“Bag of Words”: when is object recognition, just texture recognition?

A quiet meditation on the importance of trying simple things first…

16-721: Advanced Machine Perception
A. Efros, CMU, Spring 2009

Adopted from Fei-Fei Li, S.c. Zhu, and L.Walker Renninger

Object → **Bag of ‘words’**

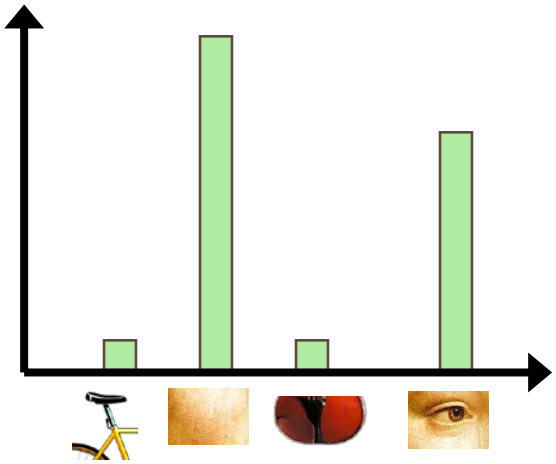
Analogy to documents

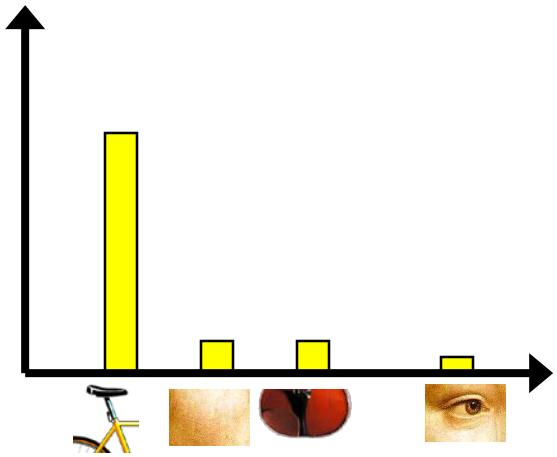
Of all the sensory impressions proceeding to the brain, the visual experiences are the dominant ones. Our perception of the world around us is based essentially on the messages that r… our eyes. For a long ti… retinal image wa… isual centers i… s a movie s… image … discove… know th… perceptio… more com… following the … ath to the various c… ortex, Hubel and Wiesel ha… demonstrate that the *message abou… image falling on the retina undergoes… -wise analysis in a system of nerve cell… stored in columns. In this system each c… has its specific function and is responsible… a specific detail in the pattern of the retinal image.*

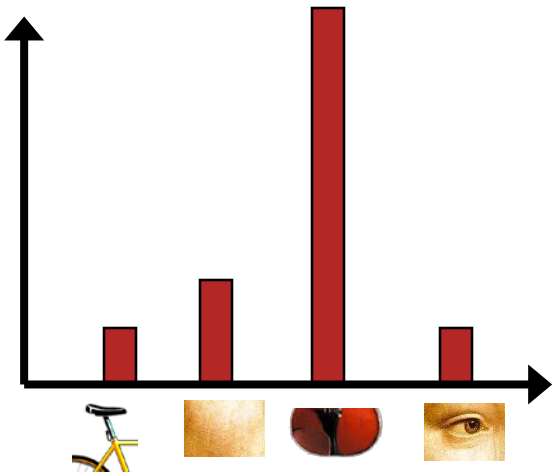
sensory, brain, visual, perception, retinal, cerebral cortex, eye, cell, optical nerve, image Hubel, Wiesel

