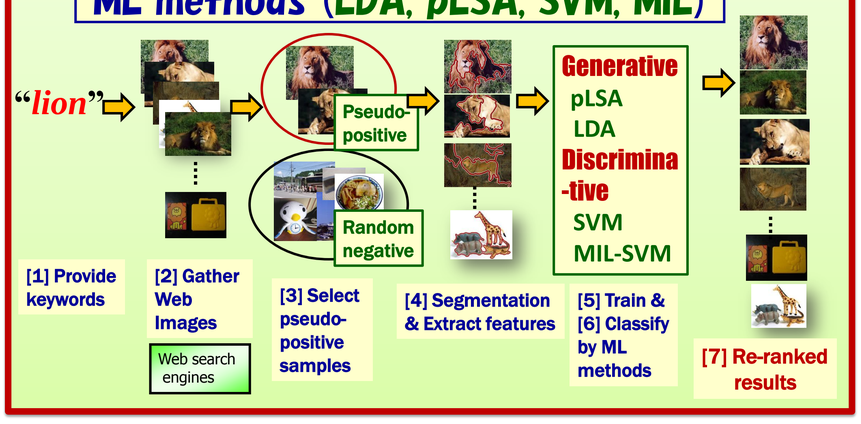
Region-based Automatic Web Image Selection

Keiji Yanai and Kobus Barnard

Univ. of Electro-Communications, Tokyo, JAPAN / Univ. of Arizona, USA

Overview

Web Image Re-ranking with region-based BoF and ML methods (LDA, pLSA, SVM, MIL)



1. Background & Objective

◆Web is the largest image DB. It is also a very noisy DB.

◆To remove noise, we apply object recognition methods.

● "Web Image Re-ranking"

- We assume no feedback and fully-automatic "Web Image Re-ranking".
- It is desirable for gathering visual knowledge of many concepts for object recognition from the Web

In this paer, we import region-based bag-of-features(BoF) to Web image re-ranking.

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

【Classifier】 (2 kinds)
- sMIL (sparse multiple instance learning) [Bunescu et al. ICML 07]
- Probabilistic latent topic methods with pLSA and LDA [Yanai et al. MIR05, Monay et al. PAMI 07]

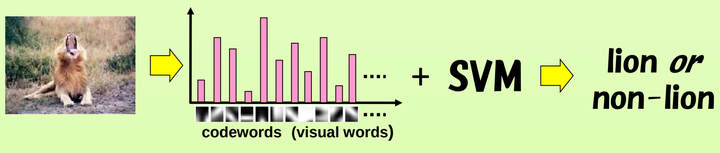
2. Related Work

◆ General Framework: Web image search + object recognition methods

key-words ⇒ Gather images from Web using search engines ⇒ image analysis: ML methods, positive img. selection ⇒ re-ranked images

◆Literature
- 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 BoF + discriminative methods (SVM / sparse MIL) or generative methods (pLSA / LDA)

3. Methods

◆processing steps :

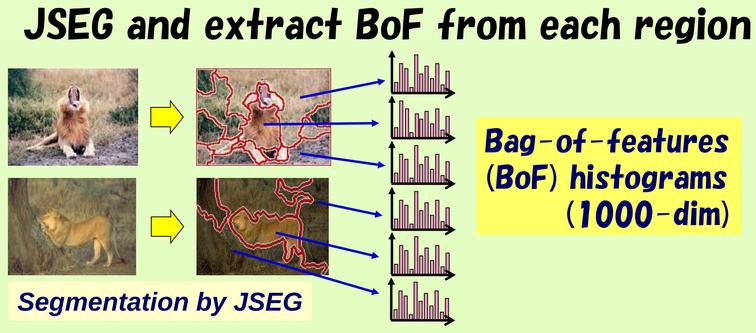
[1] Prepare keywords
[2] Gather Web images using several Web search engines
[3] Select pseudo-positive images by HTML analysis automatically
[4] Extract region-based BOF
[5] Train a model with pseudo-positive samples and random negative samples
[6] Evaluate all the images with the trained model
[7] Re-rank all the images according to the output values

