This is a post-peer-review, pre-copyedit version of a paper published in Information and Communication Technologies in Tourism 2017. The final authenticated version is available online at: https://doi.org/10.1007/978-3-319-51168-9_4

Reviewing Geotagging Research in Tourism

Elise Wong [a],
Rob Law [a], and
Gang Li [b]
[a] School of Hotel and Tourism Management
The Hong Kong Polytechnic University, Hong Kong
[a]elise.wong@outlook.com, rob.law@polyu.edu.hk

[b] School of Information Technology
Deakin University, Australia
[b] gang.li@deakin.edu.au

Abstract

Advanced medium-sharing service and mobile technologies create a large volume of geotagged data online. The characteristics of geotagged data provide a new method for tourism and hospitality researchers to analyze tourist movement and behavior. To extend knowledge on utilizing geotagged data in the tourism and hospitality industry, this study aims to review existing geotagging research in tourism and hospitality and thus identify a potential research topic in this area. Five research categories and future geotagging research topics in tourism and hospitality are identified and discussed.

Keywords: geotagged data; hospitality; tourism; literature review.

1 Introduction

After the launch of commercial Internet in the 1990s, information technologies are an indispensable resource for the tourism and hospitality organization to maintain competitive advantage (Musante et al., 2009). Currently, the research on information technologies in the tourism and hospitality sector is shifted into utilizing geotagged data to examine tourist movement and behavior. The emerging georeference resources are from the advanced medium-sharing services and mobile technologies (Zheng et al., 2012; Zheng et al., 2011). This type of data is generated by users (Vu et al., 2016) and available on the Internet for free. Users generally share and record georeference resources for their personal reasons, such as sharing experience, life logging, managing multimedia content, or sports activity analysis (Zheng et al., 2009). The content of georeferenced resource is controlled by the user, and the user can add diverse information to the original resource to enhance its expression, such as geographic location, time, visual image, and textual description (Zheng et al., 2012). Owing to the characteristics of the data, most of the existing research has named them as geotagged data or volunteered geographic information; in this study, we refer this type of information as geotagged data.

The characteristics of geotagged data create a new method for research to observe, record, and analyze human movement and behavior (O'Neill et al., 2006), allow research to further examine and identify landmarks (Girardin, Dal Fiore, Ratto et al., 2008), and become an emerging research topic in computer science and knowledge discovery from database communities (Zheng et al., 2011). In the study of Zipf and Malaka (2001), they mentioned the importance of geographic information in every tourism application presumed that mobile technologies would continuously influence the tourist behavior and cause tourists to seek for more personalized information and

service. Meanwhile, the geotagging research in the tourism and hospitality field remains in the early stage, and further research in this area is needed to enhance and extend the existing knowledge. To further understand the research trend and the potential of geotagged resource in tourism and hospitality, this study aims to provide a comprehensive overview of existing geotagging articles in the tourism and hospitality.

2 Geotagged Data

Geotagging is a process of adding geospatial information, temporal information, and/or textual description of medium resources (Dickinger et al., 2008; Kádár & Gede, 2013). In addition to textual description, four other methods are also available for users to annotate geospatially and temporal information into medium resources (Dickinger et al., 2008): 1) annotated by devices, such as car navigation system, global positioning system (GPS) device, and GPS-embedded smartphone; 2) determined by users manually; 3) identified by website servers; and 4) expounded by existing documents automatically. Although most of the existing geotagging research in tourism and hospitality has focused on geotagged photos, different types of geotagged data are available on the Internet, including videos (e.g., videos posted on YouTube) and documents (e.g., articles in Wikipedia, blogs, and Twitter) (Zheng et al., 2011). However, most of the geotagging research in tourism and hospitality has focused on utilizing geotagged photo. Until recently, Guo et al. (2015) utilized data from structured tourism blogs to analyze tourist movement and to identify interesting point and user service in a designed area. Nonetheless, no study has utilized geotagged videos in tourism and hospitality yet; thus, further research by utilizing different types of geotagged resource is needed.

2.1 Advantages of geotagged data

Tourism organizations should understand the tourist preference and their behavior. However, collecting tourist movement or behavior data was not an easy task prior to the emergence of geotagged data. Previous data-collecting techniques in tourist movement studies were limited to face-to-face interview (Forer & Simmons, 2002), questionnaire (Lau & McKercher, 2006; McKercher & Lau, 2008; Xia et al., 2009), secondary data published by the government or a tourism organization (Forer & Simmons, 2002; Li et al., 2008; Wu & Carson, 2008), observation (Xia et al., 2009), GPD system (Shoval et al., 2011; Zheng et al., 2009), and mobile device (Asakura & Iryo, 2007). These traditional methods are costly (for both monetary and time budget), and the volume of the data collected is usually much smaller.