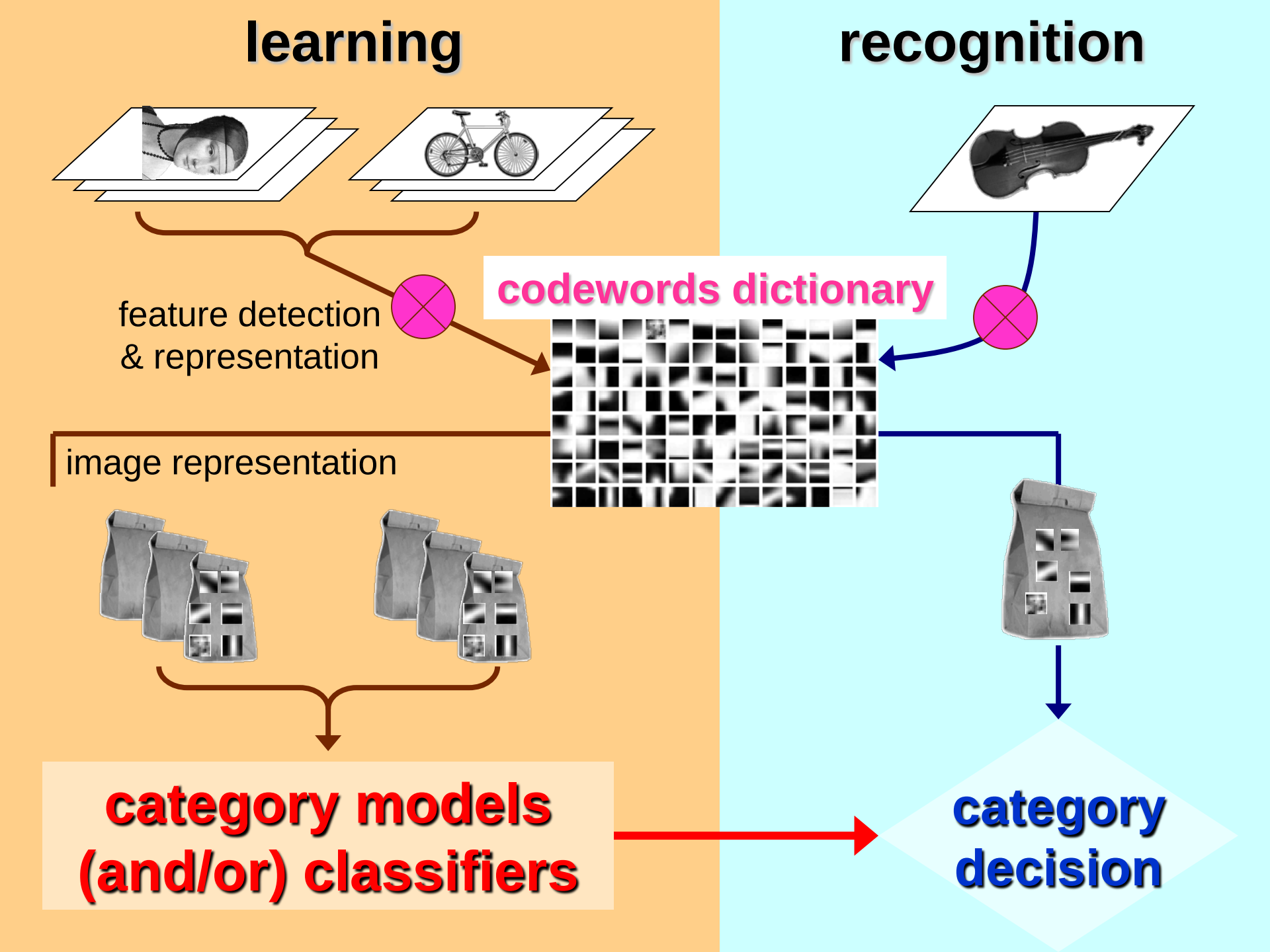
China is forecasting a trade surplus of $90bn (£51bn) to $100bn this year, a threefold increase on 2004's $32bn. The Commerce Ministry said the surplus would be created by a predicted 30% … $750bn, compared wi… $660bn. T… annoy th… China's … deliber… agrees … yuan is … governo… also need… demand so … country. China … e yuan against the dol… nd permitted it to trade within a narrow … but the US wants the yuan to be allowed … de freely. However, Beijing has made it cl… t it will take its time and tread carefully be… allowing the yuan to rise further in value.

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

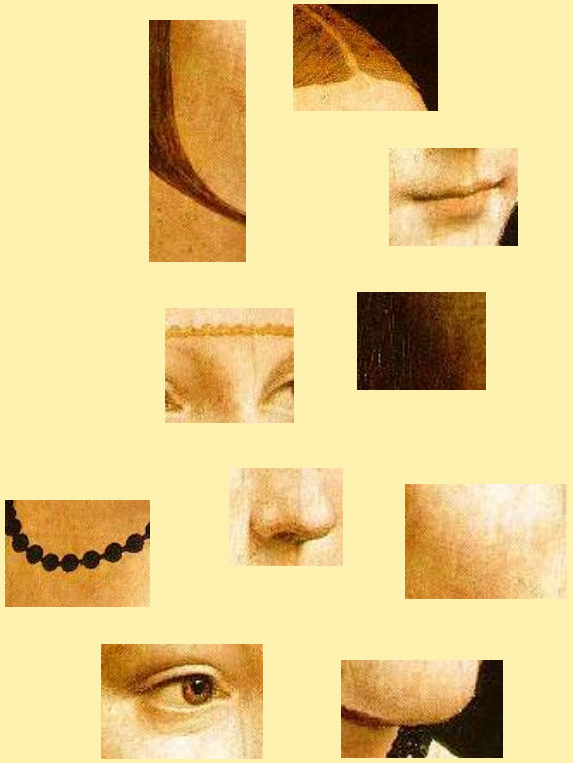







learning
recognition
feature detection
& representation
codewords dictionary
image representation
category models
(and/or) classifiers
category
decision

