Learning Object Categories from Google’s Image Search

R. Fergus et al
Present by Jie Xiao

Dept. of Computer Science
Univ. of Texas at San Antonio
Outline

- Motivation
- “Bag of words” Model
- Approaches (pLSA, ABS-pLSA, TSI-pLSA)
- Dataset
- Experiment
- Conclusion
Motivation

- Current approaches of object categorization require manual labeled dataset as training set.
- Collecting data is time-consuming, involved in numerous human work.
- Finding good examples is another concern.
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 our brain via our eyes. For a long time, it was assumed that the visual impression was immediately transmitted as a movie screen to the cerebral cortex, but recent discoveries of Hubel and Wiesel have demonstrated that this is not the case.

Through their experiments, Hubel and Wiesel were able to show that the visual perception in the brain is not just a simple projection of the image in the eye, but a considerably more complicated process. They demonstrated 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 $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 needs to do more to boost domestic demand so the yuan can rise in value.

China, trade, surplus, commerce, exports, imports, US, yuan, bank, domestic, foreign, increase, trade, value
Bag of Words Model

- LSA:
- U and V are orthonormal matrices
  \[ X = U \Sigma V^T \]
- A singular value decomposition (SVD) process
- pLSA
Bag of Words Model -- pLSA

- D: set of documents
- W: visual words
- Z: topics
- Latent variable z is associate with w and d.

Matrix $N_{M \times N}$: co-occurrence of words and doc

- $N_{(w,d)}$: the number of word w appears in document d.
Bag of Words Model – pLSA (Cont.)

\[ P(w|z) \]  co-occurrence of words within a topic

\[ P(z|d) \]  density of topic on a given document

\[ P(w_i|d_j) = \sum_{k=1}^{K} P(z_k|d_j) P(w_i|z_k) \]

\[ P(w, d) = \sum_{z=1}^{Z} P(w|z) P(z|d) P(d) \]  \hspace{1cm} (1)
**Bag of Words Model – pLSA (Cont.)**

- $P(w|z)$  
  *topic specific word distribution*

- $P(z|d)$  
  *document specific mixing proportion*

\[
P(w_i|d_j) = \sum_{k=1}^{K} P(z_k|d_j) P(w_i|z_k)
\]

![Diagram showing the relationship between words, documents, and topic labels.](image)
Bag of Words Model – pLSA (Cont.)

\[ P(w, d) = \sum_{z=1}^{Z} P(w | z) P(z | d) P(d) \]
Calculating by EM

E step: \( P(z|w, d) \)

M step: \( P(d) \ P(z|d) \ P(w|z) \)

\[
P(w, d) = \sum_{z=1}^{Z} P(w|z)P(z|d)P(d)
\]

\[
L = \prod_{d=1}^{D} \prod_{w=1}^{W} P(w, d)^{n(w,d)}
\]
Bag of Words Model (Cont.)

Object → Bag of words

Slide credit: Rob Fergus
1. Representation

feature detection & representation

codewords dictionary

2. Image representation

3. 

jxiao@cs.utsa.edu

Slide credit: Rob Fergus
Approach

ABS-pLSA
Quantize the location within the image into one of X bins

Use

\[ P(w, x, d) = \sum_{z=1}^{Z} P(w, x | z) P(z | d) P(d) \]

Instead of

\[ P(w, d) = \sum_{z=1}^{Z} P(w | z) P(z | d) P(d) \]
TSI-pLSA

Introducing latent variable, $c$, represents the centriod of the object.

$$X_{fg} \quad \text{foreground bins}$$

$$x_{bg} \quad \text{background bin}$$

$$X = X_{fg} + 1$$

$$P(w, x | z) = \sum_c P(w, x, c | z) = \sum_c P(w, x | c, z) P(c)$$
Approach (Cont.)

(a) 

(b) 

(c) 

(d)
Datasets

- PT: prepared training set, manually gathered
- P: prepared test set
- G: raw download data from Google image.
  - Good image: good examples, related to keyword category
  - Intermediate images: related to keyword category, low quality than good image
  - Junk images: totally unrelated to the keyword category
V: Google validation set.

◆ Assume the images from first pages are positive examples.
◆ Cross language collections
Datasets (Cont.)
### Datasets (Cont.)

**statistics**

<table>
<thead>
<tr>
<th>Category</th>
<th>Size of Dataset</th>
<th>Distrib. of Google Images (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PT</td>
<td>P</td>
</tr>
<tr>
<td>Airplane</td>
<td>400</td>
<td>400</td>
</tr>
<tr>
<td>Cars Rear</td>
<td>400</td>
<td>400</td>
</tr>
<tr>
<td>Face</td>
<td>217</td>
<td>217</td>
</tr>
<tr>
<td>Guitar</td>
<td>450</td>
<td>450</td>
</tr>
<tr>
<td>Leopard</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Motorbike</td>
<td>400</td>
<td>400</td>
</tr>
<tr>
<td>Wrist watch</td>
<td>180</td>
<td>181</td>
</tr>
<tr>
<td>PASCAL Cars</td>
<td>272</td>
<td>275</td>
</tr>
<tr>
<td>PASCAL Cars Bg.</td>
<td>412</td>
<td>412</td>
</tr>
<tr>
<td>PASCAL Motorbike</td>
<td>214</td>
<td>202</td>
</tr>
<tr>
<td>PASCAL Motorbike Bg.</td>
<td>570</td>
<td>754</td>
</tr>
<tr>
<td>Caltech Bg.</td>
<td>400</td>
<td>400</td>
</tr>
<tr>
<td>Cars Rear Bg.</td>
<td>400</td>
<td>400</td>
</tr>
</tbody>
</table>
Experiments

- Region detectors:
  - Convert to grayscale
  - Resize to a moderate size
  - Detect region
  - Represent by SIFT descriptor
  - Quantize descriptor vector
Experiments – region detector

Region detectors:
- Kadir & Brady saliency operator
- Multi-scale Harris detector
- Difference of Gaussian
- Edge based operator
Experiments (Cont.)

\[ p(z_1|d) = 0.990 \quad p(z_2|d) = 0.010 \]
Experiments (Cont.)

\[ p(z_1|d) = 1.000 \quad p(z_2|d) = 0.000 \]
Experiments (Cont.)

\[ p(z_1|d) = 0.250 \quad p(z_2|d) = 0.750 \]
Experiments (Cont.)

- Red: pLSA
- Green: ABS-pLSA
- Blue: TSI-pLSA

Solid line: performance of automatically chosen topic within model

Dashed line: performance of best topic within model
Discussion

- Limited categories
- Prior knowledge about number of categories
- Image background
- Similar visual word
Conclusion

- Introduce spatial information in pLSA.
- Learn object category by category name.
Thank you!