

# Practical Course WS12/13

## A Quick Introduction to Bag-of-Words

Cyrill Stachniss and Luciano Spinello

---



# Motivation: 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 the brain through the eyes. For a long time it was thought that the retinal image was a direct projection of the visual world onto the cerebral cortex. Hubel and Wiesel, however, showed that the eye, cell, optical nerve, image, Hubel, Wiesel, 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% jump in exports compared with a 18% increase in imports. The figure of \$660bn. The US Commerce Department says that China's trade surplus with the US is a deliberate policy to boost domestic growth. China's exports are valued at \$660bn, while imports are valued at \$570bn. The yuan, bank, domestic, foreign, increase, trade, value, exports, imports, US, yuan, bank, domestic, foreign, increase, trade, value.

# Bag of Visual Words

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



object



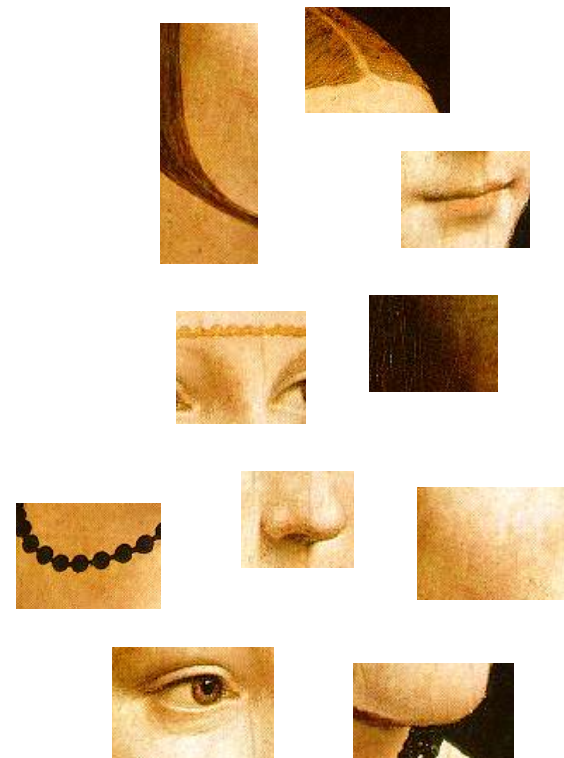
bag of  
"visual words"

# Bag of Visual Words

- Visual words = independent features



face



features

# Bag of Visual Words

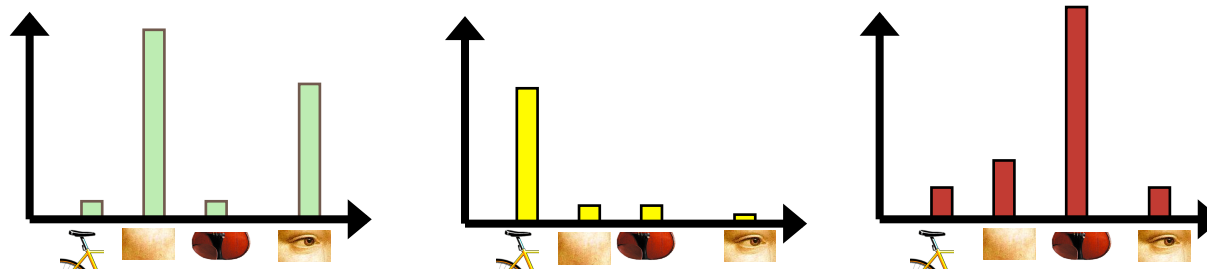
- Visual words = independent features
- Construct a dictionary of representative words

