https://doi.org/10.1016/j.tourman.2014.07.003

EXPLORING THE TRAVEL BEHAVIORS OF INBOUND TOURISTS TO HONG KONG USING GEOTAGGED PHOTOS

ABSTRACT

Insight into tourist travel behaviors is crucial for managers engaged in strategic planning and decision making to create a sustainable tourism industry. However, they continue to face significant challenges in fully capturing and understanding the behavior of international tourists. The challenges are primarily due to the inefficient data collection approaches currently in use. In this paper, we present a new approach to this task by exploiting the socially generated and user-contributed geotagged photos now made publicly available on the Internet. Our case study focuses on Hong Kong inbound tourism using 29,443 photos collected from 2,100 tourists. We demonstrate how a dataset constructed from such geotagged photos can help address such challenges as well as provide useful practical implications for destination development, transportation planning, and impact management. This study has the potential to benefit tourism researchers worldwide from better understanding travel behavior and developing sustainable tourism industries.

Keywords: Data Mining, Geotagged Photo, Travel Behavior, Global Positioning System.

1. INTRODUCTION

The tourism industry plays a major role in economic development for many countries and regions, helping to improve the livelihoods of residents (Ashley, Brine, Lehr, & Wilde, 2007). Both public and private sector tourism operations are highly dependent on sustainable development planning, which aims to create appropriate employment, maintain the natural environment, and deliver a high-quality visitor experience. Due to the perishable nature of the tourism industry, an accurate understanding of travel behaviors is crucial (Edwards, Griffin, Hayllar, Dickson, & Schweinsberg, 2009).

Tourism managers have long been seeking insights into travel behavior for the purposes of destination management, product development, and attraction marketing (Li, Meng, & Uysal, 2008). For instance, good transportation planning enables tourism providers to meet tourists' needs and coordinate their travel with local transportation flows. However, it requires planners to know about tourist preferences, daily itineraries, and the factors that influence them. Movement information can also be used to identify bottlenecks and unnecessary barriers in the flow between accommodation and attractions, or any other tourist destinations (Prideaux, 2000). In tourism location development, knowledge about travelers' location preferences can be used to redefine existing attractions, plan new ones, and market both more effectively (Lew & McKercher, 2006). The identification of the actual routes taken by tourists during their trips can help define the boundaries of districts and nodes, as well as the most appropriate gateways. This information can be used to develop new attractions and products along common routes as well as in district and destination nodes (Chancellor, 2012). In terms of impact management, it is important for tourism managers to identify the time and space characteristics of the routes and destinations which tourists visit most frequently in order to develop appropriate plans to prevent capacity overload, which has the potential for negative social, environmental, or cultural impacts (Lew & McKercher, 2006).

Some prior research focuses on developing techniques for analyzing tourist travel behaviors. For instance, Xia, Spilsbury, Ciesielski, Arrowsmith, and Wright (2010) introduce a method for tourist market segmentation based on dominant movement patterns in a case study of Phillip Island, Australia (Xia et al., 2010). Deng and Athanasopoulos (2011) utilize an anisotropic dynamic spatial lag panel Origin-Destination (OD) travel flow model to understand Australian domestic and international inbound travel patterns. Leung, Wang, Wu, Bai, Stahura, and Xie (2012) use content and social network analyses to examine online trip diaries and map overseas tourist movement patterns during the Beijing Olympics. Such growing academic attention being paid to travel behavior indicates that this is an interesting topic which is important in the planning and decision-making processes of tourism managers.

Despite these efforts, both researchers and managers still find it difficult to fully capture and understand the travel behavior of international travelers. The difficulties are due to the following barriers:

Information Capture: Survey and opinion polls are popular methods to collect travel data from tourists (Lew & McKercher, 2006; Asakura & Iryo, 2007; McKercher & Lau, 2008). However, these approaches are usually time consuming and limited in terms of the number of responses as well as the scale of the information captured. The data are normally unable to reflect actual travel patterns closely (Zheng, Zha, & Chua, 2011). There is considerable need for a more efficient method of capturing comprehensive travel behavior data from tourists.

Travel Preferences: It is a natural assumption that travel behavior will vary among different groups of tourists (Batra, 2009). For instance, tourists from different countries may have different preferences for their length of stay and the attractions they want to visit, and this will, in turn, affect their travel activities (Leung et al., 2012). There have been limited attempts to analyze travel behaviors in a way which takes these preferences into consideration.

Recently, advances in multimedia and mobile technologies have allowedlarge volumes of user-generated data, such as travel photos, to be created and shared. Many photo-capturing devices, like smartphones and tablets, now have built-in global positioning systems (GPS) technology which enable geographical information (latitude and longitude coordinates) to be stored as metadata with each photo a user takes. Those geotagged photos, with time and geographical information embedded, allow the spatial-temporal movement trajectories of the user to be inferred. One of the most popular online resources for people to share their travel experiences by uploading photos is Flickr (*www.flickr.com*). Containing millions of geotagged photos, Flickr is a rich data source for mining tourist travel patterns (Lee, Cai, & Lee, 2013; Zheng, Zha, & Chua, 2012). Research into the nature and applications of georeferenced multimedia is an emerging topic in Computer Science (Zheng et al., 2011), with many attempts having recently been made to develop tools and techniques for data mining (Yanai, Yaegashi, & Qiu, 2009; Yanai, Kawakubo, & Qiu, 2009; Kalogerakis, Vesselova, Hays, Efros, & Hertzmann, 2009). It will, therefore, be advantageous for tourism researchers to adopt these advanced technological developments when studying travel behavior.

In this paper, we attempt to address the shortcomings in the existing literature on tourist travel behavior by utilizing the geotagged photos that are available on social networking sites. Firstly, we present a method for constructing the data collection process which captures travel information from geotagged photos on Flickr. Then, we describe two relatively new data-mining techniques for processing and analyzing these data to yield information about the travel behaviors of tourists. This case study focuses on inbound tourism in Hong Kong, a major tourism destination in the Asia Pacific region. The study aims to discover the locations in which tourists are most interested and the routes they take when visiting Hong Kong. This method and the associated findings are of potential benefits to tourism researchers worldwide who are interested in travel behavior, and will also help tourism managers in Hong Kong and likely elsewhere to construct plans for sustainable development.

Having thus set the context for undertaking this work, the rest of the paper is organized as follows. In Section 2, we review existing work, which uses global positioning technology such as Geographic Information Systems (GIS) to analyze travel behavior. We then provide a summary of the techniques developed for processing geotagged photos in travel analysis, and define our research objectives. Section 3 presents the methods for extracting and analyzing geotagged photos. We demonstrate their effectiveness in a case study presented in Section 4. Finally, Section 5 concludes our paper and offers future research directions.

2. RELATED WORK

2.1. Travel Behavior Analysis using Geographical Information

Since GIS was firstly introduced, many studies have used it to explore the movement patterns of tourists (Lau & McKercher, 2006). For instance, Wu and Carson (2008) use GIS to identify multiple destination travel behavior for travelers in South

Australia. McKercher and Lau (2008) apply it to examine the movements of tourists within an urban destination in Hong Kong, and identify 78 discrete movement patterns. Other researchers use GPS to explore tourists' experiences and mobility (Zakrisson & Zillinger, 2012), or to chart visitor movement patterns in natural recreational areas (Orellana, Bregt, Ligtenberg, & Wachowicz, 2012). Additionally, McKercher, Shoval, Amit, and Birenboim (2011) compare and contrast travel behaviors between first-time and repeat visitors in Hong Kong using both GPS and GIS.

