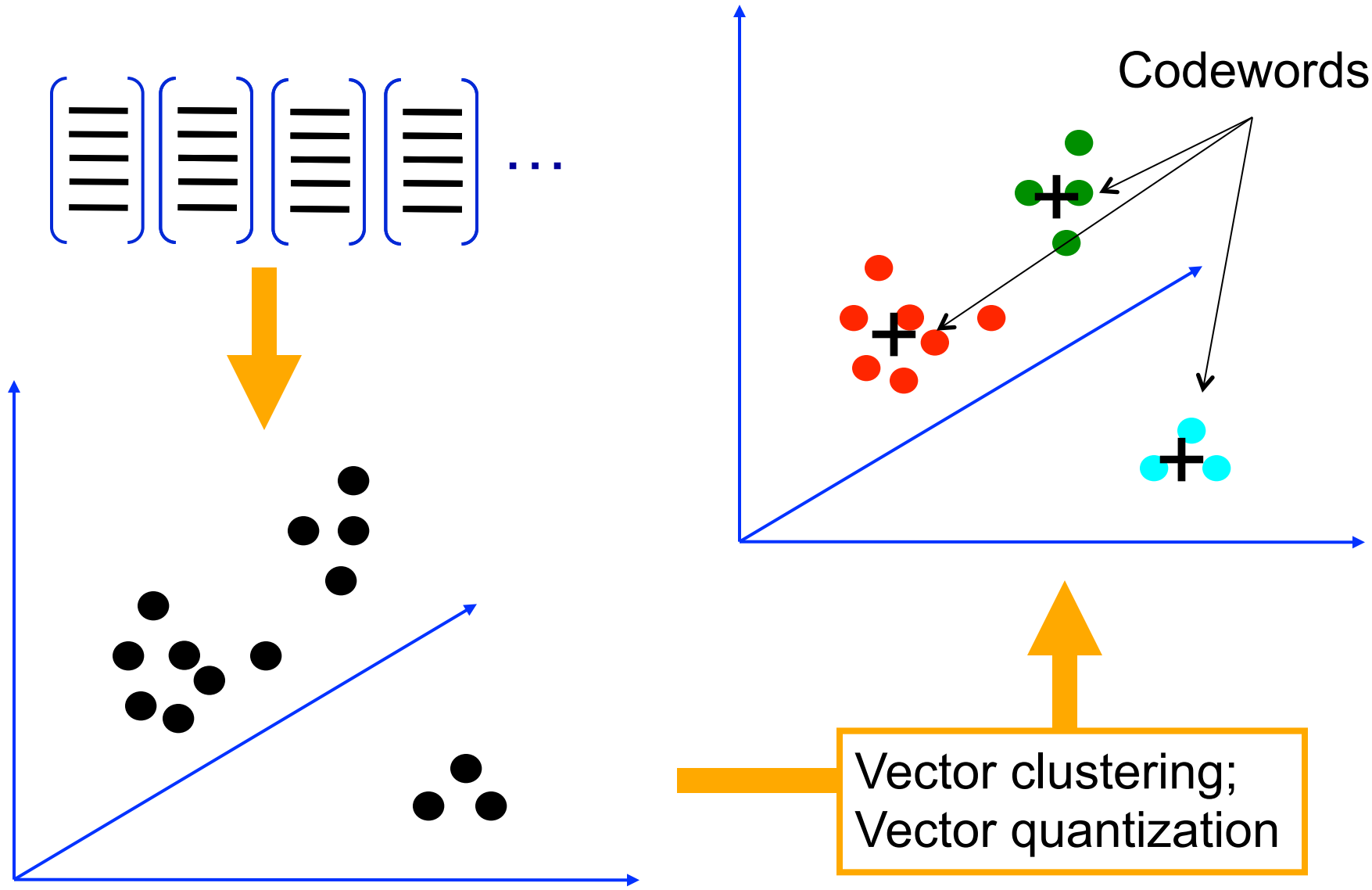


2. Codewords dictionary formation



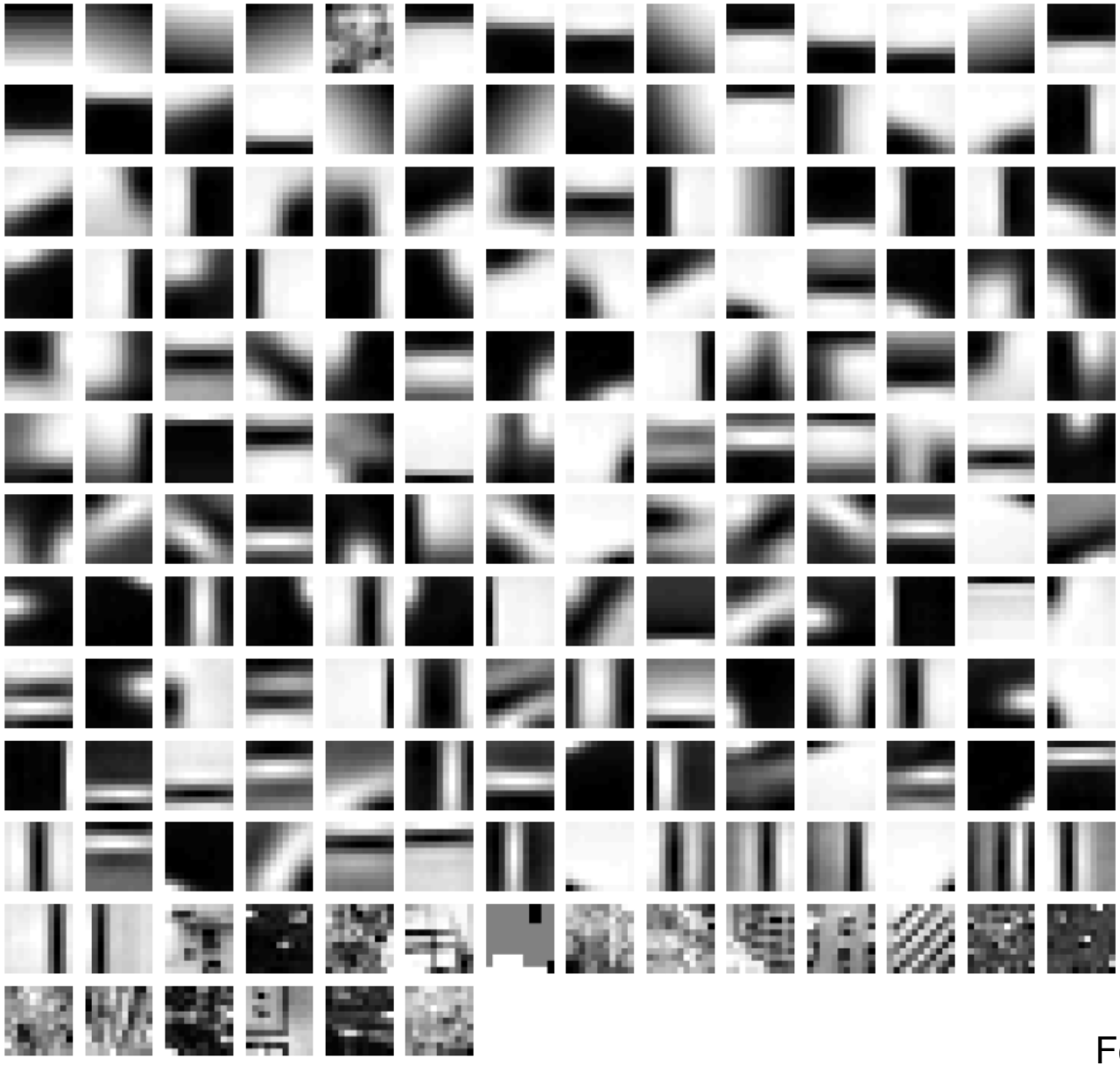
Slide credit: Josef Sivic

2. Codewords dictionary formation



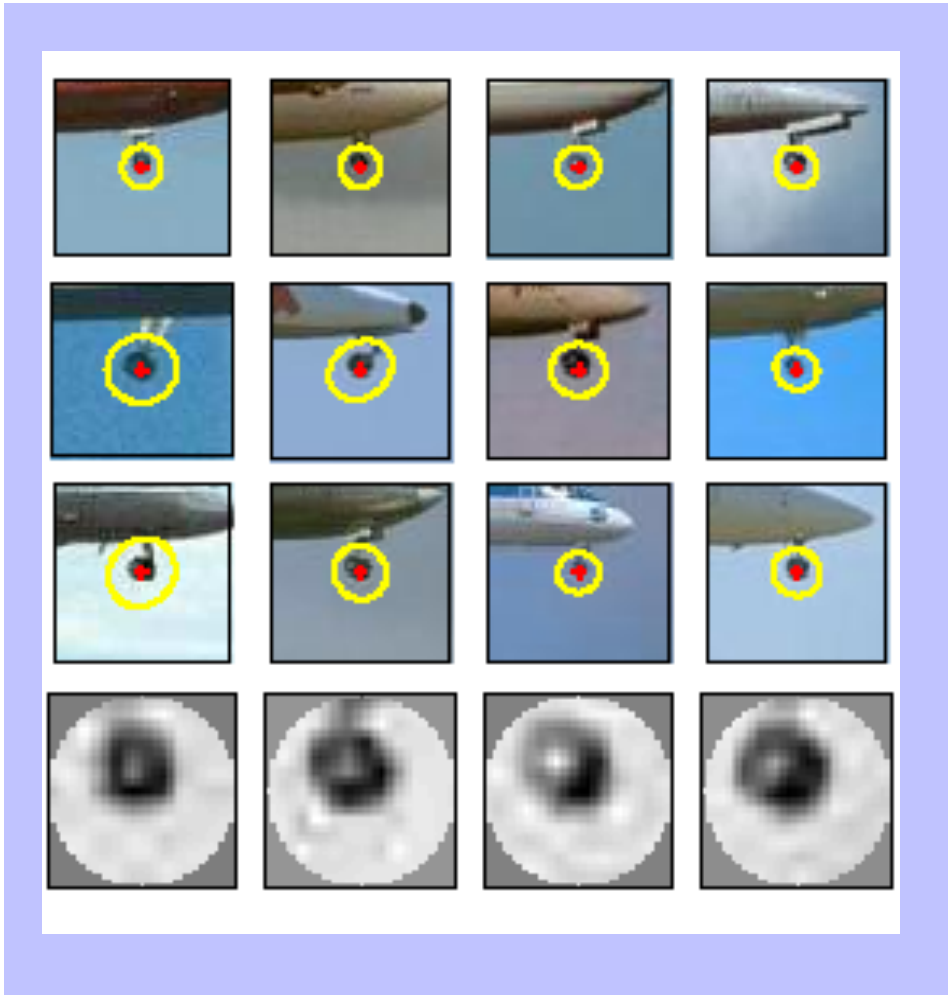
Slide credit: Josef Sivic