1.Feature detection and representation



Feature detection

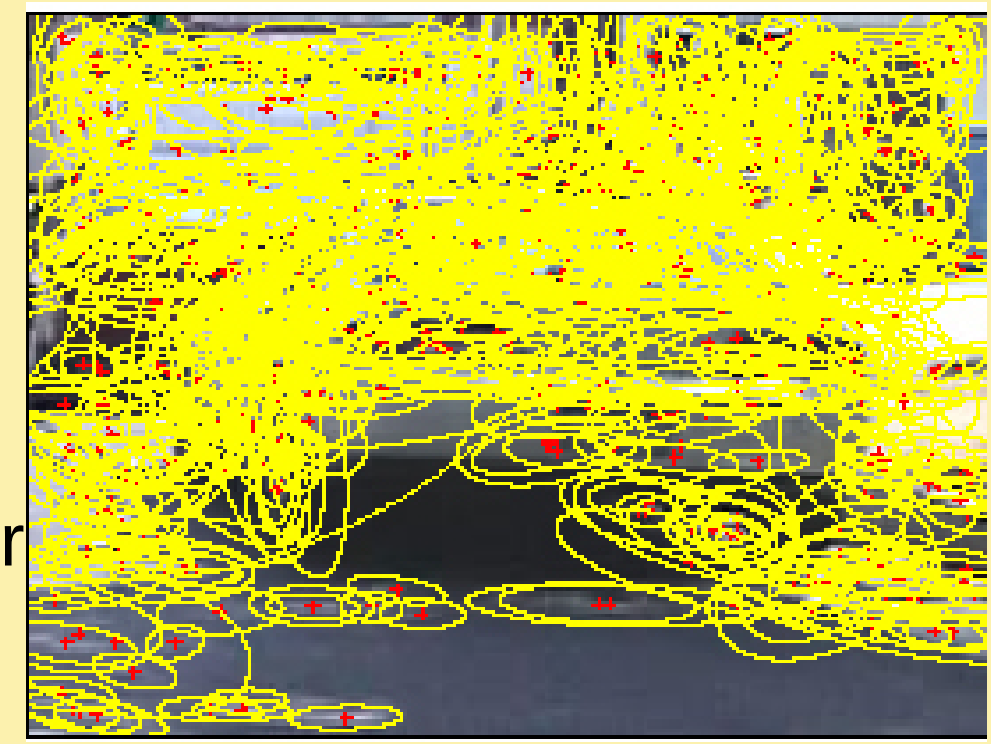
- Sliding Window
  - Leung et al, 1999
  - Viola et al, 1999
  - Renninger et al 2002

Feature detection

- Sliding Window
  - 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
  - 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
  - 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 (Ullman et al. 2002)
  - Segmentation based patches
    - Barnard et al. 2003, Russell et al 2006, etc.)

Feature Representation

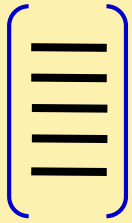
Visual words, aka textons, aka keypoints:

K-means clustered pieces of the image

- Various Representations:
  - Filter bank responses
  - Image Patches
  - SIFT descriptors

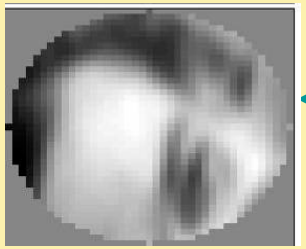
All encode more-or-less the same thing…

Interest Point Features

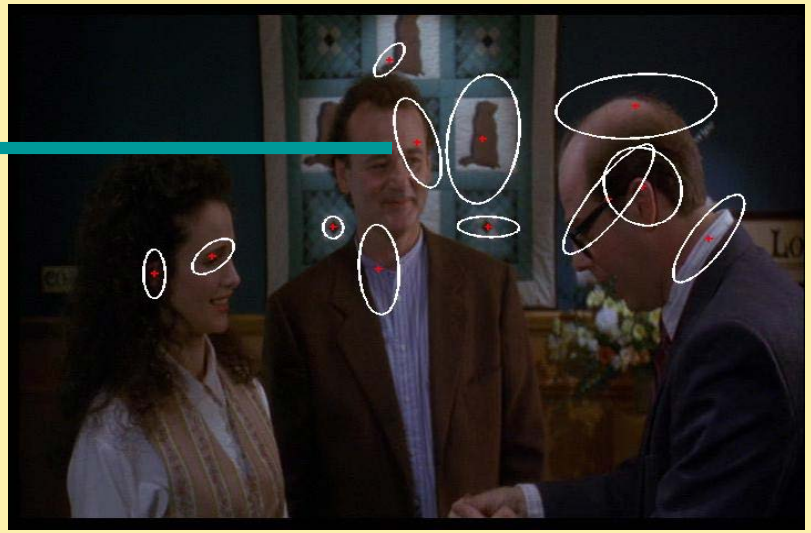


Compute SIFT descriptor

[Lowe'99]



