Construction of a Local Attraction Map According to Social Visual Attention

Ichiro Ide, Jiani Wang, Masafumi Noda, Tomokazu Takahashi, Daisuke Deguchi, and Hiroshi Murase

Abstract. Social media on the Internet where millions of people share their personal experiences, can be considered as an information source that implies people's implicit and/or explicit visual attentions. Especially, when the attentions of many people around a specific geographic location focus on a common content, we may assume that there is a certain target that attracts people's attentions in the area. In this paper, we propose a framework that detects people's common attention in a local area (local attraction) from a large number of geo-tagged photos, and its visualization on the "Local Attraction Map". Based on the framework, as a first step of the research, we report the results from a user study performed on a Local Attraction Map browsing interface that showed the representative scene categories as local attractions for geographic clusters of the geo-tagged photos.

Ichiro Ide · Masafumi Noda · Hiroshi Murase
Nagoya University, Graduate School of Information Science, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan
e-mail: ide@is.nagoya-u.ac.jp, mnoda@murase.m.is.nagoya-u.ac.jp, murase@is.nagoya-u.ac.jp

Jiani Wang
Nagoya University, Graduate School of Information Science, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan
e-mail: jwang@murase.m.is.nagoya-u.ac.jp
Currently at Oki Data Corporation

Tomokazu Takahashi
Gifu Shotoku Gakuen University, Department of Economics and Information, 1-38 Naka-Uzura, Gifu, 500-8288, Japan
e-mail: ttakahashi@gifu.shotoku.ac.jp

Daisuke Deguchi,
Nagoya University, Information and Communications Headquarters, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan
e-mail: ddeguchi@nagoya-u.jp

T. Watanabe et al. (Eds.): Intelligent Interactive Multimedia: Systems & Services, SIST 14, pp. 153–162.
springerlink.com

1 Introduction

Following the recent trend of the diffusion of social media on the Internet where millions of people share their personal experiences as digital contents, we can easily obtain thousands of photos tagged with geographic coordinates (geo-tags) indicating where they were taken. We are focusing on the contents of such photos from the point that they imply people's implicit and/or explicit visual attentions.

Especially, when the attentions of many people around a specific geographic location focus on a common content, we may assume that there is a certain target that attracts people's attentions in the area. In this paper, we call such a target a "local attraction".

A local attraction could be a static phenomenon such as an object; artificial or natural, or a scenery observed from the area. Such local attractions are not easy to infer from objective data such as satellite images and maps, since they can be anything from a small statue located exactly on the spot to a panoramic view observable from the spot which may contain geographic objects located miles away.

On the other hand, a local attraction could also be a non-static phenomenon that reflects a common activity in the area, such as shopping, eating, playing, and so on. These are even more difficult to infer from objective data, since they need to be recognized in the context of human activity observed on the ground.

Moreover, without the help of social media, it would be difficult to infer what attracts people only from objective data. We are not interested in providing users with information on interesting spots located in the middle of a desert where no one actually visits, but instead, with information on attractive spots where many other people have also visited and showed their interest by taking a photo and sharing it on the Internet.

Meanwhile, traditional media such as travel guides and maps cover both types of local attractions, but their contents are not necessarily updated frequently. In order to cope with the rapidly changing modern society, we considered that the continuously updated information provided from a large number of people through social media should be useful to construct a map that reflects the most up-to-date local attractions.

In this paper, we propose a framework that automatically detects people's common attention in a local area (e.g. local attraction) from a large number of geo-tagged photos, and its visualization on a map called the "Local Attraction Map". Based on the framework, as a first step of the research, we report the results from a user study performed on a Local Attraction Map browsing interface that showed the representative scene categories as local attractions for geographic clusters of the geo-tagged photos (Figure 1).

The paper is organized as follows: Section 2 introduces related works on landmark detection and travel planning based on social media. Section 3 introduces the proposed method to construct the Local Attraction Map based on a large number of geo-tagged photos. Section 4 reports the result of the user study, and finally Sect. 5 concludes the paper.

