Mining Cultural Differences from a Large Number of Geotagged Photos

World Wide Web 2009

Keiji Yanai and Bingyu Qiu
Department of Computer Science,
The Univ. of Electro-Communications, Tokyo, Japan

Background: geotagged photos

- The number of geo-tagged photos on the Web grows rapidly: Flickr, panoramio
  - Flickr has 40,000,000 geotagged photos. *(in the beginning of 2008)*

A "geo-tag" represents the coordinates (latitude,longitude) of a location where a photo are taken.

Motivation : Objects over the World

- So with such geotagged photos, we like to discover specific objects over the world.
  - ✓ Do you know all kinds of famous "noodles" in the world?
    - ✓ "Ramen" and "Soba" in Japan, "Thai noodle" in Thailand, "Chinese noodles", "rice noodle" Taiwan, "Spaghetti" in Italy…







  - ✓ How do such scenes as "beach", "waterfall", "mountain" look like in different areas in the world?
  - ✓ How about other objects such as "flower", "castle" …

As a result, we can discover cultural differences on specific concepts over the world !

Objective of our work

0011

- To mine representative photos for representative areas from geotagged photos related to generic concepts.

noodle, wedding cake, waterfall, car, castle, beach, flower

Raw geo-tagged photos on Flickr

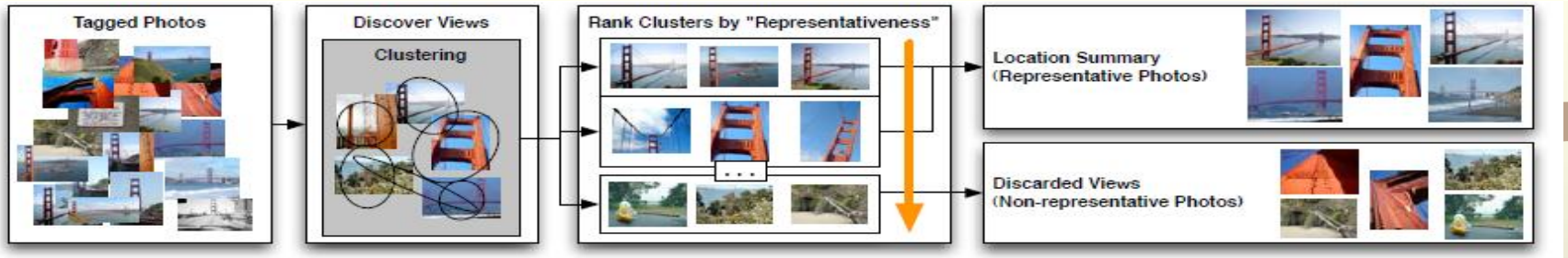
Relevant photos after noise image removal

Representative photos for typical regions

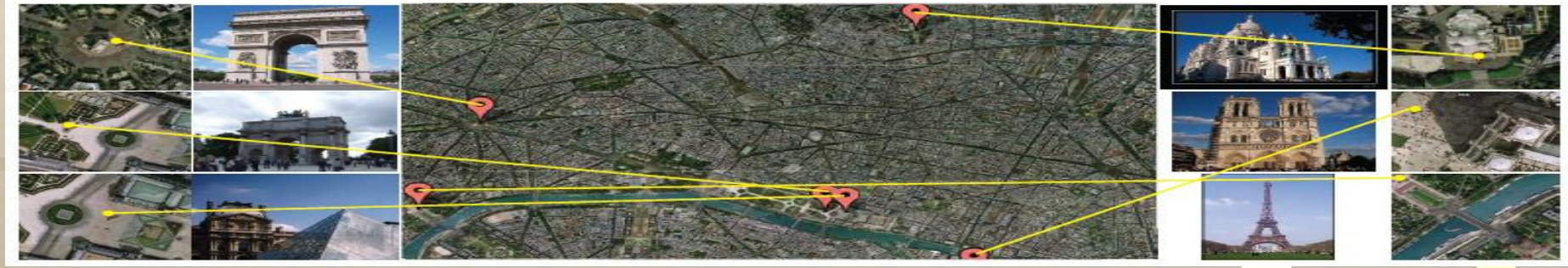
Related Work

· Automatic selection of landmarks.