Normalize patch



Detect patches

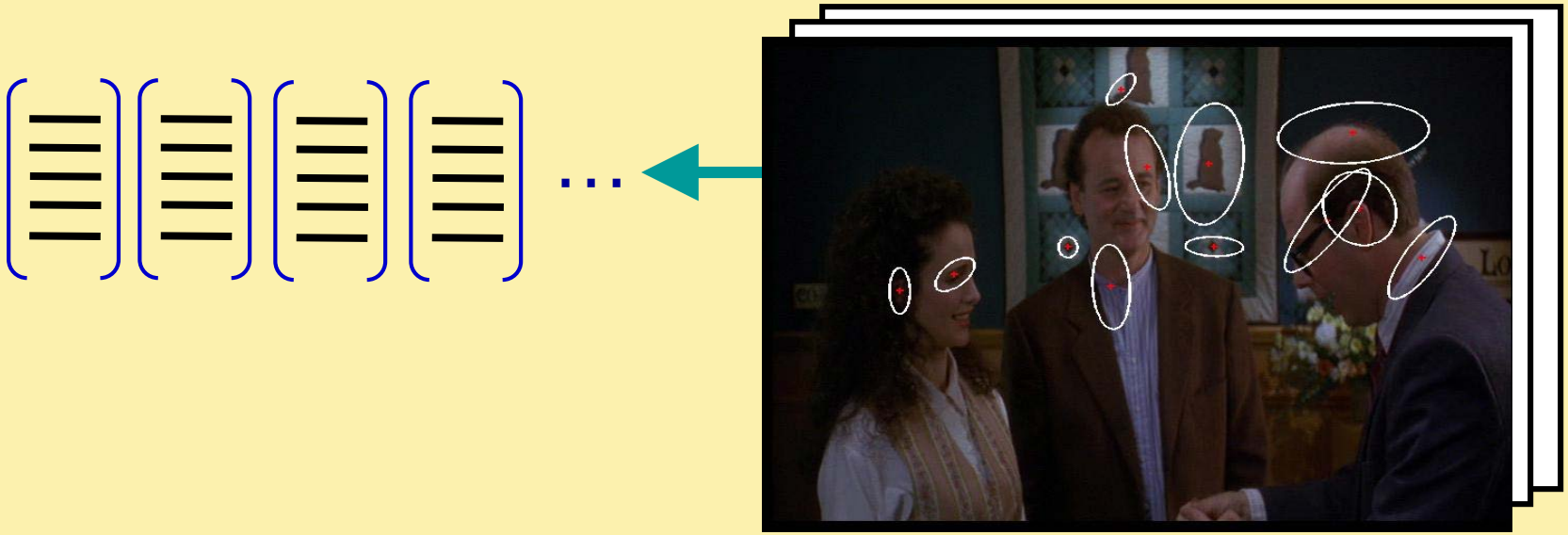
[Mikojaczyk and Schmid '02]

[Matas et al. '02]

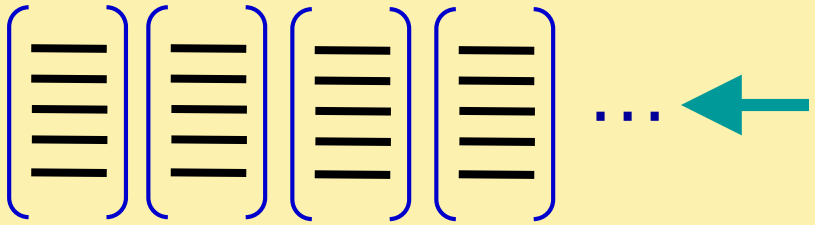
[Sivic et al. '03]