[1] Prepare Keywords
e.g. "sunset", "lion" and "apple fruit"

[2] Gather Web images
- Send the given keyword to several search engines, and gather thousands of Web images (Google Image Search, Google Text Search, MSN, ask.com, Yahoo Image, Yahoo Search)

[3] Select pseudo-positive images
- Evaluate HTMLs containing images based on the simple heuristics [Yanai ACMMM03] and select pseudo-positive images automatically

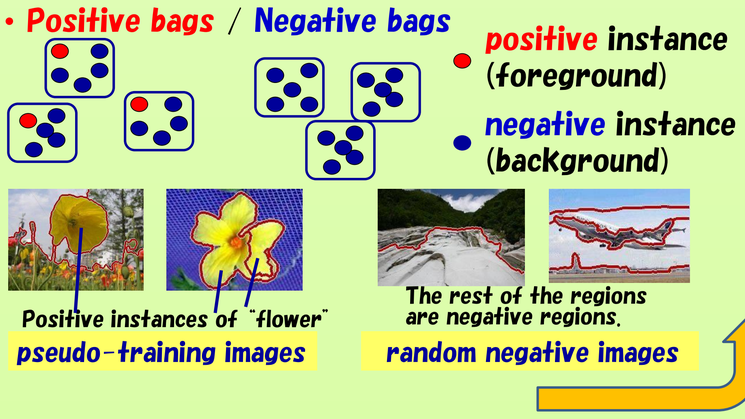
[4] Extract region-based BOF vectors
- Carry out region segmentation with JSEG and extract BoF from each region



[5] Train a model
- Two discriminative and two generative

[5-1] standard SVM
Regard all the regions from pseudo-positive images as positive samples

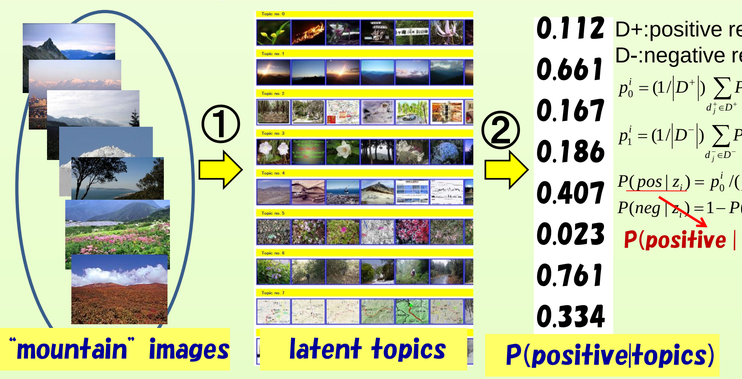
[5-2] sparse MIL [Benesku et al. ICML07]
Sparse MIL is an extension of a SVM for multiple instance (MI) settings.
- Positive bags / Negative bags



3. Methods (cont.)

[5-3] pLSA (probabilistic Latent Semantic Analysis)
[5-4] LDA (Latent Dirichlet Allocation)

① Do probabilistic topic clustering(PLSA/LDA) for all the regions with the topic number K. Obtain P(topic | region) for all the regions



② Estimate P(positive | topic) for each topic using positive regions from pseudo-positive images and negative regions from negative randomly-sampled images

③ Compute P(positive | region) for each region:

$$P(\text{positive} \mid \text{region}) = \sum_{i=1}^{k} P(\text{positive} \mid \text{topic}_i)\, P(\text{topic}_i \mid \text{region})$$

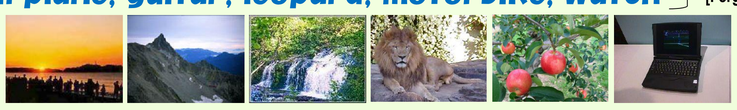
[6] Evaluate an image with the model
Regard the maximum output values of regions within an image as the output of the image (max strategy)