codeword dictionary



# 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)

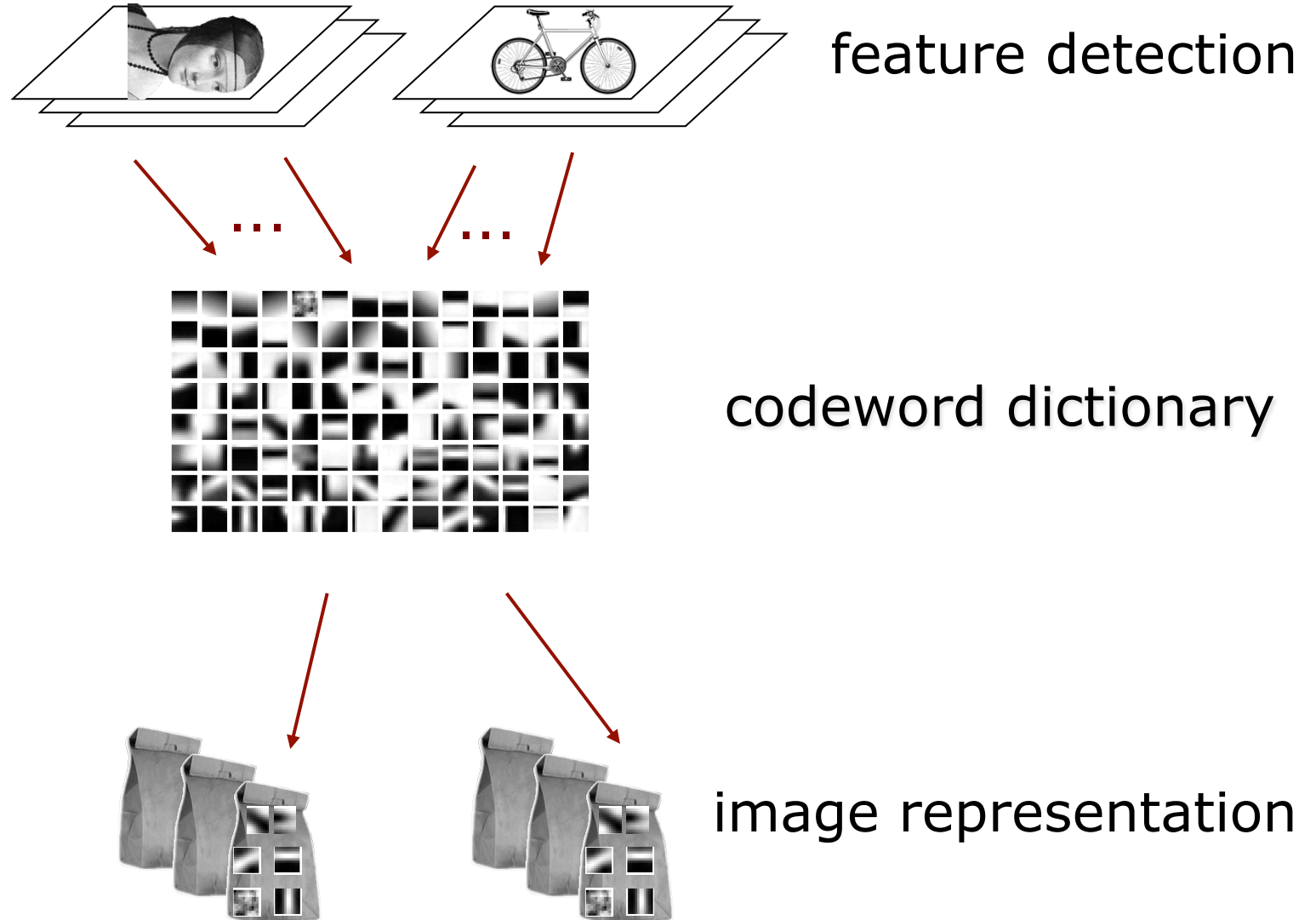


Each detected feature is assigned to the closest entry in the codebook

