Pattern recognition (3)
Things we have discussed until now

- Statistical pattern recognition
  - Building simple classifiers
    - Supervised classification
      - Minimum distance classifier
      - Bayesian classifier
      - Building discriminant functions
    - Unsupervised classification
      - K-means algorithm
Discriminants

- A function used to test the class membership is called a discriminant.
- Construct a single discriminant $g_i(x)$ for each class $\omega_i$, and assign $x$ to class $\omega_i$ if $g_i(x) > g_j(x)$ for all other classes $\omega_j$.
Last lecture we stopped at:
Discriminant functions for Bayes classifier in the 1D case

Today:
Extension to multiple classes
Naïve Bayes: considers that features are statistically independent

This lecture is based on a tutorial on object classification techniques presented at ICCV 2005.

Optional reading:
- Wikipedia article of the Bag-of-Words concept for Computer Vision applications
Naïve Bayes

○ Derivation
○ How do we represent objects through features that are statistically independent?
○ The bag-of-words model
What is the Bag-of-Words model?

- Initially used in Natural Language Processing (NLP)
- It is a method for document representation which ignores the word order
- Example: ‘a good book’ and ‘a book good’ are the same object
- The model is dictionary based
- Each document looks like a ‘bag’ containing some words of the dictionary
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 us through our eyes. For a long time it was thought that the retinal image was transmitted point by point to visual centers in the brain, as a movie screen upon which the image was projected. Through the discoveries of Hubel and Wiesel we now know that the visual perception in the brain is more complicated. 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% increase in exports to $750bn, compared with a rise in imports of $660bn. This is likely to 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 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.
How to create a bag of words in NLP

Two simple documents:
- John likes to watch movies. Mary likes too.
- John also likes to watch football games.

Based on these two text documents, a dictionary is constructed as:
```
dictionary={1:"John", 2:"likes", 3:"to", 4:"watch", 5:"movies", 6:"also", 7:"football", 8:"games", 9:"Mary", 10:"too"},
```
which has 10 distinct words.

Using the indexes of the dictionary, each document is represented by a 10-entry vector:
- [1, 2, 1, 1, 1, 0, 0, 0, 1, 1]
- [1, 1, 1, 1, 0, 1, 1, 1, 0, 0],
- where each entry of the vectors refers to count of the corresponding entry in the dictionary (a histogram-based representation).

Source:
How to create a Bag-of-Words representation in Computer Vision

- An image can be treated as a document
- Features extracted from the image are considered words
- “word” in images does not have the same straightforward meaning as in documents
Steps in generating a bag of words for images

- Feature Detection
- Feature Representation
- Generation of codebook (equivalent to the dictionary in text documents)
1. Feature detection

- extracts several local patches (or regions) which are considered as candidates for basic elements (words)
Feature selection

- Regular grid
  - Good for natural scene categorization
  - Disadv.: very little information on the image itself
Feature selection (cont’d)

- Detection of interest points (Harris corner detector, SIFT)
  - Extract salient patches (considered more important than other patches; they also attract human attention)

Example of a visual word (airplane wheel) occurring in different images
- elliptical region is superimposed.
- bottom row shows normalized regions for the top row of images.
- normalized regions appear quite similar – this is why they represent the same word.

Feature representation

- We represent patches as numerical vectors
- Normalization before representation
- Popular representation: SIFT (Scale-Invariant Feature Transform, Lowe 1999) converts each patch to a 128 vector.
Codebook generation

Final step

- Converts vector-represented patches to ‘codewords’ (analogy to words in text documents)
- The codebook = the equivalent of dictionary in text documents
- Codeword = a representative of several similar patches
  - Can be performed by performing K-means clustering over all vectors
  - The codewords are the centers of the learned clusters
2. Codewords dictionary formation

Slide credit: Josef Sivic
all patches detected for this image.

patches from two selected clusters occurring in this image (yellow and magenta ellipses).

2. Codewords dictionary formation

Fei-Fei et al. 2005
3. Image representation

![Image representation diagram]

- Frequency
- Codewords
Representation

1. feature detection & representation

2. codewords dictionary

3. image representation
Learning and Recognition

codewords dictionary

category model(s) (and/or) classifiers

category decision
Case study

• Naïve Bayes classifier
• Area: Content-Based Information Retrieval
• Question asked: Given an image, what category of objects does it contain?

Notations:
• $w_n$: each patch in an image
  – $w_n = [0,0,...1,...,0,0]^T$
• $w$: a collection of all N patches in an image
  – $w = [w_1,w_2,...,w_N]$

• $c$: category of the image
Case study: the Naïve Bayes model

\[ c^* = \arg \max_c p(c \mid w) \propto p(c) p(w \mid c) = p(c) \prod_{n=1}^{N} p(w_n \mid c) \]

Object class decision

Prior prob. of the object classes

Image likelihood given the class

Csurka et al. 2004
Our in-house database contains 1776 images in seven classes\(^1\): faces, buildings, trees, cars, phones, bikes and books. Fig. 2 shows some examples from this dataset.
Table 1. Confusion matrix and the mean rank for the best vocabulary ($k=1000$).

<table>
<thead>
<tr>
<th>True classes</th>
<th>faces</th>
<th>buildings</th>
<th>trees</th>
<th>cars</th>
<th>phones</th>
<th>bikes</th>
<th>books</th>
</tr>
</thead>
<tbody>
<tr>
<td>faces</td>
<td>76</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>buildings</td>
<td>2</td>
<td>44</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>trees</td>
<td>3</td>
<td>2</td>
<td>80</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>cars</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>75</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>phones</td>
<td>9</td>
<td>15</td>
<td>1</td>
<td>16</td>
<td>70</td>
<td>14</td>
<td>11</td>
</tr>
<tr>
<td>bikes</td>
<td>2</td>
<td>15</td>
<td>12</td>
<td>0</td>
<td>8</td>
<td>73</td>
<td>0</td>
</tr>
<tr>
<td>books</td>
<td>4</td>
<td>19</td>
<td>0</td>
<td>6</td>
<td>7</td>
<td>2</td>
<td>69</td>
</tr>
</tbody>
</table>

Mean ranks: 1.49, 1.88, 1.33, 1.33, 1.63, 1.57, 1.57
More results

Examples of images correctly classified, even if most of the keypoints were present on the background rather than on the objects
More results

Examples of images correctly classified, in the presence of partial view and varying viewpoint
Summary

• We have studied one possible extension of the Bayes’ classifier to multiple dimension
• This extension, called Naïve Bayes classifier, considers all features of an object as independent random variables
• We can build object and image representations (example: Bag of Words) that respect this assumption in the Naïve Bayes Classifier
• Next: the general Bayes Classifier
Multivariate normal Bayesian classification

- For multiple classes in a p-dimensional feature space, each class $\omega_i$ has its own mean vector $m_i$ and covariance matrix $C_i$.
- The class-conditional probabilities are:

$$p(x|\omega_i) = (2\pi)^{-d/2} |C_i|^{-1/2} e^{-\frac{1}{2}(x-m_i)^T C_i^{-1}(x-m_i)}$$
Mahalonobis distance

- The expression \((x - m_i)^T C^{-1}(x - m_i)\)

- can be considered as a distance between feature vector \(x\) and class \(i\). \(C\) is the covariance matrix computed for class \(i\).

- It can be proven that the Mahalonobis distance satisfies all properties of a distance function
Equivalence between classifiers

- Pattern recognition using multivariate normal distributions and equal priors is simply a minimum Mahalanobis distance classifier.