Detection of Similar Geo-Regions based on Visual Concepts in Social Photos

Hiroki Takimoto†, Magali Philippe††*, Yasutomo Kawanishi‡, Ichiro Ide‡, Takatsugu Hirayama‡, Keisuke Doman‡‡, Daisuke Deguchi‡, and Hiroshi Murase‡

† Graduate School of Information Science, Nagoya University, Japan
†† ENSEEIHT, France
‡ Graduate School of Informatics, Nagoya University, Japan
‡‡ School of Engineering, Chukyo University, Japan
takimotoh@murase.m.is.nagoya-u.ac.jp

Abstract. Travel destination recommendation is useful to support travel. Considering the recommendation of regions within the destination area to visit, it could be difficult for the users to explicitly indicate their preference. Therefore, we considered that it would be more intuitive to recommend regions in the destination area that are similar to a region already well-known to the user. Thus, in this paper, we propose a method for the detection of similar geo-regions based on Visual Concepts in social photos. We report experimental results and analyses by applying the proposed method to the YFCC100M dataset.

Keywords: Travel; support; recommendation; photo

1 Introduction

According to WTTC, the contribution of Travel and Tourism to world GDP rises to a total of 9.8% and the sector supports 284 million people in employment [1]. Because of such large effects on World economy and employment, promotion of the sector is very important. As one type of promotion, travel destination recommendation systems have been studied. Examples of these are systems which recommend travel routes, and those which detect and recommend landmarks to visit. We considered that it would be useful to recommend regions within the destination area to visit. However, in some cases it could be difficult for the users to explicitly indicate their preference. In order to handle such cases, we considered that it would be easier to recommend regions in the destination area that are similar to a region already well-known to the user. Thus, in this paper, we propose a method to detect similar regions (hereafter referred to as "geo-regions").

To do this, we have to define "geo-regions" and somehow describe each geo-region's feature, and then calculate the similarity between a pair of geo-regions.

* Currently at SONY Electronics.

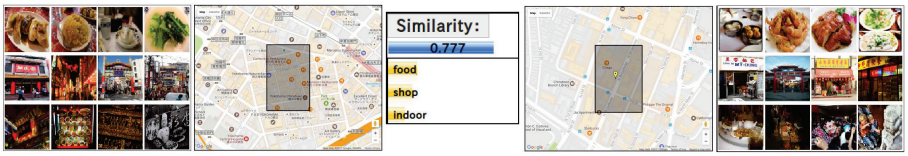

Fig. 1. Example of a detected similar geo-region pair. Both are geo-regions where people focus on Chinese-style streetscapes and Chinese foods; Chinatowns in Yokohama, Japan, and Los Angeles, USA.

Here, we focused on the recent trend of the diffusion of Social Network Services (SNS). When people travel, many of them take photos depending on their interests and post them to SNS. These photos contain rich information on the objects and activities that take place in the regions that they were taken. In order to extract the targets-of-interest from them, we decided to use Visual Concepts to represent the features of a geo-region. Focusing on the interests of the crowd instead of objective information such as that observed from a map or a satellite image, we expect to obtain more intuitive similarities.

The main contribution of this paper is the proposal of a method for the detection of similar geo-regions based on Visual Concepts in social photos. By applying the method to the YFCC100M dataset [8], we could detect similar geo-regions anywhere around the World (Fig. 1).

The remainder of this paper is organized as follows: Section 2 introduces related work. Section 3 introduces the proposed method for detecting similar geo-regions based on Visual Concepts detected from social photos. Section 4 reports and analyzes the results of the detected similar geo-regions, and finally Section 5 concludes the paper.

2 Related Work

As examples of travel recommendation systems, there are route recommendation systems, and landmark detection and recommendation systems.

As a route recommendation system, Kurashima et al. [3] proposed a method that makes use of the photographers' action history from geo-tagged photos on Flickr. Although they used geo-tagged photos to construct the photographer's behavior model, they focused especially on location information and the movement of each photographer, but not on detailed visual information in the photos. As a landmark detection and recommendation system, Shi et al. [6] proposed a personalized system. They used category information extracted from Wikipedia and recommended landmarks to users. They focused on landmarks and their predefined categories, whereas our method focuses on geo-regions and do not predefine categories.

Min et al. [4] proposed a probabilistic topic model called Multimodal Spatio-Temporal Theme Modeling (mmSTTM) and analyzed themes discovered from

twenty popular landmarks around the World. They focused only on popular landmarks and also did not consider similarities between themes, whereas our method focuses on similarities of pairs of arbitrary geo-regions.

As for work which focuses on the similarity between geo-regions, there are two major approaches: Those that refer to texts and those that refer to images.

