



Culture vs. distance: comparing the effects of geographic segmentation variables on tourists' destination images based on social media data

Mengtong Wu^{1,2}, Chao Jiang^{1,2,3*} , Yi Zhang^{1,2}, Jingjing Cao⁴, Ying Cheng³ and Yu Liu^{1,2,3}

Abstract

Culture and distance are two major factors for geographically segmenting tourists in tourism marketing and advertising. Previous empirical studies on the destination image, however, have examined extensively the effect of the culture while inadequately the effect of the distance, let alone comparing the effects of the two variables. Using social media data, this study compares the effect of distance-based segments of tourists with that of culture-based segments in producing diverse perceived images of a destination. From Sina Weibo data, 282,532 Chinese mainland tourists who visited Suzhou, China during 2012–2016 and their perceived destination images are extracted and analyzed. The main results include: 1) for distance-based segments, the image differences increased with distance and the short-haul tourists perceived a more comprehensive image than the long-haul tourists; 2) for culture-based segments, the image differences were clear and relatively complex, while tourists from Wuyue cultural region had similar image perceptions with the local visitors; 3) **the q-statistic of the Geodetector method shows that the culture-based segmentation can explain 65.8% of image variations while the distance-based segmentation can explain 46.6% of image variations, suggesting that culture is a more appropriate variable to segment the tourism market.**

Keywords: Tourism destination image, Distance, Culture, Geographic segmentation, Big data, Weibo

1 Introduction

The destination image has been an important topic in tourism research over the past 40 years and is found to have a great impact on tourists' decision-making and behaviors (Crompton, 1979; Mackay & Fesenmaier, 1997; Pike, 2002). Given a tourism destination, the physical environment and intangible atmosphere can be perceived as a generalized mental picture, which is the synthesis of one's beliefs, ideas, and

impressions of the destination (Crompton, 1979). Destination image can influence a tourist's destination choice, length of stay, consumption level, travel satisfaction, intention to revisit, and willingness to recommend (Tasci, Gartner, & Tamer Cavusgil, 2007). Hence, maintaining an interesting destination image is critical for attracting tourists and promoting tourism.

As different tourists may have different image perceptions toward a destination, one common way for tourism marketing is to group tourists based on where they live, i.e., geographic market segmentation (Kaplanidou, 2009; Obenour, Lengfelder, & Groves, 2005). It provides an effective approach for tourism managers and related practitioners to satisfy the demands of specific groups of tourists. For example, geographic segmentation is widely

* Correspondence: jiangchao2018@pku.edu.cn

¹Institute of Remote Sensing and Geographical Information Systems, School of Earth and Space Sciences, Peking University, Beijing, China

²Beijing Key Laboratory of Spatial Information Integration and Its Applications, Peking University, Beijing, China

Full list of author information is available at the end of the article

used in tourism advertising for devising area-specific media promotional campaigns (Moscardo, Pearce, & Morrison, 2001).

Many researchers have found that the destination image varies among cultural regions. While the definition of culture remains ambiguous, the geographical region is often used as a cultural indicator (Kaplanidou, 2009; Moscardo et al., 2001). For example, Lee and Lee (2009) compared the images of Guam perceived by Korean and Japanese visitors; McCartney (2008) examined the images of Macao for Chinese tourists from Mainland, Hong Kong, and Taiwan. These researches discovered the image diversification between different cultural groups. Generally, previous research findings support that the cultural region is an important factor in forming the diverse images of a destination.

Meanwhile, travel distance is also generally regarded to be important in producing diverse images, despite the inconsistent effects among the limited empirical studies. When constructing a framework for tourism destination image (TDI) studies, Gallarza, Saura, and García, H. C. (2002) listed the influence of travel distance as one of the seven main topics. Besides, according to the first law of geography, “everything is related to everything else, but near things are more related than distant things” (Tobler, 1970). In terms of tourism, it implies that tourists from near regions would perceive a more similar image with that of local tourists than those from far regions. In empirical studies, Crompton (1979) discovered that tourists perceived a more positive image toward Mexico as the distance of origin increases, while Leisen (2001) drew an opposite conclusion that tourists from the nearby region had a more favorable image toward New Mexico than those from far regions. These controversial findings suggest that more empirical work is needed to verify the influence of travel distance on the image perceptions of a destination.

Given that both culture and distance are important factors for image perceptions, an open question is which factor can better explain the variety of diverse destination images among geo-segmented tourists. In practice, an equivalent question is which variable should be used to geographically segment the tourists to yield a better benefit in tourism marketing. According to Morrison (1996), the top criteria for effective market segmentation is to make each group homogeneous. Therefore, to yield a better benefit from taking targeted marketing measures to different tourist groups, the principle of tourist segmentation is to make the image difference within a tourist group as small as possible and the image difference between groups as large as possible. For tourism organizations, managers, and authorities, selecting an appropriate variable for geographic segmentation can make their

marketing efforts more effective while otherwise would cause huge waste as the tourism marketing resources are devoted to the tourists who do not have much interest in it.

To answer the above question, social media services in the big data era provide an ideal data source that has many advantages over traditional survey data. Most tourism research used the survey data, suffering issues like high cost and small sample size. Social media offers a platform where users can share their tourism experiences and post their feelings in the form of texts, pictures, tags, and emojis. Each person can be regarded as a “sensor” that can feel the culture, history, and many other aspects (Liu, 2016; Liu et al., 2015), and a user’s city of residence can be obtained from the user profile and historical posts. This not only makes extracting diverse destination images workable but also provides a cheap way to acquire a large number of samples that would better represent tourists’ characteristics in multiple geographic segments. For example, Arefieva, Egger, and Yu (2021) applied machine learning approaches to extract the destination image of Austria from photos posted on Instagram. Peng, Bao, and Huang (2020) abstracted topics regarding different geographical hotspots in the city across different groups of individuals and found that text mining on geotagged big data makes it possible to understand the perceptions and preferences of different groups of people, thereby to understand the social, cultural and functional characteristics of a city.

Under the above background, this research aims to compare the effect of distance-based segments of tourists with that of culture-based segments in producing diverse perceived images of a destination by analyzing the social media data. Specifically, this research contributes to the literature from three aspects: 1) A methodological framework for extracting the destination image from the social media data is presented; 2) The diverse images of Suzhou, China among distance-based segments of tourists are examined to enrich the empirical work on the relationship between travel distance and destination image; 3) The effects of the culture- and distance-based segments are compared using several quantitative indexes and a Geodetector method.

2 Literature review

2.1 Impact of distance on the destination image

Distance has been identified as a factor influencing the destination image in tourism activities. A Tourist often spends considerable efforts (physical strength, time, money) to overcome the distance barrier between a destination and his/her place of origin, and the distance affects a tourist’s understanding and evaluation of the destination (Mckercher, 2008). Also, the distance decay

effect (Tobler, 1970) might be able to explain the phenomenon to some extent.

Existing studies tend to choose several specific regions of origin and have found inconsistent conclusions. For example, Crompton (1979) selected a sample of 617 students from 12 universities in different parts of the US and discovered that the farther the distance between one's residence and the destination country, the better the destination image; while a study by Leisen (2001) showed that people from nearby states had a more favorable image of New Mexico than those from distant states.

Based on a common division of the distance, tourists can also be divided as local and non-local. The nonlocal group and local group are important market segments of a destination. Researchers have paid much attention to the image differences between these two groups. In the case of Simla, a northern Indian city, Jutla (2000) found that natural and cultural landscapes are more important factors for nonlocal tourists, while familiarity with the city is a more important factor for residents. Bonn, Joseph, and Dai (2005) compared local visitors' and foreign tourists' image perception of Florida and discovered that the two groups rated the destination environment in a significantly different way. Many other empirical studies supported the conclusion that residents and nonlocal tourists have different perceptions toward a destination (Ji & Wall, 2011; Lee, Chang, Hou, & Lin, 2008).

Overall, the impact of distance on the destination image has long been a concern, but there is still a lack of empirical studies in this field, and previous studies produced conflicting conclusions. Besides, most city-scale studies chose several cities of origin as the research area. The emerging of big data makes it easier to study a continuous area covering the whole geographical region, such as a country.

