Bag of Words
or
Bag of Features
Lecture-17
(Slides Credit: Cordelia Schmid
LEAR – INRIA Grenoble)
Contents

• Interest Point Detector
• Interest Point Descriptor
• K-means clustering
• Support Vector Machine (SVM) classifier
• Evaluation Metrics: Precision & Recall
Image classification

• Image classification: assigning a class label to the image

Car: present
Cow: present
Bike: not present
Horse: not present

...
Image classification

- Image classification: assigning a class label to the image

  ![Image with car and cow]

  - Car: present
  - Cow: present
  - Bike: not present
  - Horse: not present
  ...

- Object localization: define the location and the category

  ![Object localization diagram]

  - Location
  - Category
Difficulties: within object variations

Variability: Camera position, Illumination, Internal parameters

Within-object variations
Difficulties: within class variations
Image classification

- **Given**
  
  Positive training images containing an object class

  ![Positive Examples](image1.png) ![Positive Examples](image2.png) ![Positive Examples](image3.png)

  Negative training images that don’t

  ![Negative Examples](image4.png) ![Negative Examples](image5.png) ![Negative Examples](image6.png)

- **Classify**

  A test image as to whether it contains the object class or not

  ![Test Image](image7.png)
Bag-of-features – Origin: texture recognition

• Texture is characterized by the repetition of basic elements or *textons*

Bag-of-features – Origin: texture recognition
The last duel
After quarrelling over a bank loan, two men took part in the last fatal duel staged on Scottish soil. BBC News's James Landale retraces the steps of his ancestor who made that final challenge.

West Bank water row
Palestinians have accused Israel of diverting water away from their towns in order to keep Jewish settlements in the occupied territories fully supplied. Israel denies the charge saying Palestinian farmers are to blame for using illegal connections to irrigate their fields.
Bag of Visual Words model

Image Dataset

Feature Detection

Local Patches
Bag of Visual Words model

Local Patches

Descriptors

Clustering

Generate the visual vocabulary

Words
Bag of Visual Words model

Represent an image as a histogram or bag of words
Bag-of-features – Origin: bag-of-words (text)

- Orderless document representation: frequencies of words from a dictionary
- Classification to determine document categories

### Bag-of-words

<table>
<thead>
<tr>
<th></th>
<th>d1</th>
<th>d2</th>
<th>d3</th>
<th>d4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>People</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Sculpture</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

...
Bag-of-features for image classification

- Extract regions or Interest Points
- Compute descriptors
- Find clusters and frequencies
Bag-of-features for image classification

Step 1
- Extract regions
- Compute descriptors

Step 2
- Find clusters and frequencies
- Compute distance matrix

Step 3
- Classification

SVM
Step 1: feature extraction

• Detect Interest Points
  – SIFT
  – Harris
  – Dense (take every nth pixel as interest point)

• Compute Descriptor around each interest point
  – SIFT
  – HOG
Dense features
Bag-of-features for image classification

Step 1: Extract regions
Step 2: Compute descriptors, find clusters and frequencies
Step 3: Compute distance matrix, classification

SVM
Step 2: Quantization

Clustering

Visual vocabulary

Clustering
Step 2: Quantization

• Cluster descriptors
  – K-means

• Assign each visual word to a cluster

• Build frequency histogram
Example: 3-means Clustering

Convergence in 3 steps from Duda et al.
## K-Means

Choose \( k \) data points to act as cluster centers

Until the cluster centers are unchanged

Allocate each data point to cluster whose center is nearest

Replace the cluster centers with the mean of the elements in their clusters.