Lyndon Kennedy and Mor Naaman: Generating Diverse and Representative Image Search Results for Landmarks, ACM WWW2008, pp.297-306, (2008).



Till Quack, Bastian Leibe and Luc Van Gool: World-scale Mining of Objects and Events from Community Photo Collections, ACM CIVR, pp.47-56, (2008).



They treated with only landmark photos, not generic photos

Approach : three steps

1) Select relevant photos and remove noise

- ✓ Generate image feature vectors
- ✓ Visual clustering
- ✓ Select most relevant clusters

2) Detect representative regions

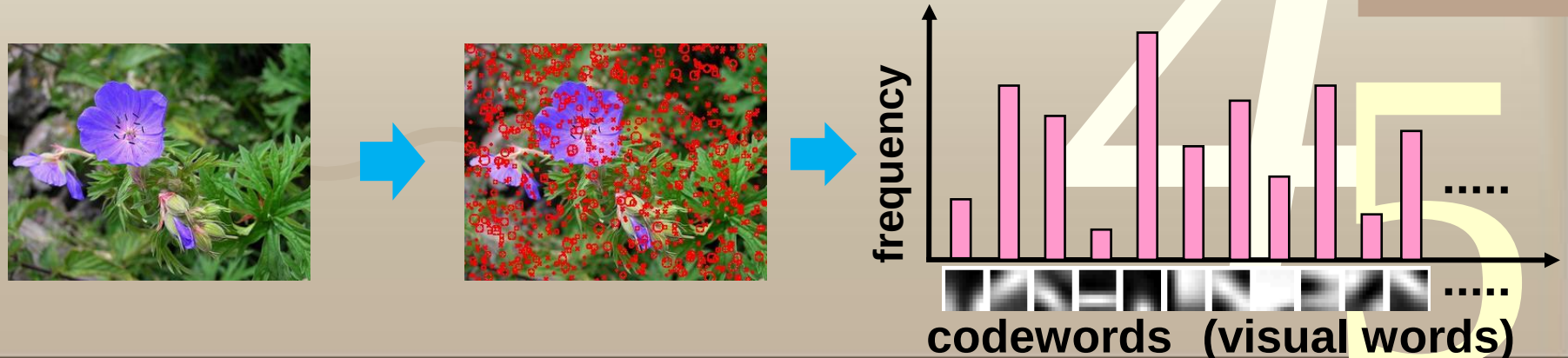
- ✓ Clustering based on geographic locations

3) Generate representative photo sets for representative regions

- ✓ Generate the PLSA topic vectors
- ✓ Aggregate photos according to the distribution of mixture topics

[Image representation] bag-of-visual-words [csu04]

- Represent an image by a set of local image features
  1. Sample keypoints randomly and compute SIFT descriptor over each keypoint
  2. Vector-quantize SIFT vectors to form visual words (codebook generation)
  3. Build histograms by mapping each feature vector to certain visual words – *regarded as image feature vectors*



Approach: (1) noise removal

1) Select relevant photos and remove noise

✓ Generate image feature vectors

✓ **Visual clustering**

k-means clustering over bag-of-visual-words vectors.
(In the experiments, we set k to 200 for 2000 images.)

✓ **Select most relevant clusters**

Evaluate the intra-cluster similarity.

$$SIM(\mathbb{C}) = \frac{\sum_{P_i,P_j \in \mathbb{C}, i \neq j} sim(P_i, P_j)}{{}_nC_2} \qquad sim(P_i, P_j) = \frac{V_i \cdot V_j}{\sqrt{\mid V_i \mid \mid V_j \mid}}$$