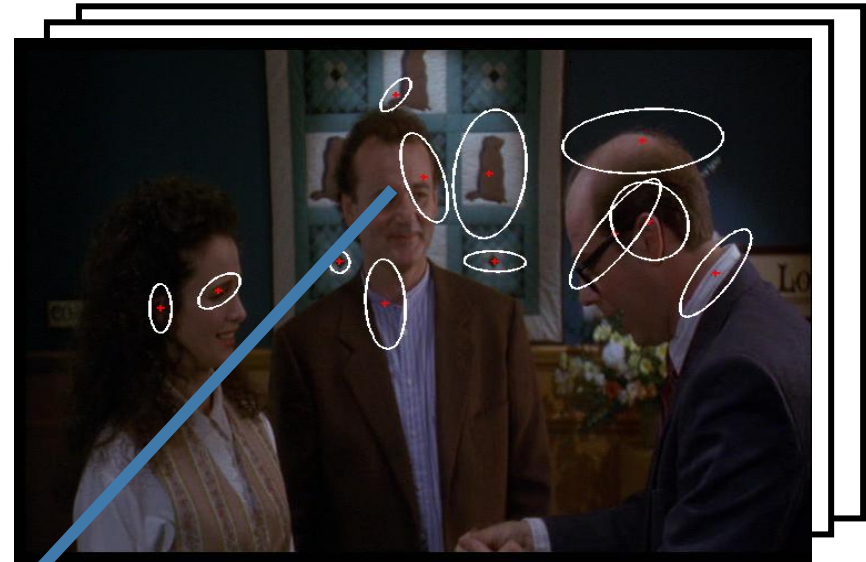


image source:  
L. Fei-Fei

# Overview



# Feature Detection and Representation

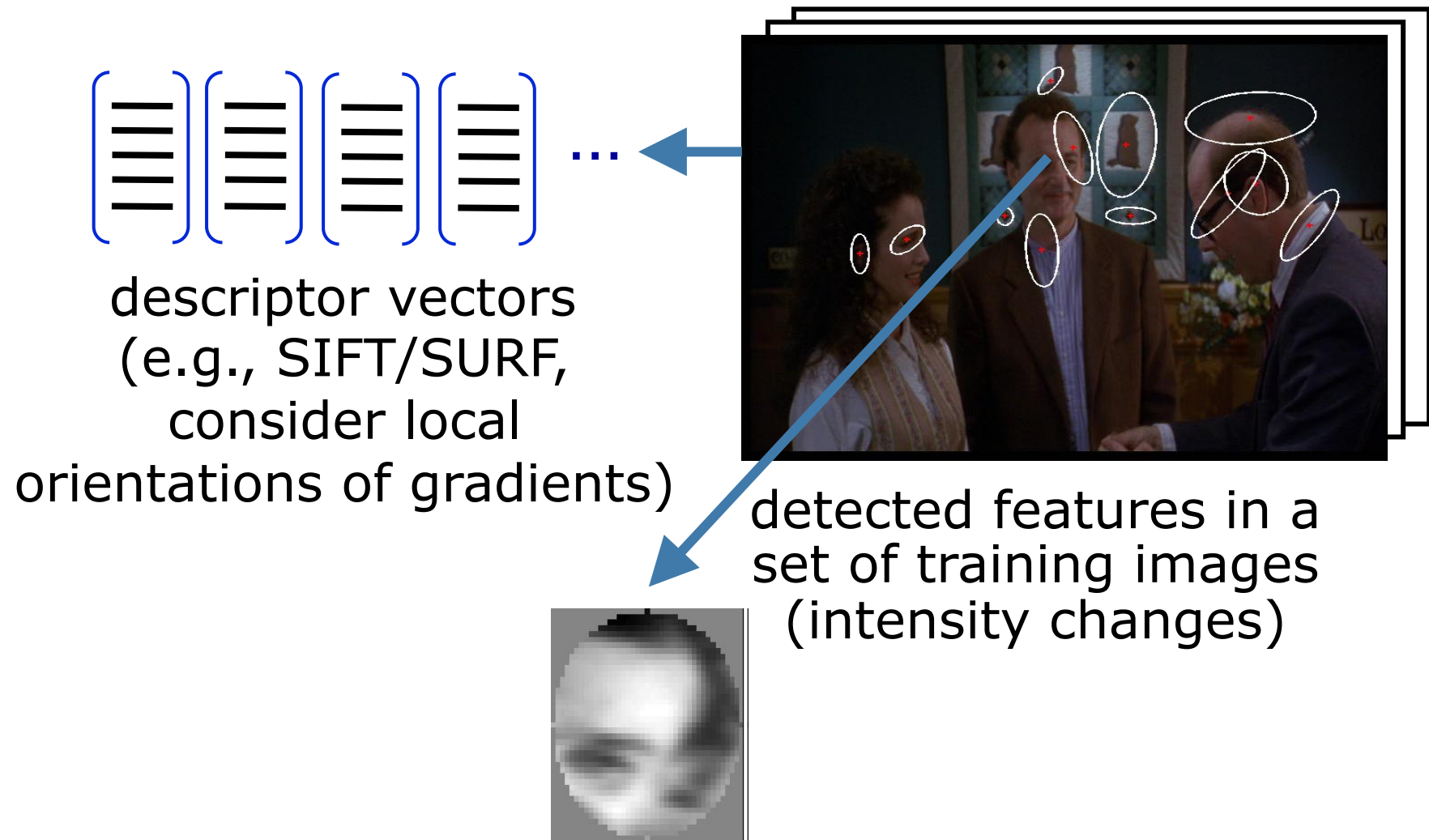


detected features in a set of training images (intensity changes)