Attempts have also been made to study tourist travel behaviors using these methods. For instance, Hwang, Gretzel, and Fesenmaier (2006) examine international tourists' multi-city trip patterns within the United States. A large-scale study of the spatial pattern of tourist flows among the selected Asia-Pacific countries over a 10-year period is presented by Li, Meng, and Uysal (2008). Asakura and Iryo (2007) look at tourist travel behavior in Kobe, Japan using tracking data collected via mobile instruments. Smallwood, Beckley, and Moore (2012) explore the movement patterns of visitors traveling within protected areas using various modes of travel through a case study of the Ningaloo Marine Park, in north-western Australia. Masiero and Zoltan (2013) analyze the factors influencing both the spatial extent of the destination visited and the selected transport mode.

2.2. Geotagged Photos for Travel Analysis

Advances in information technology, especially the introduction of GPS technology and mobile photo capturing devices, now allow for geographical information to be available and stored in a photo tag. Computer Science researchers have therefore started to develop techniques for using these geotagged photos given that they are now made publicly available on many photo-sharing sites such as Flickr and Paranomia (Kennedy, Naaman, Ahern, Nair, & Rattenbury, 2007). For instance, Jaffe, Naaman, Tassa, and Davis (2006) attempt to develop a technique that automatically summarizes and visualizes a large collection of georeferenced photographs. Also, Snavely, Seitz, and Szeliski (2008) propose an approach called Photo Tourism which enables the three-dimensional (3D) modeling and reconstruction of world sites from user-generated photos available on Google Maps. Quack, Leibe, and Gool (2008) mine events from community photo collections on a worldwide scale using geotagged photos on Flickr. Hays and Efros (2008) propose a method for automatically estimating geographical location for photos without geotags, a technique which has the potential to generate even more data for geographic-related research. Kalogerakis et al. (2009) introduce another way to infer geographic location using sequences of time-stamped photos.

Different attempts have been devoted to develop intelligent tourism recommendation system using geotagged photos. Namely, Kurashima, Iwata, Irie, and Fujimura (2010) propose a probabilistic behavior model for recommending travel path based on the past travel information of tourists on Flickr. Okuyama and Yanai (2013) make use of actual travel paths extracted from a large number of online geottaged photos to develop a travel planning system. Majida and colleagues propose methods for tourism location recommendation that are relevant to user preference and travel context (Majida et al. 2012, Majida et al. 2013). Shi, Serdyukov, Hanjalic and Larson (2013) propose a method for personalized landmark recommendation. Yin (2012) facilitates tourist trip planning by a travel path search system using geotagged photos. Similarly, Li (2013) proposes a travel planning system to design multi-day and multi-stay travel plans.

Due to the demand for knowledge about travel behavior, some methods have also been proposed to mine travel patterns from geotagged photos. Mamei, Rosi, and Zambonelli (2010) put forward a technique for the automatic analysis of geotagged photos for Intelligent Tourist Services. Rugna, Chareyron, and Branchet (2012) show that the country of origin of photo owners can be identified based on the photos they post online. Other researchers have mined tourists' travel patterns such as traffic flow between different locations within a destination (Zheng, Zha, & Chua, 2011; Zheng, Zha, & Chua, 2012). Lastly, Lee, Cai, and Lee (2013) use clustering and association rule techniques to identify tourist attractions and their associated movement patterns.

2.3. Research Objectives

Understanding travel behavior is important for tourism practitioners seeking to predict the market or to provide suitable trip recommendations. The behavior pattern discovery process can be supported by different types of data, which have traditionally been collected using surveys, questionnaires, and opinion polls. These approaches are usually time consuming and unable to accurately reflect the travel behavior of tourists. Even with the support of technology such as GIS, participants are usually required to carry mobile devices for recording their movements while traveling. The data collected are, therefore, limited in terms of the number of responses or the scale of the geographical areas included. When the collected data are not reliable or adequate, it is necessary to resort to data that are available online. The development of Computer Science research on geotagged photos presents a new direction for capturing tourist movement data in an efficient and timely manner. With millions of geotagged and time-stamped photos available on a global scale, photo-sharing sites are potential gold mines for researchers seeking to extract data and study travel behavior.

Although prior efforts have been made in utilizing geotagged photos to support travel planning and recommendation of tourists, limited work focused on analyzing travel preference of tourists to support for strategic planning and decision making for tourism managers. Some examples of potential questions ask by managers are "*What are the preferred locations for each group of tourists when visiting a tourist destination*?", "*When do they prefer to visit such locations?*" and "*What travel routes are they likely to take when visiting different locations*?". Unfortunately, none of the existing works attempted to discover such insightful knowledge. As such, it is still a challenging task for tourism managers to develop effective destination management and transportation plans to accommodate the increasing demand of international tourists.

This paper aims to address such challenges by exploiting the socially generated and user-contributed geotagged photos that are publicly available on the Internet. Our specific objectives can be defined as follows:

- to introduce a framework for effectively extracting geographical information from geotagged photos posted online and using this to analyze tourist travel behavior;
- to identify the attractions of interest to tourists with different profiles who are visiting a tourist destination such as Hong Kong; and
- to identify the travel behaviors of tourists, travel route and travel time, in order to support traffic management and the product development of tourism businesses.

3. METHODOLOGY

In this section we firstly outline the data-collection process involved in extracting geotagged photos from Flickr. Then, we discuss a number of specific challenges which need to be addressed in order to analyze this kind of data. While the massive volume of shared photos available on Flickr is a comprehensive resource for studying travel behaviors, the data can be noisy or misleading, especially when many photos have been taken in transit rather than at the attractions themselves. In addition, the sequence of photos, in some situations where tourists move from one attraction to another, implies mobility information, which is a key to inferring personal activities, intuitions, and goals. However, photos are static media, while travel behaviors are dynamic. The photos should therefore be transformed into a suitable representation for the travel analysis task. Therefore, we introduce two relatively new data mining techniques, based on density clustering and the Markov Chain, to tackle these issues.

3.1. Geographic Data Extraction from Geotagged Photos

Geotagged photos are available for public view through web applications such as Flickr, but are not directly downloadable. They must be accessed via Flickr's Application Programming Interface (API), documentation for which is available at *www.flickr.com/services/api*. One of the challenges in extracting data from this source is that it is impossible to identify individual owners whose photos should be downloaded. Therefore, we propose to address this task by firstly searching for geotagged photos taken in the environs of Hong Kong and then extracting the owners' information in order to retrieve the photo information.

We define a bounding box for the region from which we want to extract geotagged photos. Let $x_{min}, y_{min}, x_{max}, y_{max}$ represent its geographical coordinates for the minimum longitude, minimum latitude, maximum longitude, and maximum latitude, respectively. A set of *M* seed photos are randomly extracted using Flickr's photo search function with the bounding box. Those photos are returned from the search engine together with their owner's identification number (ownerID). It is possible for multiple photos to be uploaded by the same owner. We then construct a list of owners $< o_1, o_2, \ldots, o_n >$, who have uploaded photos taken in the target area. Their ownerIDs are used to retrieve the user demographic information such as country of origin using the Flickr user search function. This helps us differentiate Hong Kong residents from international travelers. Based on the ownerIDs of the targeted tourist group and the bounding box coordinates, we retrieve the entire collection of their shared photos to ensure that their travel activities are captured completely. It should be noted that in addition to the spatial information, each photo is usually stamped with the time and date on which it was taken. This temporal information is also useful in constructing a movement trajectory for travel pattern analysis. The Flickr photo search function also allows the user to specify temporal information to limit the search. If values for the minimum (t_{min}) and maximum time (t_{max}) are provided, only photos taken between that period are returned.