The emergence of geotagged data has provided a more effective and efficient method to analyze tourist movement and preference. Unlike raw data, geotagged data contain proliferation information that reflects the surrounding area and the owner of resources (Kisilevich et al., 2010). The recorded and shared geotagged media by users will become digital footprints of the users, and a sequence of a geotagged medium posting on a website can reveal the spatiotemporal movement of the users in a specific area (Jiang et al., 2011; Zheng et al., 2012). Analyzing tourist movement provides useful information to researchers. Researchers also can utilize geotagged data to analyze user movement and behavior for multiple users to identify the point of interest, attractive place, landmark, new attraction, and tourist preference in a research location (Kisilevich et al., 2010). This information is useful for tourism suppliers, such as travel agents and destination management organizations, to improve their tourism plan or product design. Moreover, Geotagged resources also benefit tourists because geotagged data can be searched by title, keywords, and/or location (Lau & McKercher, 2006); therefore, users with similar interests can be identified and matched, and the tourist searching time can be reduced (Dickinger et al., 2008).

2.2 Challenges

As mentioned above, geotagged data are useful for researchers to analyze tourist movement and behavior. However, several challenges have also been identified from reviewing existing geotagging literature and been discussed by a few studies.

First, the potential and uniqueness of geotagged data are mainly accounted for the embedded geographic, temporal, and textual information, but some of the geotagged resources might contain incorrect information because of human or device error (Girardin, Calabrese, Dal Fiore, et al., 2008); hence, considerable geotagging research needs to conduct further processes to eliminate those incorrectly tagged data from data sets.

Second, geotagged data (for textural information) are multilingual in nature. The resource developed by Zheng et al. (2011) indicated that such characteristic is one of the major challenges in analyzing textual information from geotagged photos. In fact, most of the existing geotagging literature reviewed in this study has analyzed only the geotagged photos in English (except Guo et al., 2015), and the geotagged data from non-English users are not covered.

Third, nearly all of the geotagging research in tourism and hospitality has collected data from Flickr and Panoramio, which might be because other medium-sharing websites, such as Instagram API, require scholars to seek for owner consent (Instagram, 2016) before downloading and utilizing photos, whereas Flickr and Panoramio API do not have this rule and requisition. However, Flickr and Panoramio have fewer numbers of users than the other mentioned medium-sharing websites. According to an Instagram press release in 2016, 500 million Instagram users exist worldwide (Instagram, 2016a). By contrast, only approximately 112 million Flickr users globally were recorded at the end of 2015 (Digital Stat Articles, 2016).Thus, examining geotagged data from Flickr and Panoramio might reflect a small proportion of Internet users to extract information on travel movement and behavior. Researchers might collect a larger number of tourist photos from Instagram and estimate the tourist movement and behavior more accurately than using data from Flickr. Moreover, given that using geotagged data from Flickr and Panoramio does not require consent from the photo owner, privacy and ethical issues might arise (Girardin, Calabrese, Dal Fiore, et al., 2008).

Handling and analyzing geotagged data need specific analysis skills and knowledge. Research needs to have extensive programming and applied mathematics knowledge to handle the large volume of data. Utilizing geotagged data might be a challenge for the tourism and hospitality scholars.

3 Research Methodology

In June 2016, all geotagging-related articles in tourism and hospitality were identified from Google Scholar (https://scholar.google.com.hk/), Science Direct (http://www.sciencedirect.com/), and EBSCOHOST (https://www.ebscohost.com/) - three most popular and the largest search engines and online database (Law et al., 2010). The keywords of “geotagged,” “geotagging,” and “geo-references,” with each of the terms “tourism,” “travel,” “hospitality,” and “hotel,” were used to search for geotagging-related articles in academic journals and conference proceedings. References cited in published articles were also traced. In total, 30 published articles were determined to be relevant for this study. Table 1 lists the research categories, description, and listed articles. In the following section, the key finding and

discussion in each research category are presented.

4 Findings

The geotagging research in tourism and hospitality remains in the nascent stage. Most of the geotagging articles reviewed in this study (24 out of 30) were published in computer science and information technology journals or conferences. These articles have concentrated on the technical aspect of analyzing and/or utilizing geotagged data in tourism, such as developing new data-mining methods for geotagged blog data to analyze tourist movement (Guo et al., 2015) or constructing a new method to provide tourism recommendation (Cao et al., 2010; Jiang et al., 2011; Majid et al., 2012). Seven geotagging articles have recently been published in tourism and hospitality journals and conferences, and these articles have concentrated on utilizing geotagged photos to understand or analyze tourist movement and behavior explained with related theories.

On the basis of the research aims and objectives, five research categories were classified, and their related articles are discussed in the following subsections.

