Tourist gaze through computer vision: Differences between Asian, North American, and European tourists

Kun Zhang
College of Tourism and Service Management, Nankai University, Tianjin, 300350, China
zxzhangkun@163.com

Zhibin Lin*
Durham University Business School, Mill Hill Lane, Durham, DH1 3LB, UK
Email: zhibin.lin@durham.ac.uk
*corresponding author

Jinyi Zhang
College of Tourism and Service Management, Nankai University, Tianjin, 300350, China
Email: zhangjinyic@163.com

Keywords: Tourist photo; Computer vision; Cross-cultural study; Royal garden; Park; Landscape.

Acknowledgement
This study was supported by grants from National Natural Science Foundation of China (No. 51608278), and the Fundamental Research Funds for the Central Universities (No. 63202076).

Declarations of interest
None.

Author bio
Dr. Kun Zhang, is a lecturer at Nankai University, China. Her research interests are tourism planning, with the focus on tourism destination image and AI technology, tourists' demands and sustainable tourism planning.

Dr Zhibin Lin, is Associate Professor at Durham University, UK. His research interests include digital marketing, transport, travel and tourism management.

Jinyi Zhang, is an undergraduate student of tourism management at Nankai University, China.

Tourist gaze through computer vision: Differences between Asian, North American, and European tourists

Highlights

- What tourists gaze on varies across the three groups.
- North American tourists prefer to gaze on water scenes.
- European tourists have a preference for the scenes of foliage and sky.
- Asian tourists like to present themselves in front of traditional buildings.
- The three groups also show differences in places of interest.

Introduction

Photography has an intimate relationship with tourism (Urry, 1995). The personal photos are a record of the images that the tourist has gazed on. As argued by Urry (1995), the essence of tourism is the visual consumption of places, and the fundamental motivation of tourism is to gaze on those iconic symbols such as landscapes, people, buildings that have been visually represented and publicized in the media (Garrod, 2009). Moreover, taking photographs is a major activity for the tourists and the photographs can serve as proof that the tourist has been there (Jenkins, 2003). Tourists seek out particular views that were considered "photogenic" or "iconic," and to reproduce these in their photographs (Balomenou & Garrod, 2019). They love to post their photos on social media accompanied by texts as a way to construct or re-construct their tourism experiences of the places and tourism activities (Lo & McKercher, 2015).

Prior studies of user-generated photographs mainly focus on the spatial-temporal patterns of tourist behavior (Önder, Koerbitz, & Hubmann-Haidvogel, 2016; Shoval, McKercher, Ng, & Birenboim, 2011; Vu, Li, Law, & Ye, 2015) and generate insights into tourists' perceived images at the macro level, such as a city destination (Zoltan & McKercher, 2015). Few studies have explored the visual elements of tourists' photos at a micro-level, such as a theme park. Cross-cultural studies of tourism have shown that tourists from different national cultural backgrounds differ in their perceptions of the benefits of a trip and travel behaviors (Pizam & Sussmann, 1995; Stepchenkova, Kim, & Kirilenko, 2015). Yet we have little knowledge about whether there are "gaze" differences of a theme park and its landscape between tourists from larger cultural categories such as Asian, North American, and European cultures (Zhang, Chen, & Li, 2019).

This study thus aims to identify the differences in the places of interest and images of a landscape garden between tourists from different regions of the world. The specific research questions are:

RQ1: Are there any differences in places of interest between Asian, North American, and European tourists?

RQ2: Are there any differences in landscape elements in the photos captured by Asian, North American, and European tourists?

Data and method

The tourist site selected for this study is the Summer Palace, a World Heritage site and a former royal garden of the Qing Dynasty in Beijing, China, which covers an area of 290 hectares (2.9 square kilometers).

We utilized the YFCC 100M dataset released by Yahoo. The dataset includes approximately 100 million photos uploaded by Yahoo users. With the help of the geographical information system, we retrieved photos of the Summer Palace based on their latitude and longitude information. We also retrieved users' home information by invoking the API in Flickr. In total, we obtained 2227 photos uploaded by 176 tourists from 42 countries in Asia, North America, and Europe (Figure 1).

