

“Bag of Words”



*A quiet meditation on the importance
of trying simple things first...*

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

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

2/27/2022

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 visual cortex, Hubel and Wiesel have demonstrated that the message about the image falling on the retina undergoes a fine-grained 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.

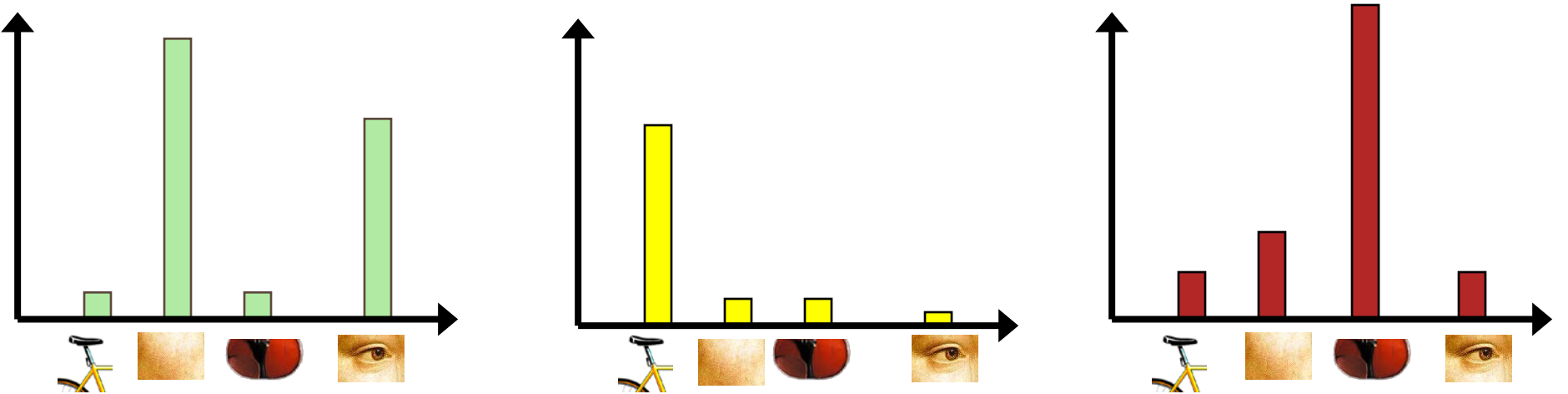
**sensory, brain,
visual, perception,
retinal, cerebral cortex,
eye, cell, optical
nerve, image
Hubel, Wiesel**

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.

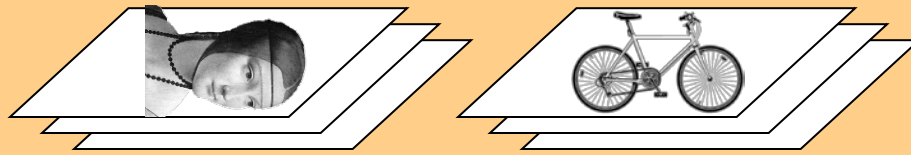
The increase in exports will be partly due to a deliberate policy to encourage exports, but also to a more competitive yuan is expected to be a factor. China's government also needs to meet the demand for foreign currency.

China's government also needs to meet the demand for foreign currency. China's government also needs to meet the demand for foreign currency. China's government also needs to meet the demand for foreign currency.

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



learning



feature detection
& representation

codewords dictionary

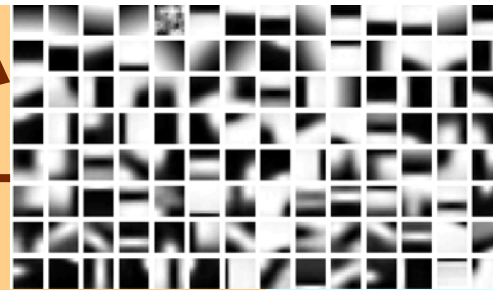
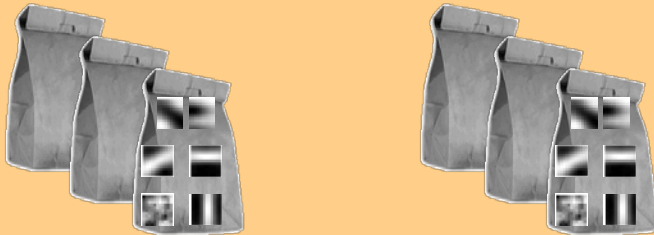