4.1 Geotagged Data Collection and Preprocessing

As mentioned earlier, geotagged data are not limited to photos but also include other media, such as video and document, that users can add with geographical, time, or/and textual description and information. Table 2 shows the data source and types of geotagged data used in the reviewed articles. To ensure that generated results are valid and reliable, most of the geotagging research has removed the noise data from the collected data; for example, a photo has incorrect geographic information and timestamp. In case of tourism, geotagging articles have also treated the photo of residents as noise data. Thus, considerable geotagging research in tourism has removed this kind of photo in the data-preprocessing stage. Two articles have suggested a method to separate photos taken by residents from those of tourists. Girardin, Calabrese, Dal Fiore et al. (2008) were the first to publish geotagging research in tourism. They used geotagged photos from Flickr with the roaming signal provided by a telecom company to determine the tourist movement in Rome and provided a method to extract tourist photo from the collected photos. The authors suggested using the presence in the area over a period of time to determine whether the photographer is resident, and this method has been further adopted by other geotagging research in tourism, such as Girardin, Dal Fiore, Ratti et al. (2008), Sun et al. (2013), and Kádár & Gede (2013). The study of Da Rugna et al. (2012) mainly introduced a method to extract tourist-geotagged photos from those of residents. They suggested that researchers should use multiple factors to separate tourist photos from resident photos. The factors included 1) the timeline of each photo taken within a location, 2) the number of place visit within a city, 3) the number of day in the country, 4) the number of the visit within a period, and 5) the number of days between two visits.

All the geotagging articles discussed above have focused on analyzing tourist movement and behavior using geotagged photos. Recently, Guo et al. (2015) developed a new data-mining method to collect geotagged data from structured tourism blogs to analyze point of interest, tourist movement, and tourist-used service in a specific area, as well as further extended the data collection technique in the geotagging research in tourism. Table 3 summarizes all the data analysis processes and methods found from the reviewed geotagging articles.

4.2 Geotagged Data Analysis

The majority of geotagging research in tourism has been concerned with developing new data analysis methods. Most of the articles have focused on developing new methods to estimate geotagged photo location or determine tourist attraction. For example, Kalogerakis et al. (2009) and Crandall et al. (2009) established a method to estimate photo location without any geospatial curve. Kisilevich et al. (2010) and Lee et al. (2013) constructed a new method to determine popular attractions from geotagged photos.

Kalogerakis and the research team (2009) utilized a modified hidden Markov chain model to identify photo locations. The model was developed based on the temporal and visual features of collected photos. With a similar research approach, Crandall et al. (2009) used visual, temporal, and textual information collected from photos to estimate the photo location and indicated that using visual and temporal information can estimate locations more accurately than using textual tag data.

Several articles have also used software to handle the data analysis of geotagged data. For example, Girardin, Calabrese, Dal Fiore, et al., (2008) and Girardin, Dal Fiore, Ratti et al. (2008) used designed software—Urban dynamic—to analyze geotagged photos based on map visualization; Garacia-Palomares et al. (2015) used a geographic information system program to analyze geotagged photos and based on density map and descriptive statistics to identify popular attraction in eight European cities.

The article presented by Kisilerich, et al. (2010) suggested a research framework to analyze geotagged photos from a visual analytic approach; the research framework contained seven tasks that aimed to identify attraction and tourist behavior at the city level. For the article that aims to develop a new method to determine popular attraction, Kisilevich et al. (2010) developed a new density-based clustering algorithm (P-DBSCAN) to determine popular locations. The P-DBSCAN was developed based on DBSCAN (Ester, Kriegel, Sander et al., 1996). Both methods are nonparameter clustering that do not need to determine the number of clusters in advance and support cluster with arbitrary shape and efficiency on large-scale data (Kisilevich et al., 2010; Zheng et al., 2012; Vu et al., 2015; Vu et al., 2016). Therefore, both methods are suitable to determine location and event from geotagged sources, but the clustering definition between the two methods is different. P-DBSCAN does not define density based on the number of photos taken in a point but uses the number of photographers to define the density that enables filtering outlier effectively; thus, it can determine popular locations accurately. Numerous geotagging articles in tourism have also used this P-DBSCAN method, such as Majid et al. (2012), Xu et al. (2015), Vu et al. (2015), Vu et al. (2016), and Leung et al. (2016). The other study presented by Lee et al. (2013) combined the DBSCAN method with the association rule-mining technique to determine point of interest; using the combined method to determine location can generate a more accurate result, but the analysis process is more complex than a single method approach.

Besides location, tourist movement has also been estimated via new methods in two articles. Zheng et al. (2012) adopted the Markov chain model to estimate tourist route and further classified tourists into busy and relax trip tourists using the modified longest common subsequence. Later on, Vu et al. (2015) also adopted the Markov chain model to estimate the tourist flow in Hong Kong. Farzanyar and Cerone (2015) used two data-mining techniques-mean-shift clustering algorithm and MapReduce Apriori-based algorithm through Hadoop-to estimate the tourist attraction and flow in Europe.

4.3 Tourism movement and tourist behavior

The tourist movement and behavior research category focused on examining tourist flow, interested attraction, selection of tourism service, and/or tourist activities. Five

geotagging articles were classified into this category, and all these articles have examined tourist movement and/or tourist behavior in city level, and no study has been conducted in any Southeast Asian cities yet.

