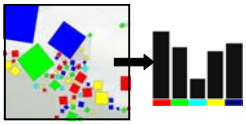
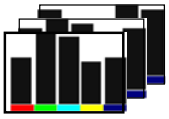


Bags of words

Thursday, February 15



Object → **Bag of 'words'**

Slide from ICCV 2005 short course, L. Fei-Fei

Analogy to documents

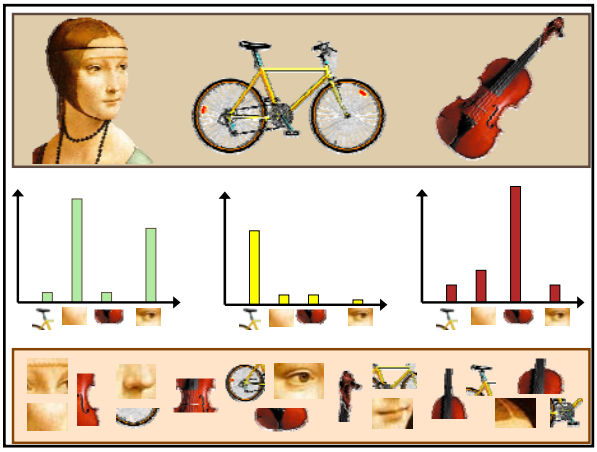
Of all the sensory impressions proceeding to the brain, the visual experiences are the dominant ones. Our perception of the world around us is based essentially on the messages that ... our eyes. For a long ti... retinal image wa... visual centers ... a movie ... image ... discov... know t... perceptio...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

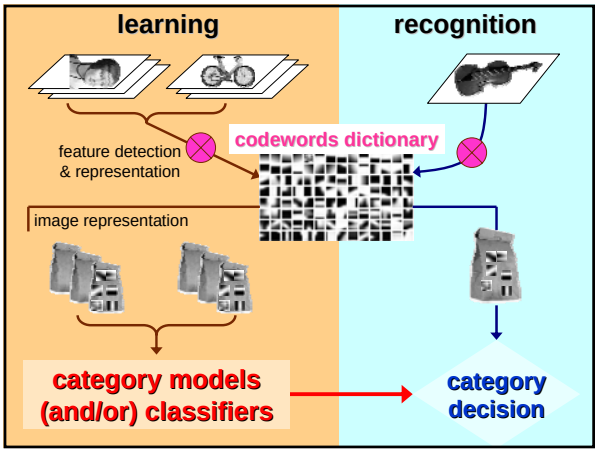
sensory, brain, visual, perception, retinal, cerebral cortex, eye, cell, optical nerve, image Hubel, Wiesel

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 ... $660bn. ... annoy t... China' ... delibe... agrees ... yuan is ... governo... also nee... demand so... country. China... yuan against the ... and permitted it to trade within a narrow ... but the US wants the yuan to be allowed ... freely. However, Beijing has made it c... it will take its time and tread carefully b... allowing the yuan to rise further in value.

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

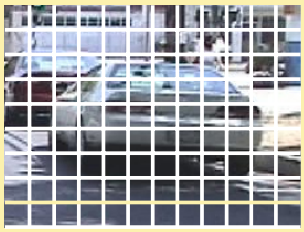
Slide from ICCV 2005 short course, L. Fei-Fei





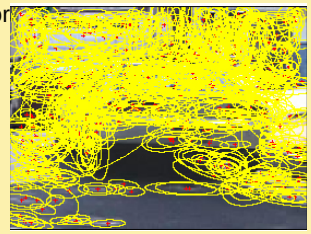
1.Feature detection and representation

- Regular grid



1.Feature detection and representation

- Regular grid
- Interest point detector



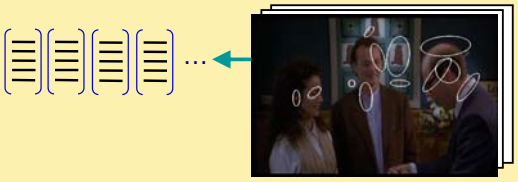
1.Feature detection and representation

- Regular grid
- Interest point detector
- Other methods
  - Random sampling (Ullman et al. 2002)
  - Segmentation based patches (Barnard et al. 2003)