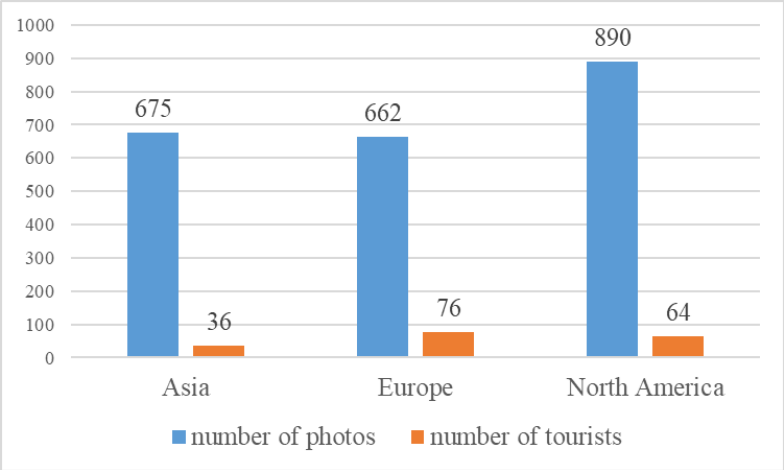

Figure 1. Statistic information about the photos and the tourists

To determine tourists' places of interest, we applied ArcGIS (a geographical information system software) to conduct a spatial analysis of tourists' itineraries. The analysis took two steps. First, we tracked tourists' trajectory by using the tool of Tracking Analyst. Second, we clustered the hotspots of moving trails by using the tool of Line Density Analysis in ArcGIS.

To determine the landscape types in tourist photos, we adopted a deep learning model of semantic segmentation for the landscape element recognition. Semantic segmentation is an important technique in computer vision, the task of which is to classify every pixel in the image to identify the objects in the photo and their location. In this study, the deep learning model's assignment is to calculate the percentage of five primary landscape elements in tourists' photos: water, building, people, sky, and foliage. People in the photos were treated as a specific component of the visual landscape, and those that could not be assorted into one of the five semantic elements were defined as background.

To train the deep learning model, we used an existing dataset of 5000 photos, each of which had a size of 512 x 512 pixels and was labeled with one of the five semantic elements. We randomly selected 4500 (i.e. 90%) of them for the training and the remaining 500 photos for validation. The state-of-the-art system, DeepLabv3 was applied to build the deep learning model structure. Proposed by Chen, Papandreou, Schroff, and Adam (2017), DeepLabv3 uses the atrous/dilated convolution and spatial

pyramid pooling strategy for the higher field of vision. Comparing with ordinary convolution, it preserves spatial resolution and makes a deeper network by capturing features at each scale and thus has more robust performance (Chen et al., 2017). Figure 2 shows three examples of the outputs.

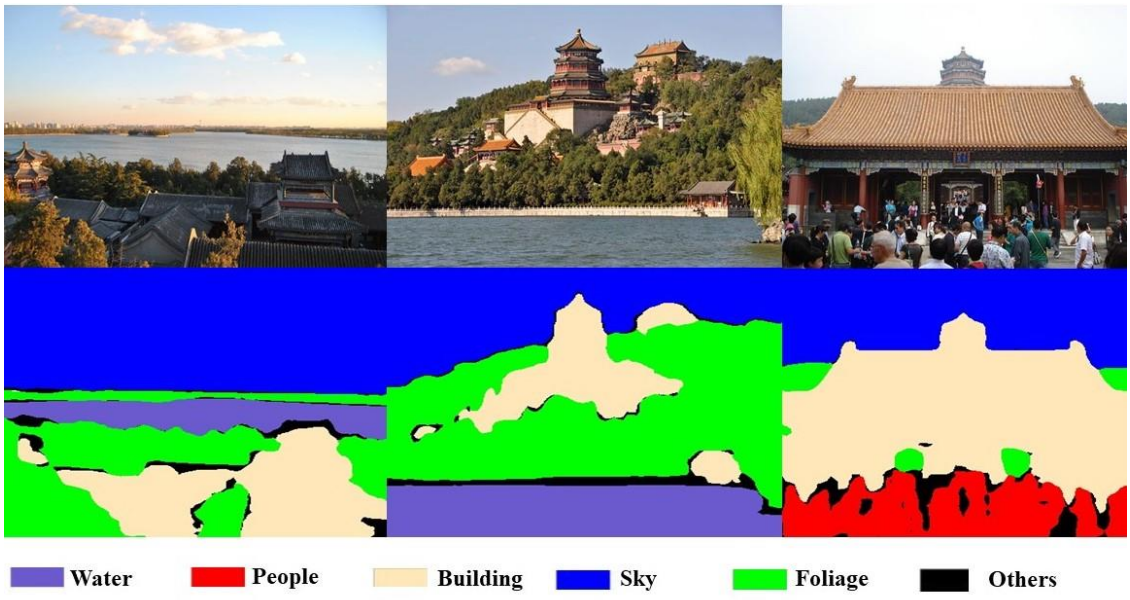

Figure 2. Examples of the outputs of the deep learning model of the semantic segmentation

Based on the outputs of the deep learning model, two further analytical steps were taken. First, all the statistical data generated by the model was exported into an Excel file, and the tool of bot plot was used for comparing the statistical difference of landscape elements. Second, the statistical data generated by the model was imported into ArcGIS and matched with the geographical coordinates.