Two studies under this research category have focused on investigating tourist attraction. Kádár and Gede (2013) identified interested locations in Budapest from geotagged photos collected from Flickr, and Leung et al. (2016) focused on investigating less popular tourist attraction in Hong Kong. Aside from tourist attraction, the study presented by Vu et al. (2015) also analyzed tourist movement and indicated that tourists tend to go nearby attraction. The study also compared the movement and behavior between Western and Asian tourists and showed the different movements and interested attractions between these two types of tourist.

Researchers also have utilized geotagged photo data to analyze tourist preferences and activities. Sun et al. (2013) analyzed the distribution of tourist-selected hotels in Vienna in each season and indicated that seasonality does influence tourist selection of hotel. Vu et al. (2016) used geotagged photo to analyze tourist activities in Hong Kong parks and used the textual information to identify tourist activities.

4.4 Tourism recommendation systems/applications

Geotagging research in this category is concerned with developing a new method to improve tourism recommendation system/application to facilitate tourist searching behavior. Eight articles were classified into this category, and those articles can further be divided into two subthemes of suggesting a location or suggesting a route. Given that this research category focused on developing a tourism recommendation method or system, the articles in this category have tended to present a detailed description of the data collection and analysis processes. Articles commonly evaluate the performance of a designed system by comparing the designed tourism recommendation system/method with another existing system.

In the early stage, articles have focused on developing a system/method to provide attraction or location suggestion to users. For example, Mamei et al. (2010) developed an application to provide location recommendation in a fixed area; this application was developed based on user travel history, but it was only suitable for the tourists who had been visiting more attractions than those who visited a few. The other study presented by Cao et al. (2010) developed a worldwide tourist recommendation system that aimed to minimize the searching time of tourists. The system enabled users to use visual images and keywords to search for tourism location and attraction.

Later on, tourism recommendation articles have focused on developing a system that provides recommendations with consideration of contextual information and emphasizes personalized suggestions. The research presented by Jiang et al. (2011) was the first article to use geotagged information with other contextual information, such as textual tag, visual image, and user similarity to provide personalized tourism recommendation. Majid et al. (2012) and Xu et al. (2015) developed tourism recommendation systems based on travel history and user current contextual information, such as weather and time, to provide a suggestion. The major difference between the two systems is that Majid et al. (2012) used another user travel history to predict tourist preference, whereas Xu et al. (2015) used the travel histories of both user and other users.

Previous tourism recommendation systems relied on the travel history of the user or other users to predict tourist preference; thus, they were only useful for the users who had rich travel experiences. The recommendation system developed by Jiang et al. (2013) overcame this problem by using user interest, current location, and time to provide relevant recommendations for the user, and it was more suitable for the user

with less travel history.

Two articles have developed route recommendation systems as well. Okuyaman and Yanai(2013) developed a tour recommendation system for a given area. Kurashima et al. (2013) established a personalized travel route recommendation system; this system also considered user current context and user interest to suggest a route with travel time and transportation method.

4.5 Others

Five articles were classified into this research category because their research theme and approach differ from those of other geotagging research in tourism. All these articles were published in tourism and hospitality-related journals and conferences, and their research theme was beyond the other geotagging research, such as tourist movement and tourist attraction. These articles have applied geotagged photos to analyze and interpret other tourism issues and phenomenon. For example, Dickinger et al. (2008) discussed and presented a method to annotate geospatial information for tourism resource automatically and indicated that geotagged tourism resource can assist the searching experience of tourists; Zanker et al. (2009) developed a method to annotate tourism and hospitality product automatically based on geotagged information; Chareyro et al. (2014) presented a conceptual paper to discuss the potential and challenge of using geotagged data to develop research in tourism. Except the works from Dickinger et al. (2008) and Zanker et al. (2009), others articles in this category did not use any data-mining method or algorithm to analyze geotagged data,which indicates that geotagged data can be handled and analyzed without programming knowledge. For example, Donaire et al. (2014) used 1786 geotagged photos from Flickr to analyze photographer view. The article used content analysis to categorize the photo scene into four types of tourist view, namely, nature, heritage, culture, and tourist service. Ward's clustering method was also used to determine four groups of tourist. Önder et al. (2014) used geotagged photos from Flickr to forecast the number of tourists spending nights in both regional and city levels. The study only used polynomial regression analysis to estimate tourist demand and found that geotagged photos are only useful for estimating the number of tourists spending night in city level instead of regional level.

5 Conclusions and future research

The present study has reviewed 30 geotagged related articles in tourism and hospitality, and provided an updated geotagged data research approach in tourism and hospitality. The study also classified existing tourism geotagged research into five categories base on the research aims and objectives, and identified majority of the tourism geotagged research was focused on the technical aspect of analysis and/or utilizing geotagged data in tourism. Only few research was concentrated on utilizing geotagged data on tourism management. Further research in utilizing geotagged data on tourism management is needed. Here, we would like to summarize the findings and to envisage the future research work along this line.