Slide credit: Josef Sivic

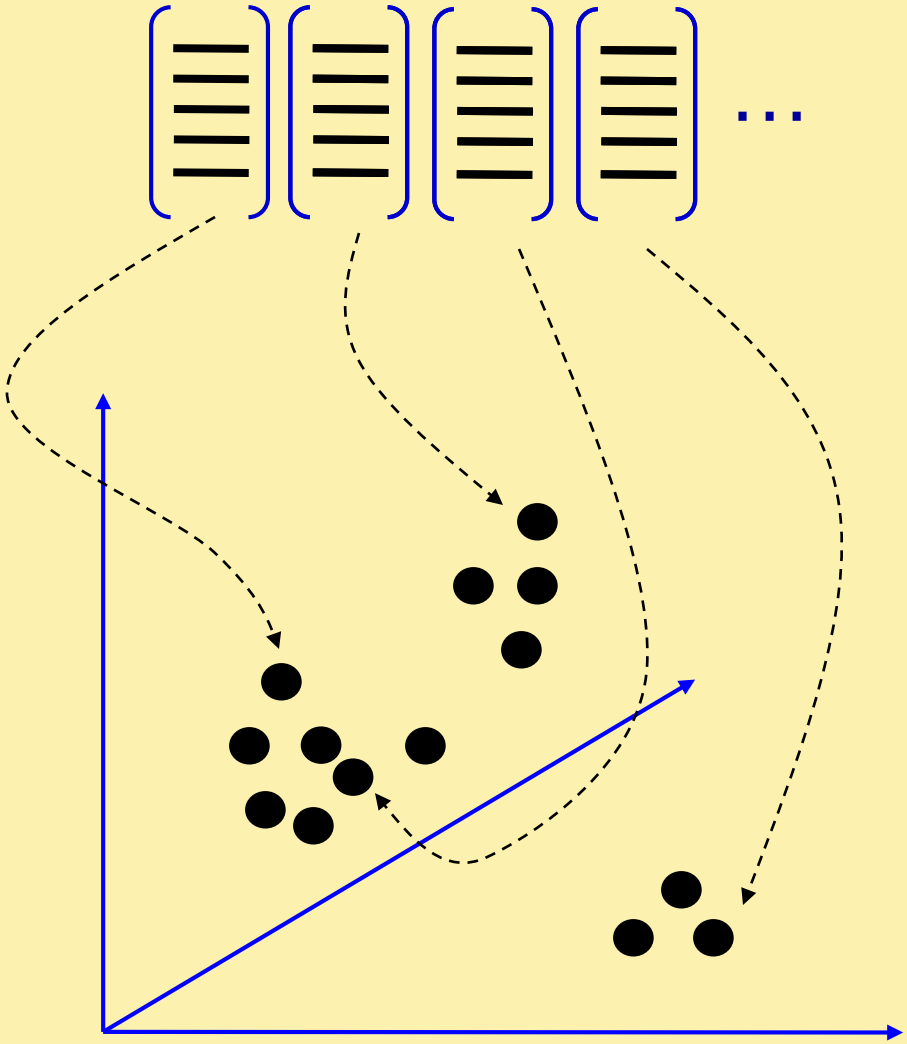
Interest Point Features



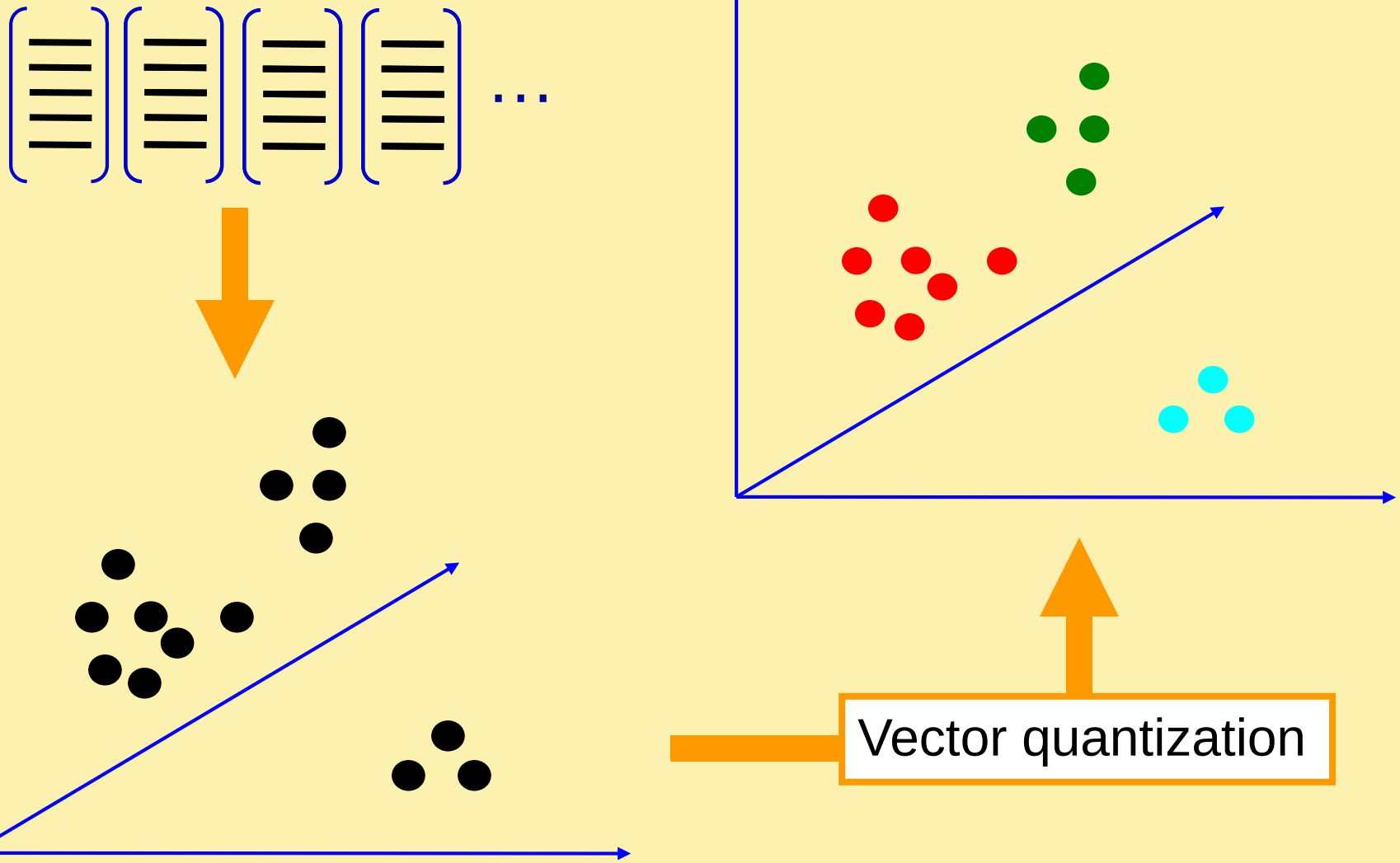