3.2. Density Clustering for Tourism Attraction Identification

Geographical location is one of the most important memory cues for recalling previous trips. Many photos taken by the tourist may be related to a single day's activity, and some may also be taken on the journey there. According to Zheng, Zha, and Chua (2012), a tourism attraction is defined as a spatial extent within a geographical location through which considerable volumes of tourist movement trajectories pass, or where many tourists visit and take photographs. Previous work proposes a Density-based spatial clustering of applications with noise (DBSCAN) (Ester, Kriegel, Sander, & Xu, 1996), to perform this task (Kisilevich, Keim, &

Rokach, 2010; Zheng et al., 2012; Lee et al., 2013). However, a shortcoming of this technique is that generic photo points are treated as being of equivalent importance, whereas the owners of photos are the main factor determining the importance of a cluster in the geotagged data analysis context. Thus, we use a new and more advanced version of DBSCAN, namely P-DBSCAN (Kisilevich, Mansmann, & Keim, 2010), to perform this task. It was specifically developed for clustering geotagged photos by taking into account information about the photo owners in this computation. The details of this technique are described below.

Suppose *D* is a collection of geotagged photos. Each photo is a point *p* referenced by a value pair $< x_p, y_p >$ for longitude and latitude coordinates. The distance between two photo points *p* and *q* is denoted by $Dist(p,q)$. The neighborhood of a photo point *p*, denoted by $N_\delta(p)$, is defined by:

$$(3.1) N_\delta(p) = (q \in D, Owner(q) \neq Owner(p) | Dist(p,q) \leq \delta)$$

where $Owner(q) = (o_i \in O)$ is an ownership function. In other words, a photo *q* is in the neighborhood of another photo *p* if they belong to different users and the location of photo *q* is within a neighborhood radius δ from photo *p*. Let $NeighborOwner(p)$ be the owner number of the neighbor photos $N_\delta(p)$, and λ be the owner number threshold. A photo *p* is called a core photo if its neighbor photos belong to at least a minimum number of owners ($NeighborOwner(p) \geq \lambda$).

The clustering process of P-DBSCAN starts with a set of unprocessed photos $P = \{p_1, p_2, \ldots\}$. For each photo p_i, if it is not a core photo it is marked as noise. Otherwise, it is assigned to a cluster c_j, and all of its neighbors $N_\delta(p_i)$ are put into a queue for further processing. Each photo $p_{ij} \in N_\delta(p_i)$ is then processed and assigned to the current cluster c_j until the queue is empty. This process repeats for the rest of the unprocessed photos in *P*. After obtaining clusters of photos, we examine their geographical coordinates to determine the name and spatial extent of the tourist attractions captured, which we define as the Area of Interest (AOI). Here, the values of δ and λ are determined based on the scale of specific applications. If the region under study is at the macro level such as a country, then the AOI may be as big as a city. Large values can be assigned to δ and λ. If the region is at a micro level such as a park, AOI may be on a much smaller scale, and δ and λ take smaller values. The implementation of this technique is discussed further in our case study in Section 4.2.1.

3.3. Using the Markov Chain for Travel Pattern Mining

Tourism managers are interested in not only where tourists travel, but also how they get there. This section describes a method, based on the Markov Chain, for mining such travel patterns and routes taken by tourists between main attractions (Xia, Zeephongsekul, & Arrowsmith, 2009).

Let $A = \{a_1, a_2, \dots, a_m\}$ as denote an AOI as identified by P-DBSCAN. The travel trajectory of a tourist from time $t = 1$ to $t = k$ is defined as $T = \{a_i^{t_1}, a_i^{t_2}, \dots, a_i^{t_k}\}$ and the probability of tourist moving to a location a_i is computed by:

$$(3.2) \qquad P\left(a_i^{t_n} \middle| a_i^{t_{n-1}}, a_i^{t_{n-2}}, \dots, a_i^{t_0}\right\} = P(a_i^{t_n} | a_i^{t_{n-1}})$$

Equation 3.2 implies the conditional independent assumption Markov Chain, which can be used to model how tourists flow from one location to another. The transition probability of a tourist moving from attraction a_i to attraction a_j ($i \neq j$), between time $t = n$ to $t = n + 1$, is computed by:

$$(3.3) \qquad P\left(a_j(n+1) \middle| a_j(n)\right) = \frac{P(a_j(n+1) \cap a_i(n))}{P(a_i(n))}$$

where the numerator is the probability of the tourist visiting both locations a_i and a_j, with a_i being visited first followed by a_j. The event $a_i(n)$ can be expressed as a combination of the mutually exclusive events $\bigcup_{j=1}^{k}(a_i(n) \cap a_j(n+1))$, thus the denominator is computed by:

$$(3.4) \qquad P\left(a_j(n)\right) = \sum_{j=1}^{m} P(a_i(n) \cap a_j(n+1))$$

The transition probability for all possible travel routes among different attractions can be computed and presented in a one-step transition probability matrix **P**:

$$\mathbf{P} = \begin{pmatrix} 0 & P\left(a_2(n+1)\middle|a_1(n)\right) & \dots & P\left(a_m(n+1)\middle|a_1(n)\right) \\ P\left(a_1(n+1)\middle|a_2(n)\right) & 0 & \dots & P\left(a_m(n+1)\middle|a_2(n)\right) \\ \vdots & \vdots & \ddots & \vdots \\ P\left(a_1(n+1)\middle|a_m(n)\right) & P\left(a_2(n+1)\middle|a_m(n)\right) & \dots & 0 \end{pmatrix}$$

where the value of each entry $P_{ij} \in \mathbf{P}$ reflects how likely tourists are to travel from attraction a_i to attraction a_j. A demonstration of this technique for travel path analysis is provided in Section 4.2.2.

4. CASE STUDY

This section presents a case study of the travel behavior of Hong Kong inbound tourists using geotagged photos. It starts with a description of our data collection, followed by the analysis and results. We then set out the practical implications of the findings to support Hong Kong tourism managers in improving their performance.

4.1. Data Collection

The dataset used in this paper is collected from Flickr following the method described in Section 3.1. Since our focus is on Hong Kong inbound travelers, we firstly search for seed photos by providing a bounding box for the Hong Kong area in the photo search function, as shown in Table 1 .

***** Please place Table 1 here *****

The coordinates of the bounding box are in a decimal degree form, which can be determined manually by using Google Maps (*maps.google.com*). We select these values to ensure that the bounding box covers the geographical area of Hong Kong entirely. We only retrieve photos taken from 2011 to August 28, 2013 in order to ensure the dataset is up to date. Accordingly, only the earliest photo taken date parameter (t_{min}) is provided.

From the seed photos, a list of owner identification numbers is extracted. Those indicating the owner's origin is not from Hong Kong are treated as labeling inbound travelers and hence used to retrieve their entire photo collection. The metadata tags of each photo contain photoID, OwnerID, owner location of origin, date and time taken, and GPS location (longitude, latitude). The GPS location indicates where the photo was taken, which reflects the tourist's travel footprint. If a user takes many photos at the same location, we keep only one of those and discard the others.

Since people from different countries tend to have different travel preferences (Leung et al., 2012), we group tourists based on their location of origin to examine the behavior of travelers with different profiles. We observe that the majority of tourists visiting Hong Kong are from countries in the Asia-Pacific, European, and North American regions. The European and North American tourists are mainly from English-speaking countries such as the U.S., Canada, and the United Kingdom. We can assume that they are more similar to one another in terms of background than tourists from Asia. Therefore, we group tourists from Europe and North America together and refer to them henceforward as Western tourists, and similarly group tourists from Asia together and label this group as Asian tourists. We arrive at a dataset containing 29,443 photos collected from 2,100 Hong Kong inbound tourists as shown in Table 2.