Existing tourist movement literature has indicated that tourist movement and behavior are influenced by different factors, such as time and expenditure budget, personal factors, and knowledge of the destination (Lew & McKercher, 2006). Therefore, analysis of the overall tourist behavior and movement in a region or city only provides general information for tourism organizations. To better understand different tourist movement and behavior, future geotagging research can classify tourists based on their demographic information, textual tag language, and visual image, as well as compare different tourist movement and behavior in a region or city.

Geotagging data can be used to analyze tourist-preferred service in a region or city in

addition to tourist attraction. For example, Sun et al. (2013) used geotagged data to investigate the distribution of hotels in Vienna. Understanding tourist service preference can assist service providers to develop and improve their products and services and help researchers further understand tourist behavior in different cities and countries. Hence, future geotagging research can further analyze tourist-preferred services, such as restaurant and transportation, in a regional or city level. Most of the geotagging research has been concerned with developing tourism recommendation systems, but those articles have only focused on providing popular attraction or route for the tourists. Information regarding popular attractions or routes should be easily access on the Internet, but tourists need to use much time to find less popular attraction and route in a destination. Accordingly, future research should try to identify less popular but unique attractions and routes from geotagged data (Chareyron et al., 2014) and to provide a diverse tourism recommendation for tourists.

Geotagged data are a potential resource for tourism and hospitality research, but not every tourist will upload a photo online or tag related information into media. Thus, geotagged data might reflect specific types of tourist behavior and movement only. As suggested by Crampton et al. (2013), researchers should use multiple sources to ensure geotagging research reliability. Therefore, future geotagging research should try to use different sources with geotagged data to analyze tourist movement and behavior. For example, researchers can cooperate with local destination management organizations and use both tourist survey and geotagged data to analyze and understand tourist movement and behavior to cross check the results between two data to ensure that the results are reliable. Moreover, most of the existing geotagged research in tourism and hospitality only used single method to collect geotagged data. Indeed, as mentioned early, there are different types of geotagged data available on the Internet, scholars should consider collecting different types of geotagged data from different sources to analyze and compare tourism movements and behavior.

Table 1: List of Geotagged Research in Tourism and Hospitality

Research Categories	Description	List of Publication
Geotagged data collection & preprocessing	concerned to develop new data collection or preprocessing method	Da Rugna, Chareyron, & Branchet (2012);Girardin, Calabrese,Dal Fiore, Ratti, & Blat (2008); Girardin, Dal Fiore, Ratti, & Blat (2008); Guo, Li, & Sun (2015).
Geotagged data analysis	concerned to develop new analysis method to analyze geotagged data	Crandall, Backstrom, Huttenlocher, & Kleinberg (2009);Farzanyar, & Cercone (2015); Garcia-Palomares, Gultiérrez, & Minguez (2015);Girardin, Dal Fiore, Ratti, & Blat (2008); Kalogerakis, Vesselova, Hays, Efros, & Hertzmann (2009); Kisilevich, Krstajic, Keim, Andrienko, & Andrienko (2010); Kisilevich, Mansmann, & Keim (2010); Lee, Cai, & Lee (2013); Zheng, Zha, & Chua (2012).
Tourism movement & Tourist behavior	concerned to examine tourist flow, interested attraction, preferred service and/or tourist activities in a region or city level	Kádár, & Gede (2013); Leung, Vu, Rong, & Miao (2016); Sun, Fan, Helbich, & Zipf (2013); Vu, Leung, Rong, & Miao (2016); Vu, Li, Law, & Ye (2015)
Tourism recommendation system/application	focused on developing a new method to improve existing recommendation system or application	Cao, Luo, Gallagher, Jin, Han, & Huang (2010); Jiang, Wang, & Yu (2011); Jiang, Yin, Wang, & Yu (2013); Kurashima, Iwata, Irie, & Fujimura (2013); Majid, Chen, Chen, Mirza, Hussain, I., & Woodward (2012); Mamei, Rosi, & Zambonelli, (2010); Okuyama, & Yanai (2013); Xu, Chen, & Chen (2015).
Others	research approach and focus are variant	Chareyron, Da-Rugan, & Raimbault, (2014); Dickinger, Scharl, Stern, Weichselbraun, & Wöber (2008); Donaire, Camprubí, & Gali (2014); Önder, Koerbitz & Hubmann-Haidvogel (2014); Zanker, M., Fuchs, M., Seebacher, A., Jessenitschnig, M., & Stromberger, M. (2009).