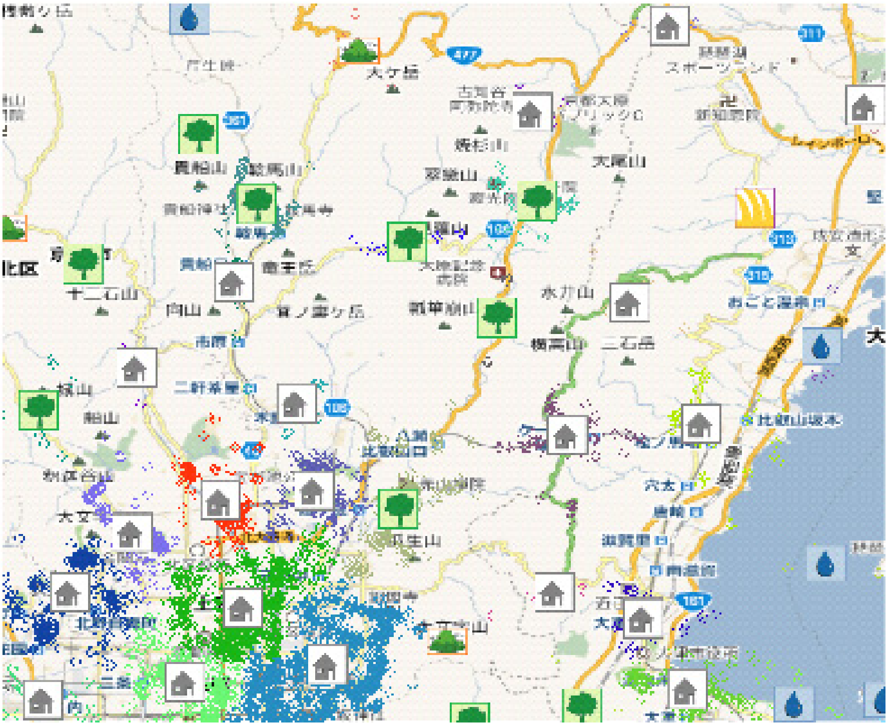

Fig. 1 A Local Attraction Map for Kyoto, Japan.

2 Related Works

Commercial softwares such as Google Map[1], Google Earth[2], and Panoramio[3] are nowadays important tools for travelers to perform a visual survey before visiting a planned travel destination. However, it is usually difficult to find sites / areas-of-interest in a destination where the user is not familiar with, by simply using these services.

Making use of the geo-tags tagged to photos is a recent trend to support travel planning for such users. This type of research can be separated into two topics: 1) travel route mining and recommendation [1, 3, 8, 9], and 2) landmark detection and representative photo selection [2, 6, 10, 12].

The travel route mining and recommendation methods analyze sequences of geo-tagged photos and propose routes that match the interests of a user. Since they focus mostly on the sequence of geo-tags, they usually do not make use of the image contents, except for Cheng et al.'s work [3] that infers user attributes from faces in the images and matches them with the user's attributes for the recommendation.

[1] `http://maps.google.com/`
[2] `http://earth.google.com/`
[3] `http://www.panoramio.com/`

Meanwhile, the landmark detection and representative photo selection methods combine geo-clustering and tag and/or image clustering in order to obtain a popular landmark and sometimes its popular angle. Weyand and Leibe's method [10] even generates a 3D model of a landmark from a large number of photos taken from different angles. Chen et al.'s method [2] is also interesting in the sense that it generates simplified icons from a representative view of a landmark, and shows them on a simplified map. These methods are useful for landmarks; mostly buildings, but not all kinds of local attractions such as those mentioned in Sect.1 can be handled.

In both research topics, providing information on actual landmarks may be too concrete for users who may not be familiar with or have a clear idea of what awaits them in the planned destination; they may simply want to know where the cultural heritages are, where they can find restaurants, or where they can do street shopping, without an idea of a particular site or shop. Thus in this paper, we will focus on providing the users with the types of the local attractions rather than concrete information on the individual local attractions.

Scene category classification itself has been a very hot research topic in the past decade, such as Xiao et al.'s work [11]. However, since it is a widely studied research topic, we will not focus much on the technology in this paper.