Patch Features



dictionary formation

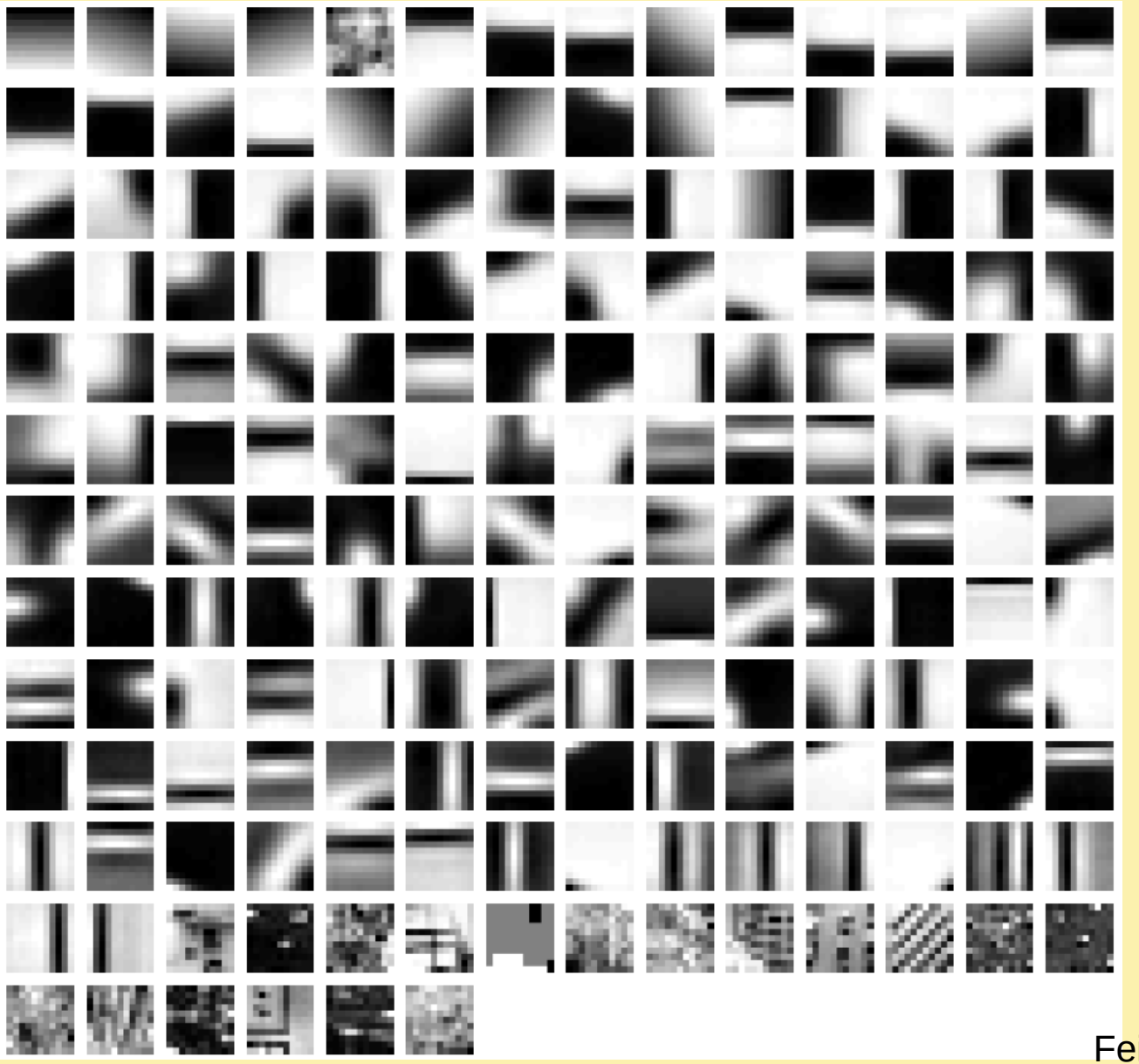


Clustering (usually k-means)



