A WORLDWIDE TOURISM RECOMMENDATION SYSTEM BASED ON GEOTAGGED WEB PHOTOS

Liangliang Cao[*], *Jiebo Luo*[♮], *Andrew Gallagher*[♮], *Xin Jin*[†], *Jiawei Han*[†] *and Thomas S. Huang*[*]

[*]Beckman Institute and Coordinated Science Lab, Dept. ECE, UIUC
[♮]Kodak Research Laboratories, Eastman Kodak Company
[†]Dept. Computer Science, University of Illinois at Urbana-Champaign

ABSTRACT

This work aims to build a system to suggest tourist destinations based on visual matching and minimal user input. A user can provide either a photo of the desired scenary or a keyword describing the place of interest, and the system will look into its database for places that share the visual characteristics. To that end, we first cluster a large-scale geotagged web photo collection into groups by location and then find the representative images for each group. Tourist destination recommendations are produced by comparing the query against the representative tags or representative images under the premise of "if you like that place, you may also like these places".

Index Terms— Tourism Recommendation, Geotagged Web Image Retrieval

1. INTRODUCTION

This paper describes a system that allows tourists to discover interesting travel destinations. In the past, people obtained suggestions for their personal tourism from their friends or travel agencies. Such traditional sources are user-friendly; however, they have serious limitations. First, the suggestions from friends are limited to those place they have visited before. It is difficult for the user to gain information from less-traveled members of the community. Second, the information from travel agencies is sometime biased since agents tend to recommend businesses they are associated with. Even worse, when users plan their travel by themselves, they often find their knowledge is too limited to produce a satisfying travel experience.

The prevalence of the Internet provides the possibility for users to learn to plan their tourism by themselves. There has been an increasing amount of visual and text information which the user can explore from various websites. However, the Internet information is too overwhelming and the users have to spend a long time finding those that they are interested in. Users desire more efficient ways to find tourism recommendations which can save time and efforts.

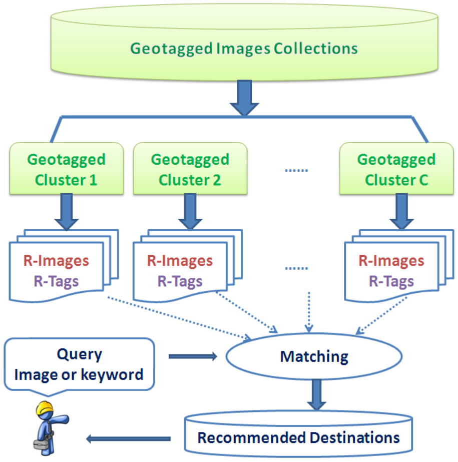


Fig. 1. Diagram of our tourism recommendation system

This paper aims to design a user-friendly and effective system for the task of tourism recommendation. We believe the most intuitive way to describe a place is to show the user images so that they know whether or not they would like such a place. We employ geotagged images to show the interesting scenes of different places in the world, and help users to find destinations which match their interests best. The geotagged images are those images taken with geographical information. With the advance in low-cost GPS chips, cell phones and cameras are becoming equipped with GPS receivers and thus able to record the location while taking the pictures. Geotagged images have become popular in recent years.

There is a huge number of geotagged images from popular websites such as Flickr and Google Earth. However, there has been no previous work studying how to use them for tourism recommendation. The difficulty lies in several aspects: First, it is not an easy task to understand a user's interests. There is always a semantic gap between the high level concept and the low level features. Second, the huge collection of online geotagged images contains many irrelevant samples, whose contents are not relevant to the geographical coordinates. Finally, an efficient tourism recommendation system demands

[*]Supported by Kodak and also CSE fellowship from UIUC.

for a fast approach to find the places with geotagged images which match user's interests.

To handle these difficulties, we propose a two step approach to build a tourism recommendation system. In the offline step, we organize the whole geotagged database and extract representative samples for future use. We develop an efficient clustering algorithm to divide the earth area into regions, which is based on not only the geographical coordinates but also the distributions of geotagged images. For each geotagged cluster, we look for the most representative images and tags. These representatives are called *R-Image* or *R-Tags*. In the online step, we allow users to input queries to describe their destinations and then search for the representative images that match the user interests. The corresponding geotagged regions are obtains as the recommended destinations. Figure 1 shows the diagram of our recommendation systems.

Our contribution is three-fold: 1. We develop an effective algorithm to cluster geographical images. Our scheme is flexible since it does not require one to specify the number of clusters. We design an efficient technique to speed up the mean shift algorithm, and our approach can cluster one million of images within 10 minutes on a typical PC. 2. We propose using representative images for tourism recommendation. The representative samples are rid of the irrelevant images and can describe well the characteristics of every locations. In addition, the number of representative images is far less than that of the whole geotagged image collection, which makes the retrieval faster. 3. We design a flexible interface which allows the user to choose either keywords or query images to describe their interests. The combination of two kinds of queries provide higher chance to the user to find a satisfying place.