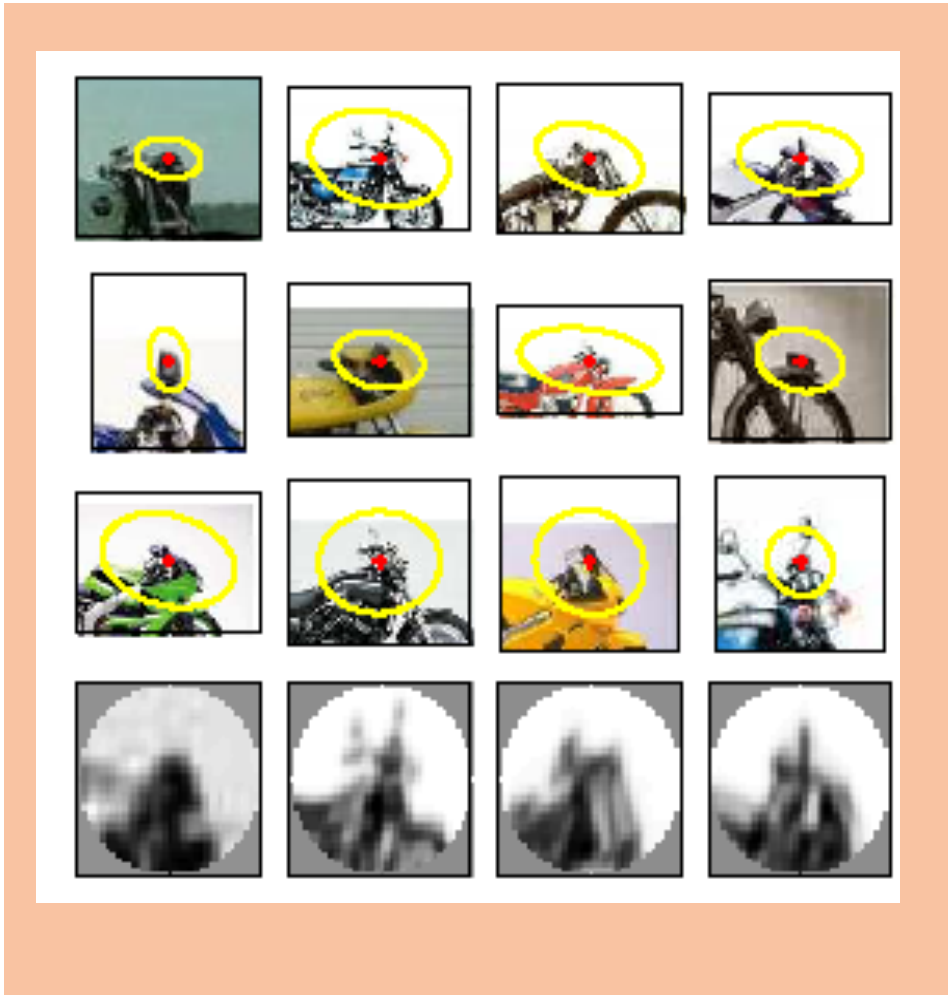
2. Codewords dictionary formation



Fei-Fei et al. 2005

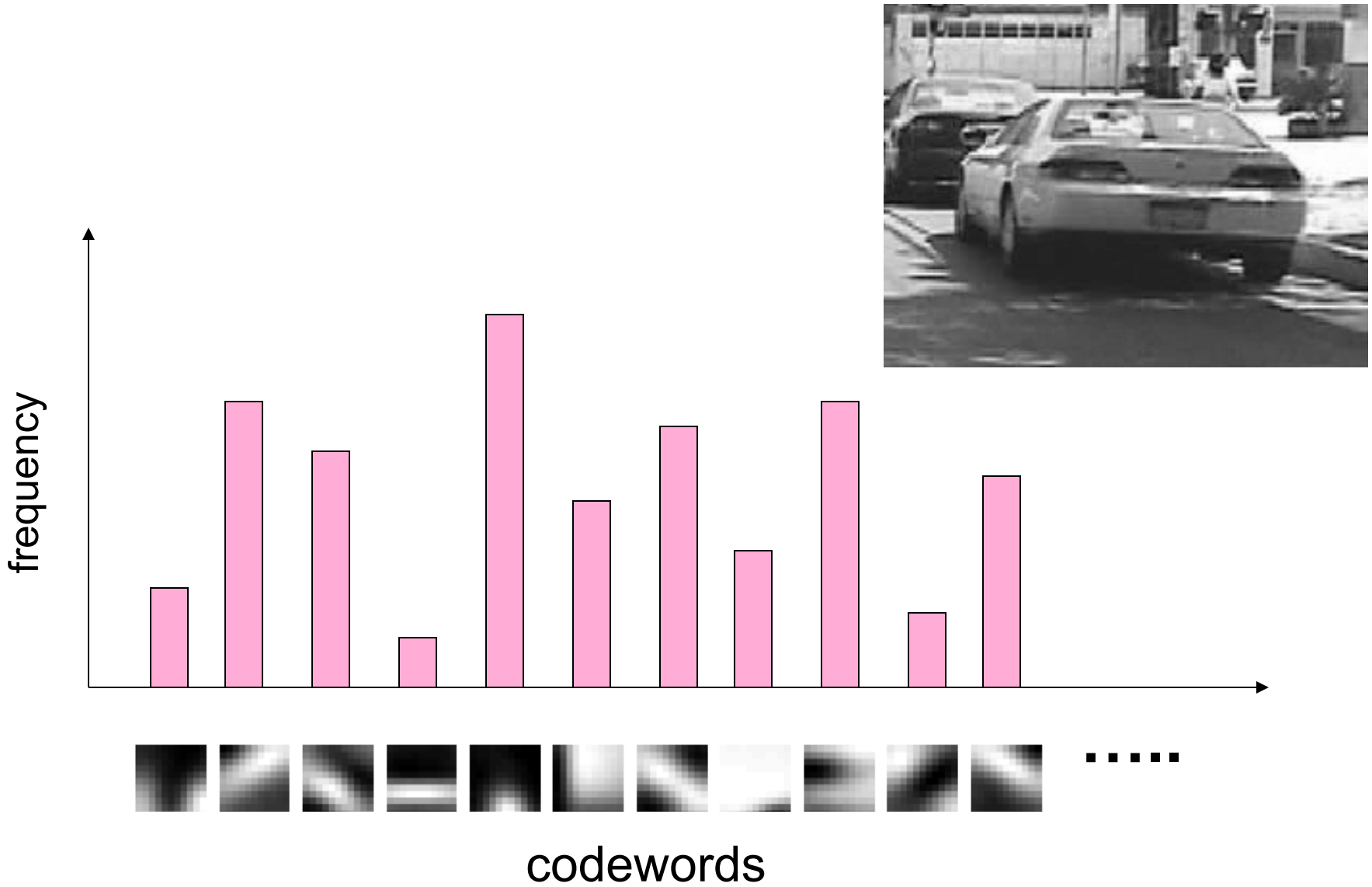
Image patch examples of codewords



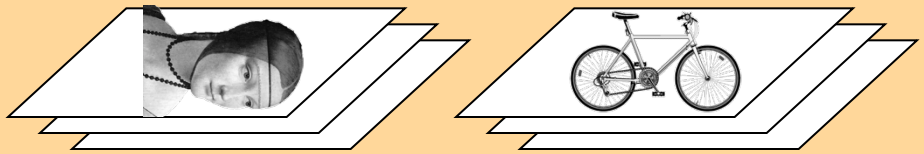


Sivic et al. 2005