As work that refers to texts, Shimada et al. [7] proposed a sightseeing spot recommendation system based on text information on the Web. It detects similar geo-regions (locations) by calculating similarities between an user-selected favorite geo-region (location) and each sightseeing spot in their database. In this system, however, similarities cannot be calculated across different linguistic areas because they cannot directly compare keywords extracted from Web sites written in different languages. Considering globalization in the travel sector, for this reason, we decided not to rely on text, but rather make good use of non-linguistic information that can be extracted from image.

As work that refers to images, Pongpaichet et al. [5] implemented a visual analytics system of social photos and made event models to detect similar events. Each model is described by Bag-of-Visual Concepts that co-occur in a pre-defined event. They focusd on the similarity between a geo-region and a pre-defined model in order to detect photos on a certain event, whereas our method focuses on the similarity between geo-regions based on arbitrary events.

Ide et al. [2] proposed a method that automatically detects people's common attention in a geo-region from a large number of geo-tagged photos, and its visualization on a map. By classifying each local area into five scene categories based on global image features extracted from social photos, they classified a geo-region into one of the five categories. In their method, however, since the category of a geo-region was decided according to the majority of photos that belonged to the scene category, other photos that did not belong to the category were ignored and visual information in them were not used to describe the geo-region. Instead, our method makes use of the information in all the photos available and describes a geo-region with combination (distribution) of objects or activities detected from them.

3 Detection of Similar Geo-Regions based on Visual Concepts in Social Photos

We propose a method for the detection of similar geo-regions based on Visual Concepts in social photos. First, geo-regions are fixed by spatially clustering a large number of geo-tagged photos. Next, for each geo-region, their features are described based on Visual Concepts detected from social photos taken there. Finally, the similarity between a pair of geo-regions is calculated and those with high similarity are output as similar geo-regions. Details of the three steps are described below.

3.1 Definition and Decision of Geo-Region

In general, a geo-region is defined as an area demarcated according to certain criteria: administrative division, climatic zone, and so on. However, since we aim to support travel, such objectively-defined areas do not necessarily satisfy the users' purposes. Instead, we subjectively define geo-regions based on the interests of the crowd.

For the decision of geo-regions, we focus on geo-tagged photos. With the diffusion of mobile devices equipped with a GPS function, more and more geo-tagged photos are posted to SNS. Since these photos directly reflect people's interests, we apply mean-shift clustering to locations where the photos were taken to decide geo-regions. Since some clusters (geo-regions) contain few photos, only those that contain sufficient numbers of photos (θ_p) to describe the geo-regions, are used.

3.2 Definition of Similarity

When a pair of geo-regions are similar, they should share something in common. Since our purpose is travel support, it should be based on subjective information in view of travelers. Therefore, as a criterion for measuring the similarity, we focus on the existence of similar objects or activities.

3.3 Description of Geo-Region's Feature

Since Visual Concepts detected from social photos indicate objects and activities that appear in photos taken by people at geo-tagged locations, they are used as a feature to describe the interests of the crowd in a geo-region. Visual Concepts are extracted and used as features to describe geo-regions as follows.

Extraction of Visual Concepts in Social Photos Recently, more and more people use SNS. They often take photos of objects and activities which aroused their interests and post to SNS. Different people have different interests: One may be interested in the landscape, but others may be interested in a landmark or food. We considered that Visual Concepts could be used to describe such diverse interests. Recent advance in deep learning methods has allowed the robust and accurate detection of Visual Concepts. Therefore, in the proposed method, we describe image contents of social photos by combinations (distributions) of Visual Concepts detected in photos.

Let the number of categories of Visual Concepts that can be detected by a detector be N, for an input image, the probability of each Visual Concept is calculated and described as an N-dimensional vector $\mathbf{v}_l$.

Description of Geo-Region's Features A geo-region's feature is described as follows. Let P_l be a set of photos taken in a geo-region c_l. For each photo

$p_{l,i} \in P_l$, the feature vector of $p_{l,i}$ is described as $\mathbf{v}_{l,i}$. The feature vector of a geo-region c_l is described as $\mathbf{V}_l = \sum_i \mathbf{v}_{l,i}$.