2.2 Impact of culture on the destination image

People's perceptions of the destination image are subjective and are related to their cultural background, which can affect their attitude and behavior (Hofstede, 2001; Kaplanidou, 2009). For a certain destination, various images can be observed among tourists with different cultural backgrounds.

In studies of comparing destination images among different cultural groups, cross-cultural research occupies a great proportion. Cross-cultural studies are mainly based on Hofstede's national cultural model (Hofstede & Hofstede, 1984), using countries of origin as cultural indicators.

Many researchers have examined cross-cultural image differences using countries of origin as cultural indicators. Chen and Kerstetter (1999) examined the image

differences among international students for rural travel in Pennsylvania and concluded that the destination image differs for tourists from different countries of origin, and the differences lie in their evaluation in different dimensions. Frías, Rodríguez, Castañeda, Sabiote, and Buhalis (2012) examined the perceptions toward Spain among tourists from several European countries and found that cultural background matters in producing different image perceptions. Ahmed (1991) and Beerli and Martin (2004) found effective relationships between perceived images and visitors' country of origin. Previous research also found that people from different cultural backgrounds focus on different themes. In a case study in Asia, Henkel, Henkel, Agrusa, Agrusa, and Tanner (2006) compared image perceptions of domestic and foreign tourists and found significant differences in the elements of Thailand's destination image, with foreign tourists paying more attention to Thailand's nightlife and recreational activities. Deng, Liu, Dai, and Li (2019) retrieved and compared the destination images of Shanghai for both the Eastern and Western inbound tourists based on a set of photo metadata from Flickr.

Intra-cultural image differences for tourists from different cultural groups within a country have also long been studied. Kwek and Lee (2008) identified the image differences between Mainland Chinese and Singaporean Chinese and found out that intra-cultural variance has a significant implication among Chinese tourist markets. McCartney (2008) examined Macau's image perceptions by four domestic cultural groups in Beijing, Shanghai, Hong Kong, and Taiwan, found that their dominant factors were not the same. By analyzing 12 tourism industry professionals and 85 Chinese visitors, Lam and Ryan (2020) found that they had different views on Macau. In image marketing, destination marketing agencies also need to develop marketing strategies based on differences in the preferences of cultural groups in different regions.

Using states as research units is insufficient to measure the similarity and difference of different cultural groups within a country, and it's necessary to compare intra-cultural groups' image perceptions. Hofstede emphasized that the concept of culture refers more to society than to the country of origin, and nationalities should not be considered equal to society (Beerli & Martin, 2004). It should be careful to use nationality as a variable in the study of cultural differences. The basic unit of cross-cultural research of national cultural groups can be further subdivided into different cultural groups within the country, such as Chinese tourists can be subdivided into different groups by generation, geographical region, dialects, etc. Nevertheless, the empirical study is still not sufficient for intra-cultural image comparison at a spatially fine-grain scale.

2.3 Impacts of both factors on the destination image

Apart from the above studies, scholars have considered both distance and culture's impact on destination image. In empirical researches, Hallab and Kim (2011) investigated the influences of geographical and cultural distance on the image of tourist destinations by studying 234 tourists from different parts of the United States. They found that environmental factors were less attractive to tourists living in the western Mississippi River, while southern tourists had a strong perception of the recreational factors in Mississippi, and generally southern tourists were less willing to travel to Mississippi than the western tourists. Tang, Choi, Morrison, and Lehto (2009) compared Macau's tourism image on Chinese and English websites from five online sources. She found that the overall images of Macau projected on Chinese and English websites were different and concluded that cultural differences, when combined with travel distance, have an impact on the image of the destination.

Despite scholars have noticed the impact of distance and culture on destination image, which factor is more important is still an intangible question. This study attempts to compare the two factors' influences with quantitative methods.

3 Methods

3.1 Analytical framework

A framework is developed to extract and analyze the diverse images of a destination from the social media data (Fig. 1). First, textual data was collected from Sina Weibo and Dianping. To distinguish different groups, cities of origin were divided into regions according to culture and distance. Through theme extraction and theme classification, we acquired the cognitive image of each group, and then several analytical methods were used to compare the differences of each group and the influence of the two factors on the image.

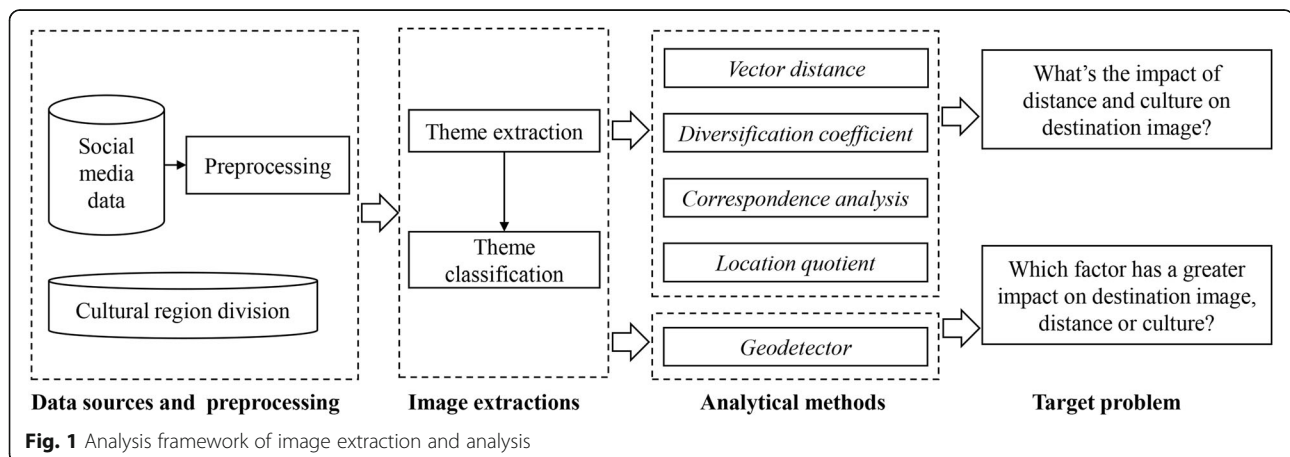
3.2 Data collection and preprocessing

The destination city in this research is Suzhou, a typical tourist city in China. Suzhou attracted more than 136 million visitors in 2019, making tourism an important part of its economy. With a history of more than 2500 years, it has many physical heritages such as ancient urban cities and towns. It also has a lot of natural landscapes and intangible cultural heritages such as Pingtan Opera, Kun Opera, and Suzhou embroidery. The main reason for choosing Suzhou is two-fold: 1) the relationship between tourists' culture/distance and cognitive images would be more prominent in such a typical tourist city than the other cities. 2) Suzhou is often regarded as a model for many cities to develop tourism, thus the research findings from Suzhou would have many implications for other cities.

Sina Weibo, an analogue to Twitter, is one of the most popular social media in China (Kim, Lee, Shin, & Yang, 2017;). With the Weibo Application Program Interface (API), the lists for all Points of Interest (POIs) in Suzhou and related users who once posted at these POIs were obtained. Then, all microblogs posted by these users were retrieved, resulting in a dataset of 5,399,161 microblogs from 470,041 Mainland Chinese users from Dec 2012 to Nov 2016. Thereafter, the city of origin for each user was identified based on Shannon's information entropy formula. For each user, we calculated the entropy value of each city k according to the following formula:

$$E_k = \sum_p -p \ln(p) = \sum_{m=1}^{12} -\frac{n_{k,m}}{N_k} \ln\left(\frac{n_{k,m}}{N_k}\right)$$

where N_k is the total amount of microblogs posted by the user in the city k ; $n_{k,m}$ is the number of microblogs that the user publishes in month m . The city corresponding to the maximum entropy value is selected as the user's city of origin. Users that posted too few blogs



(such as less than two in each city) were removed because there will be no valid result in this step.

In this research, a tourist was defined as a user who visited a tourism-related site (Xue & Zhang, 2020). All microblogs that were posted in Suzhou by the identified tourists were retained. As a result, 1,611,269 microblogs from 282,532 users were identified as tourism-related posts.