3. Image representation as BoW



Representation



1. feature detection & representation

2. codewords dictionary

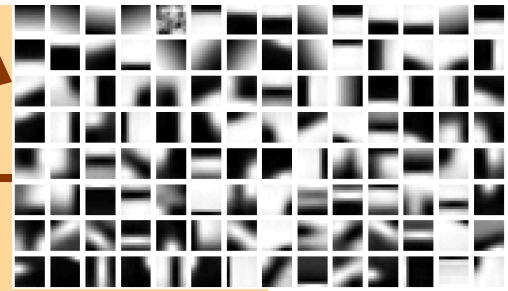
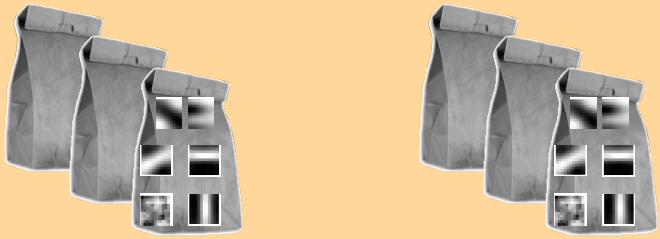


image representation

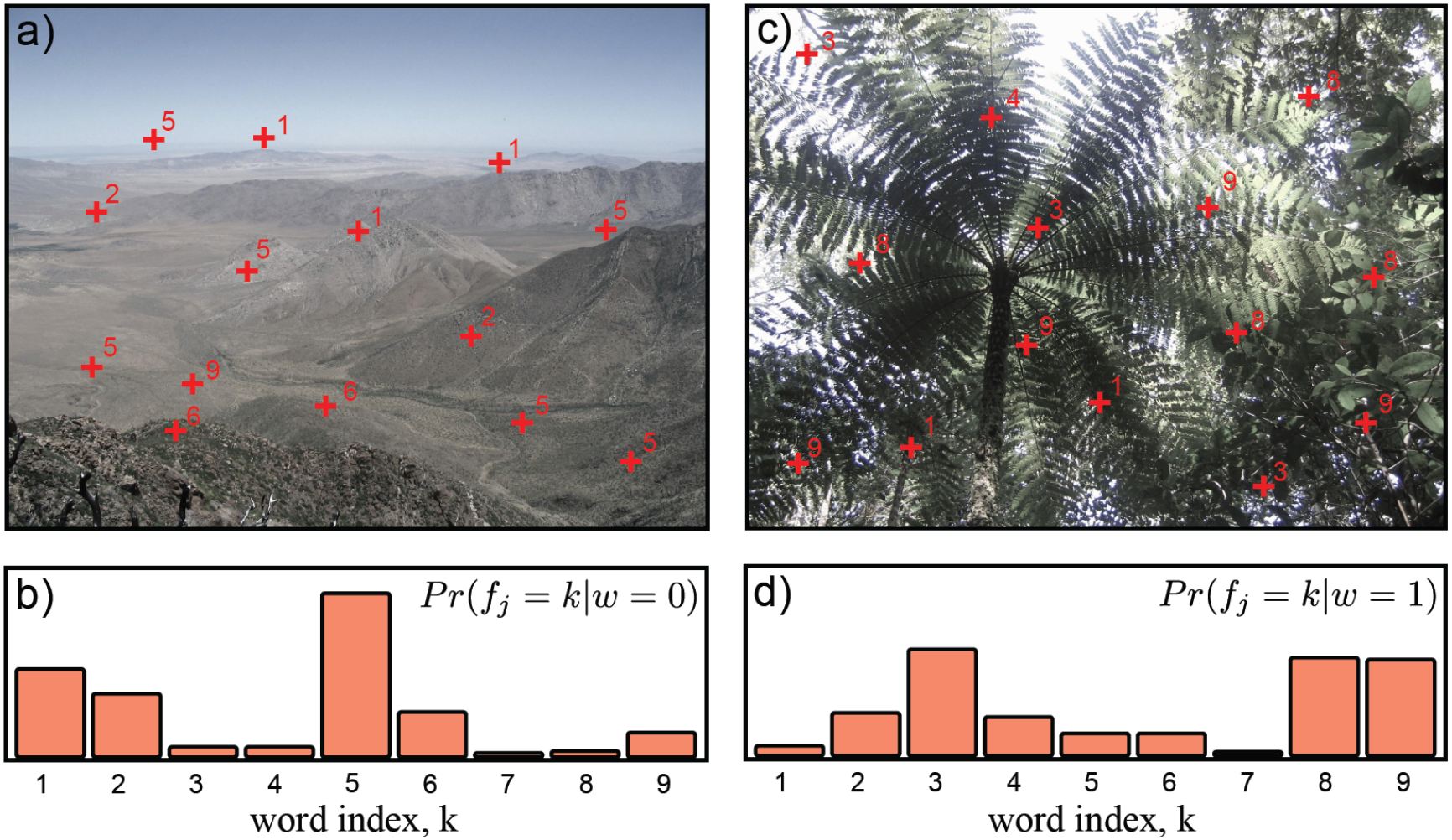
3.