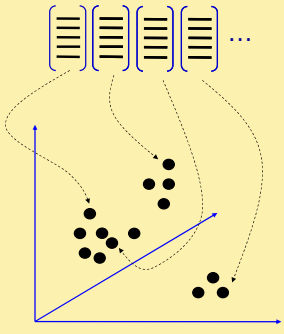
1.Feature detection and representation

Slide credit: Josef Sivic

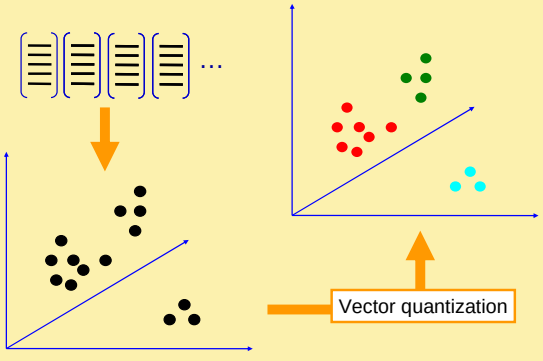
1.Feature detection and representation



2. Codewords dictionary formation

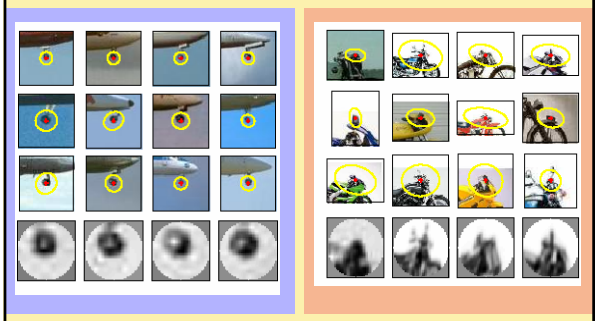


2. Codewords dictionary formation



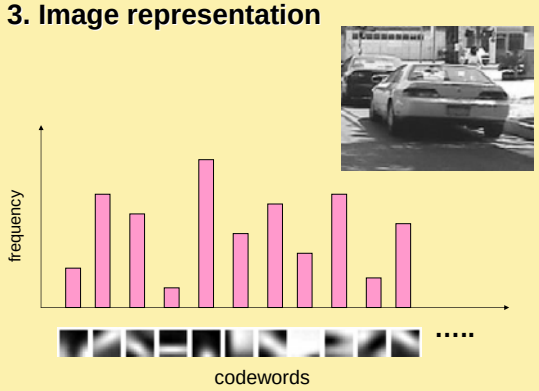
Slide credit: Josef Sivic

Image patch examples of codewords

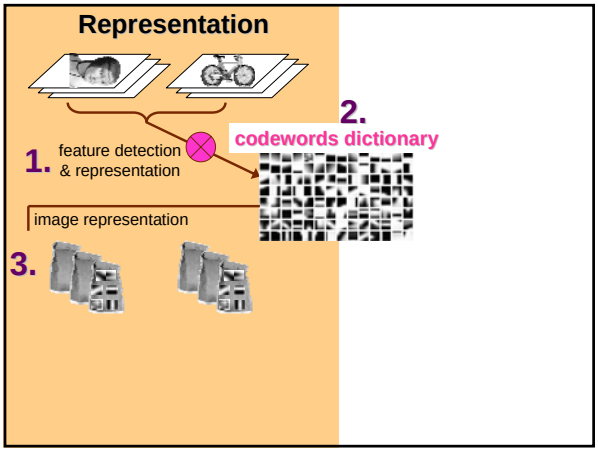


Sivic et al. 2005

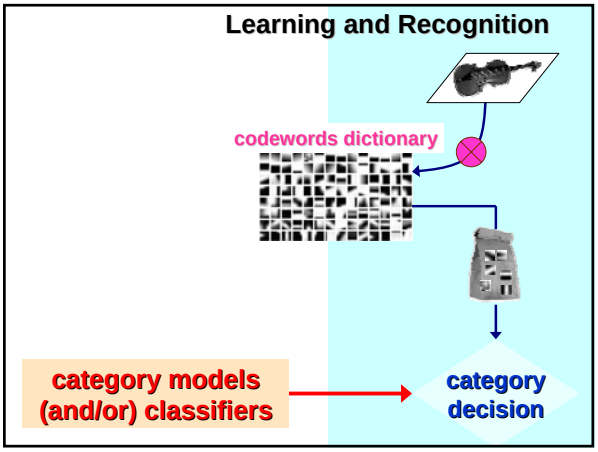
3. Image representation



Representation



Learning and Recognition



Models and classifiers

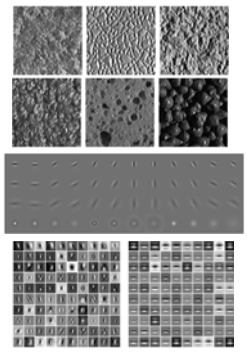
Various approaches explored with bag-of-words representation:

- Naïve Bayes classifiers
- Support Vector Machines (SVM)
- Hierarchical Bayesian text models (pLSA and LDA)
- Distance-based comparisons (chi-squared, EMD, dot product)

For textures, materials

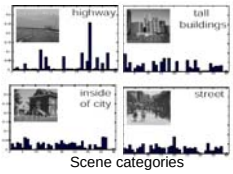
- *Texton* = cluster center of filter responses over collection of images [Leung and Malik, 1999]
- Represent texture or material with histogram of texton occurrences (or prototypes of whatever feature type employed)
- Compare histograms with chi-squared or other distribution distance

Leung and Malik, 1999.
Varma and Zisserman, 2002
Lazebnik, Schmid, and Ponce, 2003
Hayman et al., 2004



For scenes

Shot matching
Sivic et al., 2003

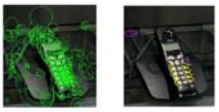


Scene categories
Fei-Fei and Perona, 2005

Scene categories
Quelhas et al., 2005

For objects

Similar ideas, but use bag of words description to build classifiers (e.g. SVM), fit statistical models (e.g. pLSA)

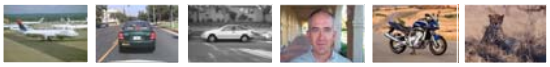


Csurka et al., 2003,
Dorko and Schmid, 2004
Sivic et al., 2003, 2005
Sudderth et al. 2005
…

Advantages

- Flexibility comes with ignoring geometry (?)
- Compact description, yet rich
- Local features → vector
  - Usable representation
  - Relatively efficient learning
- Yields good results in practice

For example…



class	ours	other results	
	Zhang et al. (2005)	Willamowski et al. (2004)	Fergus et al. (2003)
airplanes	**98.8**	97.1	90.2
cars (rear)	98.3	**98.6**	90.3
cars (side)	**95.0**	87.3	88.5
faces	**100**	99.3	96.4
motorbikes	**98.5**	98.0	92.5
spotted cats	**97.0**	—	90.0

bag of features

constellation model

See also recent object recognition challenge results.

Issues

- Flexibility comes with ignoring geometry (!)
- Background/foreground treated at once
- Vocabulary formation
  - Number of words/clusters?
  - Universal, or dataset specific?
  - May be expensive
- How to localize/segment object?

Next week: ideas for vocabulary formation and introducing spatial constraints into a bag of words.