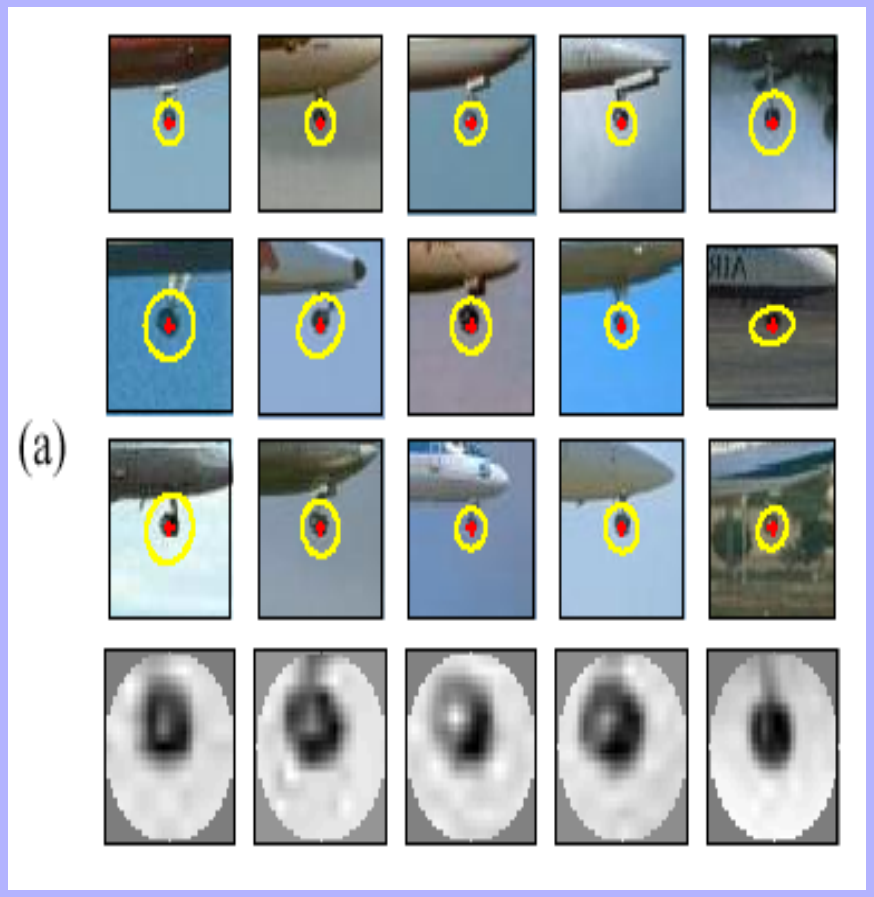
Slide credit: Josef Sivic

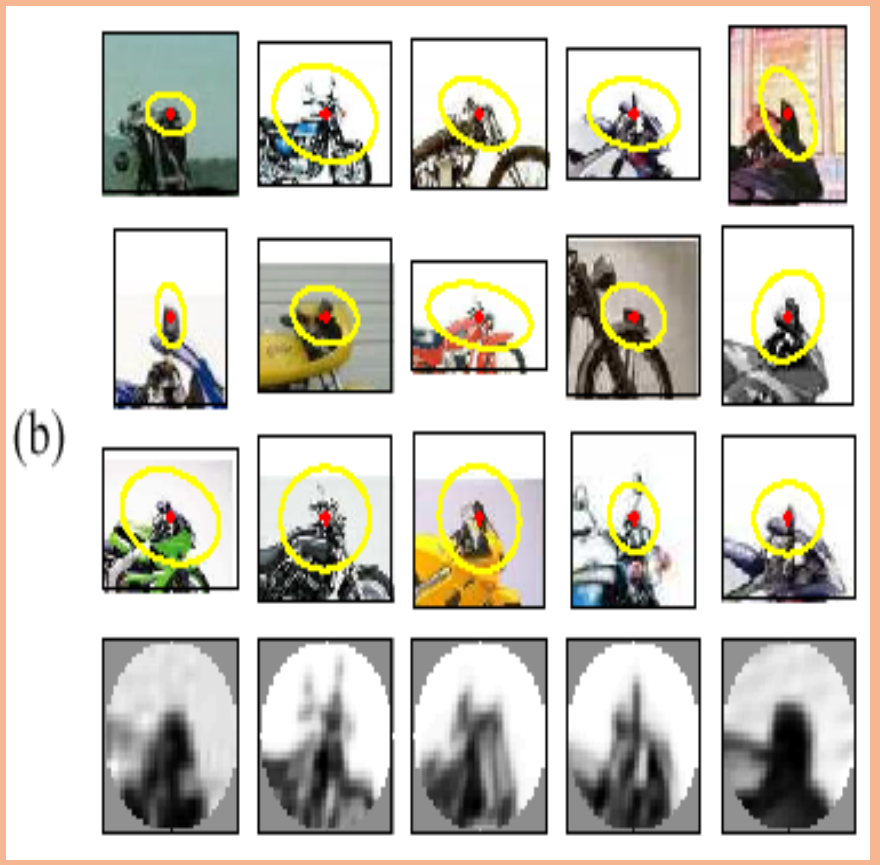
Clustered Image Patches



Fei-Fei et al. 2005

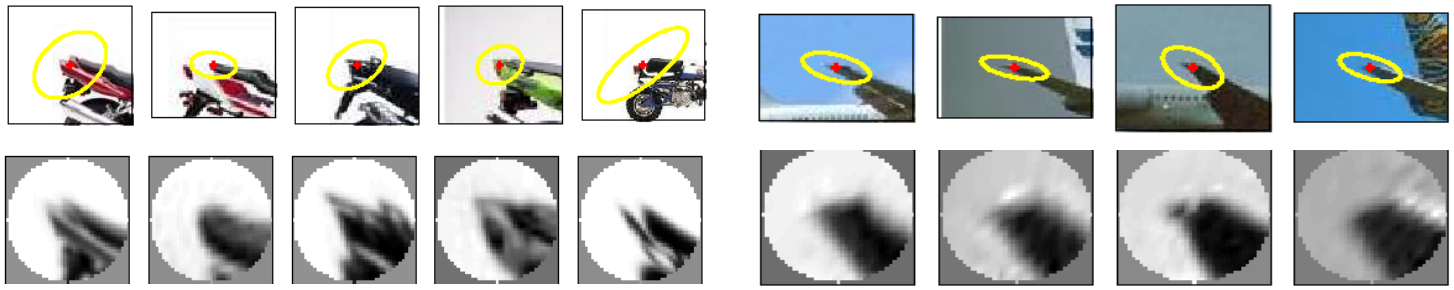
Image patch examples of codewords



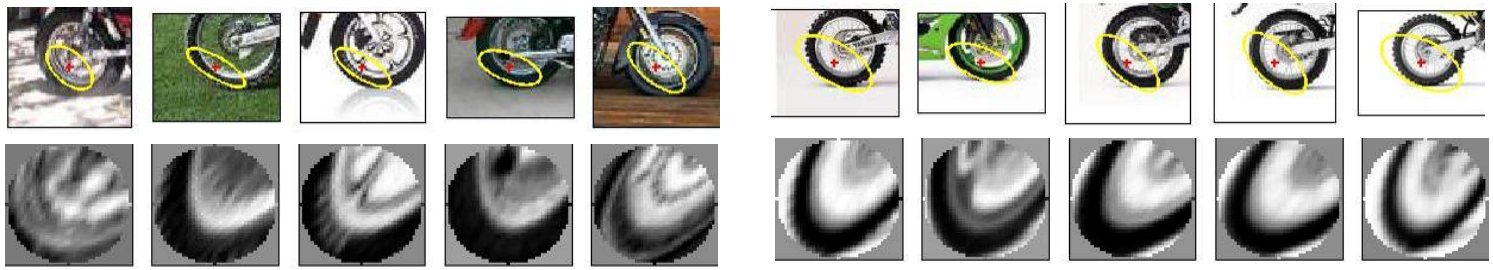


Sivic et al. 2005

Visual synonyms and polysemy

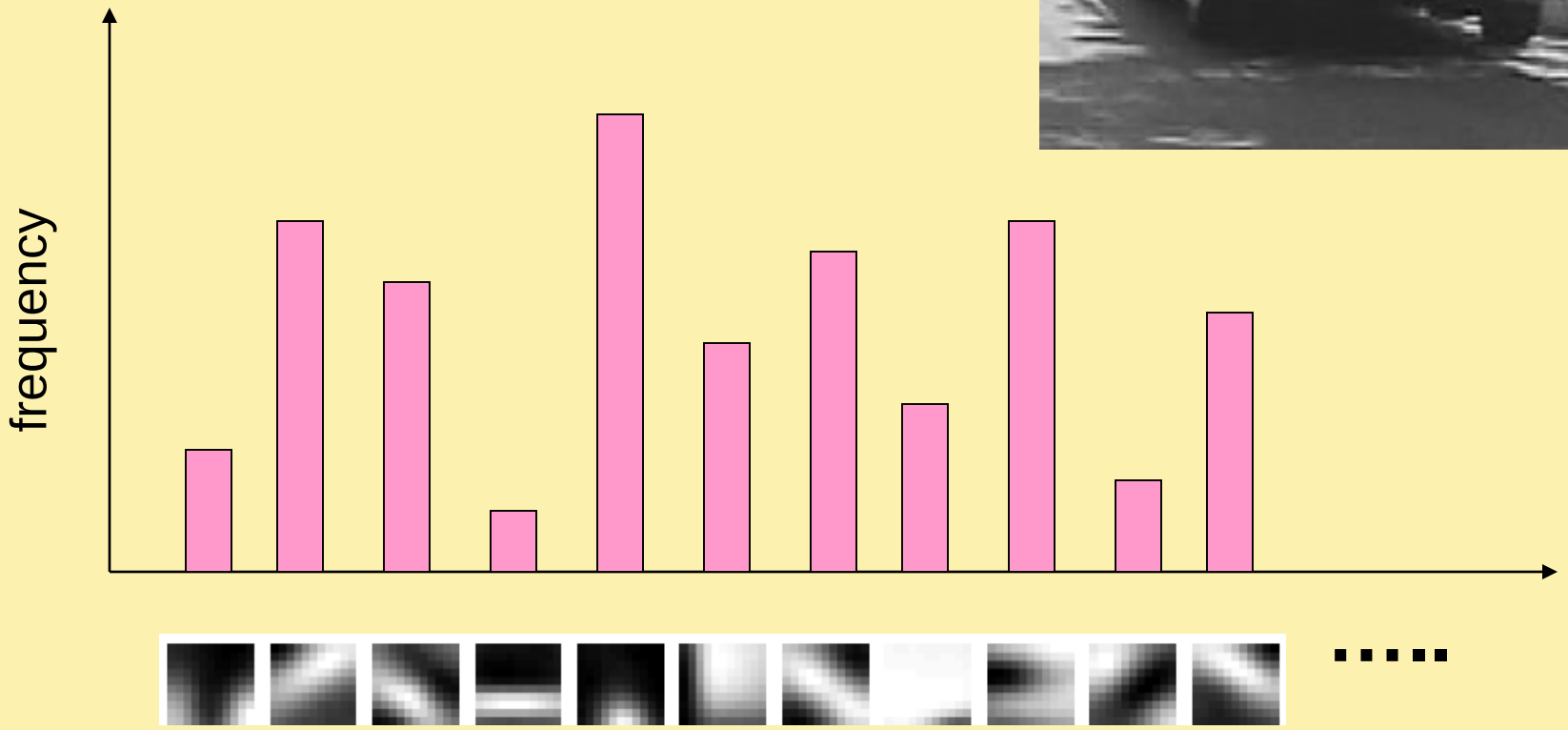


Visual Polysemy. Single visual word occurring on different (but locally similar) parts on different object categories.



Visual Synonyms. Two different visual words representing a similar part of an object (wheel of a motorbike).

Image representation



Scene Classification (Renninger & Malik)

beach

mountain

forest

city

street

farm

kitchen

livingroom

bedroom

bathroom