category models (and/or) classifiers

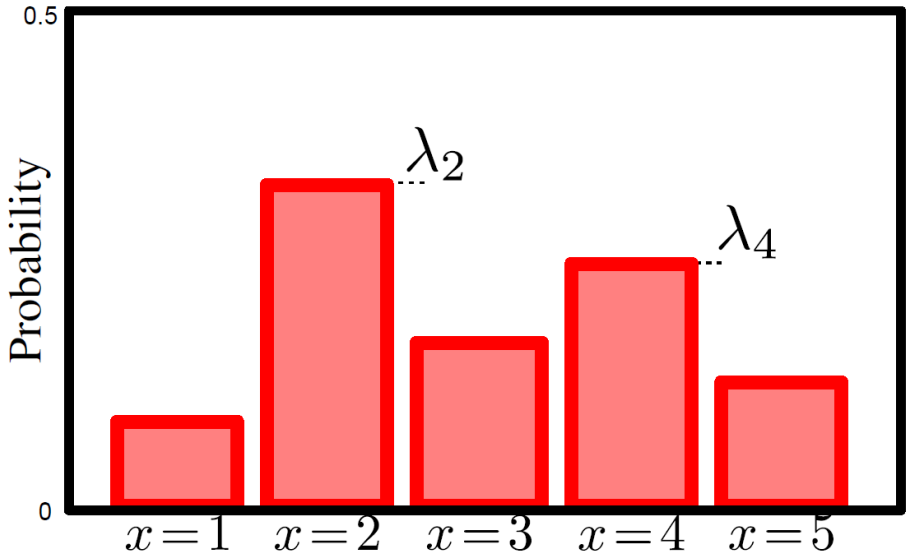
Categorical Distribution
MLE / MAP estimation

Bag of Words – Categorical Distributions



Categorical Distribution

Review



$$Pr(x=k)=\lambda_k$$

or can think of data as vector with all elements zero except k[th] e.g. [0,0,0,1 0]

$$Pr(\mathbf{x}=\mathbf{e}_k)=\prod_{j=1}^{K}\lambda_j^{x_j}=\lambda_k$$

For short we write:

$$Pr(x)=\mathrm{Cat}_x\left[\boldsymbol{\lambda}\right]$$

Categorical distribution describes situation where K possible outcomes y=1… y=k. Takes K parameters $\lambda_k \in [0,1]$ where $\sum_k \lambda_k = 1$

Dirichlet Distribution

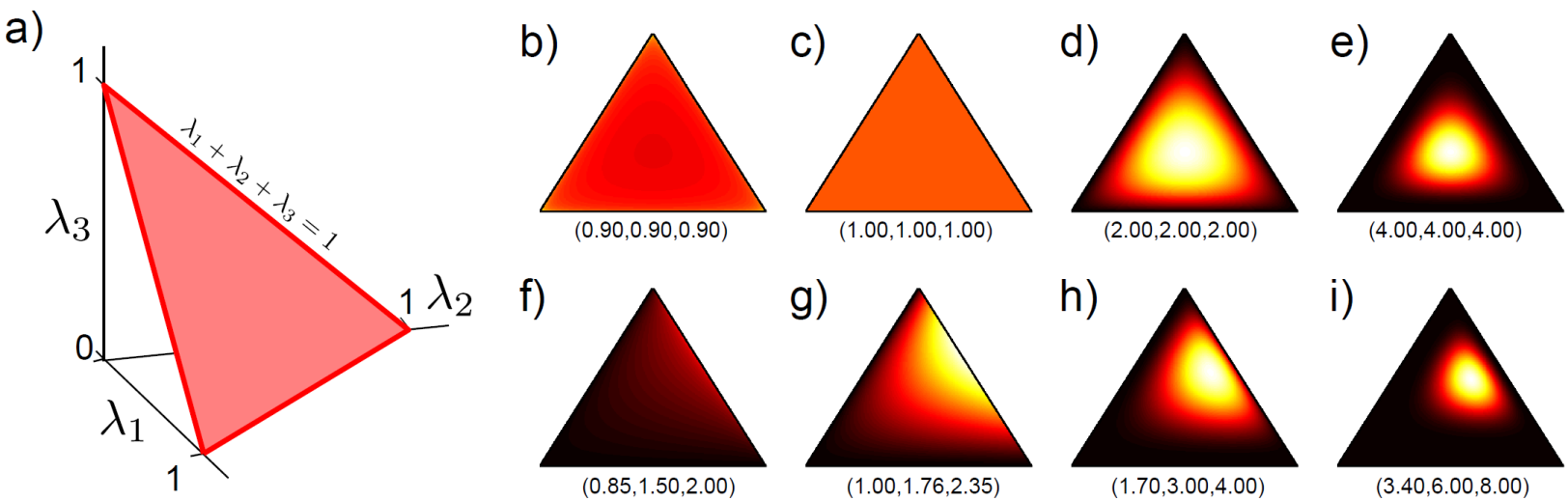
Review

Defined over K values $\lambda_k \in [0,1]$ where $\sum_k \lambda_k = 1$

$$Pr(\lambda_1 \ldots \lambda_K) = \frac{\Gamma[\sum_{k=1}^{K} \alpha_k]}{\prod_{k=1}^{K} \Gamma[\alpha_k]} \prod_{k=1}^{K} \lambda_k^{\alpha_k - 1}$$

Has k parameters $\alpha_k > 0$

Or for short: $Pr(\lambda_1 \ldots \lambda_K) = \mathrm{Dir}_{\lambda_{1\ldots K}}[\alpha_1, \alpha_2, \ldots, \alpha_K]$



Categorical distribution: MAP

Review

Take derivative, set to zero and re-arrange:

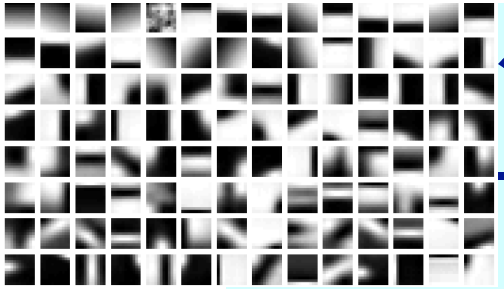
$$\hat{\lambda}_k = \frac{N_k + \alpha_k - 1}{\sum_{m=1}^{6}(N_m + \alpha_m - 1)}$$