Considering that Weibo allows a user to post various topics with at most only 140 words, it seems difficult to obtain a list of themes that are all about tourism from such short text. Dianping, akin to yelp, is a platform for users to share comments about local restaurants, businesses, and services (Xue & Zhang, 2020). It has a “tour around” module, which contains tourism-related POIs that can provide tourism services, such as attractions, hotels, restaurants, etc. In Dianping’s “tour around” module, users are expected to share their own travel experience, and their comments are mostly tourism-related; By using Weibo, users can share their lives, and their posts are diverse in themes, including travel experience, personal feeling, residents’ daily activities, etc. (Zhang, Yang, Zhang, & Zhang, 2020). Therefore, the themes extracted by Dianping should be more concentrated on tourism. We obtained the tourism-related POIs and comments in Suzhou, and a total number of 119,807 comments on 1842 POIs were retrieved at last.

In the data cleaning process, we removed duplicate data, converted traditional characters into simplified characters, and deleted specific formats in Weibo blogs such as check-in tags, URLs, at (“@”) other users. An example of data cleaning is shown in Table 1.

3.3 Image extraction

This section describes how to extract the destination image from the social media data. It involves two stages:

- 1) Extracting themes from the Dianping comments; 2) Classifying the Weibo blogs into the identified themes.

3.3.1 Themes extraction

For the Dianping data, the *pkuseg* toolkit, a Chinese tokenization toolkit that includes a tourism-domain model (Luo, Xu, Zhang, Ren, & Sun, 2019), was used to segment sentences into words. Two thousand eight hundred seventy-eight words were extracted, of which nouns were used to represent the destination image (Deng et al., 2019). Tencent Artificial Intelligence Lab provides large-scale, high-quality Chinese word vector data, with each word mapped to a vector. However, as not all words are equally important, meaningless words like “we” “here” needed to be filtered out. According to Dianping, POIs are classified into different categories. The comments in each category are merged into one document. We calculated the term frequency (TF), which is the percentage of times a given word appears in a document, and the inverse document frequency (IDF), which is the reciprocal of the ratio for the number of documents containing a given term in all documents. By ranking the maximum TF-IDF value of each word, the words with the lowest 40% value were removed. At this stage, 1694 words from the Dianping comments were retained. By clustering these word vectors, a series of themes can be obtained. Each theme consists of a group of words. According to the hierarchical clustering algorithm, all the word vectors were clustered from bottom to top. Based on the elbow method we obtained 72 clusters. By manual evaluation, 31 categories with clear topics were reserved and were regrouped into 9 themes, where similar categories are merged into one theme.

3.3.2 Themes classification

For the Weibo data, the theme for each microblog was identified with the following steps. The first step was

Table 1 Examples of preprocessing the social media data

Data source	Original text	Text after preprocessing
Weibo	(Chinese) 我在这里#陈记农家乐# 吃吃农家乐, 哈哈~ 开车一早上, 终于到了。[哈哈]@王小凡er @好好生活_天天想瘦的小青年 http://t.cn/zOo1JMX	吃吃农家乐, 哈哈~ 开车一早上, 终于到了。
	(English) I'm here: #Chen Ji Farmhouse# eat at the farmhouse, haha - drive all morning, finally arrived. [Haha] @Wang Xiaofaner @Good life_ want to be thin every day http://t.cn/zOo1JMX	eat at the farmhouse, haha - drive all morning, finally arrived.
Dianping	(Chinese) 裏面不大但是空氣風景還不錯, 湖里全是荷花不過沒開應該是季節已經過了。裏面有座橋然後在湖中有個私人會所, 中午的時候就去吃了感覺不錯, 古色古香的在盛澤地區應該算比較好的地方了	里面不大但是空气风景还不错, 湖里全是荷花不过没开应该是季节已经过了。里面有座桥然后在湖中有个私人会所, 中午的时候就去了感觉不错, 古色古香的在盛泽地区应该算比较好的地方了
	(English) The inside is not big but the air scenery is good. The lake is full of lotus but not blooming. It should be the season has passed. There is a bridge inside and there is a private club in the lake. I went to eat at noon and felt good. The antique place is good in Shengze area.	Same meaning as the original text, but converted traditional characters to simplified characters.

word segmentation and word embedding (Same as the methods for processing the Dianping data). Words that were not in the word set of the above section were filtered out. The second step was to categorize each microblog based on two rules: 1) if a microblog contains any word of theme A, then A is marked as one of its' candidate themes; 2) if a microblog has more than one candidate theme, then calculate the mean center of the word vectors it contains, and compare the mean center with that of each candidate theme; The nearest theme to the average center of its word vectors is regarded as this microblog's theme.

3.4 Geographic segmentation

As introduced in the above sections, both culture and distance can be used as the variable for geographic segmentation.

For the distance variable, the distance between the government locations of each city and Suzhou was calculated based on their latitudes and longitudes. As the city of origin for each Weibo user had been obtained, a scatter plot can be drawn with distance as the X-axis and the accumulated number of microblogs as the Y-axis. Each point in the plot represents a city. The distance values for the scatter points with abrupt Y-value changes were used as segmentation thresholds.

For the culture variable, this research adopted a generally accepted Chinese cultural regionalization. According to Wu (1996), a cultural region is a geographical area with a continuous spatial range, relatively consistent natural environment characteristics, approximate historical process, similar national tradition and population action process, and a certain common cultural landscape. In this research, the 31 provincial administrative regions in China's mainland were classified into eight cultural regions. Hong Kong, Macao, and Taiwan were not included in the study area due to a lack of data.

3.5 Measuring the differences among diverse images

Three quantitative indexes were used to measure the image differences for different segments of tourists. Each image of Suzhou can be described as a vector:

$$V_j = (v_{1j}, v_{2j}, v_{3j}, \dots)$$

$$v_{ij} = \frac{N_{ij}}{N_j}$$

where, $i = 1, 2, \dots, n$ denotes the theme; $j = 1, 2, \dots, m$ denotes the geographic segment; N_{ij} is the number of Weibo blogs that are classified with theme i and posted by tourists from segment j . The Vector Distance (VD) was used to assess the difference between the compositions of every two images. The smaller the distance, the more similar the two images:

$$VD(V_1, V_2) = \sqrt{\sum_{i=1}^n (v_{i1} - v_{i2})^2}$$

The Diversification Coefficient (DC) measures the diversification of themes for the images perceived by the tourists in a geographic segment (Marine-Roig & Clavé, 2016). The DC value varies from 0 to 1, where 0 means the proportions of different themes in the image are equal and 1 means the image focuses on only one theme:

$$DC_j = 1 - \frac{(\sum_{i=1}^n N_{ij})^2}{n \sum_{i=1}^n N_{ij}^2}$$

The Location Quotient (LQ) was used to measure the uneven distribution of the image themes for tourists in different geographic segments. It compares the proportion of microblogs that belong to an image theme for tourists in a specific segment with the proportion of microblogs that belong to the same theme for tourists in the whole study region. If the LQ_{ij} value is greater than 1, it indicates that tourists from segment j focus more on theme i than the average level. The formula is as follow:

$$LQ_{ij} = \frac{\frac{N_{ij}}{N_j}}{\frac{N_i}{N}}$$

where N is the number of total Weibo blogs; N_i is the number of Weibo blogs that belong to theme i .

The Correspondence Analysis (CA) was used to examine the similarities and differences among images of different groups. The CA can describe the relationship between two variables in a low-dimensional space as well as between the categories for each variable. For each variable, the distance between category points in a plot shows the relationship between categories, where similar categories are plotted close to each other. The relationship between variables can be described by projecting points for one variable on the vector from the origin to a category point for the other variable. In the field of tourism marketing, it is used to identify the category similarities and differences among the destination images perceived by different groups (Sun, Liang, & Chang, 2020) and to detect relatively homogeneous tourist groups (Hoffman & Franke, 1986). For each group, the number of microblogs on the nine image themes was used in CA.

3.6 Quantifying the effects of distance- and culture-based segments

The *Geodetector* was used to quantify the effects of distance- and culture-based segments. The factor detector in *Geodetector* assesses the extent to which factor *X* explains the spatial difference in attribute *Y* and has been applied in many fields of natural and social sciences (Wang & Xu, 2017; Wang, Zhang, & Fu, 2016). As shown in the following expression, the factor detector is measured by *q*:

$$q = 1 - \frac{\sum_{h=1}^L N_h \sigma_h^2}{N \sigma^2} = 1 - \frac{SSW}{SST}$$

$$SSW = \sum_{h=1}^L N_h \sigma_h^2, SST = N \sigma^2$$

where $h = 1, \dots, L$ represents stratum of *Y* or factor *X*; N_h and N represent the number of units in stratum h and the entire region; σ_h^2 and σ^2 represents the variance of *Y* in stratum h and the entire region; *SSW* is the within-group sum of squares; *SST* is the total sum of squares. The value of *q* varies from 0 to 1. It means factor *X* can explain $100 \times q\%$ of *Y*. We adopted the factor detector to estimate and compare the extent to which two factors affected the image.