***** Please place Table 2 here *****

The following sections demonstrate how the spatial and temporal information in geotagged photos can help to address the challenges of travel analysis. More specifically, we perform three major analyses to examine the behavior of inbound tourists in Hong Kong, namely AOI Identification, Tourist Movement Analysis, and Tourist Activity Analysis.

4.2. Findings and Analysis

4.2.1 *AOI Identification*

This section presents our findings on the identification of the AOI for Hong Kong inbound travelers to support managers in developing their tourism locations. Firstly, we inspect our data collection by visualizing the photo locations by incorporating their GPS information into Google Earth software, as shown in Figure 1. Each photo is indicated by a white dot on the satellite image.

***** Please place Figure 1 here *****

The geotagged photos appear all over the regions of Hong Kong, which indicate that the raw geographical information is noisy and may be misleading if directly used for analysis. We therefore apply the P-DBSCAN clustering algorithm to remove the noise and identify the locations in which tourists are most interested. An advantage of this density clustering approach is that users are not required to provide the number of clusters in the algorithm; the clusters are automatically determined based on the density of the points and number of visitors. We set the neighborhood radius value to $\delta = 0.002$, which is equivalent to approximately 150 meters. The minimum owner (λ) is set to a value of 10% of the total number of tourists.

A total of seven clusters are found, indicating there are seven areas of interest for Hong Kong inbound tourists, as shown in Figure 2. Four of them are in the Hong Kong metropolitan area, including Center Mong Kok, the Tsim Sha Tsui area, Hong Kong Central, and Times Square Towers. Two areas of interest are found in the countryside, namely the Peak Tower and the Tian Tan Buddha Statue. Hong Kong International Airport is also an AOI, as indicated by the many photos taken within its terminals. These findings are consistent with the fact that the metropolitan area of Hong Kong is well known as a major destination for visitors. P-DBSCAN is shown to be effective in removing noise from geotagged photos and identifying AOI to inbound tourists.

***** Please place Figure 2 here *****

In order to identify the location preferences of different groups of tourists, we also take their profile into account. The clusters are examined with respect to the two groups, Asian and Western tourists, and the proportion of holiday makers visiting each area is used as a measure of its popularity, as shown in Table 3. The AOI are presented in a ranking order from the most to the least popular.

- Hong Kong Central and the Tsim Sha Tsui area are the two most popular destinations for tourists in both groups, as shown by their ranking as first and second. Hong Kong Central is the central business district and the “heart” of Hong Kong, while the Tsim Sha Tsui area is a major tourist hub in the metropolitan area with many shops, restaurants, and a good view of Victoria Harbor.
- Western tourists show more interest in the Peak Tower (third) and Center Mong Kok (fourth) than in Times Square Towers (fifth) and Hong Kong International Airport (sixth). Asian tourists show the opposite trend, as indicated by their ranking of Times Square Towers and Hong Kong International Airport as more popular than the Peak Tower and Center Mong Kok.
- It is interesting to see that the Tian Tan Buddha Statue is identified as an AOI for Western but not Asian tourists. A possible explanation is that most Asian countries are likely to have somewhat similar cultural or religious sites such as temples, shrines, or Buddha statues. Asian visitors may, therefore, not be interested in seeing similar things when visiting Hong Kong. On the other hand, Western people are likely to be interested in exploring Asian culture, and the Tian Tan Buddha Statue is one such symbol.

***** Please place Table 3 here *****

4.2.2 *Tourist Movement Analysis*

Knowledge of tourist movements is important in transportation planning and traffic management, especially for metropolitan areas with a high density of traffic.

An overview of how tourists move from one area to another and what routes they prefer to take will be beneficial from developing appropriate travel management plans. To address such challenges, we therefore focus this part of the analysis on tourist movements between the AOI in the metropolitan district. To represent travel behavior, we construct the travel trajectories of tourists corresponding to spatial and temporal sequence. The GPS information from the photos is concatenated according to the time taken on a daily basis. Figure 3 represents the movement trajectory of tourists as generated from their geotagged photos between five areas of interest in Hong Kong city center. Each trajectory is displayed as a thin white line. Regions and paths with many lines passing through them indicate a high density of movement. It appears that tourists tend to travel between areas that are close to each other, such as from Center Mong Kok to Tsim Sha Tsui, Tsim Sha Tsui to Hong Kong Central, or Hong Kong Central to Times Square Towers.

***** Please place Figure 3 here *****

To gain more detailed knowledge of the travel flow and directions of different tourist groups, we analyze tourist trajectories between AOI by applying the Markov Chain technique. A one-step transition probability matrix is computed for each travel group as described in Section 3.3. A high value for $P(a_j|a_i)$ suggests that tourists are likely to visit a_j right after a_i. To make interpretation easier, we display in Figure 4 the transition probability and transition direction for Asian and Western tourists. Only $P(a_j|a_i) \geq 0.3$ are shown. Our findings can be summarized as follows:

- Figure 4a shows that Asian tourists are likely to flow to Hong Kong Central from the surrounding areas, as shown by the dark arrows and a high probability value of around 0.6. Tourists in Center Mong Kok tend to visit Tsim Sha Tsui next (dark arrow with a probability of above 0.5), while some chose to travel on directly to Hong Kong Central. Considerable tourist flow is found from Hong Kong Central and the Times Square Towers to Tsim Sha Tsui with probabilities of 0.44 and 0.307, respectively.

- Figure 4b shows some differences between Western and Asian tourists. Namely, Tourists from Asian countries tend to visit Hong Kong Central right after Center Mong Kok (dark arrow with a probability of 0.678), while some travel in the opposite direction. Western tourists are more likely to visit Tsim Sha Tsui (probability of 0.579) than Hong Kong Central (probability of 0.316) after Times Square Towers. Some tourists visited Tsim Sha Tsui immediately after the Peak Tower.

***** Please place Figure 4 here *****

We further perform z-tests to verify the statistical significance for the common flows between Asian and Western tourists at p-*value* <0.05, as shown in Table 4. Base on Table 4 and Figure 4, the flows of Western tourists from Center Mong Kok to Hong Kong Central and from Times Square Towers to Tsim Sha Tsui are significantly higher than the flows of Asian tourists (flows F_3 and F_6). However, Asian tourists are more likely to flow from Times Square Towers to Hong Kong Central than Western tourists (flow F_4). The other flows of Asian and Western tourists are not significantly different from each other as shown by *p-value >0.05* (flows F_1, F_2 and F_5).

***** Please place Table 4 here *****

Although Figure 4 indicates the travel flow of tourists from one location to another, it is not necessarily the actual routes which they take. Such routes can be revealed by analyzing the photos taken while traveling between AOI. To demonstrate this, we inspect the common travel routes taken by tourists between more widely separated areas such as Center Mong Kok and Times Square Towers. Firstly, we identify tourists who visited both locations on a single day. Then, we retrieve and plot the photos taken while traveling from one place to another in Figure 5. The white dots denote photos which represent the footprint of tourists during their trip. It can be seen that most tourists traveled to and from these locations via Tsim Sha Tsui and Hong Kong Central. To be more specific, they travel along Nathan Road between Centre Mong Kok and Tsim Sha Tsui. Direct ferry services are used to travel between Tsim Sha Tsui and Hong Kong Central. To go from Hong Kong Central and Times Square Towers tourists use Hennessy Road. We also note an interesting finding about tourist preferences. There is a direct ferry service which departs from Wan Chai, which is near Times Square Towers and goes to Tsim Sha Tsui, but most tourists did not choose this option and preferred to travel to Tsim Sha Tsui via the ferry line leaving from Hong Kong Central.