Findings

The differences in places of interest

The differences in places most visited between the three groups were visualized in Figure 3. North American tourists had the most extensive area of footprint, while Asian tourists had the smallest. For the most visited places, the tourists from North America showed an interest in the eastern part of the Long Corridor and the Buddhist Pavilion, and it is apparent that most of them used the water routes in the northeast of the lake. The hotspots for European tourists are the Buddhist Fragrance Tower, the western part of the Long Corridor, and the area around the Marble Boat. In contrast, Asian tourists were most enthusiastic about the eastern part of the Long Corridor, the Emperor's Administration Area, and the Seventeen Arch Bridge.

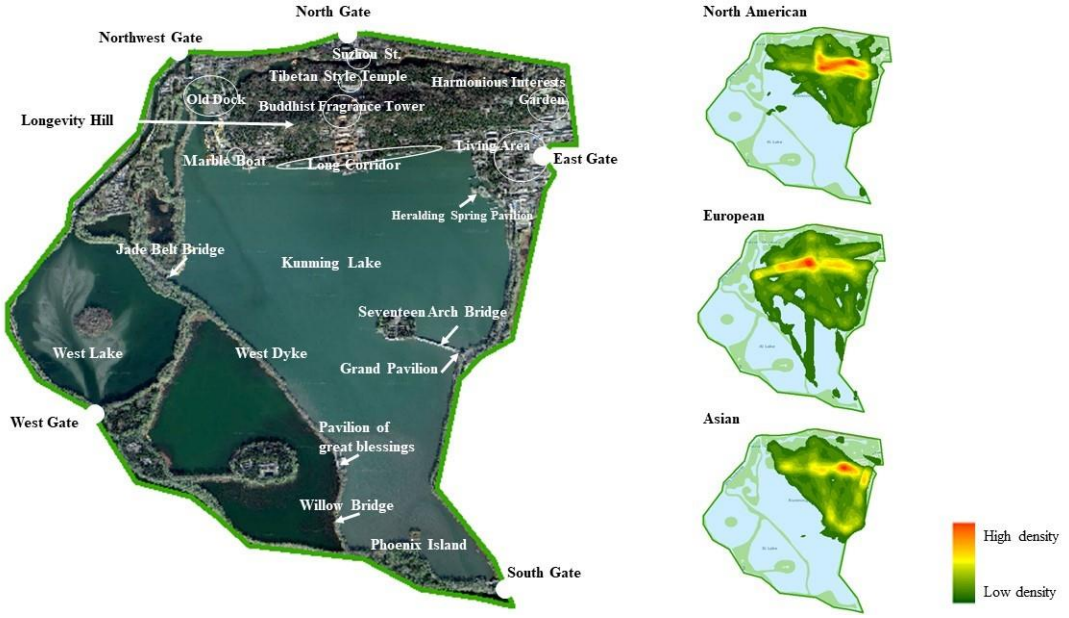

Figure 3. The spatial distribution of the tourists' hotspots in the summer palace

The differences in landscape elements

The differences in landscape elements are shown in Figure 4. North American tourists showed a preference for the water scenes in the Summer Palace. European tourists seemed to have a love with the sky and foliage, while Asian tourists seemed to enjoy taking pictures of themselves, families, or friends in front of the traditional buildings.

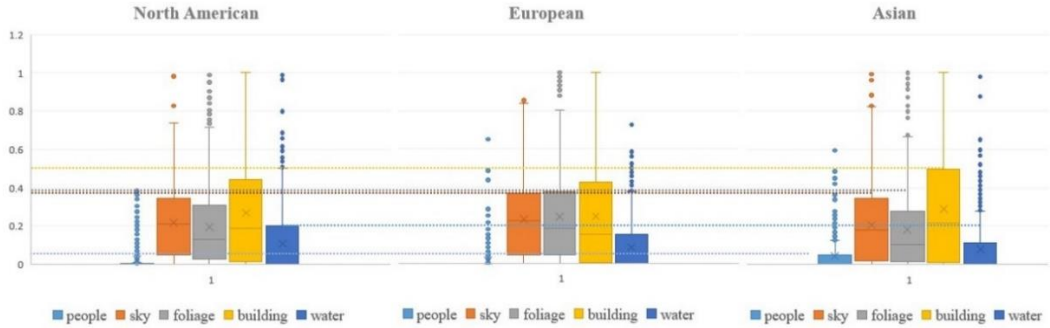

Figure 4. The statistical differences in landscape elements

Within the five landscape elements, the three groups showed several differences with some similarities.

Water. North American tourists presented pictures of the waters near the Old Dock; European tourists preferred the water landscape in the Marble Boat area; Asian tourists' photos showed the water in the area of Suzhou Cultural Street.