2. RELATED WORK

The traditional image retrieval systems consider the problem of searching in large databases for images similar to the query [10], which are usually limited to professional database such as COREL database. images accessible online, there has been an increasing interest in studying web images for the retrieval tasks. Since web images are often accompanied by text such as image title, surrounding text or user annotation, many web image retrieval works [6] consider both visual and tag features for their problems. Motivated by these works, this paper will also employ these two features for recommendation task.

Geo-tagging has been an emerging phenomenon in photo-sharing web sites. There has been much work employing geographic annotation to help image annotation [2], and image summarization and management [7]. Quack *et al.* consider the problem of object/event retrieval with the help of Wikipedia [9]. These works are different from our tourism recommendation problem. The work by Hays and Efros [5] might be the one closest related to this paper, which estimates the geographic location of an image by searching for those visually similar samples in the given dataset in a nearest fashion. However, we argue that it is possible for two images to be visually similar to each other even though they correspond to different locations (e.g., two buildings or two beaches), and both locations can be good candidates for tourism recommendations.

3. DATASET

We collect a geo-tagged database by collecting 1,123,847 images with GPS records from Flickr [1]. The GPS location for each image is represented by a two dimensional vector of latitude and longitude. Each image is also associated with user-provided tags, of which the number varies from zero to over ten.

Figure 2(a) shows the distribution of GPS locations. It can be seen that geotagged locations are not evenly distributed. We argue that the image density at a location is related with the potential for that location to be of interest to a tourist. The next section will discuss how to use clustering to avoid these regions as candidates of tourism recommendation.

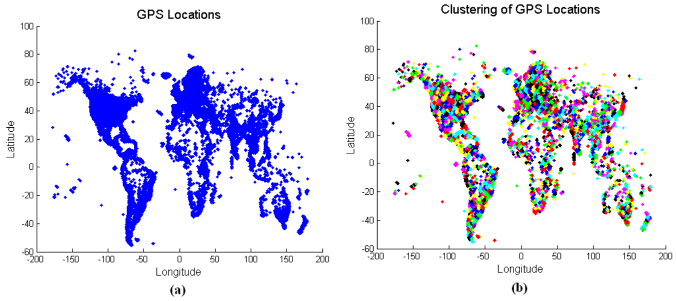


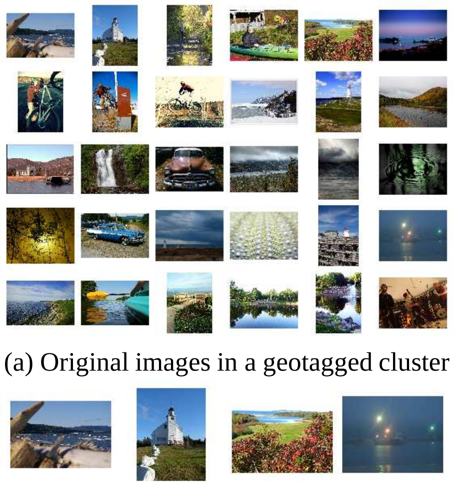
Fig. 2. The clustering results of geotagged images. (a): the distribution of 1.1 million geotagged images. (b): geo-clustering of geotagged images, where clusters are marked with different colors.

4. EFFECTIVELY CLUSTERING THE GEOTAGGED PHOTOS

To cluster the geotagged photos, we consider the mean shift algorithm [4] for the GPS coordinates. Mean shift clustering is a nonparametric method which does not require to specify the number of clusters, and does not assume the shape of the clusters. Starting from a given sample x, Mean shift looks for the vector

$$\mathbf{m}(\mathbf{x}) = \frac{\sum_i \mathbf{x}_i g_i}{\sum_i g_i} \tag{1}$$

where g_i is the local kernel density function in the form of $g_i = g(||(\mathbf{x} - \mathbf{x}_i)/h||^2)$, where g should be a nonnegative, nonincreasing, and piecewise continuous function. Fukunaga and Hostetler [4] proved that the mean shift vector $\mathbf{m}(\mathbf{x}) - \mathbf{x}$ is in the direction of the maximum increase in the density.



(a) Original images in a geotagged cluster

(b) Selected representative images

Fig. 3. Example of finding representative images.