3 Construction of a Local Attraction Map According to Social Visual Attention

In this Section, we describe how the proposed method analyzes the social visual attention and construct the Local Attraction Map.

3.1 Analyzing the Social Visual Attention

In order to analyze the social visual attention from a large number of geo-tagged photos taken within a specific region, they are first clustered according to their geographic location. Next, for each cluster, a representative scene category is decided by image classification. Details of the two steps are described in this Section.

3.1.1 Clustering of Geo-tagged Photos

Since the local attraction could be any phenomenon from an object located exactly on the spot to a terrain that covers a wide area, the size and the shape of the area that covers a local attraction should be flexible. Thus, we decided to extract clusters from the distribution of the geo-tagged photos instead of using fixed-sized shapes (most likely, blocks).

For the clustering, we used the nearest neighbor method with a restriction that if the geographic distance between two clusters are larger than θ [km], they are not merged.

Figure 2 shows the result of the clustering for constructing the Local Attraction Map shown in Fig. 1.

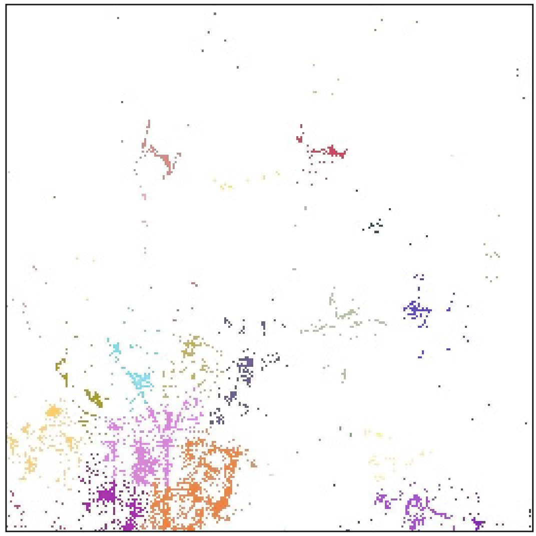

Fig. 2 Example of the clustering of geo-tagged photos. Each dot represents a photo taken at the location. Different colors indicate different clusters.

3.1.2 Decision of the Type of Local Attraction

After the clustering, for each cluster, the type of local attraction is decided. As mentioned in Sect. 1, at this moment, we have only implemented certain static phenomena, namely, scene categories as local attractions. We defined five scene categories as shown in Fig. 3; "city", "forest", "water", "flatland", and "mountain".

For the scene category classification, we implemented a five-class support vector machine (SVM) classifier with a bag-of-features (BoF) representation of local image features obtained by the scale-invariant feature transform (SIFT) algorithm [7] and a normalized HSV color histogram as image features.

The classifier was trained by 16,689 categorized photos from the SUN database [11]. We manually selected 39 categories from the SUN database and mapped them onto the five scene categories as listed in Table 1. For reference, the classifiers had

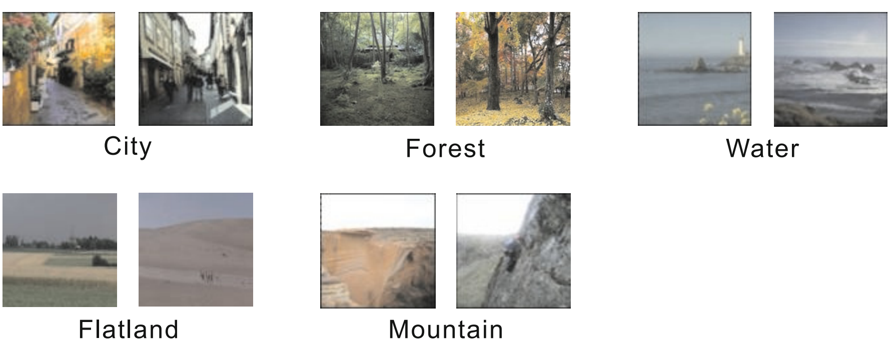

Fig. 3 Example of photos that belong to each scene category.

Table 1 Correspondence of the scene categories and the SUN database categories.