image representation



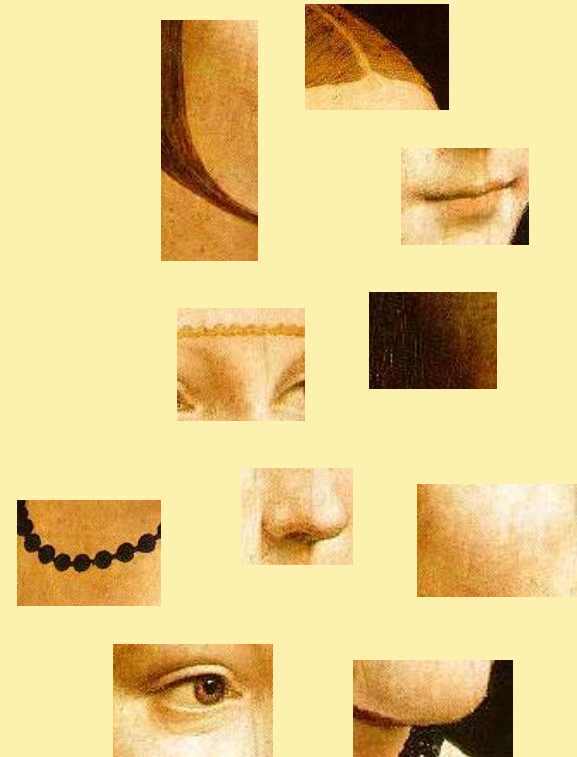
**category models
(and/or) classifiers**

recognition



**category
decision**

1.Feature detection and representation



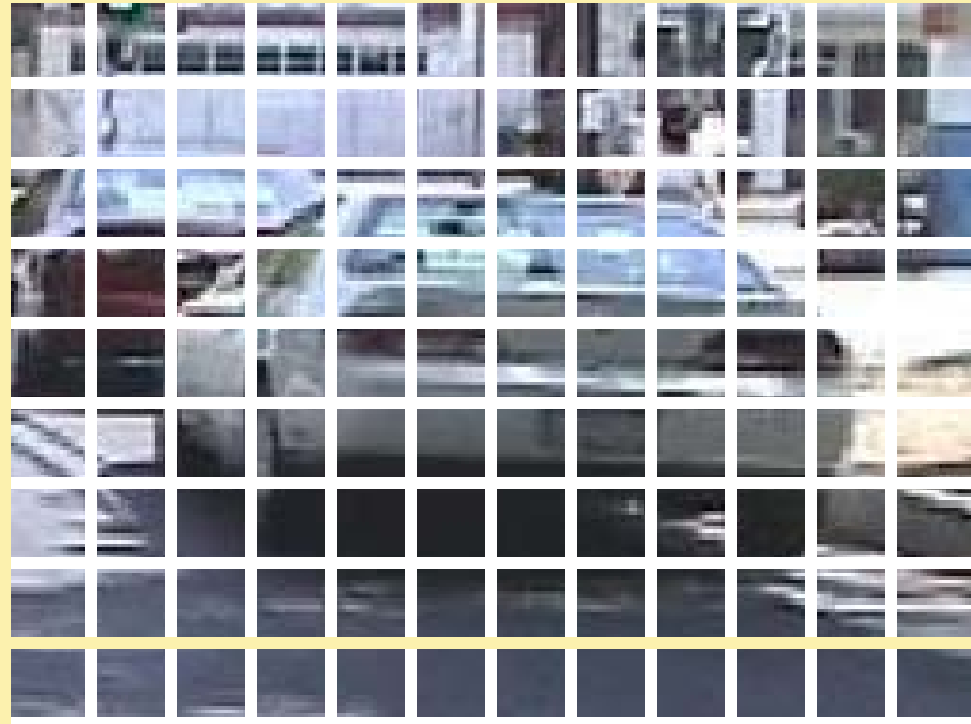
Feature detection

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



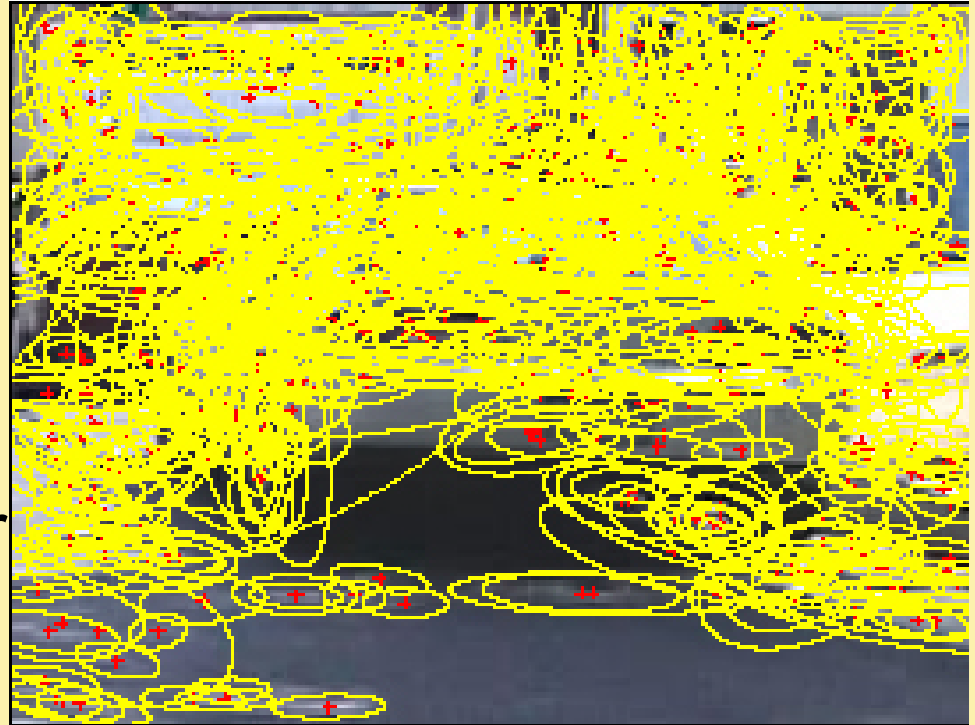
Feature detection

- 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

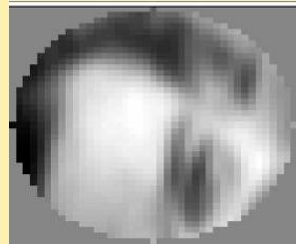
Visual words, aka textons, aka keypoints:
K-means clustered pieces of the image

