Mining Cultural Differences from a Large Number of Geotagged Photos

Keiji Yanai
The University of Electro-Communications
Chofu, Tokyo 182-8585 JAPAN
yanai@cs.uec.ac.jp

Bingyu Qiu
Beijing University of Posts and Technology
Beijing, 100876, China
bigyuqiu@bupt.cn

ABSTRACT

We propose a novel method to detect cultural differences over the world automatically by using a large amount of geotagged images on the photo sharing Web sites such as Flickr. We employ the state-of-the-art object recognition technique developed in the research community of computer vision to mine representative photos of the given concept for representative local regions from a large-scale unorganized collection of consumer-generated geotagged photos. The results help us understand how objects, scenes or events corresponding to the same given concept are visually different depending on local regions over the world.

Categories and Subject Descriptors: H.3.3 Information Search and Retrieval: Miscellaneous

General Terms: Algorithms, Experimentation, Measurement

Keywords: geotag, object recognition, representative image, Flickr

1. INTRODUCTION

Recently, consumer-generated media (CGM) on the Web has become very popular. Especially, photo sharing sites such as Flickr and Picasa are representative CGM sites, which store a huge number of consumer-generated photos people uploaded, and make them accessible via the Web for everyone. Photo sharing sites collect not only photos but also metadata on uploaded photos. As metadata users add to photos, textual information such as keywords and comments is common. Recently, in addition to texts, some users attach "geotags" to their uploaded photos. Note that a "geotag" means metadata which represents a location where the corresponding photo was taken, which is expressed by a set of a latitude and a longitude.

An accurate geotag can be obtained with a GPS device or a location-aware camera-phone. However, since it forces us to use relatively special devices, GPS-based geotags have not been common so far. Instead, map-based geotags have become common, after Flickr, which is the largest photo sharing site in the world, launched an online geotagging interface in 2006. Then, Flickr also became the largest "geotagged" photo database in the world. According to [3], there are currently over 40,000,000 public geotagged photos on Flickr, and 100,000 geotagged photos have been added every month. These geotagged photos would be valuable not only for browsing and finding individual concepts, but also for helping us understand how local specific objects or scenes over the world are different.

WWW2009, April 20-24, 2009, Madrid, Spain.
ACM 978-1-60558-487-4/09/04.

Our objective is thus to facilitate a system which can automatically select relevant and representative photographs for the general object or scene concepts in the worldwide dimensions. In particular, we consider geotagged photos on Flickr, identify the representative image groups, and generate an aggregate representation based on locations that allows navigation, exploration and understanding of the differences of general concepts depending on local locations in the world visually. We employ the state-of-the-art object recognition technique developed in the research community of computer vision to mine representative photos of the given concept for each region from a large-scale unorganized collection of consumer-generated geotagged photos. The results help us understand how objects, scenes or events corresponding to the same given concept are visually different depending on local regions over the world.

Figure 1: After collecting geotagged photo related to the given concept by the text-tag-based search, we remove noise images, cluster regions and select regional representative images.

2. OVERVIEW OF OUR APPROACH

Our approach for selecting the representative images for representative local regions from geotagged images consists of three main stages (Figure 1): (1) removing irrelevant images to the given concept, (2) estimating representative geographic regions, and (3) selecting representative images for each region.

First, we apply clustering techniques to partition the image set into similar groups, based on bag-of-visual-words feature vectors [1]. By evaluating the intra-cluster densities as well as the cluster member numbers, we discard most of the irrelevant images and obtain a reduced set of images which are visually similar to each other. This stage could be regarded as the "Filtering Stage".

Then, we geographically cluster the reduced set of images and select large geographic clusters as representative regions. Here we use the k-means clustering algorithm based

on the geographic latitude and longitude of photos to obtain representative regions in the world for the given concept.

Finally, for each representative region, we perform the Probabilistic Latent Semantic Analysis (PLSA) [2] to identify the distinct "topics", do additional clustering on the entire topic vectors, and select the "significant" cluster as the representative results for this geographic region. In addition, with the help of on-line map service, we have implemented a map-based browser to show selected representative photos for understanding of differences regarding appearances of generic object concepts over the world.

Please refer to [4] for the detail.