***** Please place Figure 5 here *****

4.2.3 *Tourist Activity Analysis*

Tourism managers are not only interested in knowing where and how tourists travel, but also when they visit places. This is crucial in transportation planning and destination management, to avoid problems of overload when too many people visit the same place in a short time frame. To address this, we also examine the visiting pattern of tourists based on the spatial and temporal information in their photos. More specifically, we compute the probability of tourists appearing at particular AOI over a 24-hour period, as shown in Figure 6. The horizontal axis represents time, and the vertical axis represents probability. Here, only six AOI common to both Asian and Western tourists are included, because these are the locations which receive many visits. We summarize the findings as follows:

- Figure 6a shows the presence of tourists at Hong Kong International Airport on a given day. Asian tourists appear to be arriving or departing during the daytime (10:00-17:00), whereas Westerners may travel at any time.
- Figure 6b shows that tourists in both groups are likely to visit Hong Kong Central between 11:00 and 16:00. In particular, the peak time for Asians is at noon, while the busiest time for Westerners is 15:00.
- Tourists are likely to visit Tsim Sha Tsui area in the late afternoon and evening, as shown in Figure 6c. Asian tourists are most likely to be there at 17:00 and 21:00 hours, While Western tourists visit most often at around 20:00. Similarly, the busy time for Times Square Towers is in the afternoon, as shown in Figure 6d.
- Similar visiting patterns to Center Mong Kok are found for both Asian and Western visitors, as shown in Figure 6e. Both groups demonstrate a peak visiting time at this location of 16:00. They also have relatively similar visiting patterns for the Peak Tower as shown in Figure 6f, where the peak for Asian tourists is 19:00.

***** Please place Figure 6 here *****

4.3 Discussion

The analysis of AOI in Section 4.2.1 highlights some differences in travel preferences for different tourist groups. Different travel packages and tour routes can be developed accordingly to meet the needs of inbound tourists. For instances, Western, but not Asian, tourists show an interest in Tian Tan Buddha Statue. Tourism managers can thus develop marketing plans and transportation arrangement to promote this attraction to Western visitors. The popularity ranking given in Table 3 can also help tourism developers to design local tour recommendations for tourists. Tours to Times Square Towers can be recommended to Asian visitors before the Peak Tower and Center Mong Kok, whereas tours to the Peak Tower and Center Mong Kok can be recommended to Westerners in preference to Times Square Towers.

The analysis of tourist movement in Section 4.2.2 can also offer some practical implication for traffic management. For example, the traffic flow captured in Figure 4 suggests that additional transportation to Hong Kong Central could be arranged for Asian visitors, as they are likely to visit this location from surrounding areas. On the other hand, more direct transportation means can be arranged for Western tourists to travel from Center Mong Kok to Hong Kong Central, or from Times Square Towers to Tsim Sha Tsui area. In addition, Figure 5 shows the current travel routes taken by tourists between Center Mong Kok and Times Square Towers. Most people tend to use the ferry service from Hong Kong Central to Tsim Sha Tsui, although there is actually a direct ferry service leaving from Wan Chai, which is near the Times Square Towers. Traffic managers can thus develop appropriate management plans to encourage tourists to use the direct service. This can reduce the traffic in Causeway Bay and the ferry line between Hong Kong Central and Tsim Sha Tsui, helping to avoid traffic congestion in the future as more visitors will travel to Hong Kong.

The tourist activity analysis presented in Section 4.2.3 reflects the visiting patterns of tourists to a particular AOI. The knowledge of where tourists are likely to be at different times of the day can support destination promotion and management. For instance, supporting service and transportation can be arranged at the Hong Kong International Airport, where many tourists are arriving and departing (Figure 6a). Different travel itineraries can be developed to avoid overcrowding in Hong Kong Central around noon (Figure 6b), or at Center Mong Kok in the afternoon (Figure 6e).

5. CONCLUSIONS

Knowledge of travel behavior is crucial to help tourism managers construct strategic plans and make decisions that will create a sustainable tourism industry. In spite of existing research efforts, managers still face significant challenges in gaining insight into tourist travel behavior due to the limitations of existing approaches on data collection such as surveys and polls. Such methods are unable to capture travel behavior comprehensively, nor to accurately reflect travel patterns. Travel behaviors vary among tourists with different profiles and so far only limited attempts have been made to take the preferences of different groups into consideration. An efficient method of comprehensively capturing the travel behavior patterns of tourists is therefore required.

To address these challenges, we approached the task of mapping tourist travel behavior by exploiting the socially generated and user-contributed geotagged photos that are publicly available on the Internet. We presented a method for constructing a travel dataset from geotagged photos on Flickr, one of the most popular websites for sharing photos. A dataset containing thousands of photos with temporal and geographic information attached enables us to capture the movement trajectories of tourists on a larger scale. In addition, we introduced two techniques, P-DBSCAN and the Markov Chain, to mine travel behavior patterns from this dataset. We demonstrated the effectiveness of these approaches by using the case study of Hong Kong inbound tourism and discovered the locations of interest to travelers, their travel patterns, and their daily activities. Practical implications are offered to support Hong Kong tourism managers in destination development, transportation planning, and impact management.

Since the collection of data from geotagged photo sets is not limited by factors of time and space, our future work will focus on analyzing tourist travel behavior on a larger scale (international or global). This will enable data from larger numbers of tourists across many years to be considered in order to model changes and trends in travel behavior for different tourism markets. Besides the spatial and temporal information, the online travel photos can provide the textual and visual information. The information would have great potential for providing insight into travelers' behavior, which can then be further explored in further research. Moreover, the analysis can be carried out with more detailed information about tourist profiles such as gender and age. Depending on practical applications, future work can reveal deeper insight into tourist behaviors.

REFERENCES

Asakura, Y., & Iryo, T. (2007). Analysis of tourist behaviour based on the tracking data collected using a mobile communication instrument. *Transportation Research Part A, 41*(7), 684-690.

Ashley, C., Brine, P. D., Lehr, A., & Wilde, H. (2007). *The role of the tourism sector in expanding economic opportunity.* Corporate Social Responsibility Initiative Report 23, Cambridge, MA: Kennedy School of Government, Harvard University.

Batra, A. (2009). Senior pleasure tourists: Examination of their demography, travel experience, and travel behavior upon visiting the Bangkok metropolis. *International Journal of Hospitality & Tourism Administration, 10(3),* 197-212.

Chancellor, H. C. (2012). Applying travel pattern data to destination development and marketing decisions. *Tourism Planning and Development, 9*(3), 321-332.

Deng, M., & Athanasopoulos, G. (2011). Modelling Australian domestic and international inbound travel: A spatial-temporal approach. *Tourism Management, 32(5)*, 1075-1084.

Edwards, D., Griffin, T., Hayllar, B., Dickson, T., & Schweinsberg, S. (2009). Understanding tourist experience and behaviour in cities: An Australian case study. Technical report. *Sustainable Tourism* Available online at: http://www.sustainabletourismonline.com/31/destination-access/understanding-tourist-experiences-and-behaviour-in-cities-an-australian-case-study.

Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density based algorithm for discovering clusters in large spatial database with noise. In *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining* (pp. 226-231), Portland, Oregon, USA.

Hays, J., & Efros, A. A. (2008). Im2gps: Estimating geographic information from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1–8), Anchorage, Alaska, USA.

Hwang, Y.H., Gretzel, U., & Fesenmaier, D. R. (2006). Multi-city trip patterns tourists to the United States. *Annals of Tourism Research, 33*(4), 1057-1078.

Jaffe, A., Naaman, M., Tassa, T., & Davis, M. (2006). Generating summaries and visualization for large collections of geo-referenced photograph. In *Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval* (pp. 89-98), Santa Barbara, California, USA.

Kalogerakis, E., Vesselova, O., Hays, J., Efros, A. A., & Hertzmann, A. (2009). Image sequence geolocation with human travel priors. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 253-260). Tokyo, Japan.