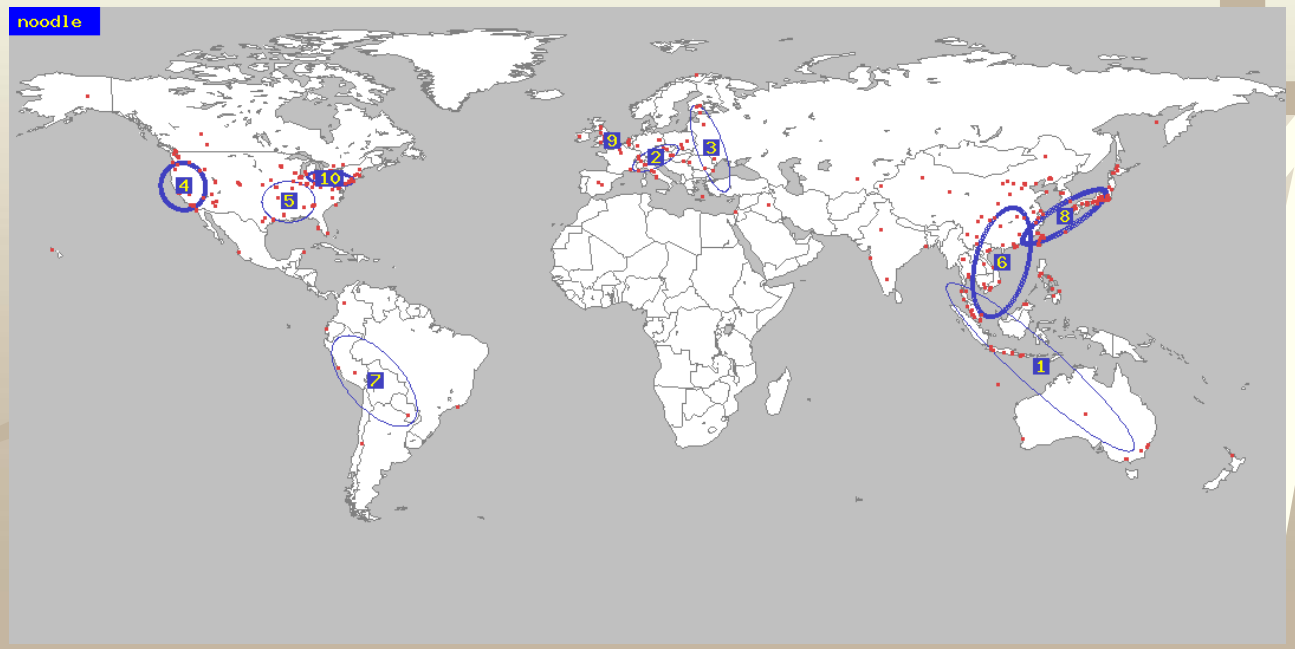
Discard the clusters with small members (m<10) or small similarity. *(select 40 clusters)*

Approach: (2) location clustering

2) Detect representative regions

✓ Clustering based on geographic locations

K-means clustering over location vectors :
(longtude, latitude) (k=5)



Approach: (3) select regional representative photos

3) Generate representative photo sets for representative regions

✓ Generate the PLSA topic vectors (vector compression)

✓ Probabilistic Latent Semantic Analysis (PLSA)

$$P(word, image) = P(image)\sum_{topic} p(word \mid topic)P(topic \mid image)$$

a method to compress a vector to extract "topic" elements like Latent Semantic Analysis

✓ Aggregate photos according to the distribution of mixture topics

K-means clustering over
PLSA topic vectors (k=5) for each region .
Select the largest cluster as a final result.

Example of a PLSA result

- *Mountain*
  10 topics

probabilistic soft clustering

P(topic|image)

Experimental Results

- "noodle", "flower", "castle", "waterfall", "beach"
  - ✓ For each concept , collect about 2000 geo-tagged photos from Flickr distributed evenly in the world wide areas
- Quantitative evaluation for the 1st step
  - ✓ Evaluation on our proposed method for extracting the most relevant photos
  - ✓ Precision and Recall
  - ✓ Color-histogram-based method for comparison
- Examples of regional representative photos