In this study, the smallest research unit is a prefecture-level city; factor *X* is a categorical culture or distance variable; *Y* is the image difference $VD(V_p, V_{local})$ compared with local visitors, and h is the number of geographic segments divided by culture or distance. N_h and N represent the number of prefecture-level cities in segment h and globally. The *q* value represents the degree to which culture or distance explains the variation of the image difference.

4 Results

4.1 Themes extraction

We extracted 2878 distinct words from 119,807 Dianping comments, and 31 categories were obtained by clustering the word vectors. Then these categories were manually regrouped into 9 themes based on existing literature (Beerli & Martin, 2004; Sun et al., 2020): natural landscape, cultural landscape, folk customs & religions, arts, history, festival & events, leisure & recreation, food, and transports, and were furthermore classified into three cognitive dimensions: landscapes, culture, and life, as shown in Table 2.

To understand each tourist's perception of Suzhou, we identified the theme of each microblog based on the method described above and extracted 111,927 Weibo blogs of the upper themes from the total 1,611,269 blogs at last. We used the proportion of Weibo posts on each theme to describe the tourism image of Suzhou (Sun et al., 2020), as shown in Table 3. Cultural landscapes (37.46%), food (24.71%) and leisure & recreation (17.21%) received the most attention from tourists, while festival & events (0.63%), transports (1.16%) and folk customs & religions (3.08%) were not so impressive.

4.2 Impact of distance and culture on Suzhou's destination image

Mainland China consists of 31 provincial administrative districts, including 293 prefecture-level cities by Dec 31, 2018. For distance-based segments, all the cities were divided into 6 groups according to the distance between its city center and the center of Suzhou (Fig. 2). Based on the scatter plot described in section 3.4, we chose Shanghai, Nanjing, Wuhan, Beijing, Guangzhou as split points. In general, the slope decreases with distance, reflecting tourists from nearby region posts more Weibo

Table 2 Tourism theme detection

Dimension	Theme	Words sample
Landscapes	Natural landscape	Ivy, plant, bonsai, azalea, tulip, petal, bird, wings, animal, bee, giraffe, stream, raindrop, hill, hillside, tree shadow, peak, et al.
	Cultural landscape	Dwellings, old street, flagstone, plaque, museum, teahouse, pavilion, porch, attic, courtyard, palace, ancient tomb, et al.
Culture	Folk customs & religions	Once upon a time, pilgrim, dragon robe, dragon boat, pagoda, monk, Buddha, Buddha statue, the god of wealth, et al.
	Arts	Silk, jade, jade sculpture, snuff bottle, Pingtan, Kunqu, melody, lyrics, embroidery, handicraft, wash painting, poet, et al.
	History culture	History, era, origins, ancestor, Guangxu, unofficial history, heritage, landlord, queen, emperor, royal, royal court, et al.
Life	Festival & events	Chrysanthemum, beer festival, carnival, arts festival, gala, flower festival, plum festival, et al.
	Leisure & recreation	Fireworks, fountain, ferris wheel, balloon, yacht, zoo, aquarium, jinricksha, poker, agricola, hot spring, racing car, club, et al.
	Food	Tsindeu, small steamed bun, chicken steak, smelly tofu, rice wine, potherb, vegetable patch, taste, crab, goose egg, et al.
	Transports	Passenger, bus stop, ticket office, train, bus, train station, et al.

Table 3 Distribution of Weibo blogs of each theme

Theme	Number of Weibo blogs	Proportion (%)
Natural landscape	5685	5.08
Cultural landscape	41,929	37.46
Folk customs & religions	3447	3.08
Arts	5615	5.02
History culture	6326	5.65
Festival & events	709	0.63
Leisure & recreation	19,257	17.21
Food	27,661	24.71
Transports	1298	1.16

blogs and vice versa. For culture-based segments, China's mainland can be divided into several cultural regions based on the regional division theory of Wu (1996), and Suzhou is geographically part of the Wuyue cultural region. Considering that there were too few microblogs in some cultural areas, the image obtained in those areas may not be representative.

To assess the distance's and culture's impact on Suzhou's tourism image, we calculated the LQ, DC, and VD, as shown in Tables 4 and 5. The VD here is the vector distance between the image perception of each group and local visitors.

The VD index shows that perceptions toward Suzhou are not homogenous among regions. For distance-based segments, the difference between image perceptions of nearby tourists and residents are

small (0.148), and tourists from 80 to 252 km (0.195), 252–777 km (0.286), 777–1165 km (0.351), and > 1432 km (0.360) have an increasing tendency in the extent of image differences, with an exception in 1165–1432 km (0.399). Figure 3 illustrates the spatial distribution of the VD index. In general, the extents of differences increase with travel distance. People from the closest area perceived the most similar image to local visitors, while farther tourists' recognition is more different. For culture-based segments, the difference between image perceptions of Wuyue tourists and residents is small ($VD = 0.149$), and Southeast tourists perceive the most different Suzhou compared to local visitors ($VD = 0.411$). There is no clear trend between distance and image difference among all the regions, but visitors from regions in the South (Jingchu: $VD = 0.337$, Bashu: $VD = 0.346$, Southeast: $VD = 0.411$) tend to have a more different perception than North regions (Huanghuaihai: $VD = 0.330$, Huangtu: $VD = 0.332$, Yungui: $VD = 0.362$, Guandong: $VD = 0.350$). Specifically, Southeast tourists perceive a more different Suzhou than Huanghuaihai and Guandong tourists, even if Guandong is farther away from Suzhou.

In distance-based segmentation, the DC increases generally with distance, being consistent with the VD. Local visitors have the most diverse cognition of Suzhou (0.510), as the quantities of Weibo blogs are almost evenly distributed across different themes. Nearby tourists from 0 to 80 km (0.519) have a slight difference compared to local visitors, and farther tourists from 80

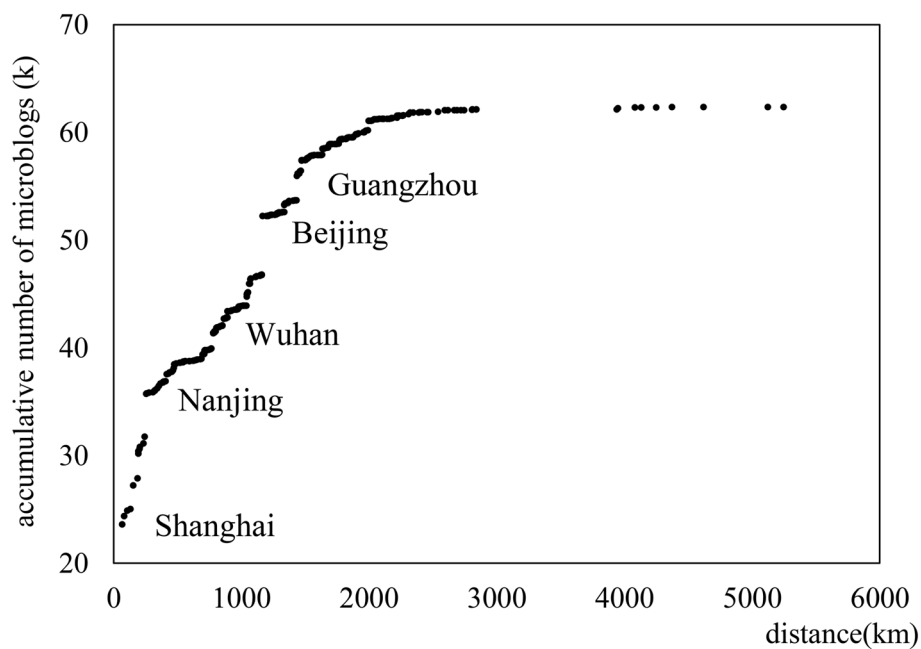
**Fig. 2** Accumulative number of Weibo blogs - distance plots