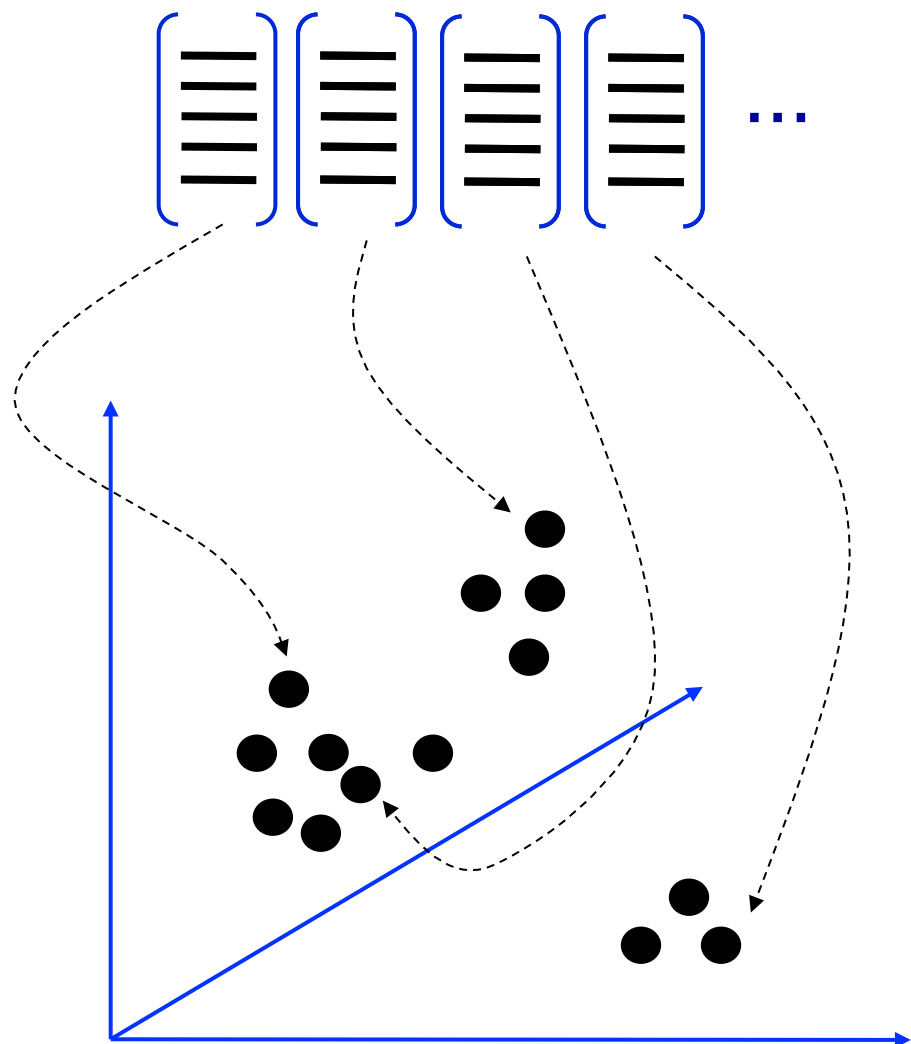




# Feature Detection

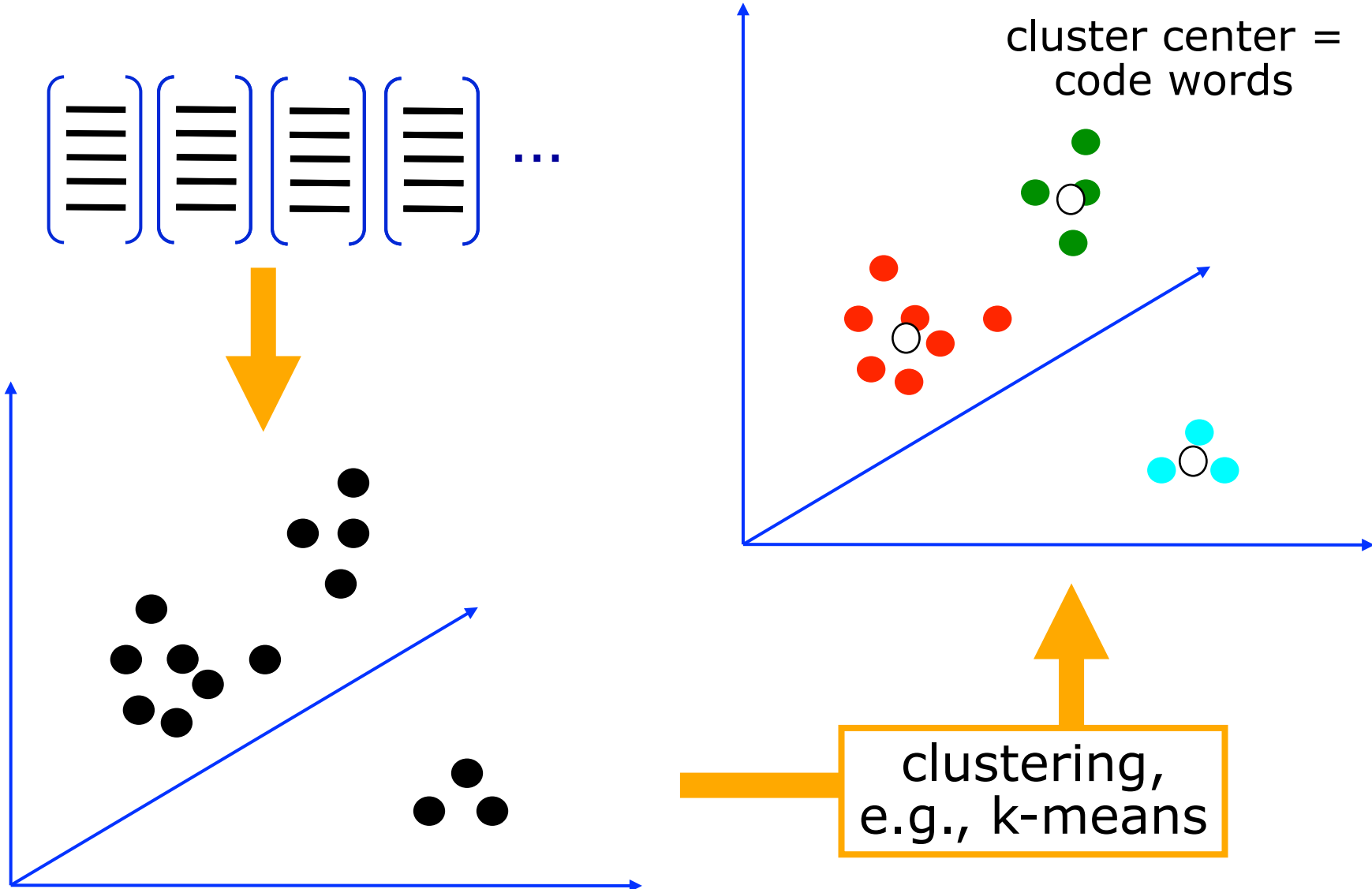


# Learning the Dictionary