[7] Re-rank all the images

4. Experiments for 15 words

◆ 4 scenes + 6 objects + 5 objects

sunset, mountain, waterfall, beach, (Same as [Yanai 05])
noodle, flower, lion, apple, baby, laptop-PC, (Same as [Fergus 05])
airplane, guitar, leopard, motorbike, watch

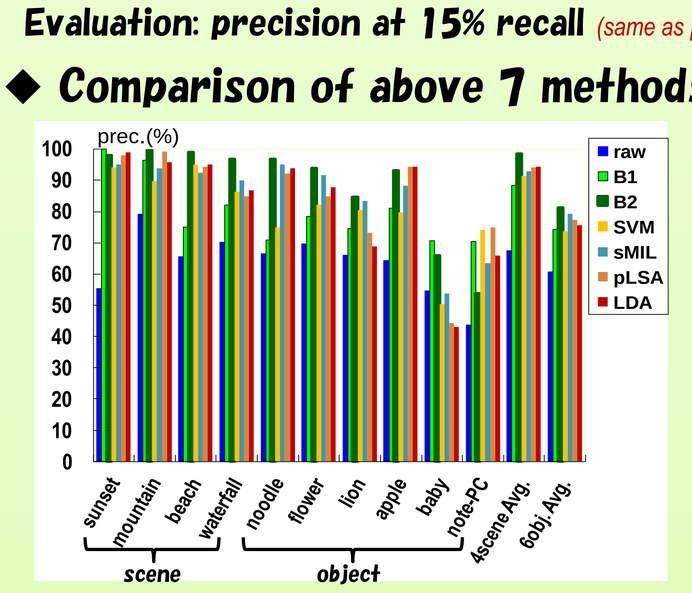


◆ Methods :

[raw data] raw (only HTML analysis) 39,143img.
[baseline1] GMM-based region prob. Model [ACM MIR05]
[baseline2] whole-image-based BoF + SVM
[SVM] region-based BoF + SVM
[sMIL] region-based BoF + sparse MIL
[pLSA] region-based BoF + pLSA
[LDA] region-based BoF + LDA

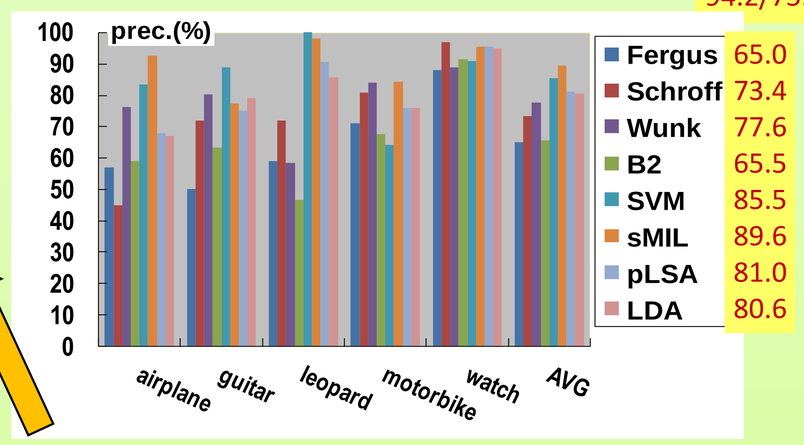
Evaluation: precision at 15% recall (same as [Schroff 07])

◆ Comparison of above 7 methods



Method Scene / object	
raw	67.5/60.8
GMM(B1)	88.3/74.3
BoF(B2)	98.4/81.3
SVM	91.3/73.7
sMIL	92.9/79.3
pLSA	94.1/77.4
LDA	94.2/75.6

◆Comparison with related work



5. Conclusions

◆We confirmed that region-based BoF is effective for "object" words.

◆S-MIL outperfomed pLSA- and LDA-based probabilistic methods.

◆ We plan to use Folksonomy, more sophisticated HTML analysis and various image features.