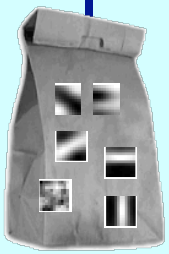
With a uniform prior ($\alpha_{1..K}$=1), gives same result as maximum likelihood.

$$\hat{\lambda}_k = \frac{N_k}{\sum_{m=1}^{6} N_m}$$

Learning and Recognition

codewords dictionary





category models (and/or) classifiers

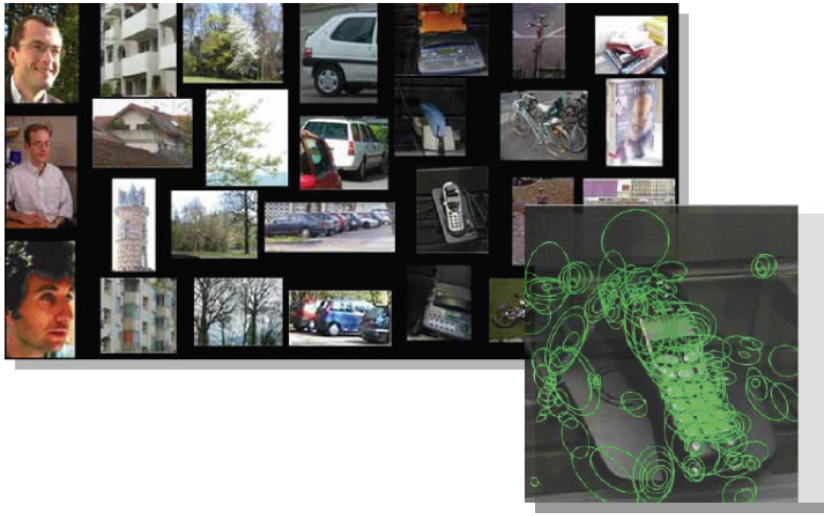
category decision

BoW as input to classifier

Now, use the image/ object class histograms as feature vectors for classification, e.g.

- Naïve Bayes classifier
- Support vector machines

Compared in Csurka, Bray, Dance & Fan, 2004



First, some notations

w_n: each patch in an image

- $w_n = [0,0,\ldots 1,\ldots,0,0]^T$

w: a collection of all N patches in an image

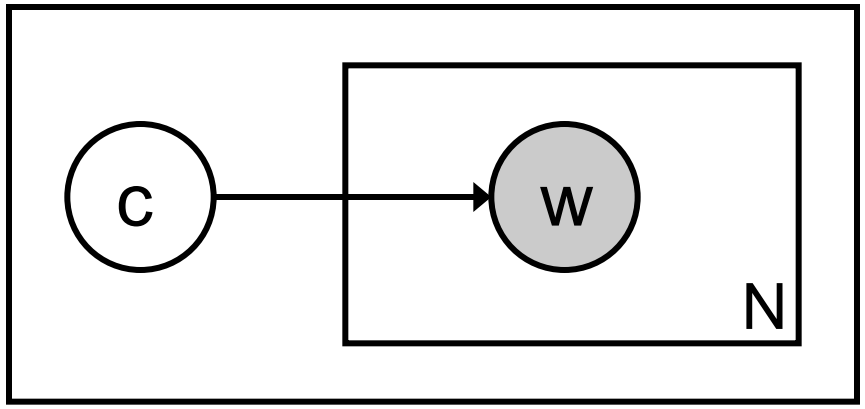
- $\mathbf{w} = [w_1, w_2, \ldots, w_N]$

d_j: the jth image in an image collection

c: category of the image

z: theme or topic of the patch (for hierarchical models)

Naïve Bayes model



$$c^* = \arg\max_c \; p(c \mid w) \propto p(c)\, p(w \mid c) = p(c) \prod_{n=1}^{N} p(w_n \mid c)$$

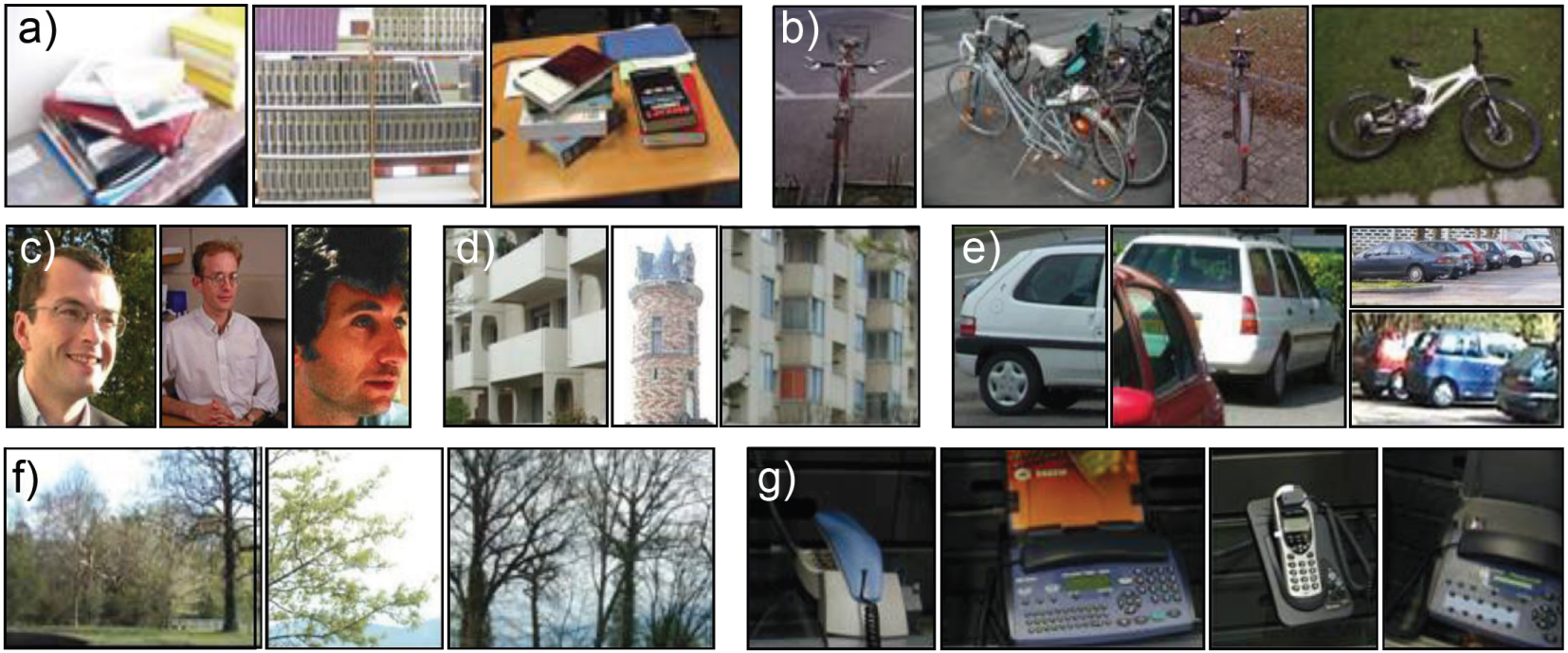
Object class decision

Prior prob. of the object classes

Image likelihood given the class

Csurka et al. 2004

Bag of words for object recognition

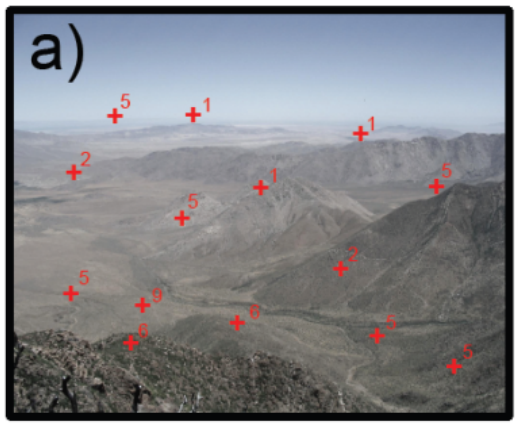


Naïve Bayes

Figure 20.3 Object recognition using bag of words. Csurka *et al.* (2004) built a generative bag of visual words model to distinguish between examples of a) books, b) bicycles, c) people, d) buildings, e) cars, f) trees and g) phones. Despite the wide variety of visual appearance within each class they achieved 72% correct classification. By applying a discriminative approach to the same problem, they managed to improve performance further.

