Practical Course WS12/13

A Quick Introduction to Bag-of-Words

Cyrill Stachniss and Luciano Spinello

Motivation: Analogy to Documents

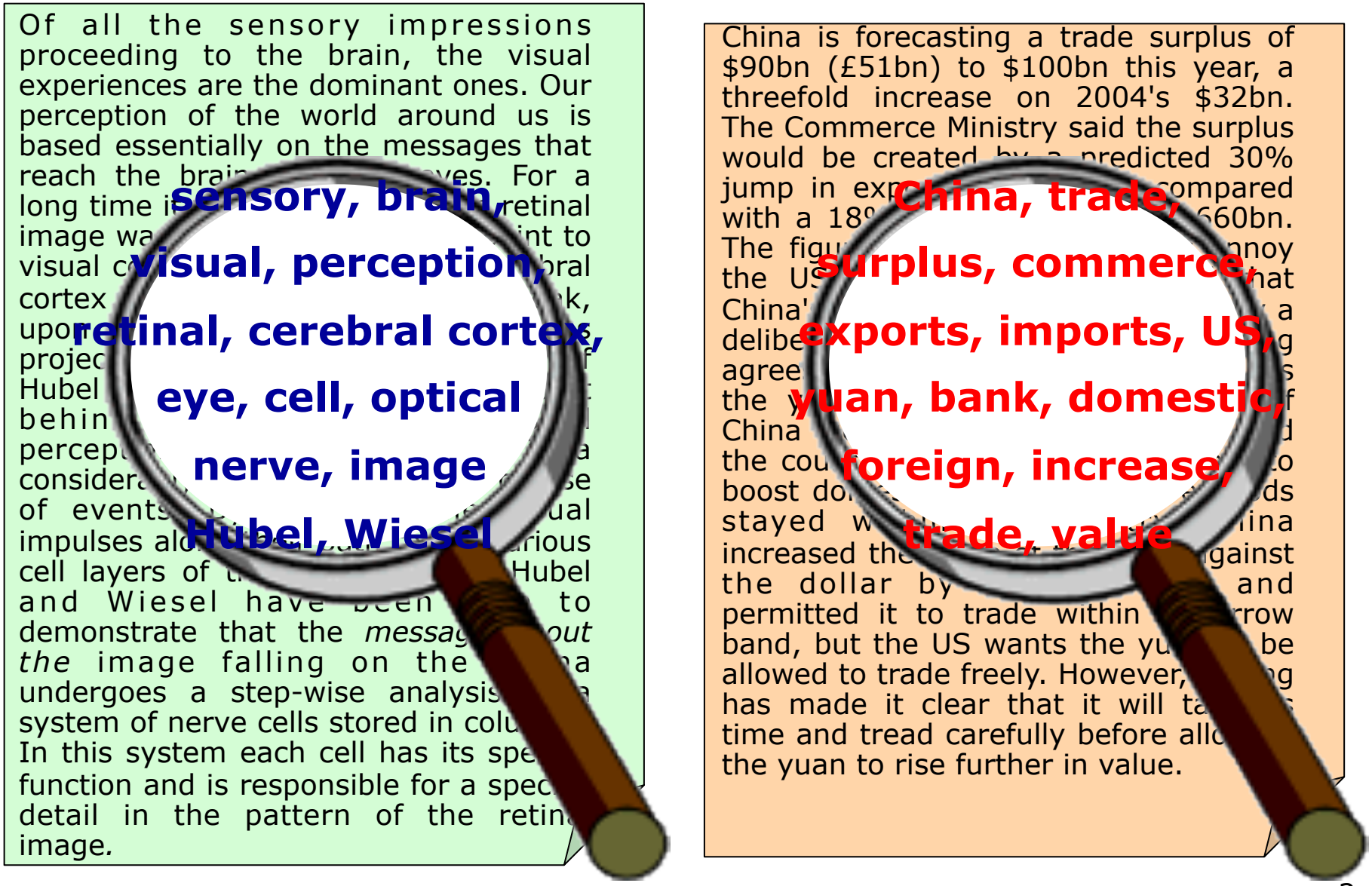


image source: L. Fei-Fei

Bag of Visual Words

- Analogy to documents: The content can be inferred from the frequency of words

object

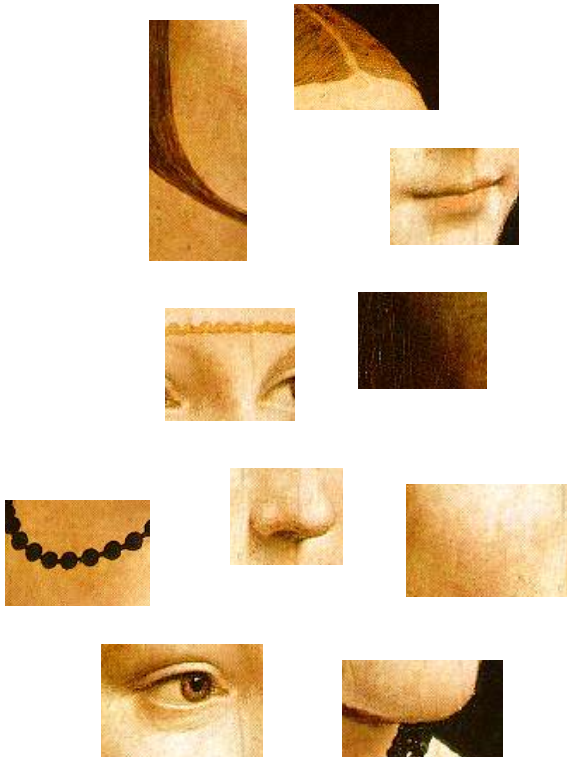
bag of
“visual words”

image source: L. Fei-Fei

Bag of Visual Words

- Visual words = independent features

