Web Image Gathering with Region-based Bag-of-features and Multiple Instance Learning


Keiji Yanai
The University of Electro-Communications, Tokyo, Japan

1. Objective & Background
2. Related Work
3. System & Methods
4. Experimental Results
1. Objective & Background
Web is the largest image DB. It is also a very noisy DB.

- To remove noise, image analysis is needed.
- Since 2001, we have been working on Web Image Gathering with image analysis
  - Non-interactive. No feedback. Fully-automatic.
- To gather visual knowledge of many concepts for object recognition from the Web
Objective of this paper

- **Import region-based bag-of-features to our Web image “gathering” system**
  [Yanai et al. ICME01, ACM MM03, ACM MIR 05, ICME08]

**Image representation**
- region-based bag-of-features
  [Ravinovich et al. ICCV 07]

**Classifier**
- mi-SVM (multiple instance learning)
  [Andrew et al. NIPS 03]

new combination!
2. Related Work
General Framework: Web image search + Object Recognition Technique

- Firstly, **gather images from the Web using Web (image) search engines such as Google, Ask.com and MSN search by providing given keywords.**

- Secondly, **re-rank the results from the Web search engines with object/scene recognition methods**
Object Recognition Technique

- Color histogram + k-means  
  [Yanai ICME01]
- Color signature + EMD + k-NN  
  [Yanai ACM MM03]
- Constellation model + RANSAC  
  [Fergus ICCV04]
- JSEG + GMM (image-word translation model)  
  [Yanai & Barnard ACM MIR 05]
- Bag-of-features (BoF) + pLSA  
  [Fergus ECCV05]
- Bag-of-features + HDP (Hierarchical Dirichlet Process) (OPTIMOL)  
  [Li CVPR07]
- Bag-of-features + SVM  
  [ICCV Schroff 07] [Yanai 07]

(This paper)
JSEG + region-based bag-of-features + mi-SVM (multiple instance learning)
**Object Recognition Technique**

- **JSEG + GMM** *(image-word translation model)*  
  [Yanai & Barnard ACM MIR 05]

- **Bag-of-features + SVM**  
  [ICCV Schroff 07] [Yanai 07]

- **[This paper]**  
  JSEG + region-based bag-of-features  
  + mi-SVM (multiple instance learning)
Contribution of this paper

- Import region-based bag-of-features to our Web image "gathering" system

[Image representation] region-based bag-of-features
[Ravinovich et al. ICCV 07]

[Classifier] mi-SVM (multiple instance learning)
[Andrew et al. NIPS 03]
3. Methods
Basic framework of our system

[ Yanai ICME01 ]

Collection stage

Gather image and HTML files using Web search engines.
Select \textit{pseudo-training images} by HTML analysis

Selection Stage

Train a classifier and rank images based on estimated relevancy

Use \textit{supervised object rec. methods with pseudo-training images}
Image features

- Divide each image into regions by JSEG (8 regions on the average)

Segmentation (JSEG)

Bag-of-features (BoF) histograms (1000-dim)
Represent an image as sets of features

1. Densely-sample points along regular grids
2. Represent local patterns around sampled points with SIFT descriptor
3. Vector-quantize SIFT vectors based on pre-computed visual words (codewords)
How to obtain visual words

- Extract many SIFT vectors from positive and negative training samples
- Perform k-means clustering

The center of clusters are “visual words”.

“Visual words” are representative local patterns.
Multiple Instance Setting

- Positive bags / Negative bags

Positive instances of "flower"  
Positive ins. (foreground)  
random images

The rest of regions are negative regions.  
negative ins. (background)
Apply soft-margin SVM iteratively

Training $\rightarrow$ classifying $\rightarrow$ training $\rightarrow$ classifying $\rightarrow$ …

During the iteration, the hyper-plane is approaching the optimal plane to discriminate positive instances from negative ones.

- positive ins. (foreground)
- negative ins. (background)
Final Image Re-ranking

- Regard the best SVM output score of the regions within an image as the score of the image
  - An image having one positive region at least is a positive image!
- Rank images based on the scores

flower
airplane
apple
4. Experimental results
Experiments for 10+5 words

- sunset, mountain, waterfall, beach, (4scenes)
- noodle, flower, lion, apple, baby, laptop-PC, (6objects)
- airplane, guitar, leopard, motorbike, watch (5objects)

Method:

- raw data raw (only HTML analysis) 39,143 images for 15 words
- baseline1 GMM-based region probabilistic model [ACM MIR05]
- baseline2 BoF + SVM
- proposed region-based BoF + SVM

Evaluation: precision at 15% recall the same as [ICCV Schroff 07]
Comparison of 4 methods (raw, GMM, BoK, reg-BoF)

<table>
<thead>
<tr>
<th>Method</th>
<th>prec. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>raw</td>
<td>62.2%</td>
</tr>
<tr>
<td>GMM</td>
<td>73.5%</td>
</tr>
<tr>
<td>BoF</td>
<td>88.17%</td>
</tr>
<tr>
<td>reg-BoF</td>
<td>87.82%</td>
</tr>
</tbody>
</table>
Pseudo-training image sets and results by perfect training set (noise-free)
Comparison with related work

- [Fergus ICCV05] Bag-of-features + pLSA
- [Schroff ICCV07] Bag-of-features + SVM
- [new] Region-based BoF + mi-SVM
Many result images

- Laptop-PC (positive and negative)
- Mountain
- Waterfall
- Flower
- Airplane
- ???

As by-products, we can estimate representative regions of images. (different from standard BoF)
Conclusion

- Import region-based bag-of-features (BoF) and mi-SVM into the Web image gathering task.

- In spite of noisy training data, the proposed method worked well.

- It was especially effective for object concepts.
Future work

- Large-scale experiments
  - More than concept for 1000 concepts

- Improve the text analysis part to obtain more accurate pseudo-training samples
  - Use co-occurrence of tags
  - Use taxonomy dic. (Wordnet, Wikipedia)
Thank you!