Table 4 Indexes of image differences among tourists from different distance

Region	Location Quotient									DC	VD
	Landscapes		Culture			Life					
	Cultural landscapes	Natural landscapes	Historical culture	Arts	Folk Customs & religions	Food	Leisure & recreation	Festival & events	Transports		
Suzhou	0.737	1.352	1.098	0.743	1.005	1.317	1.054	1.355	0.718	0.510	0.000
0–80 km	0.857	0.882	0.875	1.017	1.023	0.885	1.550	0.936	0.818	0.519	0.148
80–252 km	1.144	0.851	0.987	0.994	0.988	0.847	0.936	0.983	1.375	0.575	0.195
252–777 km	1.355	0.646	0.942	1.212	0.915	0.703	0.719	0.656	1.511	0.642	0.286
777–1165 km	1.496	0.587	0.974	1.498	0.979	0.664	0.384	0.441	1.675	0.686	0.351
1165-1432 km	1.621	0.632	0.903	1.355	0.819	0.591	0.342	0.250	1.093	0.722	0.399
> 1432 km	1.518	0.559	0.973	1.449	1.122	0.633	0.404	0.554	1.222	0.692	0.360
Std.	0.318	0.259	0.067	0.255	0.087	0.232	0.412	0.348	0.326	–	–

to 252 km (0.575), 252–777 km (0.642), 777–1165 km (0.686), and > 1432 km (0.692) seems to be impressed by some certain themes and ignore the others. Also, in culture-based segmentation, visitors from all the other regions tend to be impressed by some certain themes and ignore the others compared to local visitors and Wuyue tourists (0.530), and Southeast tourists have the most concentrated recognition toward Suzhou (0.727).

We further used the LQ index to estimate the theme preferences of each group. Each theme receives varying degrees of attention from visitors in different regions, and these themes can be divided into three patterns according to hierarchical clustering in both segmentations (Fig. 4).

Among all the themes, natural landscape, food, festival & events, and leisure & recreation belong to Pattern I, the short-haul preference or Wuyue preference pattern. Compared with the average value, only the local visitors show obvious concern for food and natural landscape ($LQ > 1$), while all other tourists are less concerned ($LQ < 1$). Festival & events also receives the most attention from local visitors ($LQ = 1.355$). For distance-based segments, non-local tourists' interests in these three themes generally decrease with distance. Leisure & recreation receives the most attention from tourists from 0 to 80 km ($LQ = 1.550$), far more than residents ($LQ = 1.054$). The distance decay effect also exists in farther areas. For culture-based segments, Wuyue tourists

Table 5 Indexes of image differences among tourists from different cultural groups

Region	Location Quotient									DC	VD
	Landscapes		Culture			Life					
	Cultural landscapes	Natural landscapes	Historical culture	Arts	Folk Customs & religions	Food	Leisure & recreation	Festival & events	Transports		
Suzhou	0.737	1.352	1.098	0.743	1.005	1.317	1.054	1.355	0.718	0.510	0.000
Wuyue	0.960	0.865	0.922	1.008	1.008	0.873	1.327	0.949	1.036	0.530	0.149
Jingchu	1.470	0.662	0.871	1.270	1.005	0.612	0.596	0.470	1.492	0.676	0.337
Huanghuaihai	1.445	0.626	1.017	1.606	0.905	0.722	0.367	0.415	1.687	0.672	0.330
Huangtu	1.456	0.625	0.953	1.398	1.036	0.664	0.499	0.757	1.267	0.673	0.332
Bashu	1.479	0.567	1.056	1.578	1.202	0.691	0.341	0.269	1.153	0.682	0.346
Guandong	1.527	0.568	0.667	1.216	1.169	0.726	0.441	0.136	0.913	0.700	0.350
Yungui	1.477	0.473	1.073	1.963	1.144	0.548	0.396	1.728	1.446	0.675	0.362
Southern East	1.644	0.513	0.928	1.309	0.935	0.536	0.370	0.434	1.308	0.727	0.411
Std.	0.281	0.255	0.125	0.335	0.098	0.225	0.331	0.500	0.286	—	—

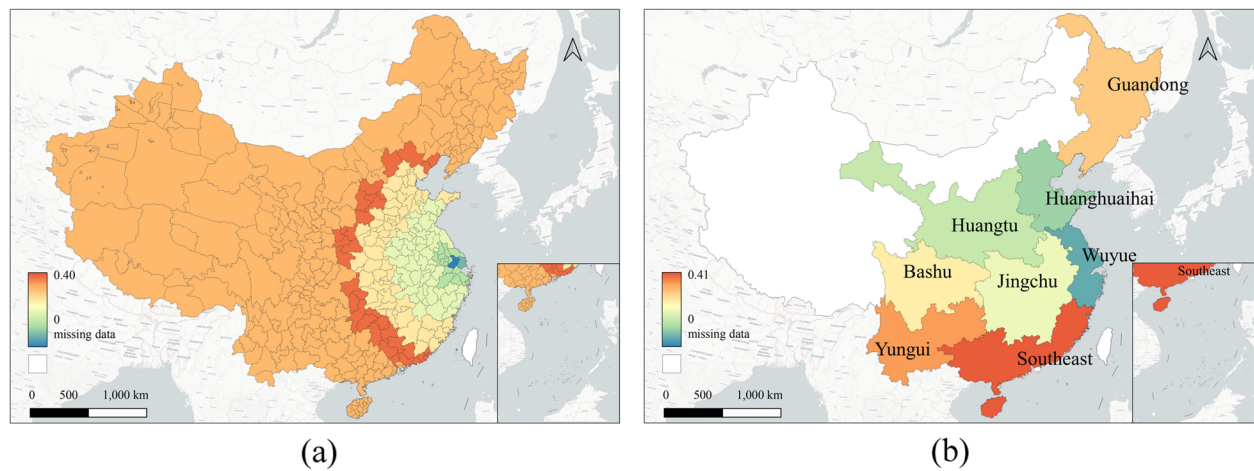


Fig. 3 VD between residents and tourists from regions of different distance (a) and cultural regions (b)

perceive the most leisure & recreation ($LQ = 1.327$) among all the tourist groups in Suzhou, while other cultural groups are indifferent to it ($LQ < 0.6$). Meanwhile, Yungui tourists are also keen on festival & events ($LQ = 1.728$), even more so than residents.

Cultural landscape, art & transports belong to Pattern II, the long-haul preference or other culture preference pattern. For distance-based segments, local visitors and nearby visitors from 0 to 80 km are not very interested in the cultural landscape (local: $LQ = 0.737$, 0–80 km: $LQ = 0.857$), and overall, as travel distance increases, visitors have a higher sense of the cultural landscape.

Long-haul visitors pay more attention to art and traffic than nearby visitors, but there is no obvious trend of increasing travel distance and increasing perception. For culture-based segments, these themes are more of a concern in other cultural areas. In addition to Suzhou and Suzhou's belonging cultural region Wuyue, visitors from other regions are more interested in Suzhou's cultural landscapes ($LQ > 1.4$), especially in the Southeast ($LQ = 1.644$). Yungui tourists are extremely interested in Suzhou's art ($LQ = 1.963$), followed by visitors from Huanghuaihai ($LQ = 1.606$) and Bashu ($LQ = 1.578$). Traffic themes are often mentioned by Huang Huaihai, Jingchu,