SVM

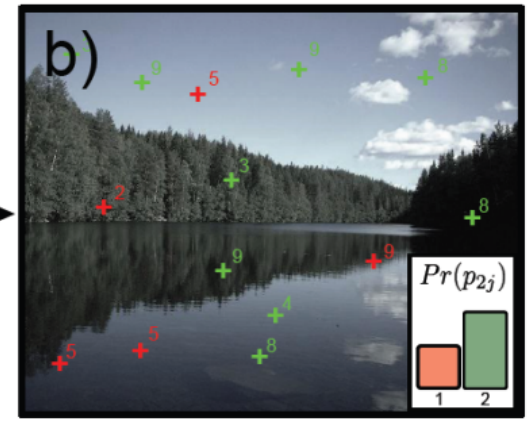
Problems with bag of words



Bag of words model

Problem 1

Does not model co-occurence of words



Solution: Hierarchical Bayesian Models (multi-topic models)

Hierarchical Bayesian Models

- Like regular BoW representation, describes relative frequency of visual words across a set of images
- However, words are not generated independently (connected by hidden variable – the topic)
- Analogy to text documents
  - Each image contains mixture of several topics (parts)
  - Each topic induces a distribution over words

Latent Dirichlet allocation

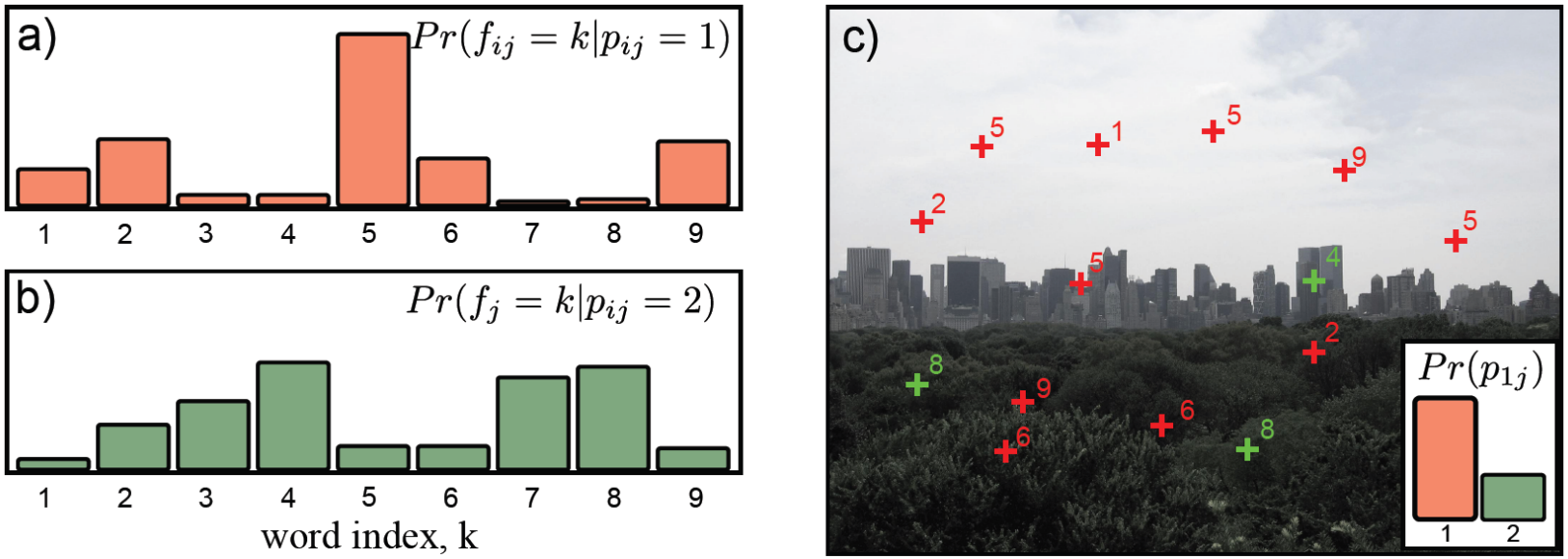
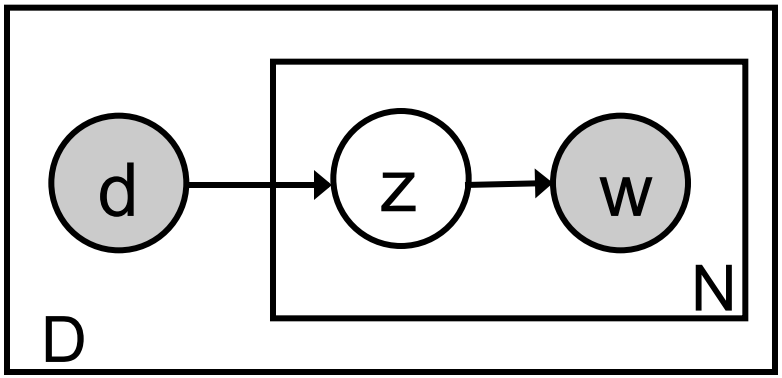


Figure 20.5 Latent Dirichlet allocation. This model treats each word as belonging to one of M different parts (here $M = 2$). a) The distribution over the words given that the part is 1 is described by a categorical distribution. b) The distribution over the words given that the part is 2 is described by a different categorical distribution. c) In each image, the tendency for each observed word to belong to each part is different. In this case part 1 is more likely than part 2 and so most of the words belong to part 1 (are red as opposed to green).

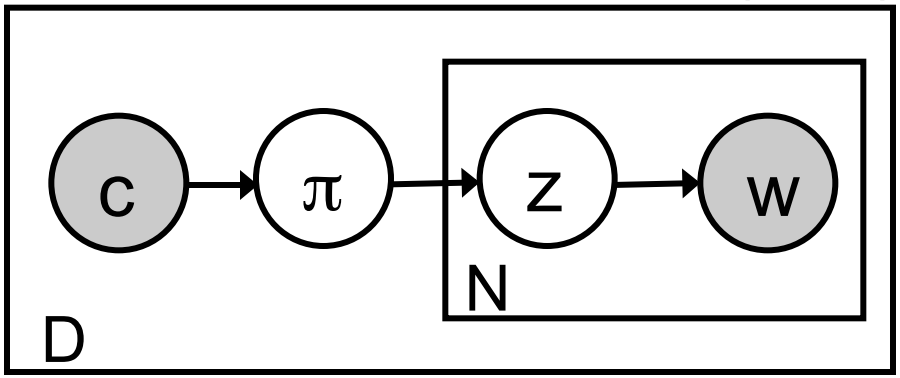
Hierarchical Bayesian models

Probabilistic Latent Semantic Analysis (pLSA)