Building. Both North American and Asian tourists focused on the iconic building-Buddhist Fragrance Tower and the Living Areas; while European tourists were not only interested in iconic buildings, but also interested in other cultural buildings in the rear area of the Longevity Hill.

People. The element of people in photos is generally similar across the three groups; however, compared with the photos by North American tourists, people in the

photos by European tourists appeared more frequently in the area of the Marble Boat, and people in the photos by Asian tourists are mostly seen on the Seventeen-Arch Bridge with a full view of the iconic buildings.

Foliage. European tourists had the widest range of plants in their photos, and the proportion of the plants in the photos was relatively large; North American and European tourists had certain similarities, but North American had fewer plant element photos in the Marble Boat area; Asian tourists had the smallest proportion of plants in their photos.

Sky. Photos containing sky elements by North American tourists appeared mostly in the Longevity Hill; European tourists had many photos containing sky elements in the Hydrophilic Area. Compared with their Western counterparts, Asian tourists covered a smaller range of the areas, mostly in the Seventeen Arch Bridge.

Discussion and conclusions

The results of this study show that there are differences in what tourists gaze on in the Summer Palace between tourists from the three continents. First, we found that North American tourists showed the most extensive area of footprint, while Asian tourists the least. This finding reflects the individualism and novelty-seeking trait of the American culture (Pizam & Sussmann, 1995), which leads to the exploration of places of interests; in contrast, influenced by their collectivism and uncertainty avoidance culture, the Asian tourists love traveling in groups and shopping of souvenirs (Pizam & Sussmann, 1995), which could lead to limited areas visited; and European cultures are somewhere between the American and Asian cultures.

Second, we found that North American tourists like to gaze on the water scenes, while European tourists on scenes of foliage and sky and Asian tourists on traditional buildings. This reflects the cultural similarity between Americans and Europeans: photos by both American and European groups are largely natural scenes, while Asian tourists' photos are mainly built environments. The empirical evidence of this study thus corroborates the argument suggested by Lo and McKercher (2015) that tourists are selective in the images to capture. These images are subjective, varying from one tourist to another, reflecting their cultural conventions (Stylianou-Lambert, 2012).

The study is limited to the dataset used, despite that, the application of the semantic segmentation model in this study offers a new possibility for advancing tourism research. The deep learning model used in this study helps to recognize the five landscape elements, an important step towards interpreting how tourist gaze is visually presented and represented.

References

Balomenou, N., & Garrod, B. (2019). Photographs in tourism research: Prejudice, power, performance and participant-generated images. *Tourism Management, 70*, 201-217.

Chen, L.-C., Papandreou, G., Schroff, F., et al. (2017). Rethinking atrous convolution

for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.
Garrod, B. (2009). Understanding the relationship between tourism destination imagery and tourist photography. *Journal of Travel Research, 47*(3), 346-358.
Jenkins, O. (2003). Photography and travel brochures: The circle of representation. *Tourism Geographies, 5*(3), 305-328.
Lo, I. S., & McKercher, B. (2015). Ideal image in process: Online tourist photography and impression management. *Annals of Tourism Research, 52*, 104-116.
Önder, I., Koerbitz, W., & Hubmann-Haidvogel, A. (2016). Tracing tourists by their digital footprints: The case of Austria. *Journal of Travel Research, 55*(5), 566-573.
Pizam, A., & Sussmann, S. (1995). Does nationality affect tourist behavior? *Annals of Tourism Research, 22*(4), 901-917.
Shoval, N., McKercher, B., Ng, E., et al. (2011). Hotel location and tourist activity in cities. *Annals of Tourism Research, 38*(4), 1594-1612.
Stepchenkova, S., Kim, H., & Kirilenko, A. (2015). Cultural differences in pictorial destination images: Russia through the camera lenses of American and Korean tourists. *Journal of Travel Research, 54*(6), 758-773.
Stylianou-Lambert, T. (2012). Tourists with cameras: Reproducing or Producing? *Annals of Tourism Research, 39*(4), 1817-1838.
Urry, J. (1995). *The tourist gaze*. Thousand Oaks, CA: Sage.
Vu, H. Q., Li, G., Law, R., et al. (2015). Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos. *Tourism Management, 46*, 222-232.
Zhang, K., Chen, Y., & Li, C. (2019). Discovering the tourists' behaviors and perceptions in a tourism destination by analyzing photos' visual content with a computer deep learning model: The case of Beijing. *Tourism Management, 75*, 595-608.
Zoltan, J., & McKercher, B. (2015). Analysing intra-destination movements and activity participation of tourists through destination card consumption. *Tourism Geographies, 17*(1), 19-35.