face



features

image source: L. Fei-Fei

Bag of Visual Words

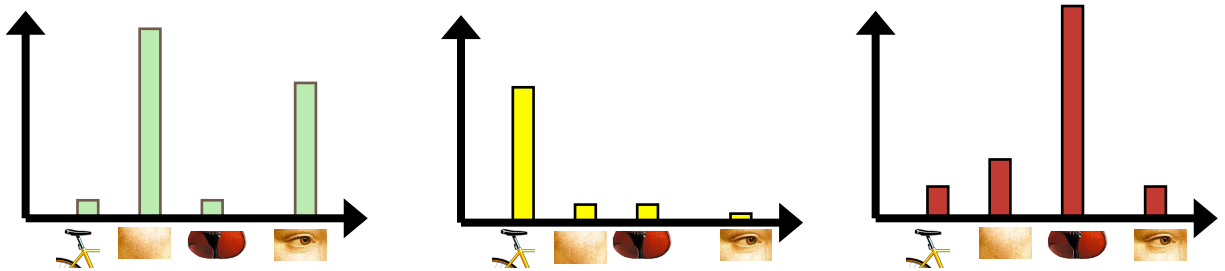
- Visual words = independent features
- Construct a dictionary of representative words

codeword dictionary

image source: L. Fei-Fei

Bag of Visual Words

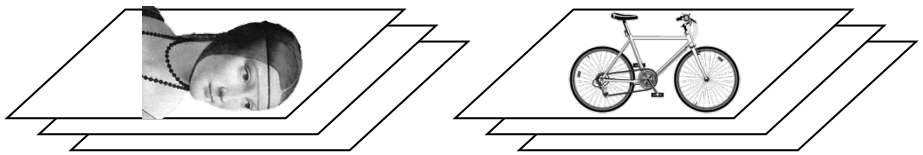
- Visual words = independent features
- Construct a dictionary of representative words
- Represent the images based on a histogram of word occurrences (bag)



