A Quick Introduction to Bag-of-Words
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 from our eyes. For a long time it was thought that the retinal image was transmitted point by point to visual centers in the cerebral cortex; the cerebral cortex was a movie screen, so to speak, upon which the image in the eye was projected. Through the discoveries of Hubel and Wiesel we now know that behind the origin of the visual perception in the brain there is a considerably more complicated course of events. By following the visual impulses along their path to the various cell layers of the retinal and cerebral cortex, Hubel and Wiesel have been able to demonstrate that the message about the image falling on the retina undergoes a step-wise 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.

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 to $750bn, compared with a 18% rise in imports to $660bn. The figures are likely to further annoy the US, which has long argued that China's exports are unfairly helped by a deliberately undervalued yuan. Beijing agrees the surplus is too high, but says the yuan is only one factor. Bank of China governor Zhou Xiaochuan said the country also needed to do more to boost domestic demand so more goods stayed within the country. China increased the value of the yuan against the dollar by 2.1% in July and permitted it to trade within a narrow band, but the US wants the yuan be allowed to trade freely. However, Beijing has made it clear that it will take its time and tread carefully before allowing the yuan to rise further in value.

image source: L. Fei-Fei
Bag of Visual Words

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

- Visual words = independent 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.
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
Feature Detection

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

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

example patch

slide adapted from: L. Fei-Fei
Learning the Dictionary

slide adapted from: L. Fei-Fei
Learning the Dictionary

cluster center = code words

clustering, e.g., k-means

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.
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

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