Table 2: Type of geotagged data used in geotagged research in tourism and hospitality

Type of geotagged data used	Source of the data	No. of publication
Geotagged photo	Flickr	20
	Panoramio	3
	Multisource	
	Flickr & Panoramio	2
	Flickr & cell phone roaming signal	1
Structured tourism blog	qunar.com	1

Table 3: Geotagged Data Analysis Processes and Methods

Data Analysis Process			
Preprocessing method	**Purpose**	**Method**	**Articles**
	Separate photo taken from resident and tourist	Girardin, Calabrese et al., (2008) advised method, using presence in the area over a period of time	Girardin, Calabrese et al. (2008); Girardin, Da Fiore et al. (2008); Sun et al. (2013); Kadar & Gede (2013)
		Used the photographer profile	Vu et al.(2015)
		Multi-factors	Da Rugna et al.(2012)
Estimate location/attraction	**Purpose**	**Method**	**Articles**
	Determine the point of interest or attraction from the collected geotagged data	DBSCAN	Kisilevich, Krstajic et al. (2010); Zheng et al. (2012)
		P-DBSCAN	Kisilevich, Mansma & Keim, (2010); Leung et al. (2016); Majid et al. (2012); Vu et al. (2016); Vu et al. (2015); Xu et al. (2015).
		K-mean	Mamei et al., (2010)
		Mean shift	Cao et al. (2010); Crandall et al. (2009); Farzany & Cercone (2015); Jiang et al. (2011); Jiang et al. (2013); Kurashima et al. (2013)
		Hierarchical clustering	Okuyama & Yanai (2013)
		Counting	Kadar & Gede (2013)
		modified Hidden Markov Model	Kalogerakis et al. (2009)
		Combined method	
		Kernel density estimation & Spatial scan statistics	Sun et al. (2013)
		DBSCAN & association rules mining	Lee et al. (2013)
		Density map & descriptive statistics	Garcia-Palomar et al. (2015)
Estimate tourist flow/movement	**Purpose**	**Method**	**Articles**
	Estimate the tourist based on the identified location and temporal information	Markov Chain	Vu et al. (2015); Zheng et al. (2012)
		Spatial statistical indicators	Garcia-Palomar et al. (2015)
		Apriori-based algorithm	Farzany & Cercone (2015)
		Combined method	
		Markov chain + topic model	Kurashima et al. (2013)
Analysis textual information	**Purpose**	**Method**	**Articles**
	Using textual data to analysis tourist behavior	GATE	Leung et al. (2016)

Data Analysis Process			
Determine representative textual tag	**Purpose**	**Method**	**Articles**
	Determine the representative textual tag which aims to reduce user search time	TF- frequency-inverse document frequency (tf-id) value	Majid et al. (2012)
		Vector space model (VSM)	Crandall et al. (2009); Jiang et al. (2011); Jiang et al. (2013)
		computing the frequency	Cao et al. (2010)
		computing the uniqueness of the tag score	Okuyama & Yanai (2013)
Determine representative photo in a location	**Purpose**	**Method**	**Articles**
	Determine the representative photo for a location	spectral clustering	Jiang et al. (2011); Jian et al. (2013)
		computing the frequency	Cao et al. (2010)
		computing the uniqueness of the photo score	Okuyama & Yanai (2013)

Acknowledgment: The work described in this paper was supported by a grant funded by the Research Grants Council of the Hong Kong Special Administrative Region, China (GRF Project Number: 15503814). The project was also funded by the Hong Kong Polytechnic University.

References

Asakura, Y. & Iryo, T. (2007). Analysis of tourist behavior based on the tracking data collected using a mobile communication instrument. *Transportation Research Part A*, 41, 684-690.

Cao, L., Luo, J., Gallagher, A, C., Jin, X., Han, J. & Huang, T. S. (2010). A Worldwide Tourism Recommendation System Based on Geotaggedweb Photos. *ICASSP*, 2274-2277.

Chareyron, G., Da-Rugan, J. & Raimbault, T. (2014). Big Data: a new challenge for tourism. In *International Conference on Big Data*, 5-7, IEEE.

Crampton, J. W., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M. W. & Zook, M. (2013). Beyond the geotag: situating 'big data' and leveraging the potential of the geoweb. *Cartography and Geographic Information Science*, 40(2), 130-139.

Crandall, D. J., Backstrom, L., Huttenlocher, D. & Kleinberg, J. (2009). Mapping the World's Photos. In *Proceedings of the 18th ACM International Conference on World Wide Web*, 761-770. ACM

Da Rugna, J., Chareyron, G. & Branchet, B. (2012). Tourist behavior analysis through geotagged photographies: a method to identify the country of origin. In *Computational Intelligence and Informatics (CINTI)*, 13th International Symposium, 347-351, IEEE.

Dickinger, A., Scharl, A., Stern, H., Weichselbraun, A. & Wöber, K. (2008). Acquisition and Relevance of Geotagged Information in Tourism. In Oconnor, P., Hpken, W. & Gretzel, U. (eds), *Proceedings of Information and Communication Technologies in Tourism 2008* (pp 545-555). Springer.

