Bag of Words Model
And
The Pyramid Match

Sid Ying-Ze Bao

This slide is partially based on
1 CVPR tutorial Bag-of-words models by Fei-fei Li
2 Spatial Pyramid Matching for Recognizing Natural Scene Categories by Svetlana Lazebnik
Image Classification

• "Person" or "Car"?
Image Classification

- Notorious Difficult Problem

![Diagram showing car, person, and house being classified by a computer system.]

- Car
- Person
- House
Image Classification

- Intra-class variation
Image Classification

- Tons of categories

- Bathroom
- Bedroom
- Bridge
- Building
- Coast
- Field
- Forest
- Street
- Kitchen
- Waterfall
Image Classification Methods

- Bag of words
- Pyramid Match

All Image Classification Methods
Bag of Words

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 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 brain. The cerebral cortex was like a movie screen upon which the image was projected. Through the discoveries of Hubel and Wiesel we now know that the perception of the visual image is a considerably more complex process involving the various cell layers of the optical cortex. By following the visual impulses along their path to the various cell layers of the optical cortex, Hubel and Wiesel have demonstrated that the message about the image falling on the retina undergoes step-by-step 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% jump in exports to $750bn, compared with a 18% rise in imports to $660bn. This is 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 to be allowed to rise 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.

China, trade, surplus, commerce, exports, imports, US, yuan, bank, domestic, foreign, increase, trade, value
A clarification: definition of “BoW”

- Looser definition
  - Independent features
A clarification: definition of “BoW”

• Looser definition
  – Independent features

• Stricter definition
  – Independent features
  – histogram representation
Learning

- Feature detection & representation
- Image representation
- Category models (and/or) classifiers

Recognition

- Codewords dictionary
- Category decision

The diagram illustrates the process of learning and recognition in computer vision, involving feature detection, image representation, and decision-making processes.
1. feature detection & representation

2. codewords dictionary

3. image representation
1. Feature detection and representation

Normalize patch

Detect patches

- [Mikojaczyk and Schmid ’02]
- [Mata, Chum, Urban & Pajdla, ’02]
- [Sivic & Zisserman, ’03]

Compute SIFT descriptor

[ Lowe’99 ]
Feature extraction

Weak features
Edge points at 2 scales and 8 orientations (vocabulary size 16)

Strong features
SIFT descriptors of 16x16 patches sampled on a regular grid, quantized to form visual vocabulary (size 200, 400)
2. Codewords dictionary formation
2. Codewords dictionary formation

Vector quantization

Slide credit: Josef Sivic
2. Codewords dictionary formation

Fei-Fei et al. 2005
Image patch examples of codewords

Sivic et al. 2005
1. feature detection & representation

2. codewords dictionary

3. image representation
3. Image representation

![Image representation diagram]

- Bars represent frequency.
- Codewords are shown below the bars.

**Diagram Description:**
- The x-axis represents codewords.
- The y-axis represents frequency.

**Image Representation:**
- The image shows a car with its rear end visible.
- The car is in a parking lot setting.

**Legend:**
- Frequency bars indicate the distribution of codewords across different patterns.
Learning and Recognition

codewords dictionary

category models
(and/or) classifiers

category decision
1. Generative method:
   - Poor Discriminative ability...

1. Discriminative method:
   - Pyramid Matching Kernel (SP)
   - SVM
Image Classification by Bag of Words

\( w_n \): each patch in an image
\( z_n \): theme or topic of the patch
\( H \): histogram of \( z_n \)
\( I \): the image
\( c \): category of the image
Image Representation by Pyramid

\( w_n \): each patch in an image

\( z_n \): theme or topic of the patch

\( H \): histogram of \( z_n \)

\( I \): the image

\( c \): category of the image

Learning

The Code book

\[ \underline{\text{SP etc..}} \]
Image representation

frequency

codewords
Pyramid
Original images

Feature histograms:
Level 3
Level 2
Level 1
Level 0
Recognition by Pyramid Matching Kernel

\( w_n \): each patch in an image
\( z_n \): theme or topic of the patch
\( H \): histogram of \( z_n \)
\( I \): the image
\( c \): category of the image

SVM will be covered by Nick
Pyramid match kernel

\( \mathcal{R}^d \bigcap \mathcal{R}^d \approx \)

optimal partial matching between sets of features

\[
K_\Delta (\Psi(X), \Psi(Y))
\]

Grauman & Darrell, 2005, Slide credit: Kristen Grauman
Pyramid Match (Grauman & Darrell 2005)

Histogram intersection

\[ I(H(X), H(Y)) = \sum_{j=1}^{r} \min (H(X)_j, H(Y)_j) \]

Slide credit: Kristen Grauman
Pyramid match kernel

\[ K_{\Delta} \left( \Psi(X), \Psi(Y) \right) = \sum_{i=0}^{L} \frac{1}{2^i} \left( \mathcal{I} \left( H_i(X), H_i(Y) \right) - \mathcal{I} \left( H_{i-1}(X), H_{i-1}(Y) \right) \right) \]

• Weights inversely proportional to bin size
• Normalize kernel values to avoid favoring large sets
Summary: Pyramid match kernel

\[ K_\Delta (\Psi(X), \Psi(Y)) = \sum_{i=0}^{L} \frac{1}{2^i} \left( I(H_i(X), H_i(Y)) - I(H_{i-1}(X), H_{i-1}(Y)) \right) \]

- Difficulty of a match at level \( i \)
- Number of new matches at level \( i \)

optimal partial matching between sets of features

Slide credit: Kristen Grauman
Object recognition results

- **ETH-80 database** object classes  
  *(Eichhorn and Chapelle 2004)*

- **Features:**
  - Harris detector
  - PCA-SIFT descriptor, $d=10$

<table>
<thead>
<tr>
<th>Kernel</th>
<th>Complexity</th>
<th>Recognition rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Match [Wallraven et al.]</td>
<td>$O(dm^2)$</td>
<td>84%</td>
</tr>
<tr>
<td>Bhattacharyya affinity [Kondor &amp; Jebara]</td>
<td>$O(dm^3)$</td>
<td>85%</td>
</tr>
<tr>
<td>Pyramid match</td>
<td>$O(dmL)$</td>
<td>84%</td>
</tr>
</tbody>
</table>

Slide credit: Kristen Grauman
Object recognition results

• Caltech objects database 101 object classes

• Features:
  – SIFT detector
  – PCA-SIFT descriptor, $d=10$

• 30 training images / class

• 43% recognition rate
  (1% chance performance)

• 0.002 seconds per match

Slide credit: Kristen Grauman
The diagram illustrates the process of image classification using a codewords dictionary. The left side, labeled "learning," shows the steps of feature detection and representation, which lead to category models (and/or) classifiers. The right side, labeled "recognition," shows how the features are used to make a decision about the category. The codewords dictionary acts as a bridge between the learning and recognition stages, facilitating the classification process.