3. EXPERIMENTAL RESULTS

To test and verify if our approach works in practice, we conducted preliminary experiments with photos collected directly from Flickr. In the experiments, we used seven "object" concepts and two "scene" concepts including "noodle", "wedding cake", "flower", "castle", "car", "waterfall" and "beach". For each concept, we collected about 2000 most relevant geotagged photos distributed evenly in the world wide areas. The precision of raw photos of these seven concepts is 43% on average, which is defined as (# relevant images)/(# all images). After "Filtering Stage", it was improved to 80%, which indicated that noise removal was effective.

We show the representative photos selected for several representative regions, while these regions were generated automatically based on geographic locations of the most relevant photos selected in the "Filtering Stage". Figure 2 and Figure 3 show the results for the concept "noodle", each of which presents the most representative photos generated for the approximate regions: Japan and Europe. Without doubt, these results can help us understand about the "noodle" in these local areas. For example, Figure 2 demonstrates many "ramen" photos in Japan and Figure 3 demonstrates "spaghetti" photos in the European area. In addition, South East Asia, Mideast US and Western US are obtained as other representative regions, representative photos of which also have characteristics such as "noodles" in the South East Asia area containing some Taiwanese style noodles and spicy Thai noodles.

Figure 4 and Figure 5 correspond to "wedding cake" in Europe and in Mid US, respectively. We can find many of the wedding cakes in Mid US are much taller than ones in Europe.

For the scene concept "waterfall", we extracted the representative photos for four large regions: Asia, Europe, North America, and South America. From the results, we can find that waterfalls in South America seem to be more powerful, while waterfalls in the Asian area are somehow more beautiful. Such kinds of implications would be helpful in guiding travels around the world.

To see more results, please visit the following website: http://img.cs.uec.ac.jp/yanai/ASRP/ .

4. CONCLUSIONS AND FUTURE WORK

In this paper, we present a novel topic which is for the purpose of generating representative photographs for typical regions in the world for mining cultural differences on the given concepts, and provide an approach to achieve it with the help of geotagged collections. The results help us understand how objects or scenes associated with the same concept are different depending on local regions in the world visually.

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

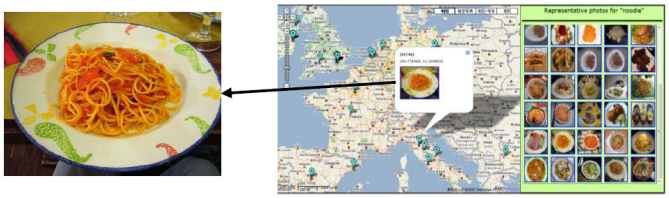

Figure 3: "Noodle" in Europe. Most of photos are "Spaghetti".

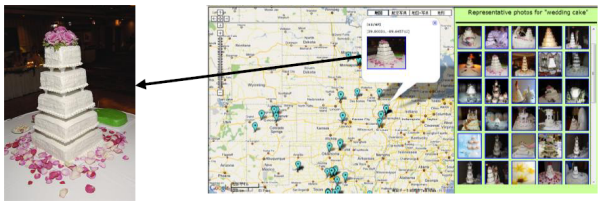

Figure 4: "Wedding cake" in Mid US. Tall cakes are common. This is five-layered.

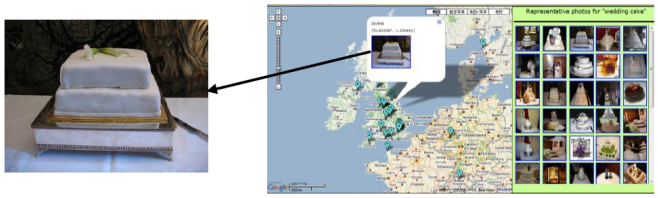

Figure 5: "Wedding cake" in Europe. They are much shorter and simpler than US.

For future work, we plan to make extensive experiments for more concepts from a larger set of photos, and think out some other strategies in detecting more representative regions with a more precise scope. In addition, we will conduct some quantitative evaluations on the representativeness of the photos selected for the corresponding region and the differences of the tendency of representative photo sets among different local regions.

5. REFERENCES

[1] G. Csurka, C. Bray, C. Dance, and L. Fan. Visual categorization with bags of keypoints. In *Proc. of ECCV Workshop on Statistical Learning in Computer Vision*, pages 59–74, 2004.

[2] T. Hofmann. Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning*, 43:177–196, 2001.

[3] L. Kennedy and M. Naaman. Generating diverse and representative image search results for landmarks. In *Proc. of the International World Wide Web Conference*, pages 297–306, 2008.

[4] B. Qiu and K. Yanai. Objects over the world. In *Proc. of Pacific-Rim Conference on Multimedia*, pages 296–305, 2008.