Digital Stat Articles. (2016). *By the Numbers: 14 Interesting flickr Stats*. [Online]. Retrieved Jul 31 2016 from: http://expandedramblings.com/index.php/flickr-stats/

Donaire, J. A., Camprubí, R. & Gali, N. (2014). Tourist clusters from Flickr travel photography. *Tourism Management Perspectives*, 11, 26-33.

Ester, M., Kriegel, H. P., Sander, J. & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Data Mining and Knowledge Discovery*, 226-231.

Farzanyar, X. & Cercone, N. (2015). Trip Pattern Mining Using Large Scale Geo-tagged Photos. *In Proceedings of the International Conference on Computer and Information Science and Technology*, Ottawa, Canada, 113.

Forer, P. & Simmons, D. (2002). Serial Experiences: Monitoring, Modelling and Visualising the Free Independent Traveller in New Zealand at Multiple Scales with GIS. In Arnberger, A., Brandenburg, C. & Muhar, A. (eds). *Monitoring and Management of Visitor Flows in Recreational and Protected Areas* (pp. 173-180). Vienna: Institute of Landscape

Architecture and Landscape Management, Bodenkultur University.
Garcia-Palomares, J. C., Gultiérrez, J. & Minguez, C. (2015). Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS. *Applied Geography*, 63, 408-417.
Girardin, F., Calabrese, F., Dal Fiore, F., Ratti, C. & Blat, J. (2008). Digital Footprinting: Uncovering Tourists with User-Generated Content. *IEEE Pervasive computing*, 7(4), 36-43. IEEE
Girardin, F., Dal Fiore, F., Ratti, C. & Blat, J. (2008). Leveraging explicitly disclosed location information to understand tourist dynamics: a case study. *Journal of Location Based Services*, 2(1), 41-56.
Guo, L., Li, Z. & Sun, W. (2015). Understanding Travel Destination From Structured Tourism Blogs. In *Proceedings of 2015 Wuhan International Conference on e-Business*, 144-151.
Instagram. (2016a). *Press News*. [Online]. Retrieved 25 Jul 16 from: https://www.instagram.com/press/
Instagram. (2016). *Instagram Today: 500 Million Windows to the World.* [Online]. Retrieved Jul 31 2016 from: http://blog.instagram.com/post/146255204757/160621-news
Jiang, K., Wang, P. & Yu, N. (2011). ContextRank: Personalized Tourism Recommendation by Exploiting Context Information of Geotagged Web Photos. In *Proceedings of IEEE16th International Conference on Image and Graphics*, 931-937. IEEE
Jiang, K., Yin, H., Wang, P. & Yu, N. (2013). Learning from contextual information of geo-tagged web photos to rank personalized tourism attractions. *Neurocomputing*, 119, 17-25.
Kádár, B. & Gede, M. (2013). Where Do Tourists Go? Visualizing and Analysing the Spatial Distribution of Geotagged Photography. Cartographica: *The International Journal of Geographic Information and Geovisualization*, 48(2), 78-88.
Kalogerakis, E., Vesselova, O., Hays, J., Efros, A. A. & Hertzmann. A. (2009). Image Sequence Geolocation with Human Travel Priors. In *Proceedings of 12th IEEE International Conference on Computer Vision (ICCV)*, 253-360. IEEE
Kisilevich, S., Krstajic, M., Keim, D., Andrienko, N. & Andrienko, G. (2010). Event-based analysis of people's activities and behavior using Flickr and Panoramio geotagged photo collections. In *Proceedings of 14th IEEE International Conference Information Visualization*, 289-296. IEEE.
Kisilevich, S., Mansmann, F. & Keim, D. (2010). P-DBSCAN: A density based clustering algorithm for exploration and analysis of attractive areas using collections of geo-tagged photos. In *Proceedings of the 1st International Conference and Exhibition on Computing for Geospatial Research & Application*, 38-41, Bethesda, Maryland, USA.
Kurashima, T., Iwata, T., Irie, G. & Fujimura, K. (2013). Travel route recommendation using geotagged photos. *Knowledge and information systems*, 37(1), 37-60
Lau, G. & McKercher, B. (2006). Understanding tourist movement patterns in a destination: A GIS approach. *Tourism and Hospitality Research*, 7(1), 39-49
Law, R., Qi, S. & Buhalis, D. (2010). Progress in tourism management: A review of website evaluation in tourism research. *Tourism Management*, 31, 297-313.
Lee, I., Cai, G. & Lee, K. (2013). Mining Points-of-Interest Association Rules from Geo-tagged Photos. In *System Sciences, 46th Hawaii International Conference*, 1580-1588, IEEE.
Leung, R., Vu, H. Q., Rong, J. & Miao, Y. (2016). Tourists Visit and Photo Sharing Behavior Analysis: A Case Study of Hong Kong Temples. In Inversini, A., & Schegg, E. (eds), *Proceedings of Information and Communication Technologies in Tourism 2016*, 197-209.
Lew, A. & McKercher, B. (2006). Modeling Tourist Movements: A Local Destination Analysis. *Annals of Tourism Research*, 33(2), 403-423.
Li, X., Meng, F. & Uysal, M. (2008). Spatial Pattern of Tourist Flows among the Asia-Pacific Countries: An Examination over a Decade. *Asia Pacific Journal of Tourism Research*,