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

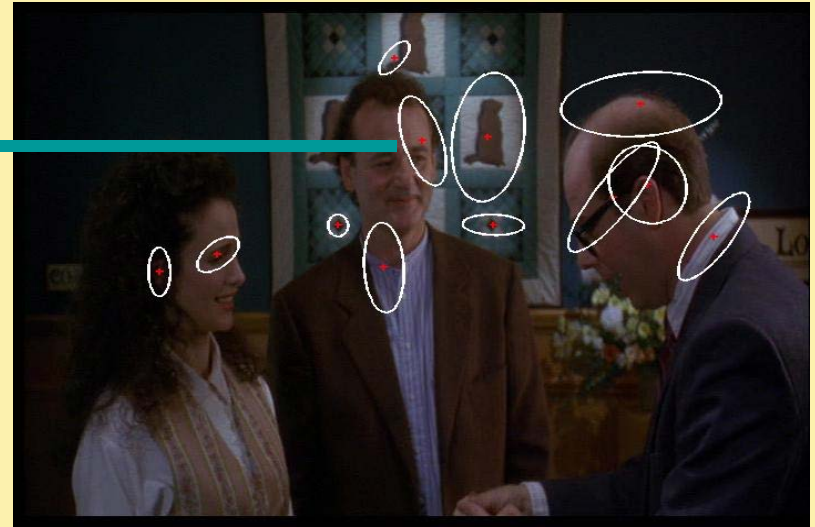
All encode more-or-less the same thing...

Interest Point Features


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



**Normalize
patch**



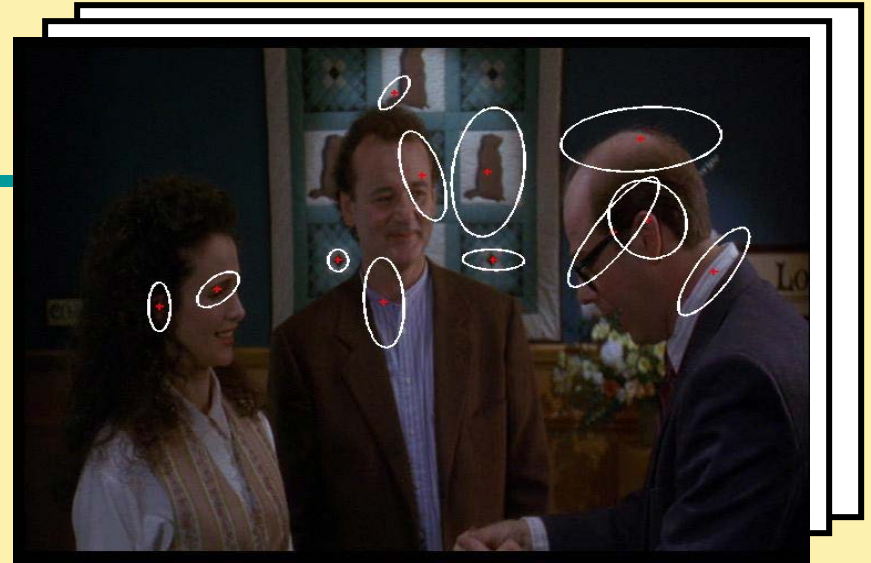
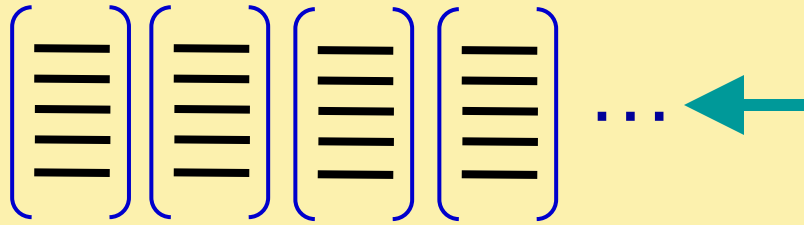
Detect patches

[Mikojczyk and Schmid '02]

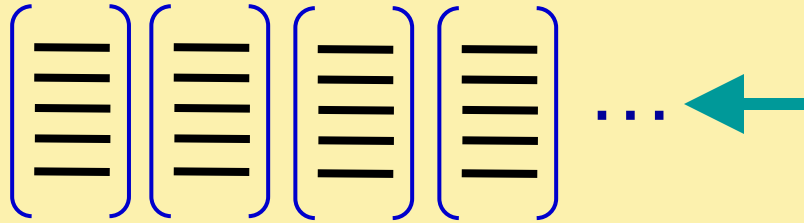
[Matas et al. '02]