The most expensive operation of the mean shift method is finding the closest neighbors of a point in the space. In this paper, we propose to formulate the kernel function g as flat kernel

$$g(x) = \begin{cases} 1 & \text{if } |x| \leq 1 \\ 0 & \text{if } |x| > 1 \end{cases}$$

we can easily obtain that $g_i \neq 0$ if and only if $||\mathbf{x} - \mathbf{x}_i||^2 < h^2$. Since each $\mathbf{x}$ is GPS coordinate in R^2, we can get the necessary condition for $g_i \neq 0$:

$$|\mathbf{x}(1) - \mathbf{x}_i(1)| \leq h, \quad |\mathbf{x}(2) - \mathbf{x}_i(2)| \leq h \tag{2}$$

With (2), we can search for the closest neighbors of a sample effectively and speed up the clustering process. Algorithm 1 describes our clustering procedure.

Algorithm 1 : Mean-shift based GPS Clustering

```
Input: GPS coordinates $\mathcal{X} = \{\mathbf{x}_l\}$, where $\mathbf{x}_l$ is a two dimensional vector denoting longitude and latitude.
Initialize center set $\mathcal{C} = \emptyset$, and non-visited set $\mathcal{U} = \mathcal{X}$.
for each $\mathbf{x}_l \in U$ do
  Set $\mathbf{x} = \mathbf{x}_l$, $\mathcal{V} = \{\mathbf{x}_l\}$.
  do
    Find $\mathbf{x}$'s neighborhood set $\{x_j\}$ using (2).
    Compute the vector $\mathbf{m}(\mathbf{x})$ using (1).
    Update $\mathbf{x} = \mathbf{m}(\mathbf{x})$ and $\mathcal{V} = \mathcal{V} \bigcup \{x_j\}$.
  until $\mathbf{x}$ converge.
  Update $\mathcal{C} = \mathcal{C} \bigcup \mathbf{x}$ and $\mathcal{U} = \mathcal{U} - \mathcal{V}$
end for
Output: The set of cluster centers $\mathcal{C}$ and the corresponding samples in each cluster.
```

Algorithm 1 works very efficiently with low dimensional data. For our dataset of more than 1.1 millions of images, the clustering procedure takes less than 10 minutes. Figure 2(b) show the clustering results, where different clusters are marked by different colors. Even from a high-level (without zoom-in), our algorithm obtains reasonable clustering results (1108 clusters).

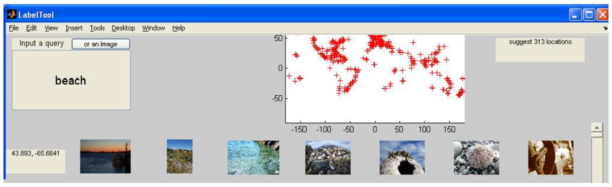


Fig. 4. The interface of the proposed system: for the query of keyword "beach". A web-based prototype system will be released.

5. FINDING REPRESENTATIVE SAMPLES

Our next step is to find the representative samples in each geotagged cluster. We consider two kinds of representatives, tags and images, which are named as *R-images* and *R-tags*, respectively. We explore the user labeled tags associated with each image to find R-tags. We compute the occurrence of each tag in each cluster, and choose the representative tags with occurrence larger than a threshold (set as 10 in our experiments).It is a non-trivial task to find the R-images. We employ the affinity propagation [3] for this task.

Given N image in a geotagged cluster, the similarity between image i and k is denoted as $s(i,k)$. In our experiments, the similarity is measured by a Gaussian function

$$s(i,k) = \exp(-||\mathbf{f}_i - \mathbf{f}_k||^2/\delta)$$

where $\mathbf{f}$ denotes the image feature, e.g., GIST [8] or color histogram. δ is set as the estimated variance of the given features. Using affinity propagation, we are looking for exemplar c_i for each image i, where $c_i = 1, ..., N$. Here $c_i = i1$ means the image i is a representative image since its exemplar is itself. Affinity propagation considers all data points as potential exemplars and iteratively exchanges messages between data points until it finds a good solution with a set of exemplars. There are two kinds of messages: *responsibility* $r(i,k)$ stands for the confidence of image i belongs to a cluster k , while *availability* $a(k,i)$ denotes the possibility of image k being the exemplar of image i. The affinity propagation algorithm updates $r(i,k)$ and $a(k,i)$ iteratively until converge. Finally The exemplar for image i is selected by $p_i = \arg\max_k[r(i,k) + a(k,i)]$.

Although affinity propagation finds the potential representative images in each geotagged cluster, not all these images are meaningful. To remove the insignificant images e.g., those without popular scenery contents, we count the popularity N_p for each potential representative images p, i.e., the number of

Table 1. Some example tourism recommendations using keyword query.