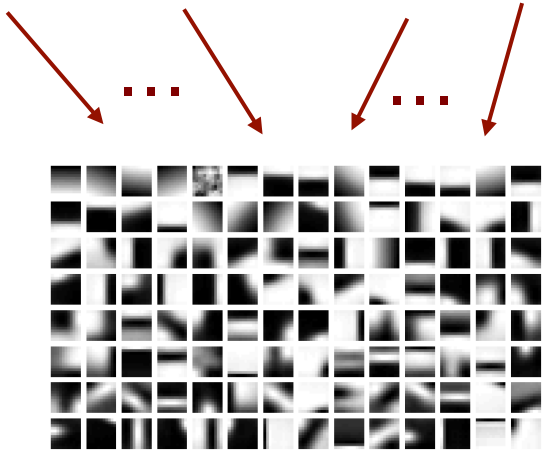
Each detected feature is assigned to the closest entry in the codebook

image source:
L. Fei-Fei

Overview



feature detection



codeword dictionary

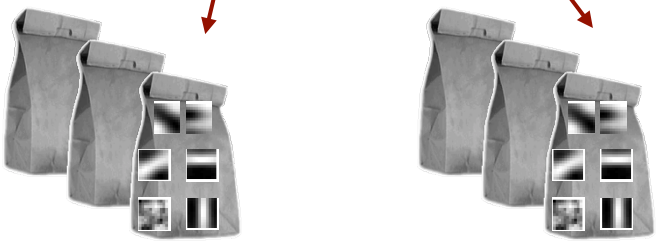
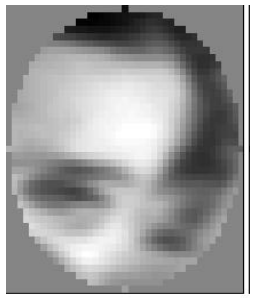


image representation

slide adapted from: L. Fei-Fei

Feature Detection and Representation

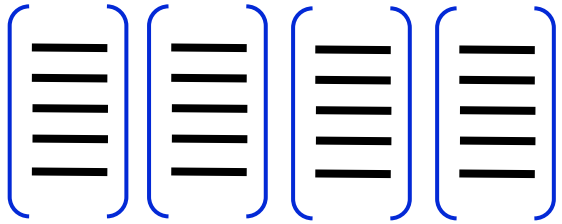
detected features in a set of training images (intensity changes)