[Sivic et al. '03]

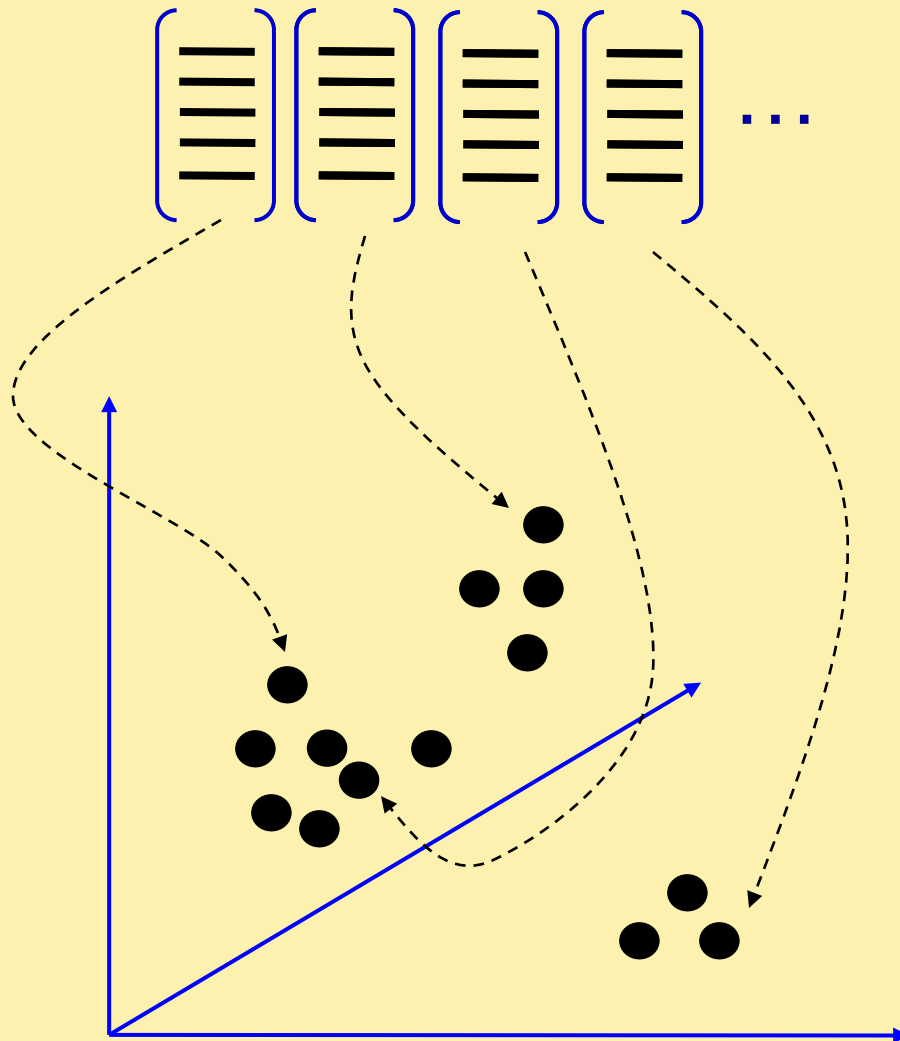
Interest Point Features



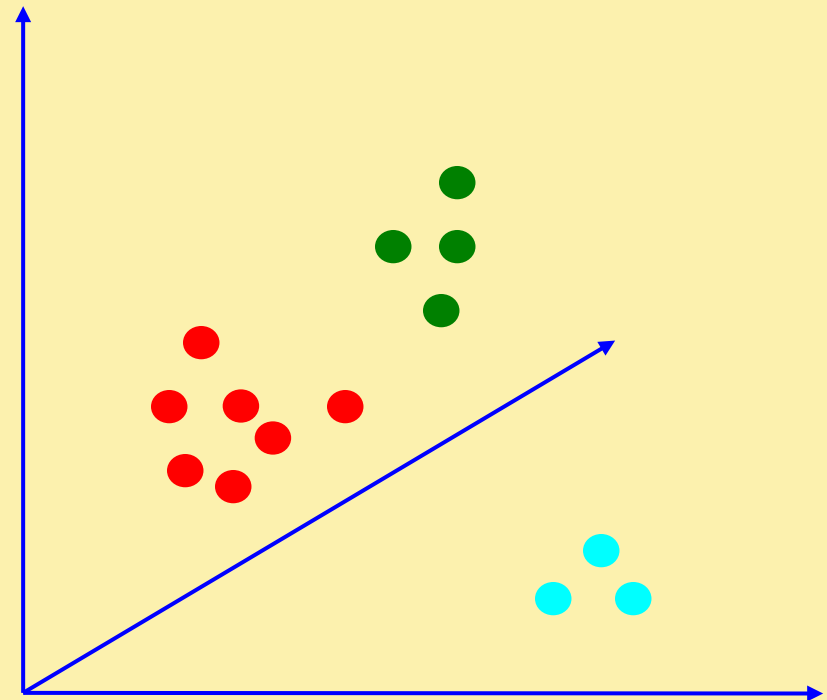
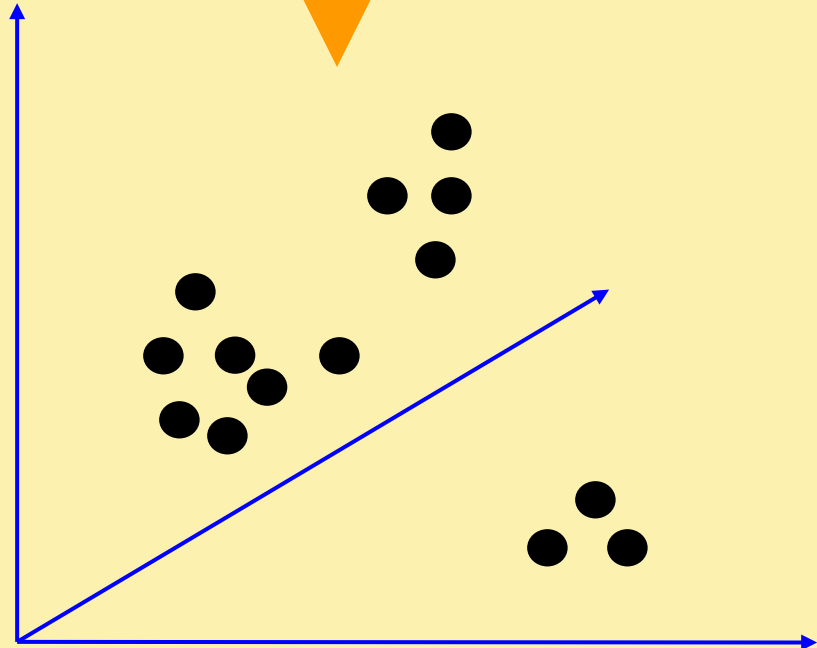
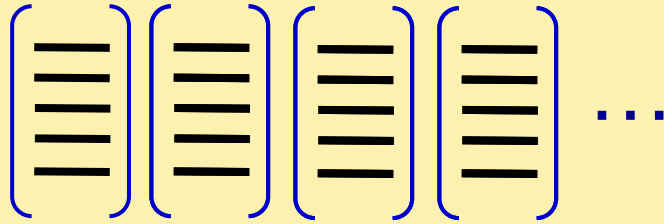
Patch Features