Kennedy, L., Naaman, M., Ahern, S., Nair, R., & Rattenbury, T. (2007). How Flickr helps us make sense of the world: Context and content in community-contributed media collections. In *Proceedings of the 15th International Conference on Multimedia* (pp. 631-640), Augsburg, Germany.

Kisilevich, S., Keim, D., & Rokach, L. (2010). A novel approach to mining travel sequences using collections of geotagged photos. In *Geospatial Thinking* (pp. 163-182). Berlin Heidelberg: Springer.

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*, Article No. 38, Bethesda, Maryland, USA.

Kurashima, T., Iwata, T., Irie, G., & Fujimura, K. (2010). Travel route recommendation using geotags in photo sharing sites. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management* (pp. 579-588), Toronto, Canada.

Lau, G., & McKercher, B. (2006). Understanding tourist movement patterns in a destination: A GIS approach. *Tourism and Hospitality Research, November, 7*, 39-49.

Lee, I., Cai, G., & Lee, K. (2013). Mining points-of-interest association rules from geo-tagged photos. In *Proceeding of the 46th Hawaii International Conference on System Sciences* (pp. 1580-1588), Grand Wailea, Maui, Hawaii, USA.

Leung, X. Y., Wang, F., Wu, B., Bai, B., Stahura, K. A., & Xie, Z. (2012). A social network analysis of overseas tourist movement patterns in Beijing: The impact of the Olympic Games. *International Journal of Tourism Research, 14*, 469-484.

Lew, A., & McKercher, B. (2006). Modeling tourist movement 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.

Li, X. (2013). Multi-day and multi-stay travel planning using geotagged photos. In *Proceeding of the 2nd ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information* (pp. 1-8), Orlando, Florida, USA.

Mamei, M., Rosi, A., & Zambonelli, F. (2010). Automatic analysis of geotagged photos for intelligent tourist services. In *Proceeding of the 6th International Conference on Intelligent Environments* (pp. 146-151), Kuala Lumpur, Malaysia.

Masiero, L., & Zoltan, J. (2013). Tourist intra-destination visits and transport mode: A bivariate probit model. *Annals of Tourism Research*, in press.

Majid, A., Chen, L., Chen G., Mirza, H.T., & Hussain, I. (2012). 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. (2013). A context-aware personalized travel recommendation system based on geotagged social media data mining. *International Journal of Geographical Information Science*, *27*, 662-684.

McKercher, B., & Lau, G. (2008). Movement patterns of tourists within a destination. *Tourism Geographies: An International Journal of Tourism Space, Place and Environment*, *10*(3), 355-374.

McKercher, B., Shoval, N., Ng, E., & Birenboim, A. (2011). First and repeat visitor behaviour: GPS tracking and GIS analysis in Hong Kong. *Tourism Geographies: An International Journal of Tourism Space, Place and Environment*, *14*(1), 147-161.

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*, Springer New York, pp. 657-670.

Orellana, D., Bregt, A. K., Ligtenberg, A., & Wachowicz, M. (2012). Exploring visitor movement patterns in natural recreational areas. *Tourism Management, 33,* 672-682.

Prideaux, B. (2000). The role of the transport system in destination development. *Tourism Management*, *21*(1), 53-63.

Quack, T., Leibe, B., & Gool, L. V. (2008). World-scale mining of objects and events from community photo collections. In *Proceedings of the 7th International Conference on Content-based Image and Video Retrieval* (pp. 47-56), Niagara Falls, Canada.

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

Shi, Y, Serdyukov, P., Hanjalic, A. & Larson, M. (2013). Nontrivial landmark recommendation using geotagged photos. *ACM Transaction on Intelligent System Technology, 4*(3), Article No. 47.

Smallwood, C. B., Beckley, L. E., & Moore, S. A. (2012). An analysis of visitor movement patterns using travel networks in a large marine park, north-western Australia. *Tourism Management, 33(3),* 517-528.

Snavely, N., Seitz, S. M., & Szeliski, R. (2008). Modeling the world from Internet photo collections. *International Journal of Computer Vision, 80*(2), 189-210.

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., Evans, F., Spilsbury, K., Ciesielski, V., Arrowsmith, C., & Wright, G. (2010). Market segments based on the dominant movement patterns of tourist. *Tourism Management, 31(4),* 464-469.

Xia, J., Zeephongsekul, P., & Arrowsmith, C. (2009). Modelling spatio-temporal movement of tourists using finite Markov chains. *Mathematics and Computers in Simulation*, *79*(5), 1544-1553.

Yanai, K., Kawakubo, H., & Qiu, B. (2009). A visual analysis of the relationship between word concepts and geographical locations. In *Proceedings of the ACM International Conference on Image and Video Retrieval*, Article No. 13, Santorini Island, Greece.

Yanai, K., Yaegashi, K., & Qiu, B. (2009). Detecting cultural differences using consumer-generated geotagged photos. In *Proceedings of the 2nd International Workshop on Location and the Web*, Article No. 12, Boston, Massachusetts, USA.

Yin, H., Wang, C., Yu, N., & Zhang, L. (2012). Trip mining and recommendation from geo-tagged photos. In *Proceedings of the IEEE International Conference on Multimedia and Expo Workshops* (pp. 540-545), Melbourne, Australia.

Zakrisson, I., & Zillinger, M. (2012). Emotions in motion: Tourist experiences in time and space. *Current Issues in Tourism, 15*(6), 505-523.

Zheng, Y.T., Zha, Z.J., & Chua, T.S. (2011). Research and applications on georeferenced multimedia: A survey. *Multimedia Tools and Applications, 51*(1), 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), 1-18.

Table 1. Parameters for Photo Search Function.

Parameter	**Value**	**Description**
x_{min}	113.887603	minimum *longitude* of the bounding box
y_{min}	22.215377	minimum *latitude* of the bounding box
x_{max}	114.360015	maximum *longitude* of the bounding box
y_{max}	22.51446	maximum *latitude* of the bounding box
t_{min}	1/1/2011	earliest photo taken date
t_{max}	-------------	latest photo taken date

Table 2. Data Collection.

Group	Number of Tourists	Number of Photos
Western Tourist	1,036 tourists	15,990 photos
Asian Tourist	1,064 tourists	13,453 photos
Total:	2,100 tourists	29,443 photos

Figure 1. Photo Locations of Hong Kong Inbound Tourists.

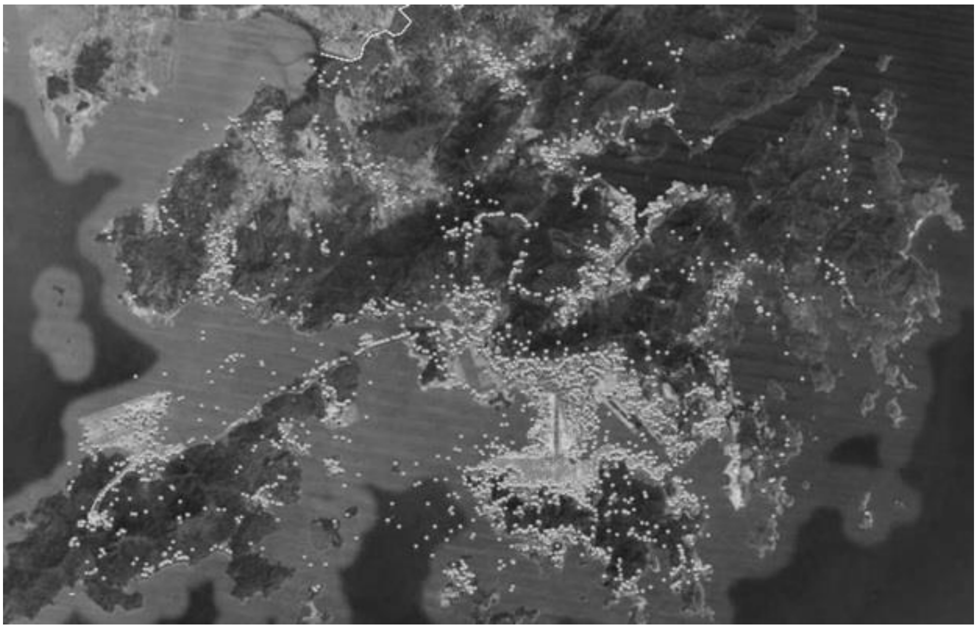


Figure 2. Areas of Interest for Inbound Tourists.