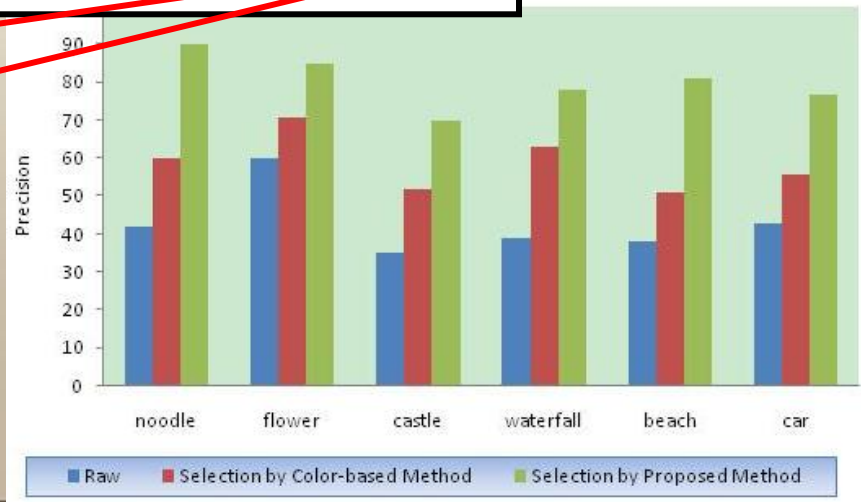
Quantitative Evaluation for (1) noise removal

- raw vs. color-based vs. BoVW (proposed)

Concepts	Raw photos From Flickr	Selection by Color-based Method	Selection by Proposed Method
noodle	2080 (42)	769 (60, 54)	752 (90, 80)
flower	2225 (60)	703 (71, 37)	705 (85, 45)
castle	1848 (35)	780 (52, 61)	761 (70, 81)
waterfall	1901 (39)	689 (63, 59)	672 (78, 70)
beach	1917 (38)	824 (51, 58)	813 (80, 90)
TOTAL/AVG.	9971 (43)	3765 (59 54)	3703 (81 73)