dictionary formation

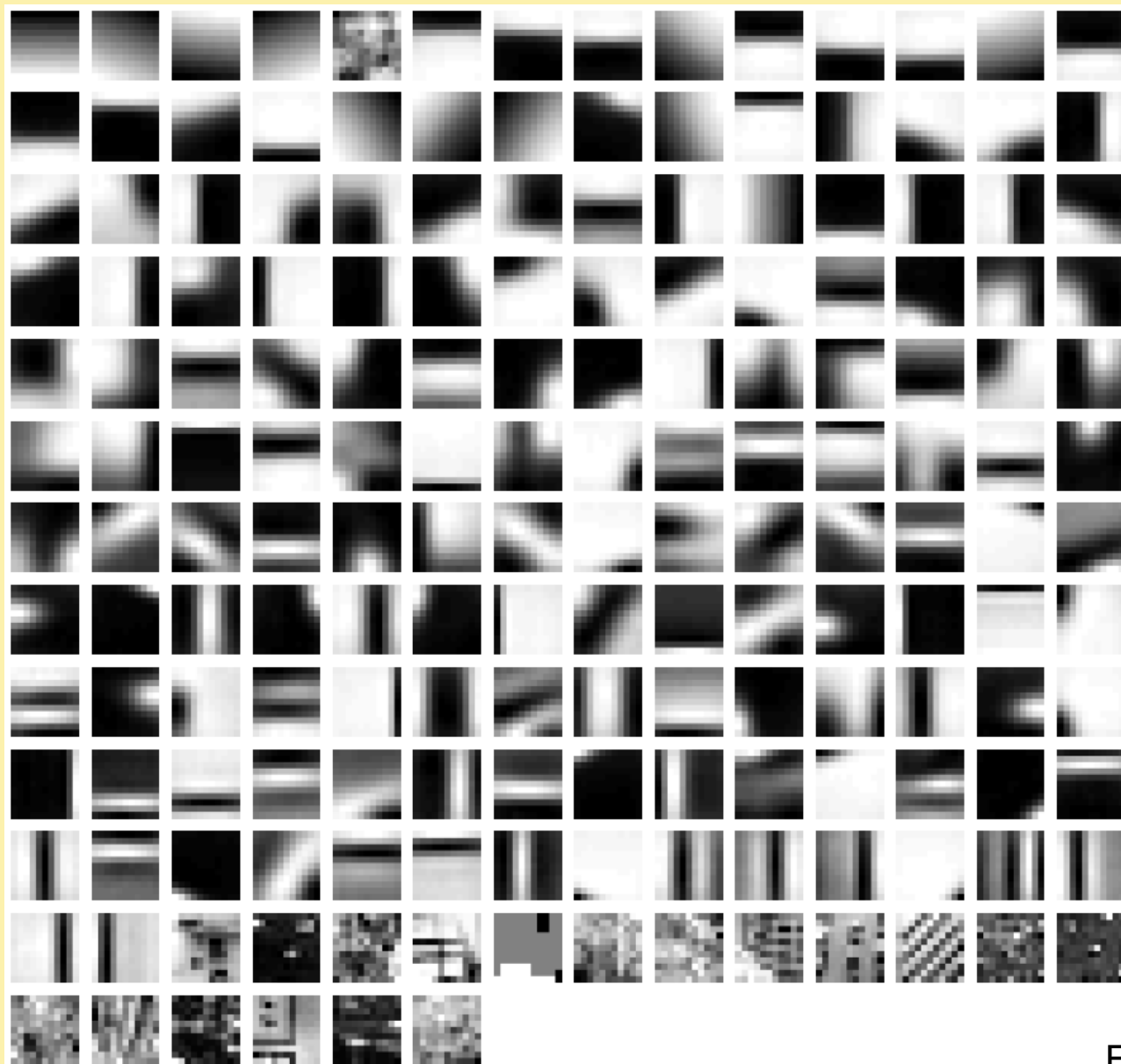


Clustering (usually k-means)