end

**Algorithm 16.5: Clustering by K-Means**
### Examples for visual words

<table>
<thead>
<tr>
<th>Category</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Airplanes</td>
<td><img src="image1" alt="Airplane Examples" /> <img src="image2" alt="Airplane Examples" /> <img src="image3" alt="Airplane Examples" /> <img src="image4" alt="Airplane Examples" /></td>
</tr>
<tr>
<td>Motorbikes</td>
<td><img src="image5" alt="Motorbike Examples" /> <img src="image6" alt="Motorbike Examples" /> <img src="image7" alt="Motorbike Examples" /> <img src="image8" alt="Motorbike Examples" /></td>
</tr>
<tr>
<td>Faces</td>
<td><img src="image9" alt="Face Examples" /> <img src="image10" alt="Face Examples" /> <img src="image11" alt="Face Examples" /> <img src="image12" alt="Face Examples" /></td>
</tr>
<tr>
<td>Wild Cats</td>
<td><img src="image13" alt="Wild Cat Examples" /> <img src="image14" alt="Wild Cat Examples" /> <img src="image15" alt="Wild Cat Examples" /> <img src="image16" alt="Wild Cat Examples" /></td>
</tr>
<tr>
<td>Leaves</td>
<td><img src="image17" alt="Leaf Examples" /> <img src="image18" alt="Leaf Examples" /> <img src="image19" alt="Leaf Examples" /> <img src="image20" alt="Leaf Examples" /></td>
</tr>
<tr>
<td>People</td>
<td><img src="image21" alt="People Examples" /> <img src="image22" alt="People Examples" /> <img src="image23" alt="People Examples" /> <img src="image24" alt="People Examples" /></td>
</tr>
<tr>
<td>Bikes</td>
<td><img src="image25" alt="Bike Examples" /> <img src="image26" alt="Bike Examples" /> <img src="image27" alt="Bike Examples" /> <img src="image28" alt="Bike Examples" /></td>
</tr>
</tbody>
</table>
Image representation

- each image is represented by a vector, typically 1000-4000 dimension,
Bag-of-features for image classification

Step 1: Extract regions
Step 2: Compute descriptors and find clusters and frequencies
Step 3: Compute distance matrix and Classification

SVM
Step 3: Classification

• Learn a decision rule (classifier) assigning bag-of-features representations of images to different classes
Training data

Vectors are histograms, one from each training image

positive

Train classifier, e.g. SVM
Support Vector Machines (SVM)
Application

- Pattern recognition
- Object classification/detection
Usage

- The classifier must be trained using a set of negative and positive examples.

- The classifier “learns” the regularities in the data

- If training was successful classifier is capable of classifying an unknown example with a high degree of accuracy.
Linear Classifier

- Binary classifier $\rightarrow$ Task of separating classes in feature space

$$w^T x + b > 0$$
$$w^T x + b < 0$$

$$f(x) = \text{sign}(w^T x + b)$$
Linear Classifier cont’d

- Which of the linear separators is optimal?
Margin

- Distance from example to the separator is (Point to Plane Distance Equation)
  \[ r = \frac{w^T x + b}{\|w\|} \]
- Examples closest to the hyperplane are support vectors.
- Margin $2\gamma$ of the separator is the width of separation between classes.
Maximum Margin Classification

- Maximizing the margin is good according to intuition.
- Implies that only support vectors are important; other training examples are ignorable.
LibSVM

SVM implementation

- http://www.csie.ntu.edu.tw/~cjlin/libsvm/
- http://www.cs.wisc.edu/dmi/svm/
Bag-of-features for image classification

- Excellent results in the presence of background clutter
Examples for misclassified images

Books - misclassified into faces, faces, buildings

Buildings - misclassified into faces, trees, trees

Cars - misclassified into buildings, phones, phones
Evaluation of image classification

- PASCAL VOC [05-10] datasets

- PASCAL VOC 2007
  - Training and test dataset available
  - Used to report state-of-the-art results
  - Collected January 2007 from Flickr
  - 500,000 images downloaded and random subset selected
  - 20 classes
  - Class labels per image + bounding boxes
  - 5011 training images, 4952 test images

- Evaluation measure: average precision
PASCAL 2007 dataset
PASCAL 2007 dataset
Evaluation Metrics

\[
\text{precision} = \frac{\text{GT} \cap \text{RM}}{\text{RM}} = \frac{\text{TP}}{\text{RM}} \\
\text{recall} = \frac{\text{GT} \cap \text{RM}}{\text{GT}} = \frac{\text{TP}}{\text{GT}}
\]