We should notice that some visual concepts are general and detected from many photos. Thus, we apply a TF-IDF-like approach to each component of $\mathbf{v}_l$. TF and IDF are calculated as follows. Let the n-th component of $\mathbf{V}_l$ and $\mathbf{v}_{l,i}$ be $V_l(n)$ and $v_{l,i}(n)$, respectively $(n = 1, 2, ..., N)$, then,

$$\mathrm{TF}_{l,i}(n) = \frac{v_{l,i}(n)}{\sum_{m=1}^{N} v_{l,i}(m)}. \tag{1}$$

Let P be all the photos in the dataset, then,

$$\mathrm{IDF}(n) = \log \frac{|P|}{|\{p_j \mid v_j(n) > \theta_{\mathrm{vc}}, p_j \in P\}|}. \tag{2}$$

Here, θ_{vc} is a threshold for judging whether we should consider a Visual Concept exists in a photo or not. Finally, by calculating the following equation, we obtain the geo-region's feature vector $\mathbf{V}'_l = (V'_l(1), V'_l(2), ..., V'_l(N))$, where

$$V'_l(n) = \sum_{i=1}^{|P_l|} \mathrm{TF}_{l,i}(n) \cdot \mathrm{IDF}(n). \tag{3}$$

3.4 Similarity Calculation

Once the geo-regions are described by feature vectors, we calculate the similarity between a pair of them. For a given pair of geo-regions c_i, c_j, their similarity is calculated by Zero-mean Normalized Cross Correlation (ZNCC) as,

$$S_{i,j} = \frac{\sum_{n=1}^{N} (V'_i(n) - \overline{\mathbf{V}'_i})(V'_j(n) - \overline{\mathbf{V}'_j})}{\sqrt{\sum_{n=1}^{N} (V'_i(n) - \overline{\mathbf{V}'_i})^2 \cdot \sum_{n=1}^{N} (V'_j(n) - \overline{\mathbf{V}'_j})^2}}. \tag{4}$$

$$\overline{\mathbf{V}'_l} = \frac{1}{N} \sum_{n=1}^{N} V'_l(n). \tag{5}$$

Finally, pairs with similarities higher than a threshold θ_s are detected as similar geo-regions.

4 Experiment

In this section, we report and analyze experimental results by showing examples of the pairs of similar geo-regions detected by actual social photos.

4.1 Dataset

We used YFCC100M (Yahoo! Flickr Creative Commons 100 Million) dataset [8], which contains 100 million photos and videos from Flickr taken all over the World. Each photo in the dataset is accompanied with metadata (tags, timespan, location and so on). Since the proposed method needs location information of photos, we only used geo-tagged photos in the dataset. In addition to metadata, each photo is annotated with visual information detected by $N = 1,570$ Visual Concept classifiers using CNN. We used them to describe the Visual Concepts in each photo.

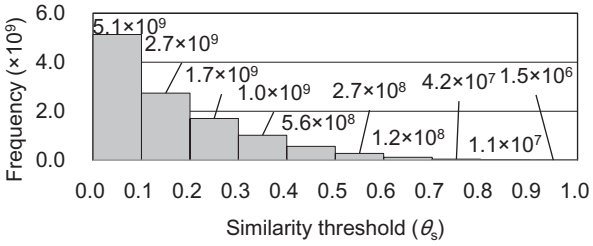

Fig. 2. Frequency of geo-regions according to their similarities.

4.2 Parameters

To decide geo-regions, we need to set the bandwidth for mean-shift clustering and the threshold θ_p for obtaining geo-regions which contain sufficient number of photos. We empirically decided bandwidth $= 0.001$ and $\theta_p = 50$.

To describe a geo-region's feature, parameters N and θ_{vc} need to be set. In the YFCC100M dataset, $N = 1,570$ and since each photo is already annotated with Visual Concepts (and their probabilities), we did not set a specific value for θ_{vc} in this experiment. For the threshold to detect similar geo-regions, we set $\theta_s = 0.7$.

4.3 Result

By applying the proposed method with the above parameters to the YFCC100M dataset, 152,409 geo-regions and 54,834,813 pairs of similar geo-regions out of ${}_{152,409}\mathrm{C}_2 = 11,606,099,193$ pairs of geo-regions were yielded. The frequency of geo-regions according to their similarities is shown in Fig. 2.

4.4 Analysis

In this section, we analyze the pairs of similar geo-regions detected in Section 4.3. For this, we analyzed the results using an interface that visualizes them as shown in Fig. 1. With the interface, we can see the range of each geo-region on a map, Visual Concepts that highly contributed to the similarity, and photos that contained the highly-contributed Visual Concepts. The contribution of Visual Concept vc_k to the similarity is calculated as,

$$C_{\mathrm{vc_k},i,j} = \frac{(V'_i(k) - \overline{\mathbf{V}'_i})(V'_j(k) - \overline{\mathbf{V}'_j})}{\sqrt{\sum_{n=1}^{N}(V'_i(n) - \overline{\mathbf{V}'_i})^2 \cdot \sum_{n=1}^{N}(V'_j(n) - \overline{\mathbf{V}'_j})^2}}. \tag{6}$$

The detected similar geo-regions can be divided into two types; One which people focused on certain objects, and one which they focused on certain activities.