Scene category	SUN Database categories
City	alley, amusement_park, bridge, building, fountain, gazebo, house, market, pagodas, place, railroad_track, shopfront, street, temple, tower, village
Forest	botanical_garden, forest, forest_path, park
Water	bridge, canal, coast, creek, dam, hot_spring, islet, lake, ocean, pond, river, sea_cliff, waterfall
Flatland	amphitheater, badlands, desert, field
Mountain	cliff, dam, mountain, sea_cliff, valley

a recognition rate of approximately 77% in a ten-fold cross validation experiment. Although this accuracy is not significantly high, we can expect the use of state-of-the-art general object recognition methods in the future.

Finally, for each cluster, the most frequent scene category is selected as the representative one; the type of local attraction.

3.2 Construction of the Local Attraction Map

Finally, both the location of each geo-tagged photo and the representative scene category for each cluster are super-imposed onto a digital map as shown in Fig. 1. In the map, each geo-tagged photo is indicated by a '○' with a color that represents the representative scene category for the cluster that it belongs to. The representative scene category for each cluster is also indicated by a scene category icon as shown in Fig. 4.

Fig. 4 Icons that represent scene categories in the Local Attraction Map.

4 User Study

In order to evaluate the usefulness of the proposed Local Attraction Map, we performed a user-study using a Local Attraction Map browsing interface as shown in Fig. 5. The interface allows users to browse the Local Attraction Map (left-hand side), and at the same time, scan through photos that belong to the representative scene category for a specified cluster displayed in a separate panel (right-hand side).

Fig. 5 The Local Attraction Map browsing interface. The picture browsing panel allows users to browse geo-tagged photos that belong to a representative scene category at a location specified in the Local Attraction Map panel.

4.1 Preparation and Experimental Conditions

For the user study, we constructed a Local Attraction Map for Kyoto, Japan (Figure 1). Table 2 shows the parameters and conditions set to create the map. From the specified area, we collected 4,536 geo-tagged photos from Panoramio. The clustering yielded 39 clusters with an average of 112 photos per cluster.

The user study was performed in the following steps:

Step 1. Asked a subject to perform a survey on a planned travel destination; Kyoto, Japan, using both Panoramio and the proposed Local Attraction Map browsing interface.

Table 2 Parameters and conditions set to create the Local Attraction Map in Fig. 1.

Parameter / condition	Value
Target area	$(+35.00244°, +135.71102°) - (+35.16454°, +135.90946°)$ A square area with a size of approximately 20 [km] $\times$ 20 [km] that includes Kyoto city and its vicinity.
Image source	Panoramio (`http://www.panoramio.com/`)
Clustering threshold (θ)	2 [km]
Subjects	25 students

Step 2. After lecturing the function of the Local Attraction Map browsing interface, asked the subject to evaluate its usefulness for the purpose of performing a survey on a planned travel destination. The subjects selected from the following five candidates for the evaluation: Useful, Relatively useful, Neutral, Relatively Useless, and Useless. They were also asked to provide reasonings for the judgments.

4.2 Result and Analysis

Table 3 shows the result of the evaluation by the users. 76% $(= 19/25)$ of the users evaluated the Local Attraction Map as useful to some extent. From this result, we consider that the proposed representative scene category visualization function combined with the geographic layouting of geo-tagged photos was useful for performing a survey on a planned travel destination.

We analyzed the reasonings for the evaluations provided by the users. First, the following are excerpts of the reasonings provided by the users who evaluated either "Useful' or "Relatively useful" (Translated from Japanese).

- The map allows me to grasp at a glance, the location and the types of scenes that I can expect at an unfamiliar travel destination.
- The map shows what (which scene category) most people shoot at at a specific location.
- Since the local information is evaluated according to the number of photos, the map may reveal hidden spots-of-interest.
- Even photos without tags can be classified and searched using the map.
- Different from Panoramio's tag-based search, the map can show various scene categories at the same time.

These reasonings matched our purpose of the proposed Local Attraction Map.

Next, the following are excerpts of the reasonings provided by the users who evaluated either "Neutral" or "Relatively useless" (Translated from Japanese).

Table 3 Result of user evaluation.