Hoffman, 2001

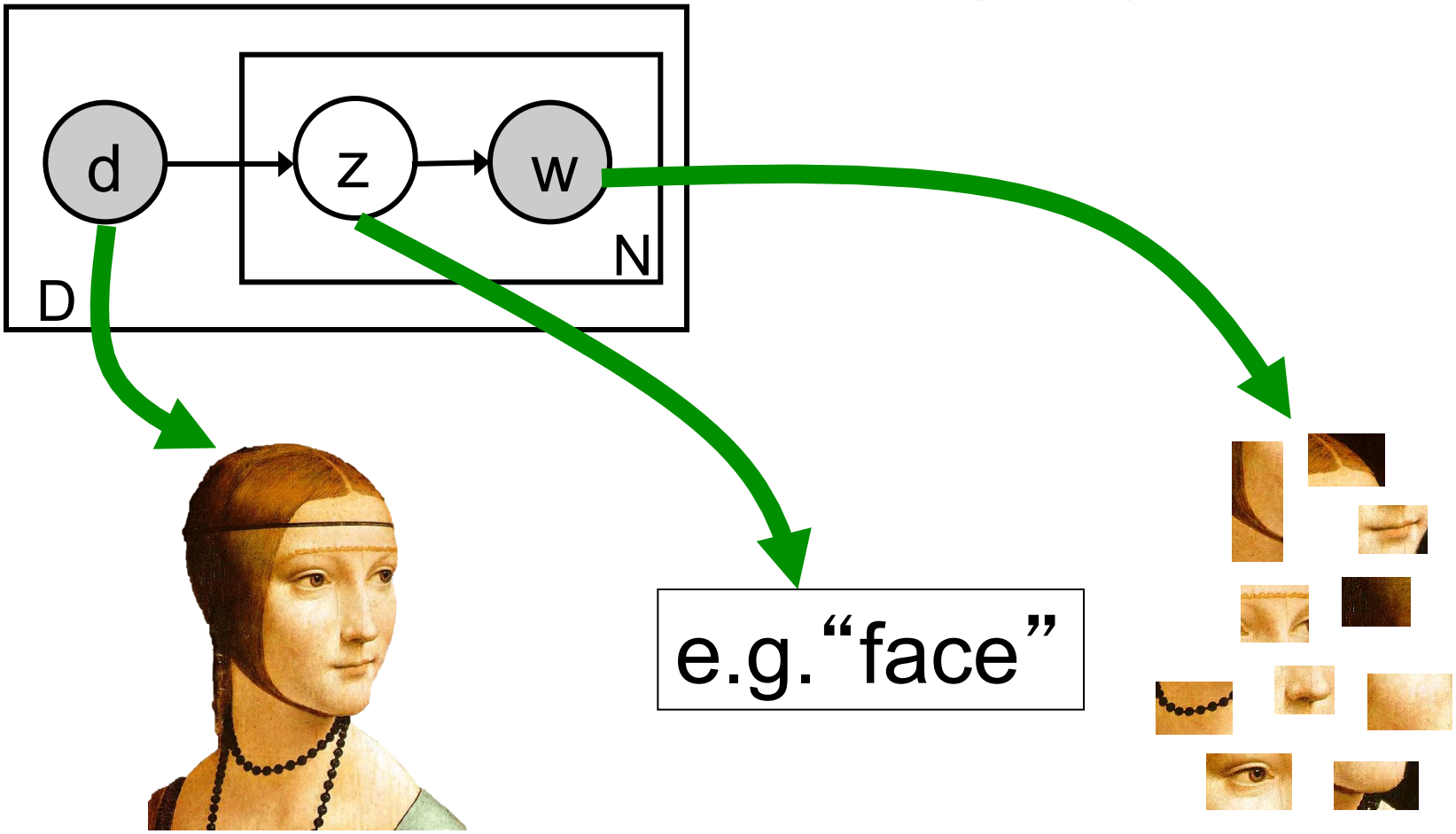
Latent Dirichlet Allocation (LDA)



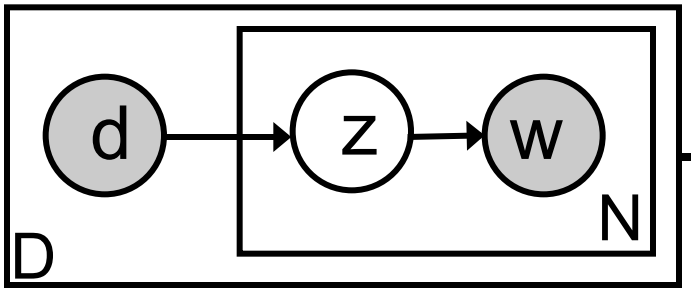
Blei et al., 2001

Hierarchical Bayesian models

Probabilistic Latent Semantic Analysis (pLSA)