13(3), 229-243.
Majid, A., Chen, L., Chen, G., Mirza, H. T., Hussain, I. & Woodward, J. (2012). A context-aware personalized travel recommendation system based on geotagged social media data mining. *International Journal of Geographical Information Science*, 27(4), 662-684.
Mamei, M., Rosi, A. & Zambonelli, F. (2010). Automatic Analysis of Geotagged Photos for Intelligent Tourist Services. In *Proceedings of IEEE 16th International Conference on Intelligent Environments*, 146-151. IEEE.
McKercher, B. & Lau, G. (2008). Movement Patterns of Tourists within a Destination. *Tourism Geographies*. 10(3), 355-374.
Musante, M., Bojanic, D. & Zhang, J (2009). An Evaluation of Hotel Website Attribute Utilization and Effectiveness by Hotel Class. *Journal of Vacation Marketing*, 15(3), 203-215.
O'Neill, E., Kostakos, V., Kindberg, T., Penn, A., Fraser, D. S. & Jones, T. (2006). Instrumenting the city: Developing methods for observing and understanding the digital cityscape. In *Proceedings of International Conference on Ubiquitous Computing*, pp 315-332. Springer Berlin Heidelberg.
Okuyama, K. & Yanai, K. (2013). A Travel Planning System Based on Travel Trajectories Extracted from a Large Number of Geotagged Photos on the Web. *The era of interactive media*, 657-670, Springer New York.
Önder, I., Koerbitz, W. & Hubmann-Haidvogel, A. (2014). Tracing Tourist by Their Digital Footprints: The Case of Austria. *Journal of Travel Research*, 55(5), 566-573.
Shoval, N., McKercher, B., Ng, E. & Birenboim, A. (2011). Hotel Location and Tourist Activity in Cities. *Annals of Tourism Research*, 38(4), 1594-1612.
Sun, Y., Fan, H., Helbich, M. & Zipf, A. (2013). Analyzing Human Activities Through Volunteered Geographic Information: Using Flickr to Analyze Spatial and Temporal Pattern of Tourist Accommodation. In *Progress in Location-Based Services*, 57-69, Springer Berlin Heidelberg.
Vu, H. Q., Leung, R., Rong, J. & Miao, Y. (2016). Exploring Park Visitors' Activities in Hong Kong using Geotagged Photos. In Inversini, A., & Schegg, E. (eds), *Proceedings of Information and Communication Technologies in Tourism 2016*, 183-196.
Vu, H. Q., Li, G., Law, R. & Ye, B. H. (2015). Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos. *Tourism Management*, 46, 222-232.
Wu, C. L. & Carson, D. (2008). Spatial and Temporal Tourist Dispersal Analysis in Multiple Destination Travel. *Journal of Travel Research*, 46(3), 311-317.
Xia, J. C., Zeephongsekul, P. & Arrowsmith, C. (2009). Modelling spatio-temporal movement of tourists using finite Markov chains. *Mathematics and Computers In Simulation*, 79, 1544-1553.
Xu, Z., Chen, L. & Chen, G. (2015). Topic based context-aware travel recommendation method exploiting geotagged photos. *Neurocomputing*, 155, 99-107.
Zanker, M., Fuchs, M., Seebacher, A., Jessenitschnig, M., & Stromberger, M. (2009). An Automated Approach for Deriving Semantic Annotations of Tourism Products based on Geospatial Information. In: W. Höpken et al. (eds), *Processings of Information and Communication Technologies in Tourism 2009*, 211-221.
Zheng, Y. T., Zha, Z. J. & Chua, T. S. (2011). Research and applications on georeferenced multimedia: a survey. *Multimedia Tools Appl*, 51, 77-98.
Zheng, Y. T., Zha, Z. J. & Chua, T. S. (2012). Mining Travel Patterns from Geotagged Photos. *ACM Transactions on Intelligent Systems and Technology*, 3(3), 56-73.
Zheng, Y., Zhang, L., Xie, X. & Man, W. Y. (2009). Mining Interesting Locations and Travel Sequences from GPS Trajectories. In *Proceedings of the 18th ACM International Conference on World Wide Web*, 791-800. ACM
Zipf, A. & Malaka, R. (2001). Developing Location Based Services for Tourism- The Service Providers View Information and Communication Technologies in Tourism. In Sheldon, P. J., Wöber, K. W. & Fesenmaier, D. R. (eds), *Proceedings of Information and Communication Technologies in Tourism 2001* (pp. 83-92). Springer.