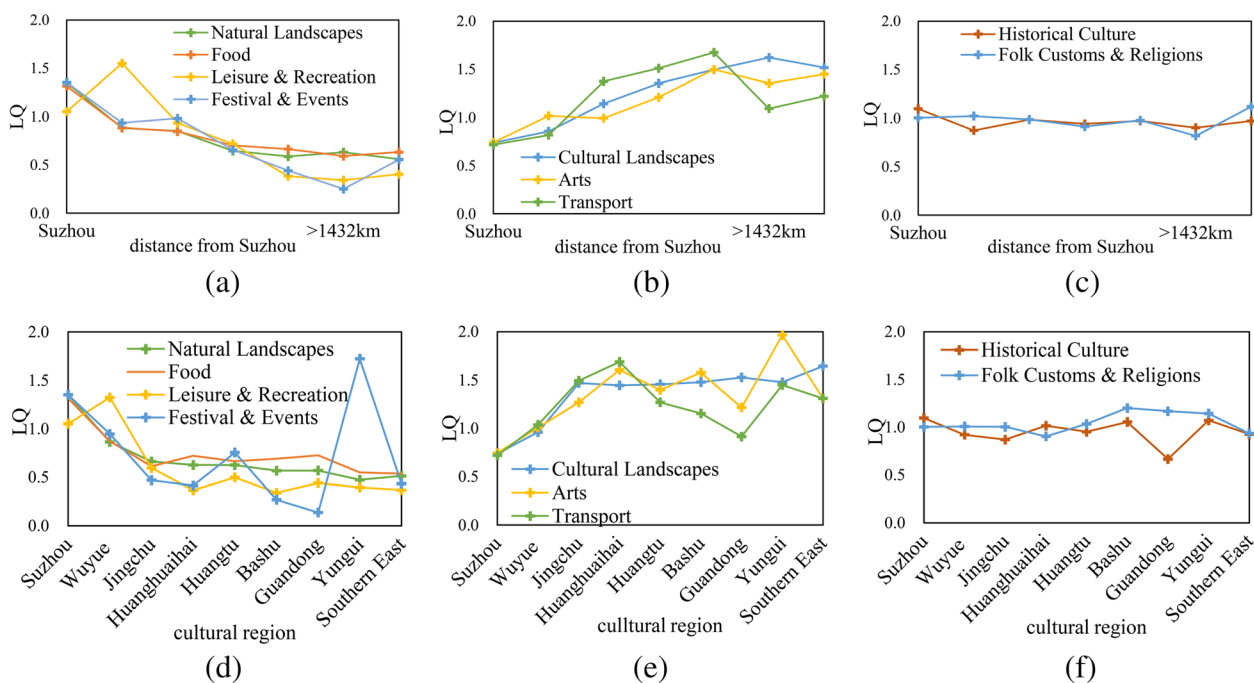


Fig. 4 LQ index of themes in three modes: a, b, c for distance-based segments, d, e, f for culture-based segments

and Yungui tourists, and local visitors and Guandong tourists rarely mention traffic-related on Weibo.

According to the Std. value, the upper themes are more focused on by some tourists than the other, and thus should be paid more attention to in strategic tourism marketing for different groups. Folk customs & religions and historical culture belong to Pattern III, with no obvious preference pattern. The Std. index shows that folk custom & religions and historical culture have no obvious difference in attractiveness to different groups, which are less important in tourism marketing.

The result of correspondence analysis in Fig. 5 visualizes the relationship among different tourist groups, among different themes, and between themes and each group's perceptions. Three patterns were identified in both segmentations. For distance-based segments, the results show that Dimension 1 divides tourists into two groups. Long-haul tourists are on the left side of the map, local visitors and short-haul tourists are on the right. The result also infers that local visitors prefer festivals & events, food, and natural landscapes, short-haul tourists prefer leisure & recreation, and long-haul tourists prefer cultural landscapes, transports, and art. Besides, the remote groups' image perceptions are relatively close to each other, local visitors' and short-haul tourists' cognitive images are more unique. Moreover, leisure & recreation has the most explicit target groups, while folk custom & religions has the least. For culture-based segments, tourists from different cultural regions are on the right side of the map, and residents and tourists from the same cultural regions (Wuyue) are on the left. Consistent with the results of the distance-

based segmentation, local visitors pay the most attention to festivals & events, food, and natural landscapes, Wuyue visitors prefer leisure & recreation, and visitors from other cultural regions prefer cultural landscapes, transports, and art. Besides, the images for tourists from different cultural regions are similar, and the cognitive images of local visitors and Wuyue visitors are unique.

4.3 Comparing impact of the two factors on image perceptions

In the Geodetector analysis, the spatial analysis unit should be defined at first. The distance-based segmentation was based on the aggregation of prefecture-level cities, while the cultural division was based on the aggregation of provinces. Therefore, we adopted the smaller unit, prefecture-level cities, for the Geodetector analysis. Considering that there were too few microblogs in some cities, the image obtained in those cities may not be representative. Finally, 74 cities were reserved as the input records.

We measured the image of each city and calculated the image VD between tourists from those cities and Suzhou visitors. The inputs of Geodetector include the VD index, the cultural region, and the distance category for each city. The outputs are exhibited in Table 6.

As shown in the factor detector result, distance-based segmentation can explain 46.6% of image variations at the 0.01 significance level, and culture-based segmentation can explain 65.8% of it at the 0.01 significance level, which means from a geographic perspective, regional culture affects the image perception to a higher extent than distance.

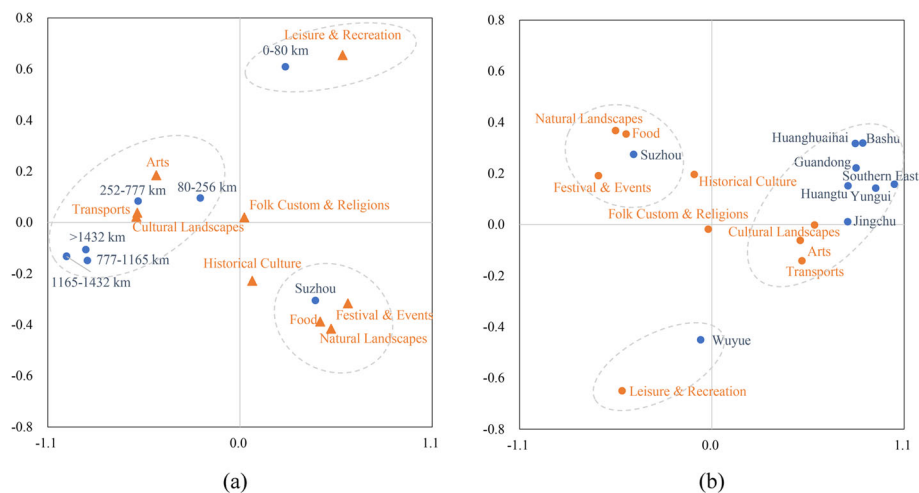


Fig. 5 CA results of the image among tourists from different distance-based (a) and cultural regions (b)

Table 6 Factor detector results

Segmentation	q-value	p-value
Distance-based (split points)		
80, 252, 777, 1165, 1432 km	0.466	< 0.001
Culture-based		
Wuyue, Jingchu, Huanghuaihai, Yungui, Huangtu, Southeast, Bashu, Guandong	0.658	< 0.001

5 Discussion and conclusions

Using social media data, we examine and compare the effects of distance- and culture-based segments of tourists in producing diverse perceived images of a destination, i.e. Suzhou, China. Specifically, an analytical method for extracting and analyzing the diverse images of a destination from the social media data is proposed. In summary, for the perceived images of Suzhou among tourists from different regions, obvious differences in several quantitative measures are observed and the culture factor is found to have a much more important effect than the distance factor.

This research divided the entire Chinese mainland market, instead of a limited number of cities, of Suzhou tourism based on distance and examined the relationship between distance and destination image. The results show that: 1) The closer the tourists are from, the smaller the difference between their image perception and the local visitors', and vice versa. 2) The diversification of cognitive dimensions shows a general trend, short-haul tourists tend to perceive a more comprehensive image than long-haul tourists, while long-haul tourists tend to be impressed by certain themes and ignore the others. 3) Image preference also exists among tourists from different distances. Residents are more interested in food, festival & events, and natural landscapes, nearby tourists tend to perceive more leisure & recreation, while remote tourists prefer culture-related themes. These findings are generally consistent with previous studies (Bao & Mckercher, 2008; Mckercher, 2008).

Different from existing cross-cultural studies, we examined the image of Suzhou perceived by different cultural groups in Mainland China. The results show that: 1) Tourists with the same cultural background have a similar perception of destination image to local visitors, and there are clear image differences among tourists from different cultural backgrounds. But the image differences of cultural regions are relatively complex and there is no trend that the image differences increased with distance. 2) Also, local visitors perceive the most food and natural landscapes; Wuyue tourists, similar to short-haul tourists, perceive more relaxation (such as leisure & recreation); other tourists, similar to long-haul

tourists, focus more on sightseeing (such as cultural landscapes).