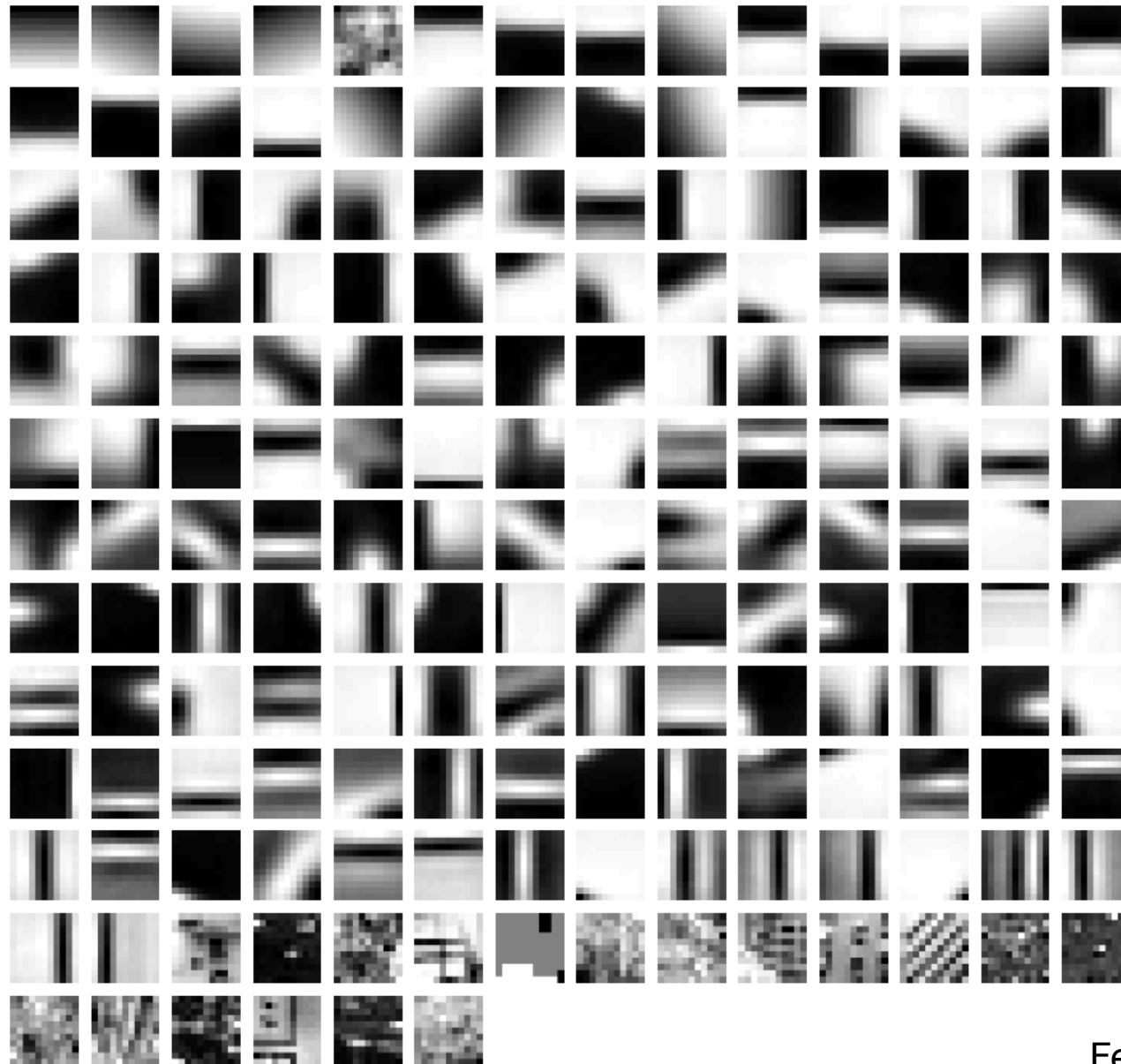
slide adapted from: L. Fei-Fei

# Learning the Dictionary



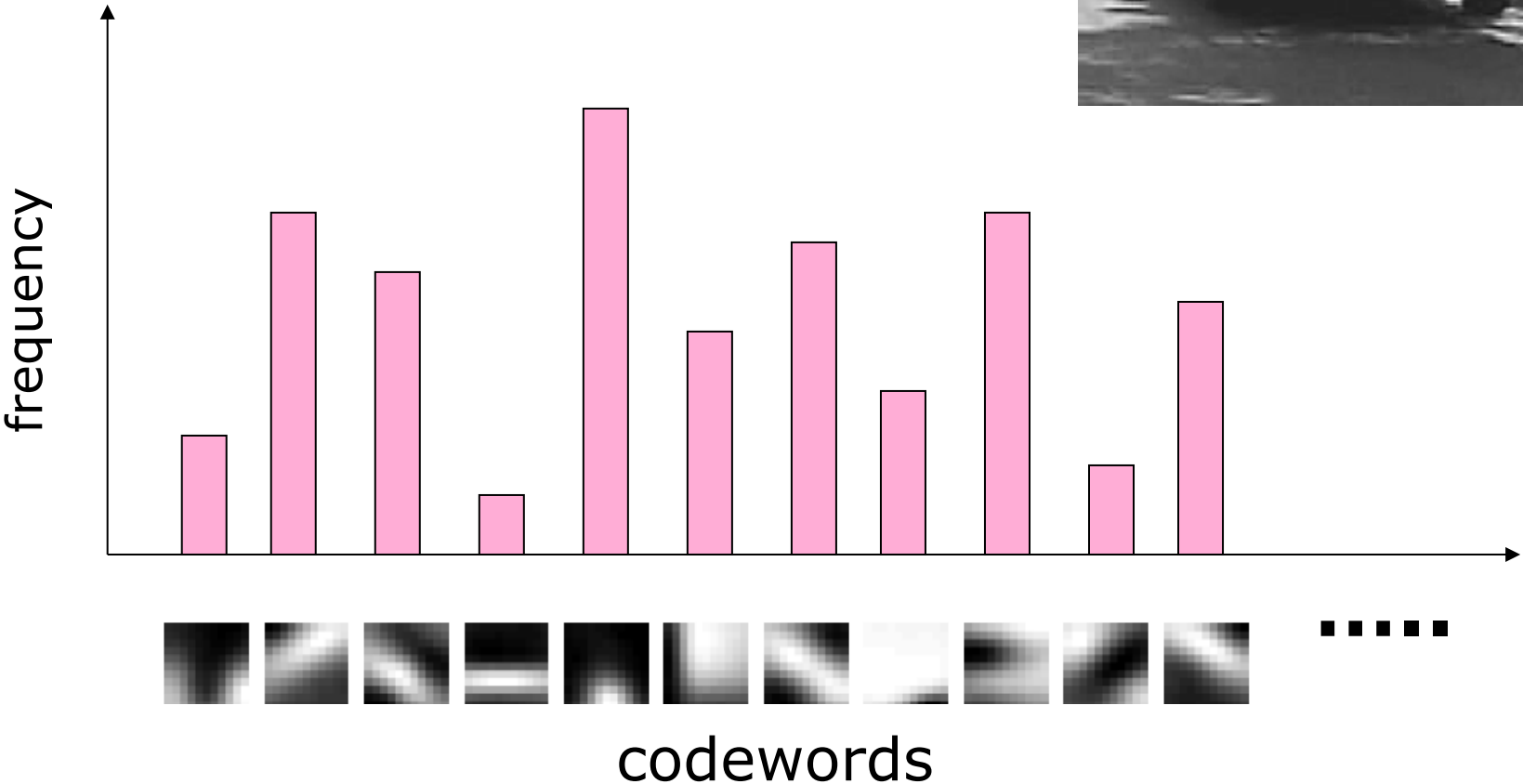
slide adapted from: L. Fei-Fei

# Example Codeword Dictionary



# 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: Dictionary and reference images
- Query: Image
- Wanted: Image from the reference images that best matches the query image