Usefulness	Useful	Relatively useful	Neutral	Relatively useless	Useless
Ratio	24% (6/25)	52% (13/25)	16% (4/25)	8% (2/25)	0% (0/25)

- It would be more useful if the map displayed not only photos from a representative scene category, but rather popular photos for all categories.
- If the scene categorization were very accurate, the map could be useful.
- The definition of the scene categories was ambiguous and difficult to understand. The map is not useful unless they are more concrete concepts.
- What happens if two different scene categories are present in a single photo?

Following these reasonings, in the future, we will consider the following points in order to improve the usefulness of the Local Attraction Map.

- Add a function that shows a ranked list of representative scene categories per cluster.
- Modify the scene category classifier so that it could handle the situation where multiple scene categories are present in a single photo.
- Improve the scene category classification accuracy by using a state-of-the-art general object recognition method, and also by developing a classification method that considers the inclusion relation between scene categories.
- In addition to the current scene categories, add those that represent non-static phenomena, such as “Eating”, “Shopping”, “Playing (sports)”, and so on.

5 Conclusion

In this paper, we proposed the framework to construct a “Local Attraction Map” by analyzing the social visual attention from a large number of geo-tagged photos. The user study showed positive results, but at the same time we found several important points that need improvement. In the future, we will work on these points, and also try to perform an experiment and an user study in a larger scale.

Acknowledgements. Parts of this work were supported by Grants-in-aid for Scientific Research from the Japanese Ministry of Education, Culture, Sports, Science and Technology.

References

1. Arase, Y., Xie, X., Hara, T., Nishio, S.: Mining people’s trips from large scale geo-tagged photos. In: Proceedings of the 18th ACM International Conference on Multimedia, pp. 133–142 (2010)
2. Chen, W.C., Battestini, A., Gelfand, N., Setlur, V.: Visual summaries of popular landmarks from community photo collections. In: Proceedings of the 17th ACM International Conference on Multimedia, pp. 789–792 (2009)

3. Cheng, A.J., Chen, Y.Y., Huang, Y.T., Hsu, W.H., Liao, H.Y.M.: Personalized travel recommendation by mining people attributes from community-contributed photos. In: Proceedings of the 19th ACM International Conference on Multimedia, pp. 83–92 (2011)
4. Crandall, D., Backstrom, L., Huttenlocher, D., Kleinberg, J.: Mapping the world's photos. In: Proceedings of the 18th International Conference on World Wide Web, pp. 761–770 (2009)
5. Csurka, G., Bray, C., Dance, C., Fan, L., Willamowski, J.: Visual categorization with bags of keypoints. In: Proceedings of the ECCV2004 International Workshop on Statistical Learning in Computer Vision, pp. 1–22 (2004)
6. Gao, Y., Tang, J., Hong, R., Dai, Q., Chua, T.S., Jain, R.: W2Go: A travel guidance system by automatic landmark ranking. In: Proceedings of the 18th ACM International Conference on Multimedia, pp. 123–132 (2010)
7. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
8. Lu, X., Wang, G., Yang, J., Pang, Y., Zhang, L.: Photo2Trip: Generating travel routes from geo-tagged photos for travel planning. In: Proceedings of the 18th ACM International Conference on Multimedia, pp. 143–152 (2010)
9. Okuyama, K., Yanai, K.: A travel planning system based on travel trajectories extracted from a large number of geotagged photos on the Web. In: Proceedings of the Pacific-Rim Conference on Multimedia (2011)
10. Weyand, T., Leibe, B.: Discovering favorite views of popular places with iconoid shift. In: Proceedings of the 13th IEEE International Conference on Computer Vision, pp. 1132–1139 (2011)
11. Xiao, J., Hays, J., Ehinger, K., Oliva, A., Torralba, A.: SUN Database: Large-scale scene recognition from abbey to zoo. In: Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 3485–3492 (2010)
12. Zheng, Y.T., Zhao, M., Song, Y., Adam, H., Buddemeier, U., Bissacco, A., Brucher, F., Chua, T.S., Neven, H.: Tour the World: Building a Web-scale landmark recognition engine. In: Proc. 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1085–1092 (2009)