example patch

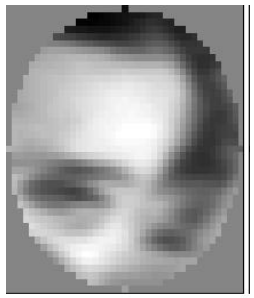
slide adapted from: L. Fei-Fei

Feature Detection



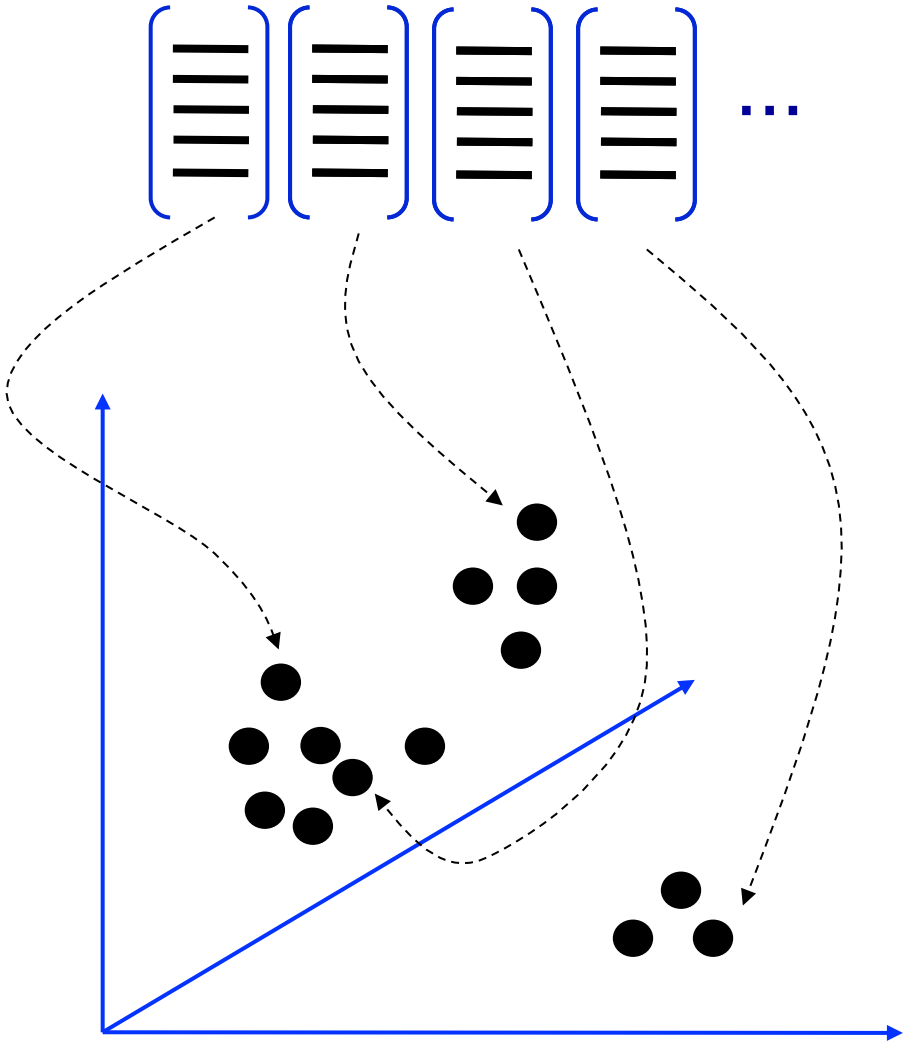
descriptor vectors
(e.g., SIFT/SURF,
consider local
orientations of gradients)

detected features in a
set of training images
(intensity changes)