Vector quantization

Clustered Image Patches



2/27/2022

17

Fei-Fei et al. 2005

Filterbank

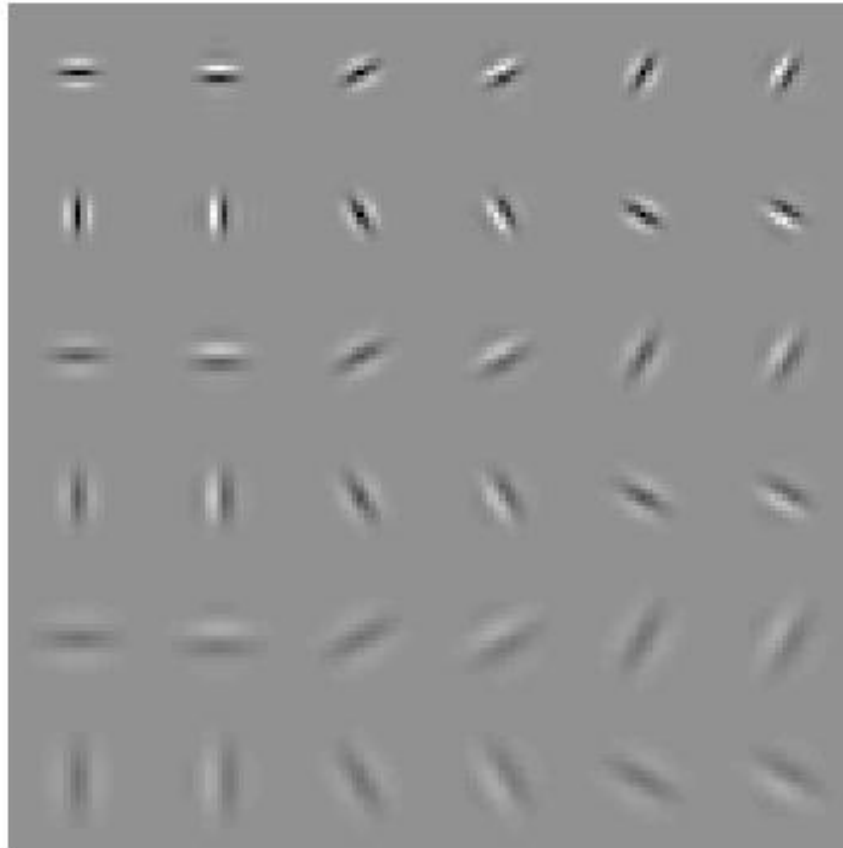