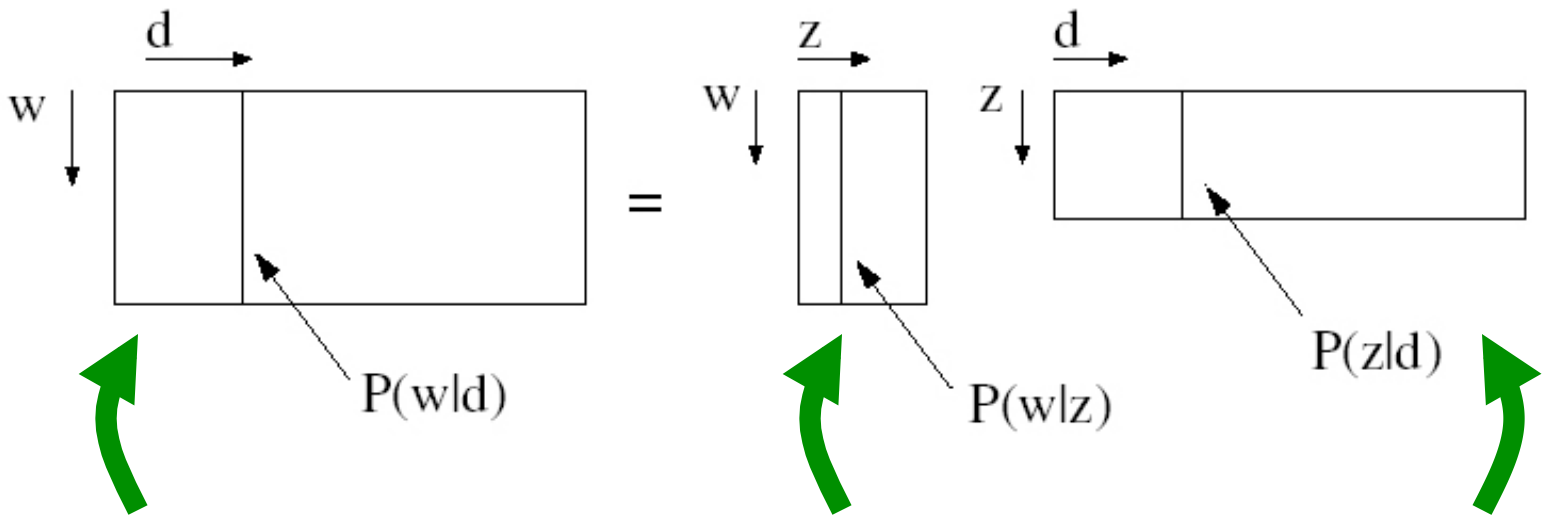


Sivic et al. ICCV 2005



The pLSA model

$$p(w_i \mid d_j) = \sum_{k=1}^{K} p(w_i \mid z_k)\, p(z_k \mid d_j)$$



Observed codeword distributions

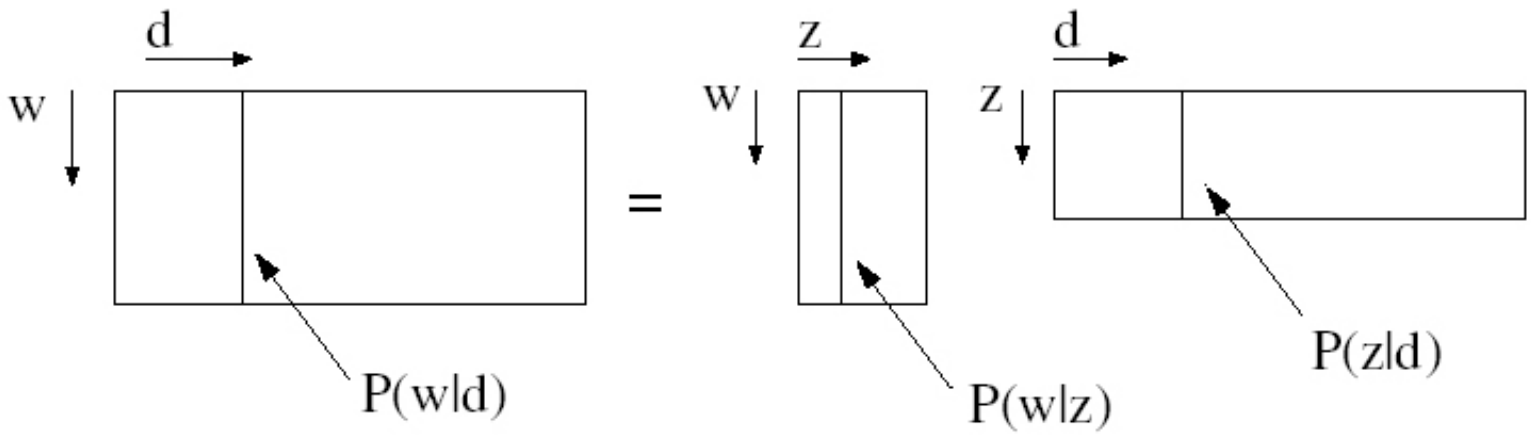
Codeword distributions per theme (topic)

Theme distributions per image

Slide credit: Josef Sivic

Recognition using pLSA

$$z^* = \arg\max_z p(z \mid d)$$



Slide credit: Josef Sivic

Learning the pLSA parameters

Observed counts of word *i* in document *j*

$$L = \prod_{i=1}^{M} \prod_{j=1}^{N} P(w_i|d_j)^{n(w_i,d_j)}$$

$$\sum_{k=1}^{K} P(z_k|d_j)P(w_i|z_k)$$

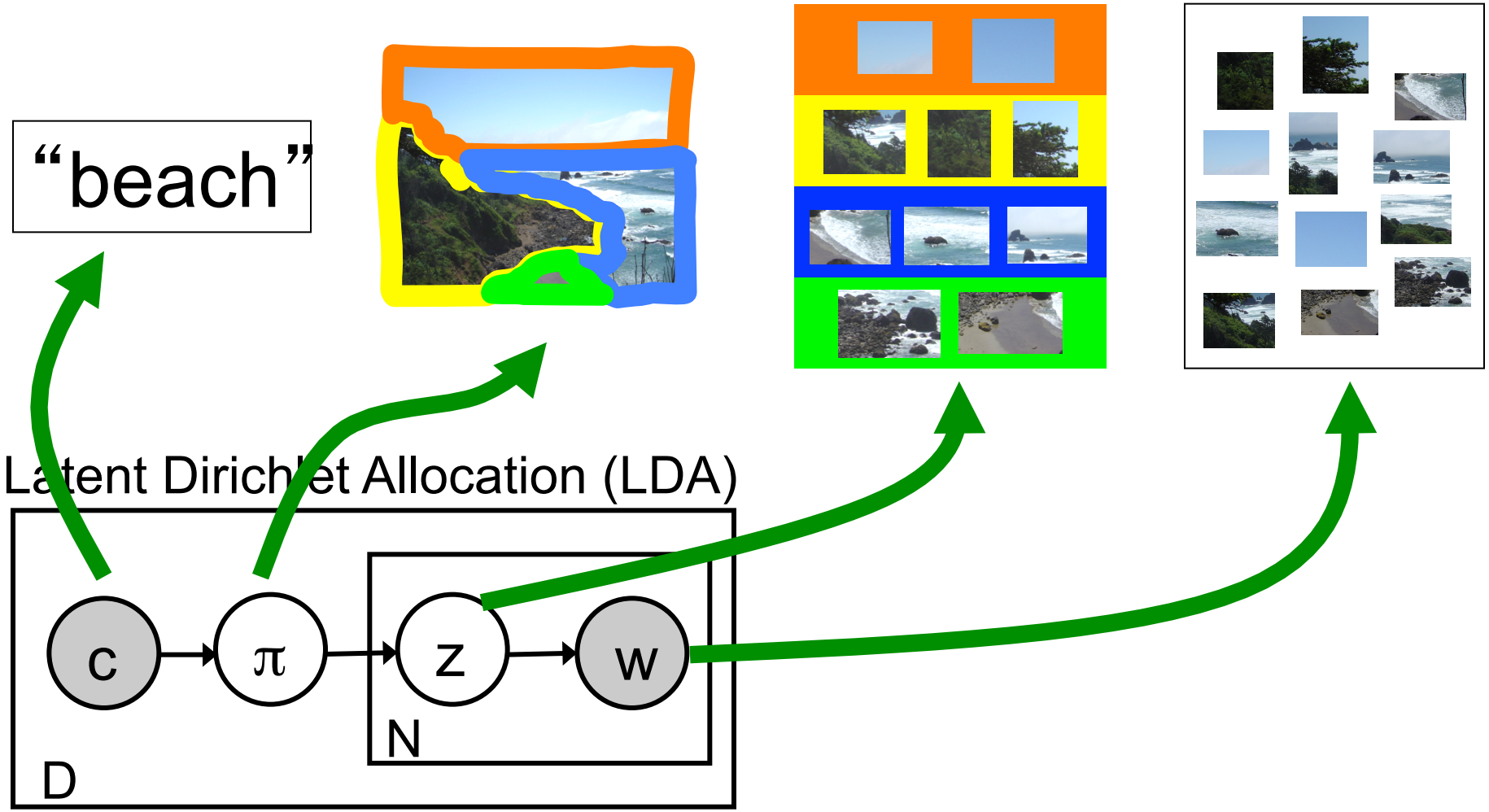
Maximize likelihood of data using EM

M … number of codewords

N … number of images

Slide credit: Josef Sivic

Hierarchical Bayesian models



Fei-Fei et al. ICCV 2005

Latent Dirichlet allocation

- Can be thought of as mixture of categorical distributions
- See Prince textbook, Sections 20.3 and 20.4

Some References

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

Project 1 Assignment

Choose Task: object recognition or scene recognition; supervised vs unsupervised

Choose Dataset: if at all possible, choose a dataset that already exists so you compare results.

Choose feature extraction method. Dense vs sparse; what is your feature descriptor

Choose a quantization method, e.g. kmeans, EM, mean-shift etc…

Choose parameter learning method: MLE or MAP, care to try pLSA or LDA?

Choose a classifier method: Naïve Bayes vs SVM or any discriminative classifier you want

Quantitative Evaluation: how will you measure how well your method works?