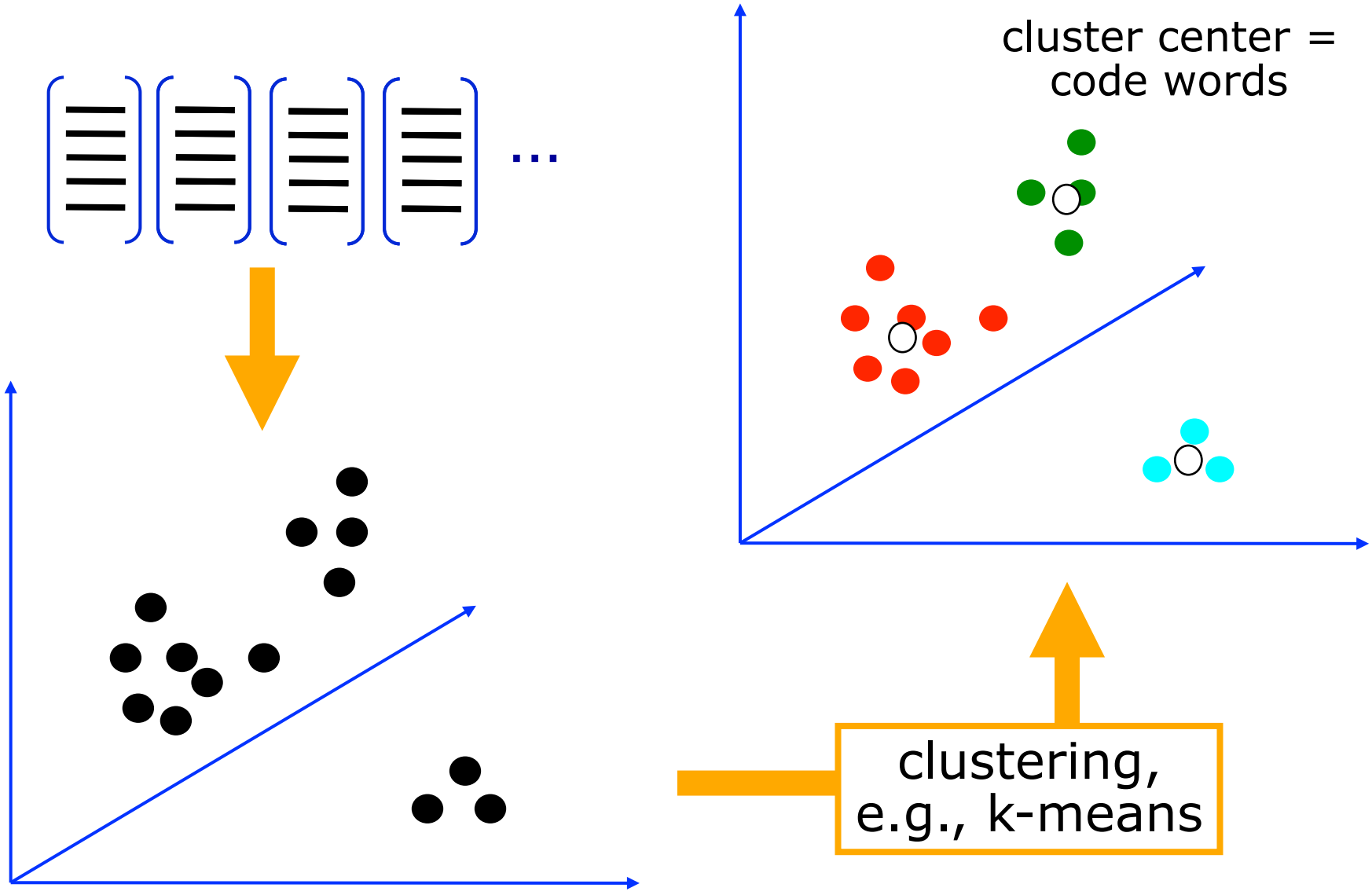
example patch

Learning the Dictionary



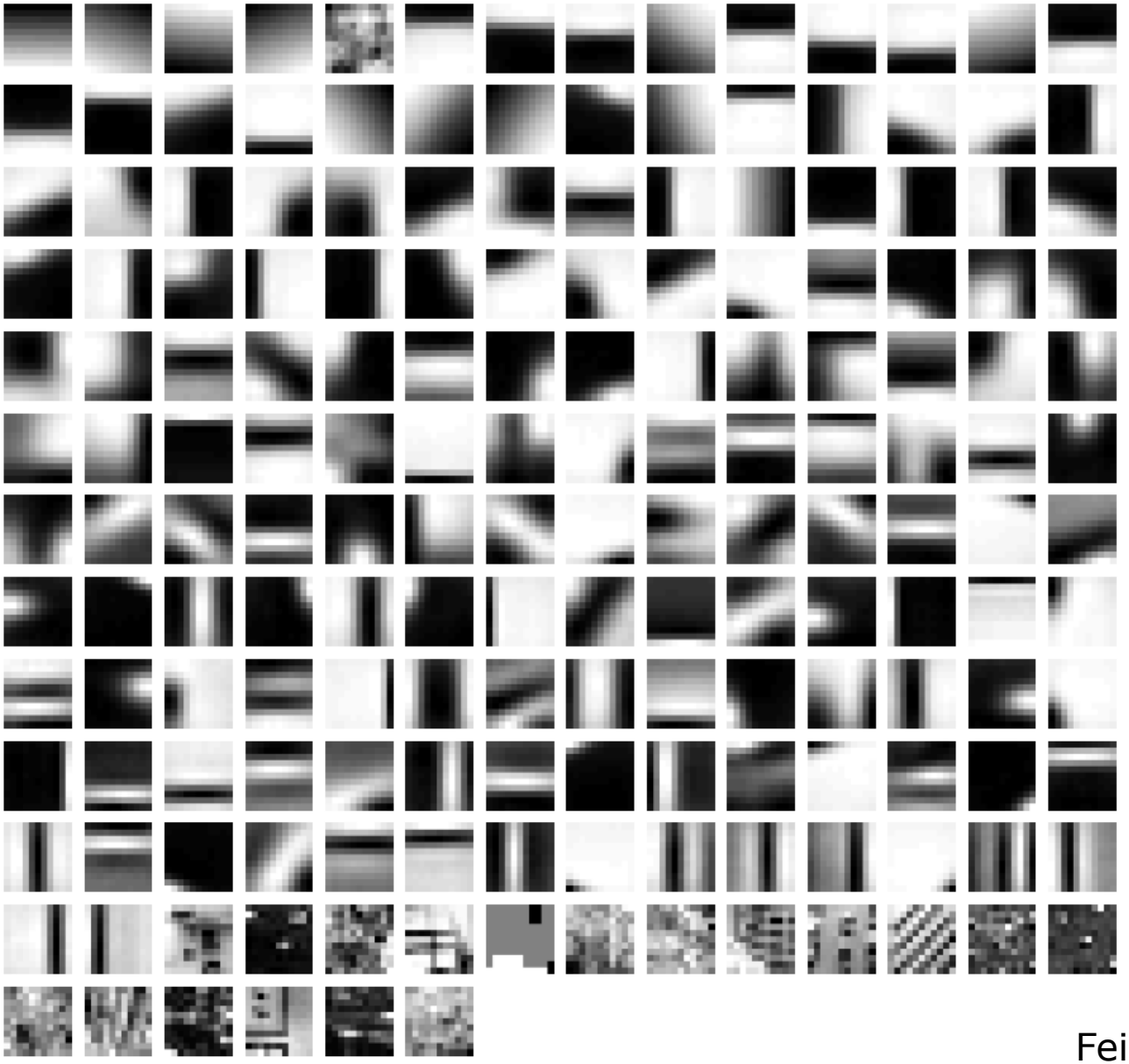
slide adapted from: L. Fei-Fei

Learning the Dictionary



slide adapted from: L. Fei-Fei

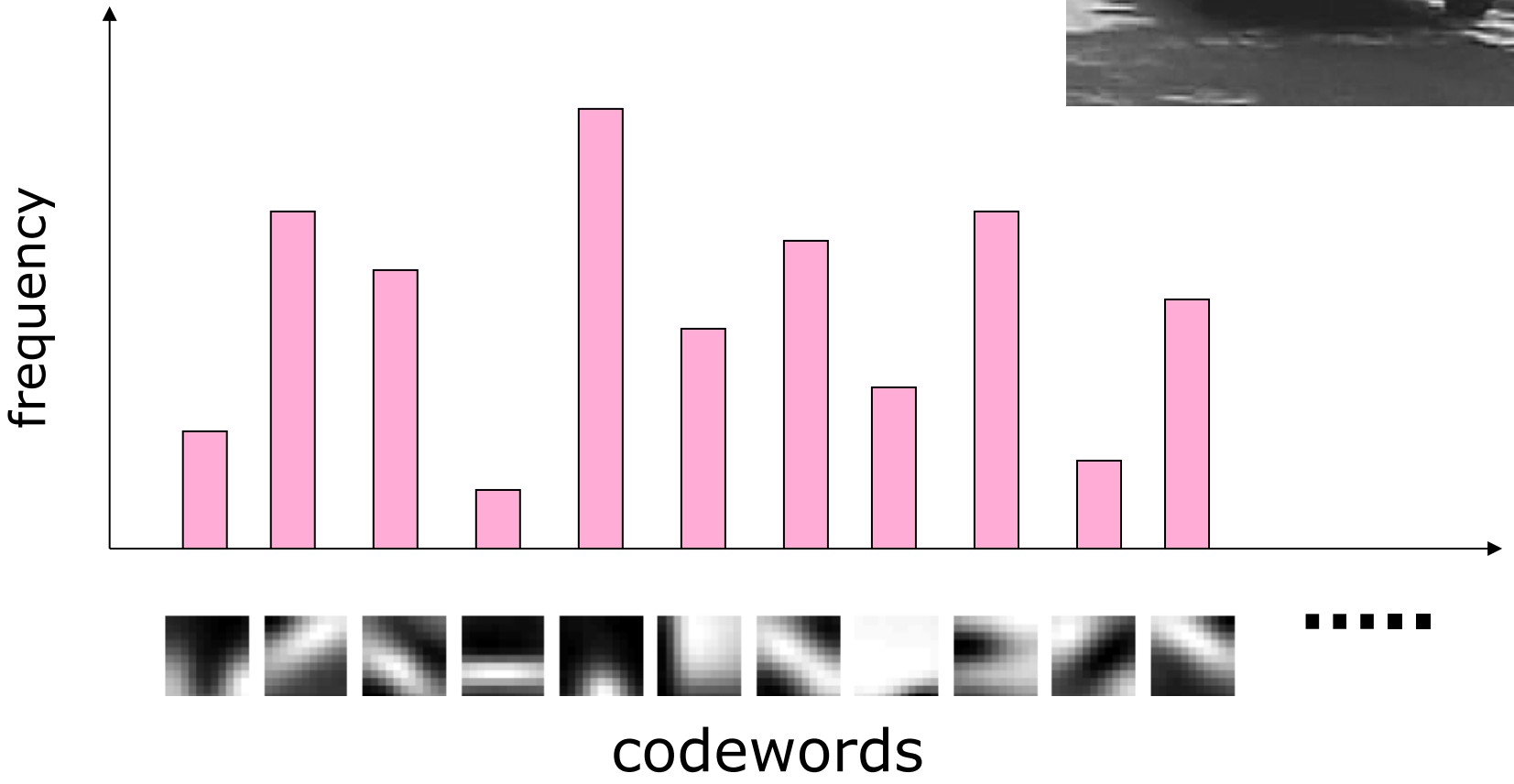
Example Codeword Dictionary



Fei-Fei et al. 2005

Example Image Representation

- Build the histogram by assigning each detected feature to the closest entry in the codebook



slide adapted from: L. Fei-Fei

Properties Bag-of-Words

- Compact summary of content
- Flexible to viewpoint and deformations
- Can be used for place recognition by comparing the histograms
- Ignores spatial arrangement
- Unclear how to choose optimal vocabulary
  - Too small: Words not representative of all patches
  - Too large: Artifacts, over-fitting

Simple Image Recognition

- Given: Dictonary and reference images
- Query: Image
- Wanted: Image from the reference images that best matches the query image