Precision

Recall

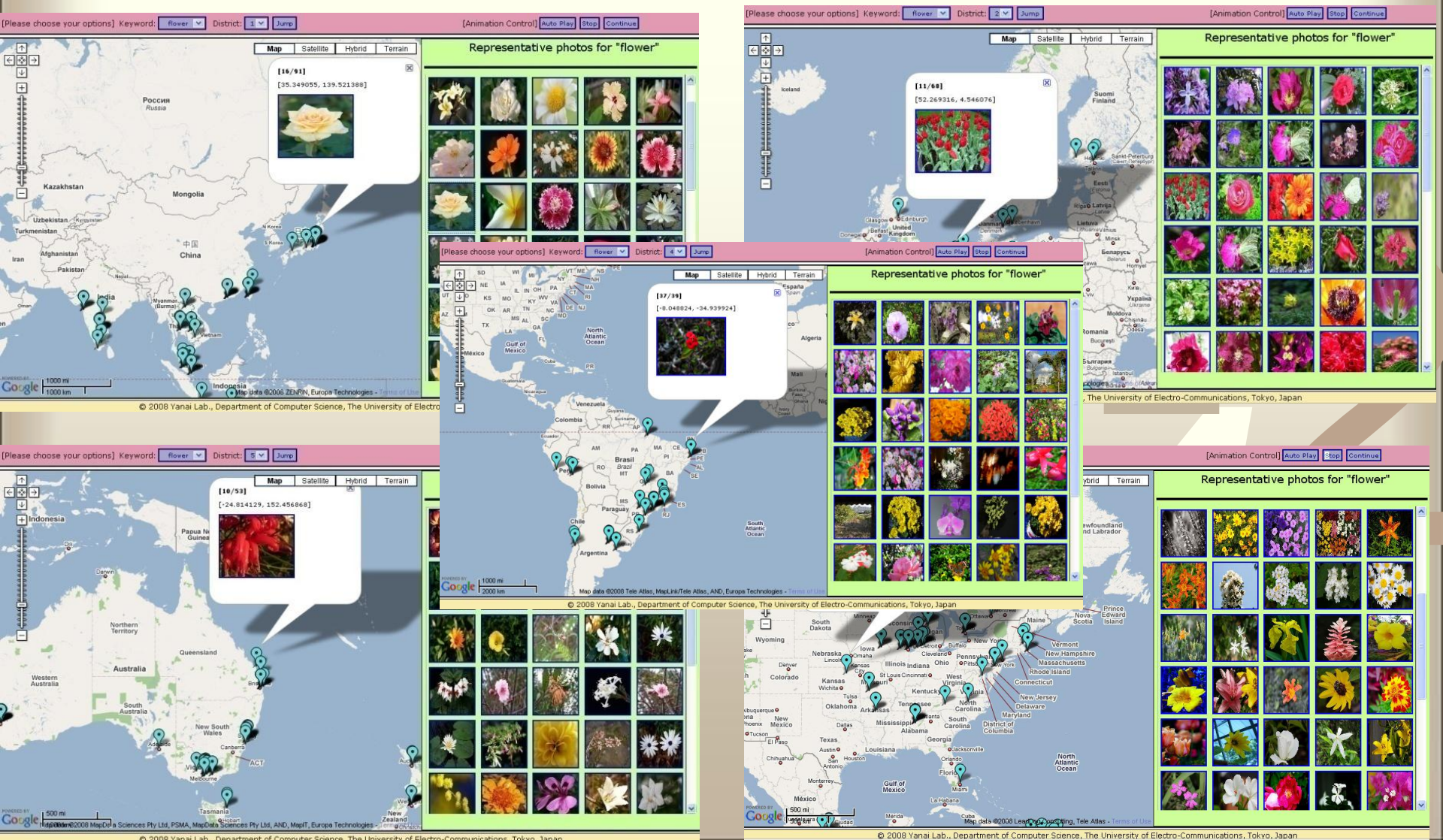


[Example of results] "noodle"

Figure 2: "Noodle" in Japan. Chinese-style noodle "ramen" is popular.

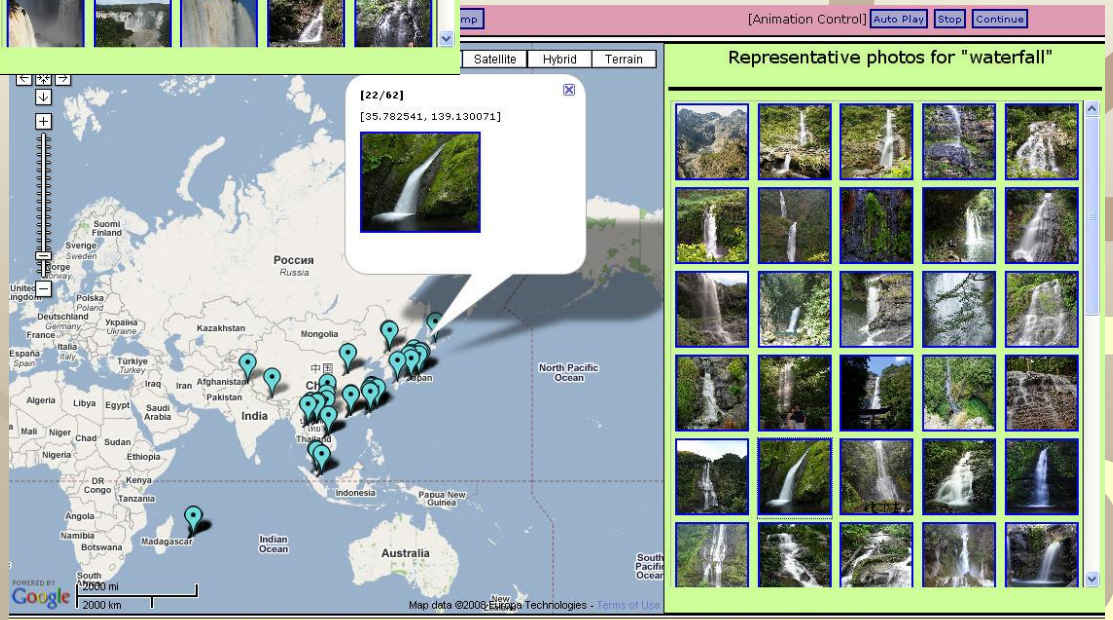
Figure 3: "Noodle" in Europe. Most of photos are "Spaghetti".

[Example of results] "flower"



[Example of results] "waterfall"

"Powerful " waterfalls in South America



"Beautiful" waterfalls in Asia

Figure 4: "Wedding cake" in Mid US. Tall cakes are common. This is five-layered.

Figure 5: "Wedding cake" in Europe. They are much shorter and simpler than US.

Conclusion

- ✓We have proposed a novel topic which is to discover canonical photos of the specific objects over the world.
  - ✓The results help discover cultural differences.

For more results, please access:
http://mm.cs.uec.ac.jp/qiu-b/ASRP/