This study further compared the influence of distance and culture on the destination image. According to *Geo-detector*, the findings demonstrated that travel distance and regional culture influence destination image to different extents, and regional culture is a more significant factor to affect image perception comparing to travel distance. Besides, in both divisions, the tourists can be divided into three groups: local visitors, short-haul tourists, and long-haul tourists. The image differences are significant between each two of them but are not so obvious within each group. Nearby tourists have a more similar perception to residents than other visitors and tend to focus more on life and relaxation themes, short-haul visitors tend to perceive more relaxation themes, and remote tourists prefer sightseeing. This phenomenon might be related to the travel motivation of different groups. Many researchers studied the impact of distance on motivation. Some discovered that short-haul tourists tend to visit the destination mainly for rest, relaxation, and escapism, while long-haul tourists mainly for improving their knowledge and learning about strange cultures (McKercher & Cros, 2003). Others found that residents travel for relaxation and shopping, long-haul tourists travel mainly for sightseeing, and short-haul tourists' motivation is a mix of the two modes above, including relaxation and sightseeing (Xue & Zhang, 2020).

This study has some practical significance for the management, operation, and marketing of Suzhou tourism. For tourism managers, using big data can allow them to segment the tourists that fully cover the market of Mainland China, enabling them to make a complete tourism marketing plan for Suzhou. As cultural factor is more important than distance factor, it might be more appropriate to divide tourist groups according to culture. After getting the destination image of each group, they can adopt different targeted marketing strategies to attract tourists from different cultural backgrounds. For example, Beijing-Tianjin-Hebei, Yangtze River Delta, and Pearl River Delta are the three major Chinese mainland tourism markets of Suzhou. Yangtze River Delta belongs to Wuyue Cultural District, just like Suzhou, they perceive Suzhou as a city of leisure & recreation. Beijing-Tianjin-Hebei and Pearl River Delta belong to the Huanghuaihai and Southeast cultural areas, Suzhou's image for them should be a city with a long history and rich in culture. Besides, residents prefer Suzhou's natural landscapes and delicacy, so marketing for local visitors should emphasize the food and natural environment. For managers of attraction, hotel, and restaurant, they can learn the tourists' preferences from the results of the study and can optimize and adapt their products to cater

to the main guests. For example, if an attraction has a lot of guests from the Yungui area who like all kinds of festivals as much as local visitors, then attraction managers should consider holding more carnival events; and if an attraction has many guests from the Bashu who prefer arts and cultural landscapes, then attraction managers should enhance the artistic atmosphere and strengthen the cultural background.

The study has some limitations. First, the user distribution of Sina Weibo is not even in different ages, cities, and so on. According to the Sina Weibo data center (2013), Sina Weibo users are mainly young people. Visitors of other ages might have different perceptions. Also, people prefer posting blogs of certain themes on social media (Yuan et al., 2020). Though image differences can be examined, the images might not be accurate. Second, it should be noted that the research object of this article is Chinese domestic tourists, and the social media of the Weibo platform is commonly used by Chinese people. Foreign tourists may post their perceptions on other social media, and the image extracted from these data might be different, which requires further empirical analysis. Third, theoretically, the destination image is usually made up of the cognitive image, affective image, and overall image. However, only recognition images can be obtained using a data-driven method, while the affective image and overall image are difficult to extract.

The impact of distance and culture on the image is a long-term topic in tourism image literature, and the emergence of big data provides new opportunities for empirical analysis and theoretical development. Other interesting patterns might appear when taking other factors into accounts, such as seasons and time. Future studies should combine more data sources and adopt better semantic processing technologies to examine the destination image more adequately and to compare the image among groups more accurately.

Code availability

The source code is available from the first author.

Authors' contributions

Conceptualization, Mengtong Wu, Chao Jiang, and Yi Zhang; Methodology, Mengtong Wu, Chao Jiang; Software, Mengtong Wu; Data Curation, Jingjing Cao; Writing-Original Draft Preparation, Mengtong Wu; Writing-Review & Editing, Chao Jiang, Yi Zhang, Ying Cheng and Yu Liu. All authors have read and agreed to the published version of the manuscript.

Funding

The research is supported by the Ministry of Science and Technology of the People's Republic of China (2018YFB0505004), China, National Natural Science Foundation of China (41830645, 42001159), China.

Availability of data and materials

Restrictions apply to the availability of these data. Data are available from the first author with the permission of the Institute of Remote Sensing and Geographical Information Systems, Peking University.

Declaration

Competing interests

The authors declare no conflict of interest.

Author details

¹Institute of Remote Sensing and Geographical Information Systems, School of Earth and Space Sciences, Peking University, Beijing, China. ²Beijing Key Laboratory of Spatial Information Integration and Its Applications, Peking University, Beijing, China. ³State Key Laboratory of Media Convergence Production Technology and Systems, Beijing, China. ⁴Management School, North Minzu University, Yinchuan, China.

Received: 5 May 2021 Accepted: 15 June 2021

Published online: 16 July 2021

References

- Ahmed, Z. U. (1991). The influence of the components of a State's tourist image on product positioning strategy. *Tourism Management*, 12(4), 331–340. [https://doi.org/10.1016/0261-5177\(91\)90045-U](https://doi.org/10.1016/0261-5177(91)90045-U).
- Arefieva, V., Egger, R., & Yu, J. (2021). A machine learning approach to cluster destination image on Instagram. *Tourism Management*, 85, 104318. <https://doi.org/10.1016/j.tourman.2021.104318>.
- Bao, Y. F., & McKercher, B. (2008). The effect of distance on tourism in Hong Kong: A comparison of short haul and long haul visitors. *Asia Pacific Journal of Tourism Research*, 13(2), 101–111. <https://doi.org/10.1080/10941660802048332>.
- Beerli, A., & Martin, J. D. (2004). Factors influencing destination image. *Annals of Tourism Research*, 31(3), 657–681. <https://doi.org/10.1016/j.annals.2004.01.010>.
- Bonn, M. A., Joseph, S. M., & Dai, M. (2005). International versus domestic visitors: An examination of destination image perceptions. *Journal of Travel Research*, 43(3), 294–301. <https://doi.org/10.1177/0047287504272033>.
- Chen, P. J., & Kerstetter, D. L. (1999). International students' image of rural Pennsylvania as a travel destination. *Journal of Travel Research*, 37(3), 256–266. <https://doi.org/10.1177/004728759903700307>.
- Crompton, J. L. (1979). An assessment of the image of Mexico as a vacation destination and the influence of geographical location upon that image. *Journal of Travel Research*, 17(4), 18–23. <https://doi.org/10.1177/00472875901700404>.
- Deng, N., Liu, J., Dai, Y., & Li, H. (2019). Different cultures, different photos: A comparison of Shanghai's pictorial destination image between east and west. *Tourism Management Perspectives*, 30, 182–192. <https://doi.org/10.1016/j.tmp.2019.02.016>.
- Frias, D. M., Rodríguez, M. A., Castañeda, J. A., Sabiote, C. M., & Buhalis, D. (2012). The formation of a tourist destination's image via information sources: The moderating effect of culture. *International Journal of Tourism Research*, 14(5), 437–450. <https://doi.org/10.1002/jtr.870>.
- Gallarza, M. G., Saura, I. G., & García, H. C. (2002). Destination image: Towards a conceptual framework. *Annals of Tourism Research*, 29(1), 56–78. [https://doi.org/10.1016/S0160-7383\(01\)00031-7](https://doi.org/10.1016/S0160-7383(01)00031-7).
- Hallab, Z., & Kim, K. (2011). The effects of nonresidents' geographical and cultural distance on a tourist destination's image. *Advances in Hospitality and Leisure*, 7(2011), 131–152. [https://doi.org/10.1108/S1745-3542\(2011\)0000007011](https://doi.org/10.1108/S1745-3542(2011)0000007011).
- Henkel, R., Henkel, P., Agrusa, W., Agrusa, J., & Tanner, J. (2006). Thailand as a tourist destination: Perceptions of international visitors and Thai residents. *Asia Pacific Journal of Tourism Research*, 11(3), 269–287. <https://doi.org/10.1080/10941660600753299>.
- Hoffman, D. L., & Franke, G. R. (1986). Correspondence analysis: Graphical representation of categorical data in marketing research. *Journal of Marketing Research*, 23(3), 213–227. <https://doi.org/10.1177/002224378602300302>.
- Hofstede, G. (2001). Culture's recent consequences: Using dimension scores in theory and research. *International Journal of Cross Cultural Management*, 1(1), 11–17. <https://doi.org/10.1177/147059580111002>.
- Hofstede, G., & Hofstede, G. H. (1984). *Culture's consequences: International differences in work-related values* (Vol. 5). Sage.
- Ji, S., & Wall, G. (2011). Visitor and resident images of Qingdao, China, as a tourism destination. *Journal of China Tourism Research*, 7(2), 207–228. <https://doi.org/10.1080/19388160.2011.576937>.
- Jutla, R. S. (2000). Visual image of the city: tourists' versus residents' perception of Simla, a hill station in northern India. *Tourism Geographies*, 2(4), 404–420. <https://doi.org/10.1080/146166800750035512>.