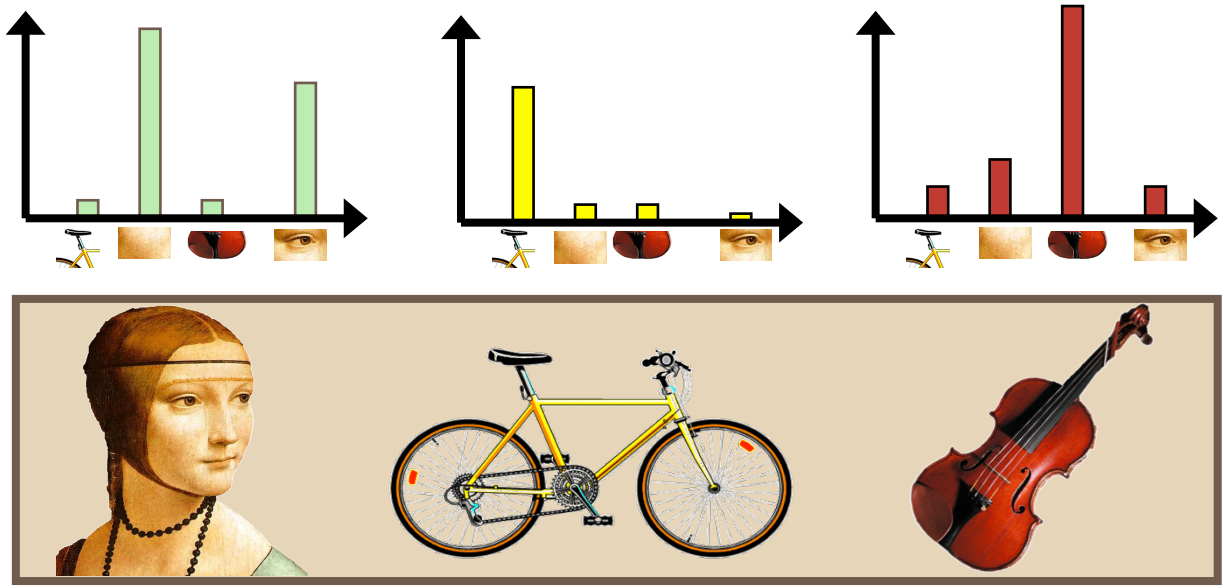
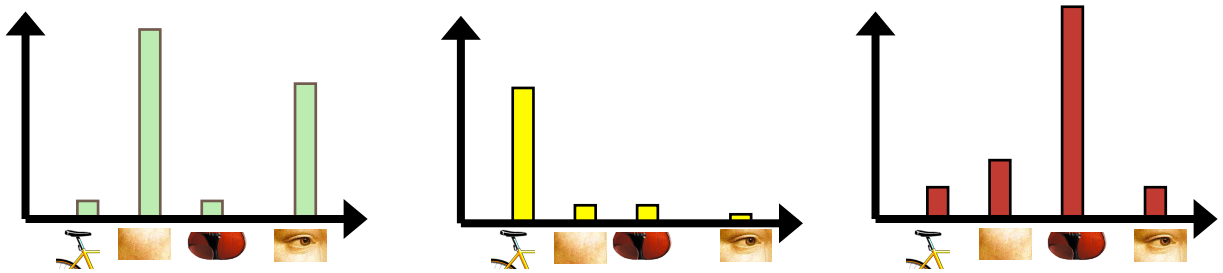


image source:
L. Fei-Fei

Histogram Comparison

- Place recognition task boils down to histogram comparisons



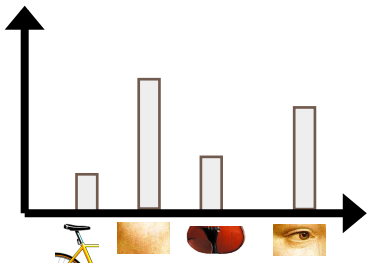


image source:
L. Fei-Fei

Histogram Comparison

- How to compare two histograms?

TF-IDF Reweighting

- How to compare two histograms?
- **First:** re-weight the bin of the histograms

$$t_i = \frac{n_{id}}{n_d} \log \frac{N}{n_i}$$

- n_{id} : #occ of word i in the image d
- n_d: total #words in the image d
- n_i: #images in which word i occurs
- N: total #images in the dataset

Cosine Comparison

- How to compare two histograms?
- **Second:** compare via cosine distance

$$similarity(A, B) = \cos(\theta) = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$

Next Steps

- Read "*Video Google: A Text Retrieval Approach to Object Matching in Videos*" by Sivic and Zisserman
- Identify components that need to be realized
- Identify dependencies between components
- Create a schedule and assign tasks
- Go!