Table 1. Top five contributed Visual Concepts for "Tokyo Tower" and "Eiffel Tower"

Rank	Visual Concept	Contribution
1	tower	0.182
2	architecture	0.104
3	building	0.048
4	steeple	0.038
5	skyline	0.003

Table 2. Top five contributed Visual Concepts for "Bali" and "Coolangatta"

Rank	Visual Concept	Contribution
1	surfing	0.175
2	water skiing	0.142
3	water sport	0.126
4	wave	0.120
5	water ski	0.117

For geo-regions where people focused on objects, we found examples where people focused on "towers", "trains", "airport", "mountains", and so on. Here, we introduce an example of similar geo-regions where people focused on a tower; "Tokyo Tower" in Tokyo, Japan and "Eiffel Tower" in Paris, France with a similarity of 0.719. From the top five contributed Visual Concepts shown in Table 1, we can see that people focused on "tower" in both geo-regions.

In addition, we found a geo-region which contained "Tokyo Skytree" in Tokyo, Japan which was similar to both of the above two geo-regions. In this case, although "Tokyo Skytree" appears visually different from them as a tower, by using Visual Concepts, we could still detect it as a similar geo-region that contains sceneries with a tower.

Meanwhile, for geo-regions where people focused on activities, we found two types; "eating" and "water sporting". As examples of the eating activities, we found pairs from Chinatowns in Yokohama, Japan and Los Angeles, USA as shown in Fig 1. In addition, we found other pairs of Chinatowns as similar geo-regions: Nagasaki and Kobe, in Japan, and San Francisco, New York, and Chicago in USA, and so on. Highly contributed Visual Concepts common in these geo-regions were "food", "shop", "indoor", "text", "meal", and "architecture". We consider that in Chinatown, people focus not only on Chinese food detected as Visual Concepts "food" and "meal", but also on Chinese-style streetscapes detected as "shop" and "architecture". It seems that combinations of these Visual Concepts indicate activities in Chinatown. However, since these Visual Concepts do not directly indicate Chinese food or streetscapes, some geo-regions which are actually not Chinatown were also detected as similar geo-regions. Although it is still fine to detect them as geo-regions with eating activities, if we can detect more detailed Visual Concepts such as oriental architecture, we could detect more specific activities, precisely.

As examples of the water sporting activities, we found pairs from Bali, Indonesia and Coolangatta, Australia with a similarity of 0.905. From the top five contributed Visual Concepts shown in Table 2, we can see that people focused on water sporting. Although various activities can take place on a seashore; sightseeing, whale watching, beach volleyball, water sports, and so on, for travel recommendation, it is important to find those where specific activities take place.

5 Conclusions

We proposed a method for the detection of similar geo-regions based on Visual Concepts in social photos. From experimental results and their analyses, we confirmed the usefulness of the proposed method. Future work includes modelling of similar geo-regions based on common Visual Concepts in them, and also temporal analysis of the similar geo-regions.

6 Acknowledgment

Parts of this work were supported by Grant-in-aid for Scientific Research from MEXT/ JSPS.

References

1. Travel & tourism economic impact 2016 annual update summary, `http://www.wttc.org/-/media/files/reports/economic-impact-research/2016-documents/economic-impact-summary-2016\_a4-web.pdf` [Accessed: May 15, 2017]
2. Ide, I., Wang, J., Noda, M., Takahashi, T., Deguchi, D., Murase, H.: Construction of a local attraction map according to social visual attention. In: Intelligent Interactive Multimedia: Systems and Services, pp. 153–162. Springer (2012)
3. Kurashima, T., Iwata, T., Irie, G., Fujimura, K.: Travel route recommendation using geotags in photo sharing sites. In: Proc. 19th ACM Int. Conf. on Information and Knowledge Management. pp. 579–588 (2010)
4. Min, W., Bao, B.K., Xu, C.: Multimodal spatio-temporal theme modeling for landmark analysis. IEEE Multimedia 21(3), 20–29 (2014)
5. Pongpaichet, S., Tang, M., Jalali, L., Jain, R.: Using photos as micro-reports of events. In: Proc. 2016 ACM Int. Conf. on Multimedia Retrieval. pp. 87–94 (2016)
6. Shi, Y., Serdyukov, P., Hanjalic, A., Larson, M.: Nontrivial landmark recommendation using geotagged photos. ACM Trans. on Intelligent Systems and Technology 4(3), 17–37 (2013)
7. Shimada, K., Uehara, H., Endo, T.: A comparative study of potential-of-interest days on a sightseeing spot recommender. In: Proc. 3rd IIAI Int. Conf. on Advanced Applied Information. pp. 555–560 (2014)
8. Thomee, B., Shamma, D.A., Friedland, G., Elizalde, B., Ni, K., Poland, D., Borth, D., Li, L.J.: YFCC100M: The new data in multimedia research. Comm. ACM 59(2), 64–73 (2016)