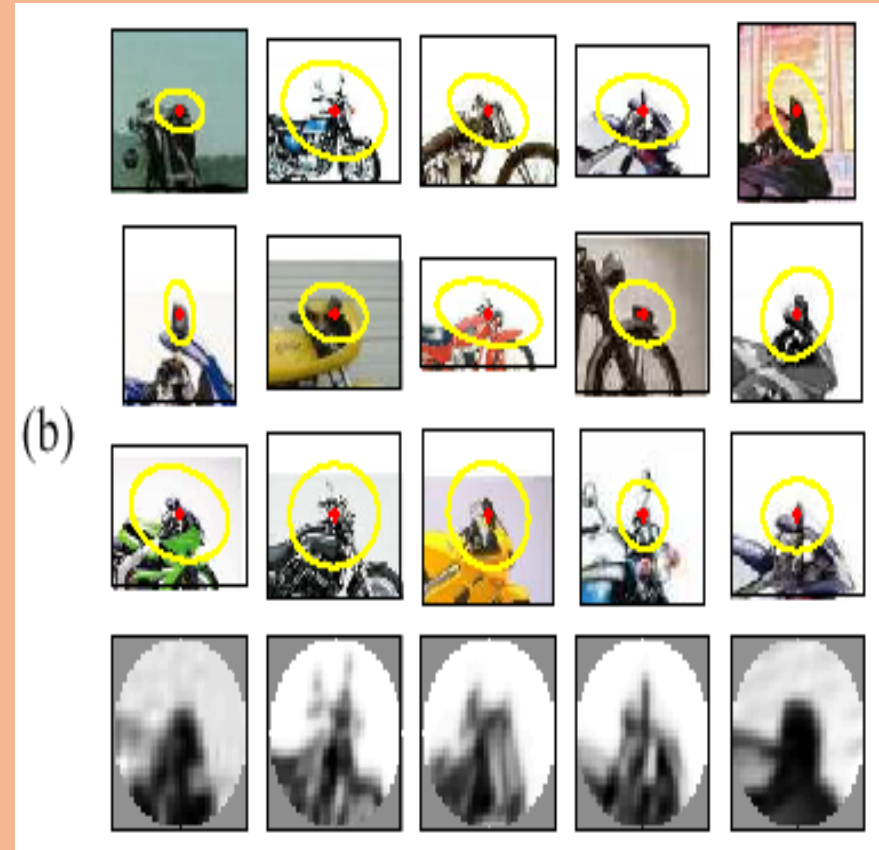
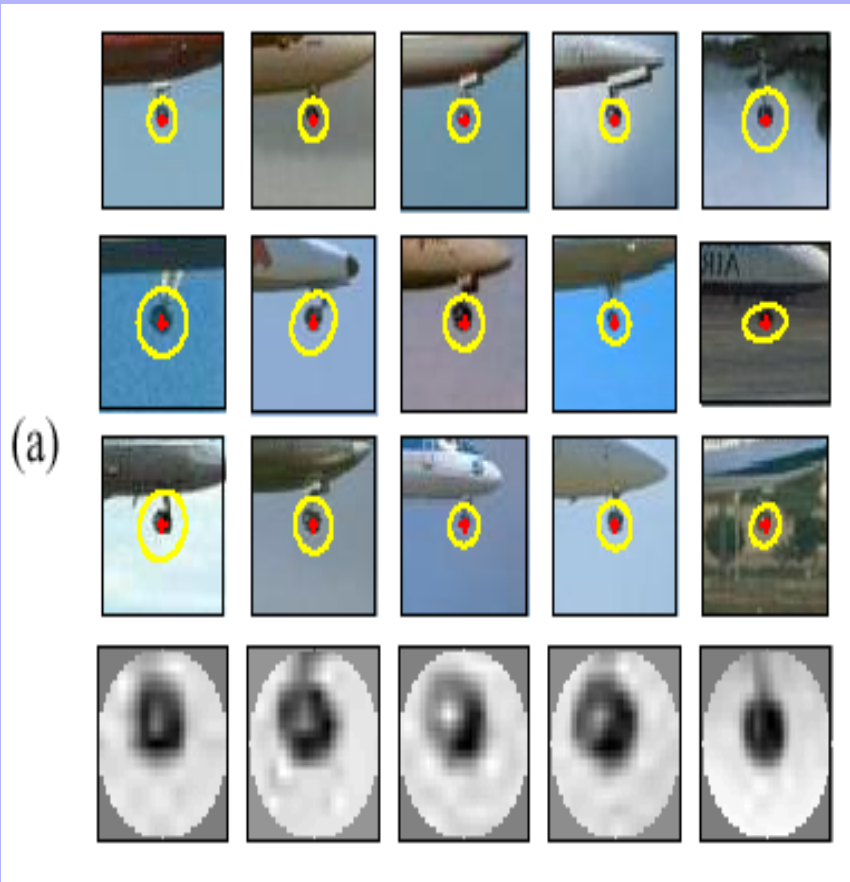
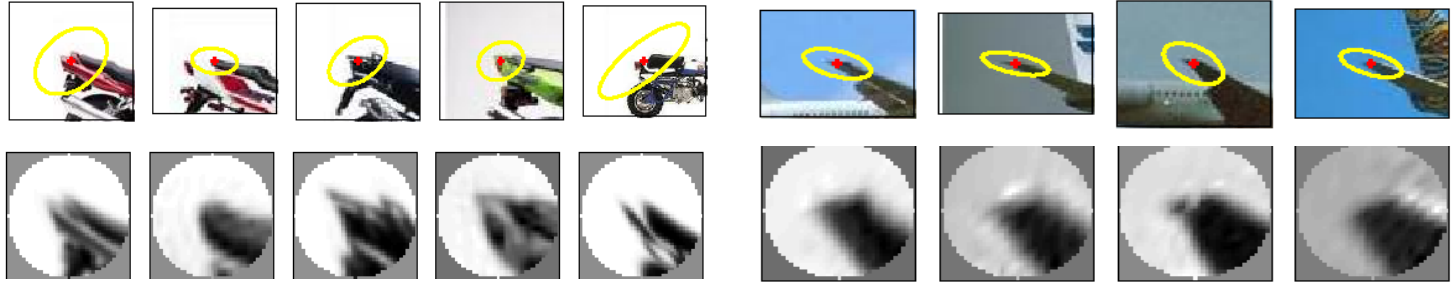


