Probabilistic Web Image Gathering

ACM SIGMM workshop MIR ’05 Nov. 11th, 2005

Keiji Yanai The University of Electro-Communications, Japan
Kobus Barnard University of Arizona, USA

1. Objective & background
2. Related work
3. Method
4. Experiments
(Long-term) Objective of our Web Image Gathering

Gather several hundreds of images associated with one concept from the Web without human intervention

- It’s not *image search*.
  - Non-interactive. No feedback. Fully-automatic.
    - We can gather 10,000 kinds of images while sleeping!!
    - We can use as much time as we need for processing.

- A large number of “X” images

  visual knowledge of “X”

Acquisition of visual knowledge for realizing generic object recognition for any concepts
There are great many kinds of images in the real world. We can understand all of these images easily. But, computer systems cannot, because they don't have enough visual knowledge to understand these images. Extracting visual knowledge from Web images is a promising approach to solve it, I believe. Gathering Web images with high precision is the first step.
What is Web Image Gathering?

Many “waterfall” images with high precision. This is beyond search results.

This is “visual knowledge” on “waterfall”.

The University of Electro-Communications
Tokyo, Japan (UEC)
Objective of this paper

- Introduce a probabilistic method to our Web image gathering system [MM03]
  - Import a simplified translation model [ICCV01]
  - Select relevant images using a probabilistic model
  - No supervision, No feedback, No human intervention, No prior knowledge

Just provide “keywords”
2. Related work
Web image search

- “Search” does not need a lot of images as output.
  - High precision and quick response are the most important.
  - Some systems assume interaction (RF).

- Commercial: apply text search methods to images
  - Google Image Search, Picsearch, AltaVista Image Search, Lycos
  - Only based on HTML analysis,
    - Image analysis for billions of images is too expensive
  - Cannot expect good (high-precision) results

**different from “gathering”**
Web image search (-2002)

- Research: combine textual and image features

- [until 2002] HTML and image analysis after Web-crawling by their original crawler
  - WebSeer [Frankel 96], WebSEEk [Smith 97], Image Rover [Scarloff 99]
  - Original Crawler: Small-scale experiments.

query keywords

Searching for images by keyword

selecting images

searching for similar images by image features

Text-based result

Final search result
Web image search (recent)

- [recently] Filtering the results of Google Image Search
  - [Feng MM04] Co-learning of textual and image features with SVM: Need (a little) supervision interactive
  - [Fergus CVPR04, ICCV05] (the latest object recognition technique) Unsupervised part-based probabilistic image classification

Similar to ours

We’re region-based!!

Part-based

Region-based
3. System overview & probabilistic method
Overview of the method

- **Collection stage**: same as our previous system [MM03]
  - Obtain URL of HTML files related to the query keywords by using text search engines
  - Gather images from WWW by using those URLs
- **Selection Stage**: Replace CBIR-based by prob.-based
  - Select images based on their image features

**Flowchart**
- **Query keywords**: sunset「夕日」
- **Text search engines**
- **Collection stage** (search by keywords)
  - **URLs**
  - **HTML files & images**
- **Selection stage** (analyzing images)
  - **Results**

University of Electro-Communications
Tokyo, JAPAN (UEC)
Collection stage [MM03]

1. Send the query to text search engines, and obtain URLs of HTML files related to the keyword.
2. Gather HTML files from WWW.
3. Extract URLs of images from HTML files.
4. Evaluate relevancy between images and the keywords, and classifying images into A, B, C.
5. Gather only group A and B images from WWW.
Selection stage [new]

(A) Segment imgs. & extract region features
(B) Learn P(X|r) with GMM and EM
(C) Select high P(X|r) regions, repeat to (B)
(D) Calculate P(X|I) and select high P(X|I) images

A: initial training set
B: unlabeled set

Repeat several times to refine “X” model
Evaluation of images by analyzing HTML files

Classifying URLs of image files

- file name, ALT tag, link words
- title tag, neighborhood of the image-tag
- Otherwise

Examples

- `<IMG SRC="lion.jpg" ALT="lion in Ueno zoo">` ➔ A
- `<A HREF="raion01.jpg">a photo of a lion</a>` ➔ A
Region features

Segmentation (JSEG)

Extract region features from regions the size of which are larger than a threshold.