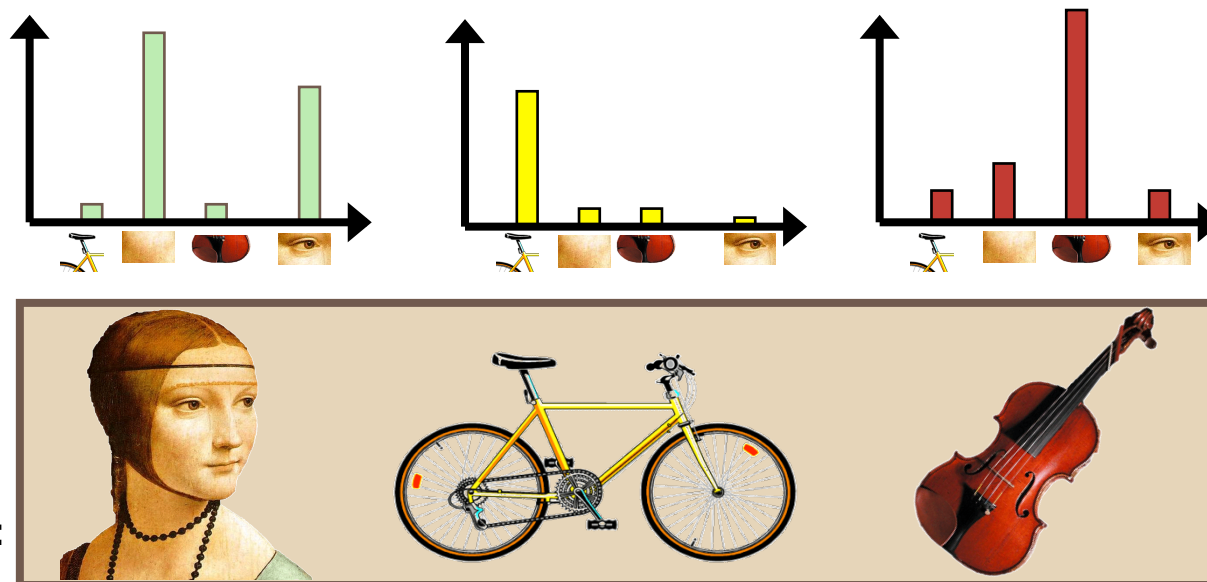
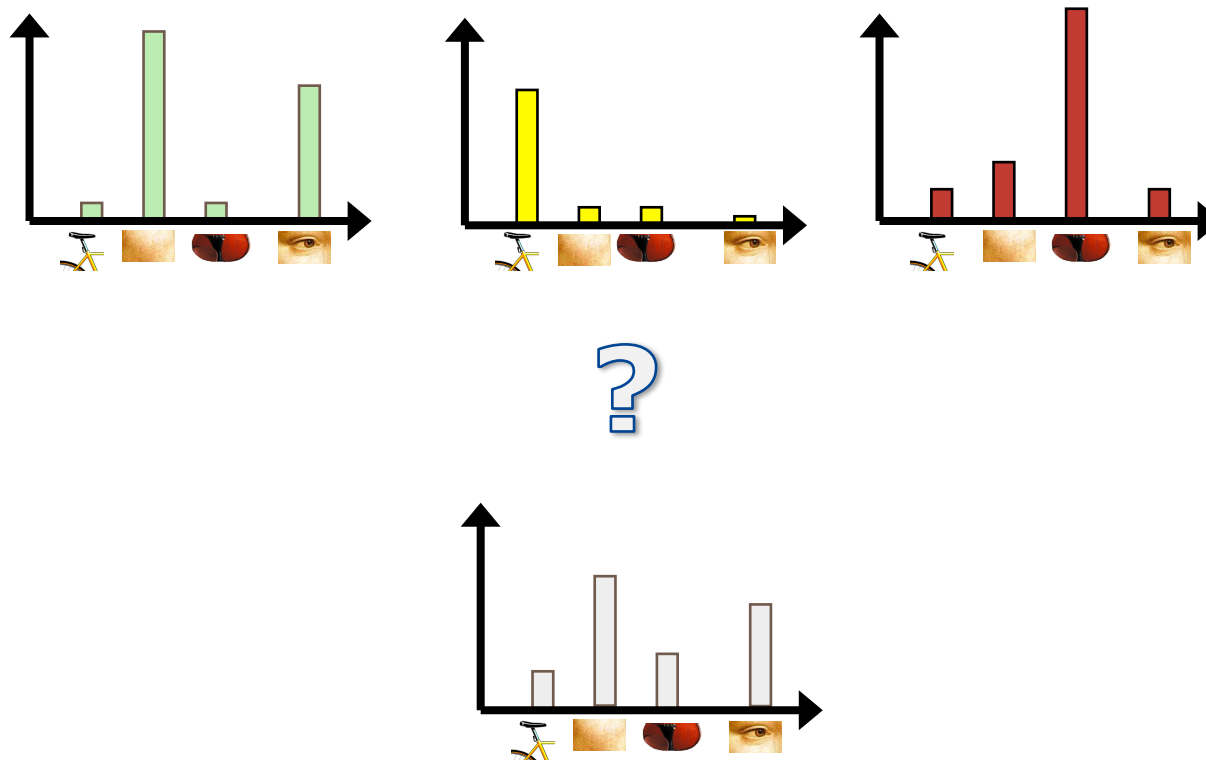


image source:  
L. Fei-Fei

# Histogram Comparison

- Place recognition task boils down to histogram comparisons





# 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

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

# 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!