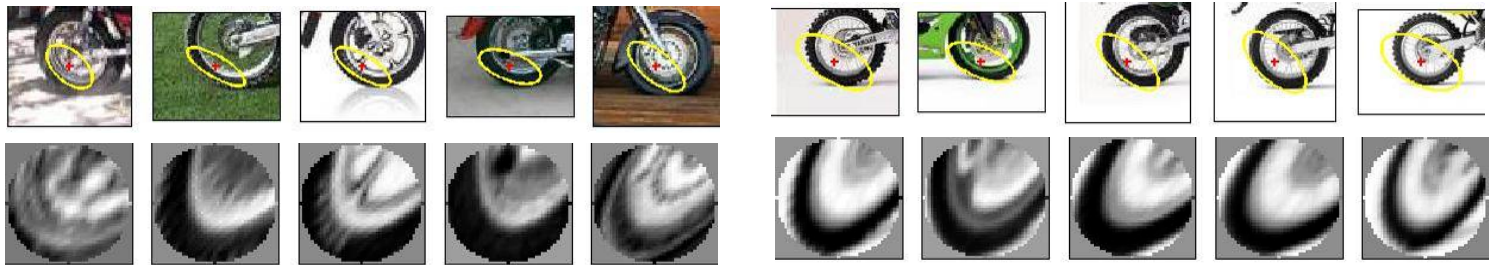
Image patch examples of codewords



Visual synonyms and polysemy

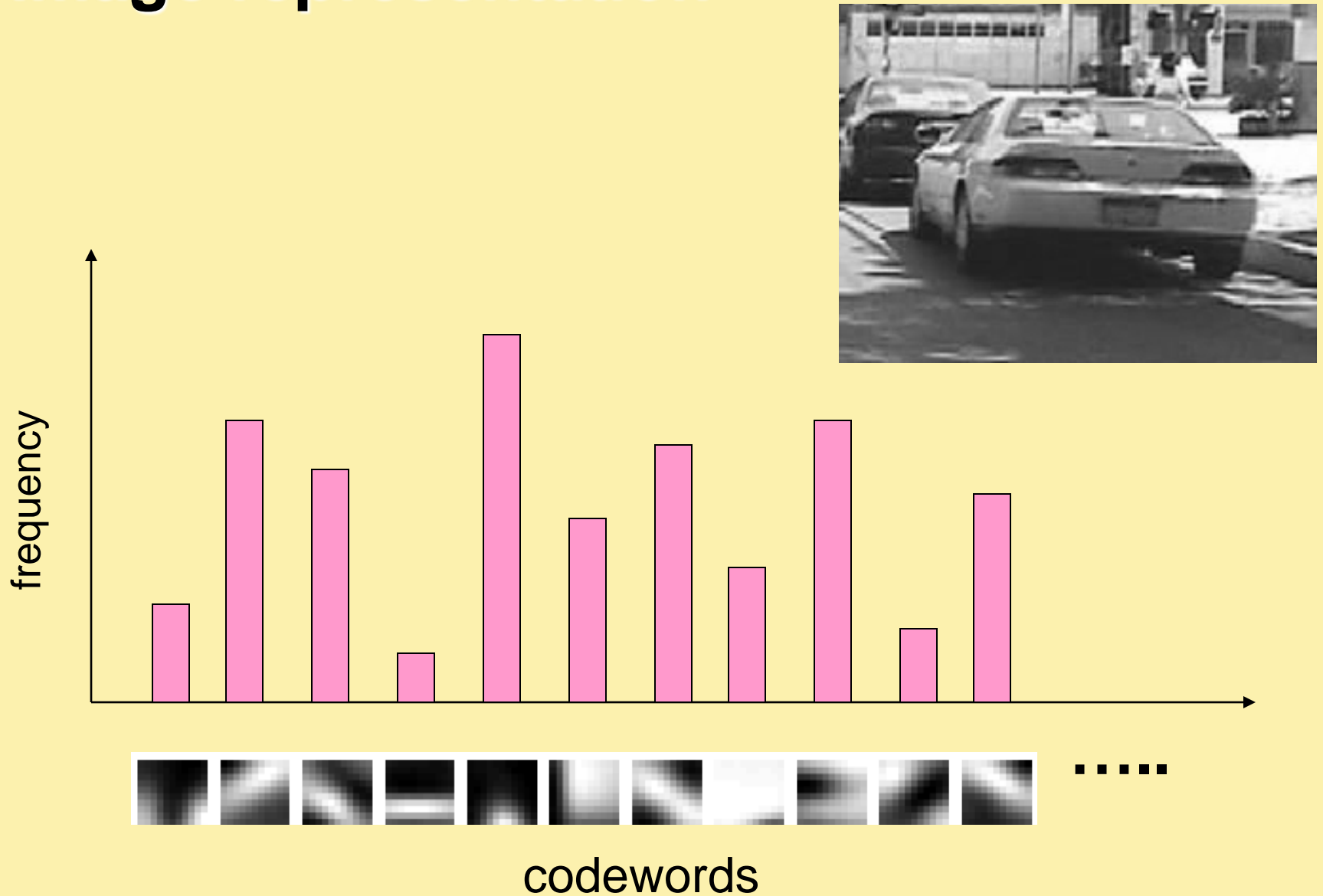


Visual Polysemy. Single visual word occurring on different (but locally similar) parts on different object categories.



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