- Kaplanidou, K. (2009). Relationships among behavioral intentions, cognitive event and destination images among different geographic regions of Olympic games spectators. *Journal of Sport & Tourism*, 14(4), 249–272. <https://doi.org/10.1080/14775080903453815>.
- Kim, S. E., Lee, K. Y., Shin, S. I., & Yang, S. B. (2017). Effects of tourism information quality in social media on destination image formation: The case of Sina Weibo. *Information & Management*, 54(6), 687–702. <https://doi.org/10.1016/j.im.2017.02.009>.
- Kwek, A., & Lee, Y. S. (2008). Intra-cultural variance of Chinese tourists in destination image project: Case of Queensland, Australia. *Journal of Hospitality & Leisure Marketing*, 16(1–2), 105–135. <https://doi.org/10.1080/10507050802096984>.
- Lam, I. K. V., & Ryan, C. (2020). Intra-cultural variation among tourists of the same cultural background – Does it matter? *Tourism Recreation Research*, 1–14. <https://doi.org/10.1080/02508281.2020.1822020>.
- Lee, G., & Lee, C. K. (2009). Cross-cultural comparison of the image of Guam perceived by Korean and Japanese leisure travelers: Importance–performance analysis. *Tourism Management*, 30(6), 922–931. <https://doi.org/10.1016/j.tourman.2008.11.013>.
- Lee, S. H., Chang, S. C., Hou, J. S., & Lin, C. H. (2008). Night market experience and image of temporary residents and foreign visitors. *International Journal of Culture Tourism and Hospitality Research*, 2(3), 217–233. <https://doi.org/10.1108/17506180810891591>.
- Leisen, B. (2001). Image segmentation: The case of a tourism destination. *Journal of Services Marketing*, 15(1), 49–66. <https://doi.org/10.1108/08876040110381517>.
- Liu, Y. (2016). Revisiting several basic geographical concepts: A social sensing perspective. *Acta Geographica Sinica*, 71(4), 564–575. <https://doi.org/10.11821/dlxb201604003>.
- Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., Chi, G., & Shi, L. (2015). Social sensing: A new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers*, 105(3), 512–530. <https://doi.org/10.1080/00045608.2015.1018773>.
- Luo, R., Xu, J., Zhang, Y., Ren, X., Sun, X. (2019). Pkuseg: A toolkit for multi-domain chinese word segmentation. *arXiv preprint arXiv:1906.11455*.
- Mackay, K. J., & Fesenmaier, D. R. (1997). Pictorial element of destination in image formation. *Annals of Tourism Research*, 24(3), 537–565. [https://doi.org/10.1016/S0160-7383\(97\)00011-X](https://doi.org/10.1016/S0160-7383(97)00011-X).
- McCartney, G. (2008). Does one culture all think the same? An investigation of destination image perceptions from several origins. *Tourism Review*, 63(4), 13–26. <https://doi.org/10.1108/16605370810912182>.
- Mckercher, B. (2008). The implicit effect of distance on tourist behavior: A comparison of short and long haul pleasure tourists to Hong Kong. *Journal of Travel & Tourism Marketing*, 25(3–4), 367–381. <https://doi.org/10.1080/10548400802508473>.
- Mckercher, B., & Cros, H. D. (2003). Testing a cultural tourism typology. *International Journal of Tourism Research*, 5(1), 45–58. <https://doi.org/10.1002/jtr.417>.
- Marine-Roig, E., & Clavé, S. A. (2016). Perceived image specialisation in multiscalar tourism destinations. *Journal of Destination Marketing & Management*, 5(3), 202–213. <https://doi.org/10.1016/j.jdmm.2015.12.007>.
- Morrison, A. M. (1996). *Hospitality and travel marketing* (2nd ed.). Delmar.
- Moscardo, G., Pearce, P., & Morrison, A. (2001). Evaluating different bases for market segmentation: A comparison of geographic origin versus activity participation for generating tourist market segments. *Journal of Travel & Tourism Marketing*, 10(1), 29–49. https://doi.org/10.1300/J073v10n01_03.
- Obenour, W., Lengfelder, J., & Groves, D. (2005). The development of a destination through the image assessment of six geographic markets. *Journal of Vacation Marketing*, 11(2), 107–119. <https://doi.org/10.1177/1356766705052569>.
- Peng, X., Bao, Y., & Huang, Z. (2020). Perceiving Beijing's "city image" across different groups based on geotagged social media data. *IEEE Access*, 8, 93868–93881. <https://doi.org/10.1109/ACCESS.2020.2995066>.
- Pike, S. (2002). Destination image analysis—A review of 142 papers from 1973 to 2000. *Tourism Management*, 23(5), 541–549. [https://doi.org/10.1016/S0261-5177\(02\)00005-5](https://doi.org/10.1016/S0261-5177(02)00005-5).
- Sina Weibo Data Center (2013). 2013 Weibo user development report. Retrieved on Feb 11 2020 from <http://data.weibo.com/report/reportDetail?id=76>
- Sun, Y., Liang, C., & Chang, C. C. (2020). Online social construction of Taiwan's rural image: Comparison between Taiwanese self-representation and Chinese perception. *Tourism Management*, 76, 103968. <https://doi.org/10.1016/j.tourman.2019.103968>.
- Tang, L., Choi, S., Morrison, A. M., & Lehto, X. Y. (2009). The many faces of Macau: A correspondence analysis of the images communicated by online tourism information sources in English and Chinese. *Journal of Vacation Marketing*, 15(1), 79–94. <https://doi.org/10.1177/1356766708098173>.
- Tasci, A. D. A., Gartner, W. C., & Tamer Cavusgil, S. (2007). Conceptualization and operationalization of destination image. *Journal of Hospitality & Tourism Research*, 31(2), 194–223. <https://doi.org/10.1177/1096348006297290>.
- Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit Region. *Economic Geography*, 46(sup1), 234–240. <https://doi.org/10.2307/143141>.
- Wang, J. F., & Xu, C. D. (2017). Geodetector: Principle and prospective. *Acta Geographica Sinica*, 72(1), 116–134. <https://doi.org/10.11821/dlxb201701010>.
- Wang, J. F., Zhang, T. L., & Fu, B. J. (2016). A measure of spatial stratified heterogeneity. *Ecological Indicators*, 67, 250–256. <https://doi.org/10.1016/j.ecolind.2016.02.052>.
- Wu, B. (1996). The formation and division of Chinese cultural district. *Academic Monthly*, 3, 12–17. <https://doi.org/CNKI:SUN:XSXK.1996-03-001>.
- Xue, L., & Zhang, Y. (2020). The effect of distance on tourist behavior: A study based on social media data. *Annals of Tourism Research*, 82, 102916. <https://doi.org/10.1016/j.jannals.2020.102916>.
- Yuan, Y., Lu, Y., Chow, T. E., Ye, C., Alyaqout, A., & Liu, Y. (2020). The missing parts from social media-enabled smart cities: Who, where, when, and what? *Annals of the American Association of Geographers*, 110(2), 462–475. <https://doi.org/10.1080/24694452.2019.1631144>.
- Zhang, X., Yang, Y., Zhang, Y., & Zhang, Z. (2020). Designing tourist experiences amidst air pollution: A spatial analytical approach using social media. *Annals of Tourism Research*, 84, 102999. <https://doi.org/10.1016/j.jannals.2020.102999>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.