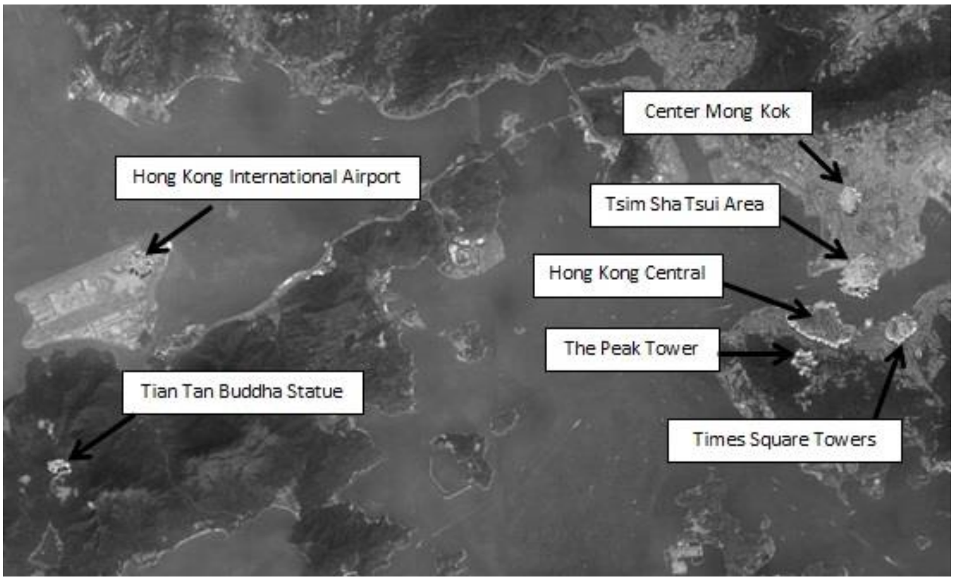


Table 3. Popularity of Areas of Interest.

Group	Area of Interest	Percentage (%)	Popularity Rank
Asian Tourist	Hong Kong Central	40.35	1
	Tsim Sha Tsui Area	38.80	2
	Times Square Towers	20.08	3
	Hong Kong International Airport	18.34	4
	The Peak Tower	14.19	5
	Center Mong Kok	10.52	6
Western Tourist	Hong Kong Central	47.92	1
	Tsim Sha Tsui Area	44.64	2
	The Peak Tower	19.74	3
	Center Mong Kok	15.14	4
	Times Square Towers	12.22	5
	Hong Kong International Airport	10.99	6
	Tian Tan Buddha Statue	10.43	7

‘

Figure 3. Movement Trajectory of Tourist Generated from Geotagged Photos.

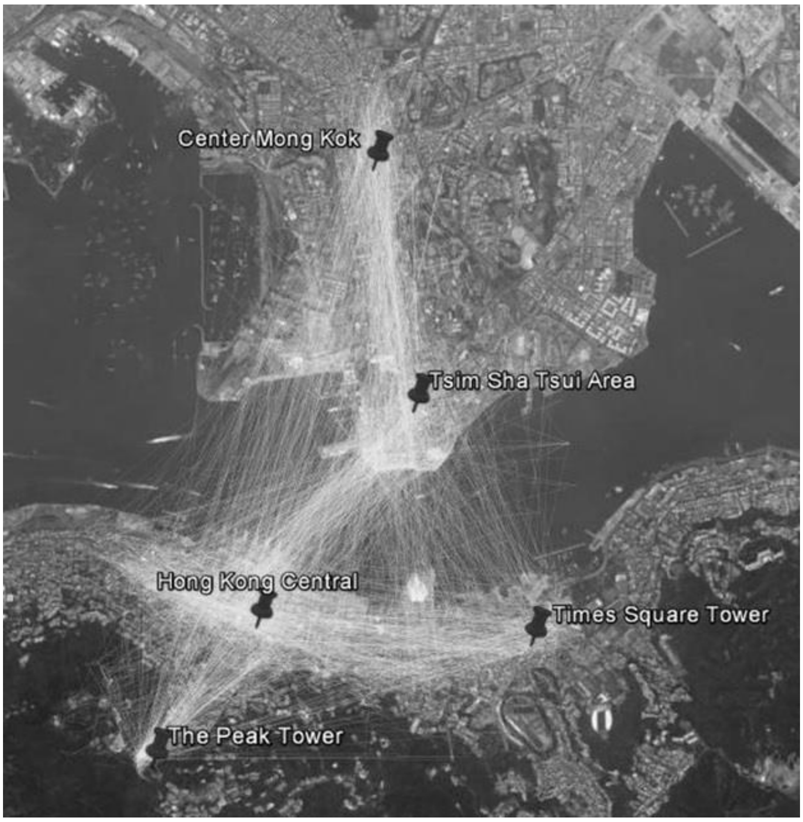
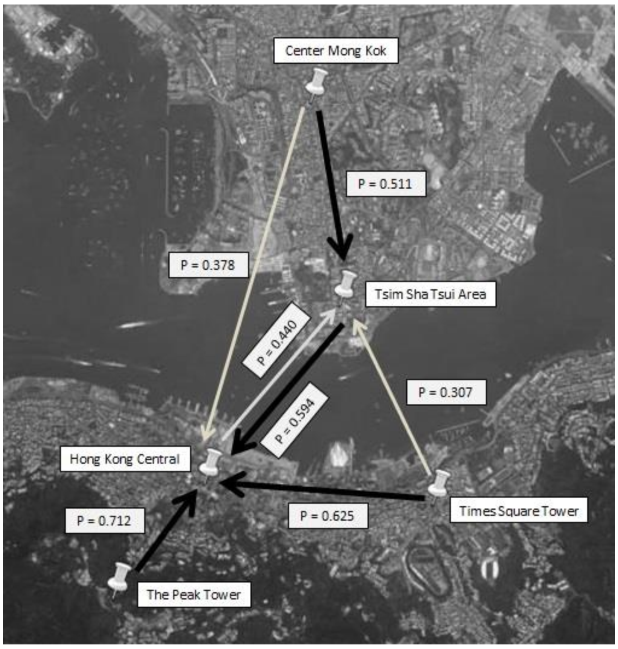
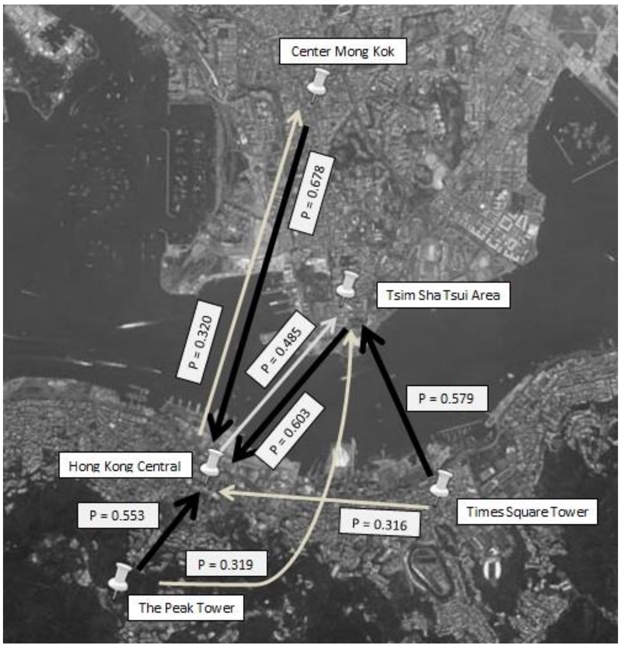


Figure 4. Tourist Traffic Flow in Hong Kong Metropolitan Area.