Query	Loc 1	Loc 2	Loc 3	Loc 4	Loc 5	Loc 6	Loc 7
Beach	(16.6, -22.9) Santa Maria	(52.0,4.8) Den Haag	(14.1, -61.0) St. Lucia	(37.0, -3.9) Malaga	(10.5,103.8) Kampong Saom	(13.5,144.7), Tamuning	(36.7,-76.1) Virginia Beach
Diving	(26.6,-78.9) Fort Lauderdale	(53.7, 10.1) Hamburg	(39.5, -0.3) Valencia	(4.3,118.6) Taungoh	(22.3, 114.1) Hong Kong	(24.5,-76.0) Nassau	(-8.7, 119.5) Indonesia
Mountain	(52.0,4.8) Driebruggen	(63.5,-149.7) Denali	(37.0,-3.9) Arenas del Rey	(38.8,-9.1) Prior Velho	(50.3,14.3) Mirejovice	(39.7, 2.9) Sencelles	(41.8,24.2) Asenovgrad

images which choose p as their exemplar. When N_p is small, it means p is probably a outlier. We only choose R-images with N_p large enough. Figure 3 shows an example of finding R-images.

6. TOURISM RECOMMENDATION SYSTEM

We build the tourism recommendation system based on the representative tags and images with corresponding GPS locations. The system interface is shown in Figure 4. The user can choose to provide a query either as a keyword or an image, then the system goes through the database and matches the representative images and tags with the given query. For a key word query, a geotagged location is chosen if the representative tags contains the query keyword. For an image query, the geotagged locations are ranked according to the similarity between the query images and the representative image feature in different clusters.

Table 1 shows some retrieval examples using keywords. We show 7 locations for each query although the total recommendations can be as many as hundred. Since it is not easy to interpret GPS coordinates directly, we list the closest city names. We can see that our travel recommendation system can provide a wide range of destinations, therefore is more appealing in the variety than those from friends or travel agencies and potentially more powerful.

We also evaluate the accuracy of using image queries. We first select 10 topics, and for each topic we randomly select 20 image queries. Of the destinations our system recommends, we evaluate how accurately the top 10 recommended places match the given topic. The topics are: architecture, beach, flower, building, island, mountain, lake, park, river, snow. The matching accuracy is measured by the precision among top 10 recommendations. Figure 5 shows our recommendation accuracy. The precisions of our system are quite satisfactory across the 10 topics.

7. CONCLUSION

This paper demonstrates our efforts to build a tourism recommendation system using large-scale geotagged images. We propose to cluster the geotagged images into clusters, and then compute the representative images for each geotagged clusters. The results shows that such a system is helpful for users to find tourism destinations of interests.

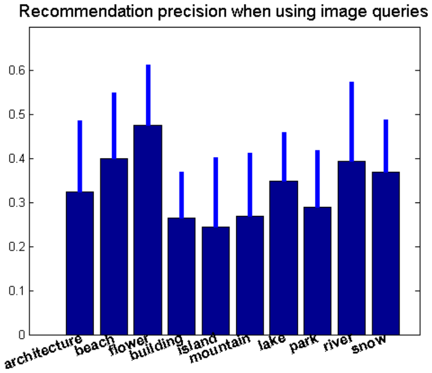


Fig. 5. The precision of our recommendation system using random select queries. The thick bars and thin bars stand for the average precision and standard deviation for top 10 recommendations in each topic.

8. REFERENCES

[1] Flickr APIs. http://www.flickr.com/services/api/.

[2] L. Cao, J. Luo, and T. Huang. Annotating photo collections by label propagation according to multiple similarity cues. In *ACM Multimedia*, 2008.

[3] B. Frey and D. Dueck. Clustering by passing messages between data points. *Science*, 315(5814):972, 2007.

[4] K. Fukunaga and L. Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Trans. Information Theory*, 21(1):32–40, 1975.

[5] J. Hays and A. A. Efros. Im2gps: estimating geographic information from a single image. In *IEEE Proc. CVPR*, 2008.

[6] J. Jia, N. Yu, and X.-S. Hua. Annotating personal albums via web mining. In *ACM Multimedia*, pages 459–468, 2008.

[7] L. Kennedy, M. Naaman, S. Ahern, R. Nair, and T. Rattenbury. How flickr helps us make sense of the world: Context and content in community-contributed media collections. In *ACM Multimedia*, 2007.

[8] A. Oliva and A. Torralba. Modeling the shape of the scene: a holistic representation of the spatial envelope. *IJCV*, 42(3):145–175, 2001.

[9] T. Quack, B. Leibe, and L. Van Gool. World-scale mining of objects and events from community photo collections. *ACM Proc. CIVR*, 2008.

[10] Y. Rui, T. Huang, and S.-F. Chang. Image retrieval: current techniques, promising directions, and open issues. *J. Visual Comm. and Image Representation*, 10:39–62, 1999.