- color
- texture
- shape

24 dim. vector

- Need to prepare negative images (backgrounds)
- Collect a lot of images from the Web

random Web images
Probabilistic region selection (1)

- Positive regions from rank-A images: $r^+_i$
- Negative regions from background images: $r^-_i$

Probabilistic clustering with GMM

Select ‘X’ and ‘non-X’ components

We set $t=0.85$
Probabilistic region selection (2)

Learning an “X” region model

- \( a < (1-t) \)
- \( a > t \)
- \( a > t \)
- \( a < (1-t) \)

‘X’ model

\[ P(X | r_i) \]

‘nonX’ model

\[ P(nonX | r_i) \]

Select “X” regions

- \( r_1, r_8, r_2 \)
- \( r_5 \)
- \( r_3, r_4, r_6, r_7 \)

Selected ‘nonX’ regions

\[ P(X | r_i) < P(nonX | r_i) \]

Repeat learning & selecting

\[ P(X | r_i) > P(nonX | r_i) \]
Final Image Selection

- Select top T regions regarding $P(X \mid r_i)$ ($r_i \in I_j$) for each image $I_j$ and average them.

$$P(X \mid r_{top1_j}) \quad P(X \mid I_j) = \frac{P(X \mid r_{top1_j}) + P(X \mid r_{top2_j})}{2}$$

Finally, images of $P(X \mid I_j) \geq th$ are selected as final results.
4. Experimental results
Experiments for 10 words

- sunset, mountain, waterfall, beach, ramen (Chinese noodle), flower, lion, apple, baby, laptop-PC

Method:
- raw (only HTML analysis) 29,944 images for 10 words
- old[MM03] (color-histogram) 16,687
- new (proposed method) 14,825

Evaluation:
- precision, # of output images

Time: about 6 hours/concepts with 10 PCs
Results: # images and precision

#clusters: 150, repeat: 2 times

Avg. prec. 66.0% -> 73.5%

1482 imgs/words
Many result images

- Sunset (positive and negative)
- Mountain (positive and negative)
- waterfall
- laptop-PC

http://img.cs.uec.ac.jp/mm05/
Results: using word vectors as well as image features
Precision of A-images

- **Avg. raw** 72%
- **old** 73%
- **new** 80%

(617.1 img)
Precision of B-images
(equivalent to the recognition results by the models)

Precision(%)

- raw
- old
- new

Avg.
- raw 56%
- old 58%
- new 67%
(865.4img)
Results in case of varying times of repeat and # GMM components.
Comparison with Google Image Search

Avg. prec. GIS top500
58.6%
ours 73.5%
(1483)
Comparison with related work

- [Feng MM04] Co-learning of textual and image features with a little supervision
  - 54.0 F-measure for 15 concepts
- [Fergus CVPR04] Unsupervised
  - 65.9% precision at 15% recall for 11 concepts
- [Fergus ICCV05] Unsupervised + improved model
  - 69.9% precision at 15% recall for 7 concepts
  - Good at “objects” such as airplane, car, and bike
- Our results: good at “scene” such as sunset and mountain
  - 63.0 F-mesures (pre.73.5%, rec.55.1%) for 10
Conclusion

- We introduced region-based probabilistic image selection method into Web image gathering.
  - Iterative algorithm with GMM and EM enabled training from imperfect data.
  - We used images evaluated highly in HTML analysis as initial training set.
- Precision 73.5%, Recall 55.1% for 10 concepts
Future work

- Improve by combining part-based and region-based approach
  - Part-based seems to be better for “object”

- Build real world image corpus for generic object recognition
  - 1000 images/concept for 1000 concepts
    We can build it while sleeping!
Thank you!

You can eat all the results at

http://img.cs.uec.ac.jp/mm05/