(A) Asian Tourist



(B) Western Tourist

Table 4. Z-test result on travel flow between Asian and Western tourists.

Tourist Traffic Flow	*Z-Score*	*p-value*	ID
Hong Kong Central → Tsim Sha Tsui	-1.000	0.317	F_1
Tsim Sha Tsui Area → Hong Kong Central	-0.162	0.873	F_2
Center Mong Kok → Hong Kong Central	*-3.490*	*0.000*	F_3
Times Square Towers → Hong Kong Central	*3.952*	*0.000*	F_4
The Peak Tower → Hong Kong Central	1.636	0.101	F_5
Times Square Towers → Tsim Sha Tsui	*-3.507*	*0.000*	F_6

Figure 5. Travel Route of Tourist between Center Mong Kok and Times Square Towers.

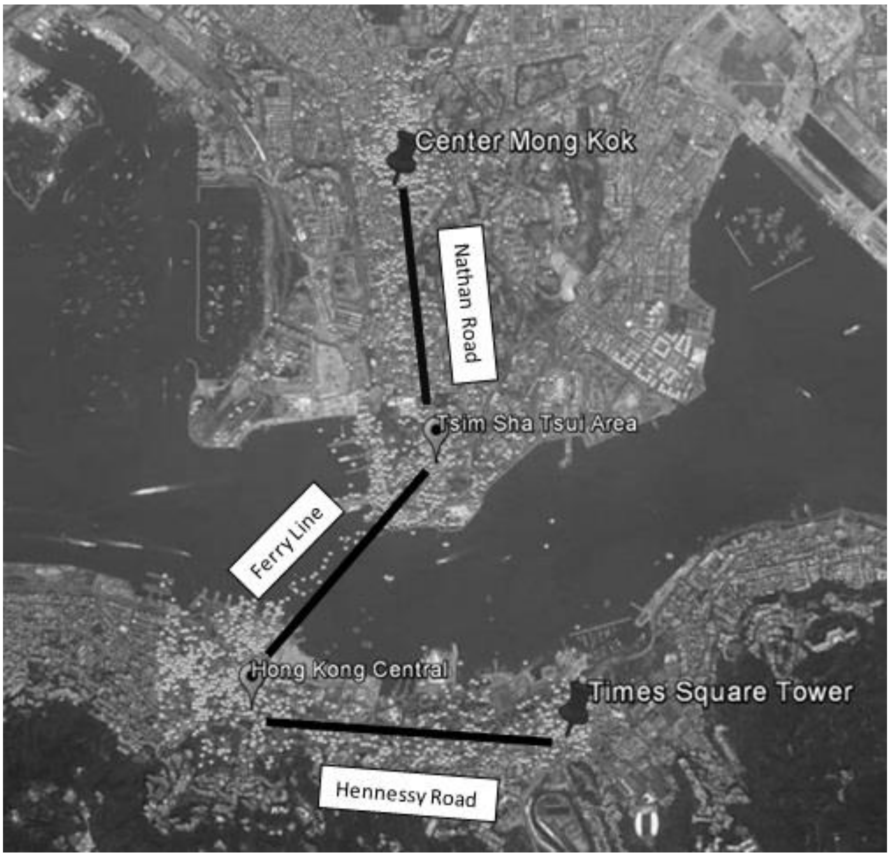
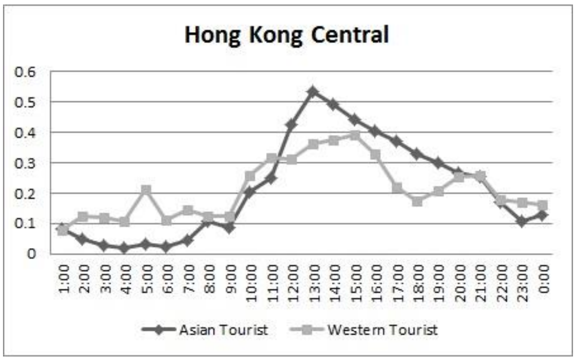
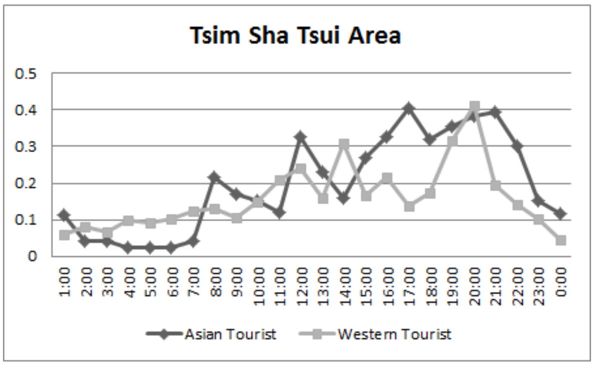


Figure 6. Tourist Activities in Areas of Interest.

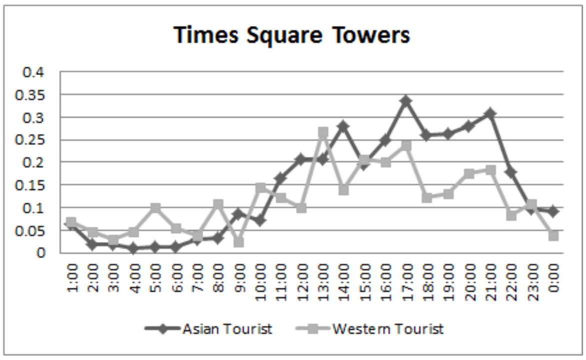
(A)



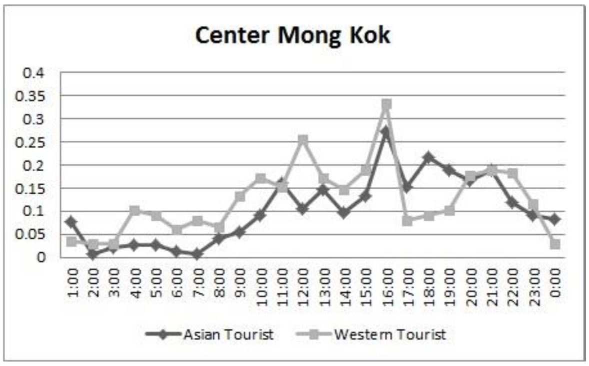
(B)



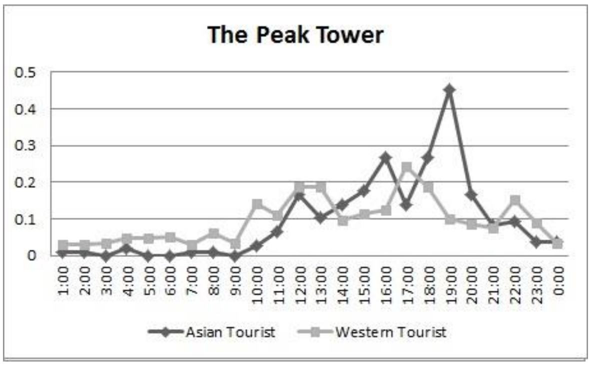
(C)



(D)



(E)



(F)