255

An Exploration of the Presence and Flows of Visitors in the City using the Social Media Data in London

YUTAO SONG

University College London
yutaosong@yahoo.com

KAYVAN KARIMI

University College London
k.karimi@ucl.ac.uk

ABSTRACT

Most global cities have a large volume of visitors every day. These visitors interact with locals and have noticeable impacts on the city. London is the type of city that actively attracts and interacts with tourists. Tourists are, almost by definition, a group of "spatial users" that do not live and work within the city in the long term. The way they move and navigate through the city relates to their plans and destinations, and it is not always easy to distinguish such a group from other spatial users. Therefore, it requires a special effort to study their movement patterns and to identify the relationship between the group and space. Instead of using traditional on-site observations, this research has explored a new method for collecting and converting social media platform data into movement patterns using the data from Flickr.

Comparing with traditional data-collecting methods, the proposed method can cover a wider territory (5-km radius) and period (5 years), and most importantly, it collects users' movement data without interfering with their regular activities, which might collect more natural movement results. The essential concept of movement data conversion is to arrange the geo-tagged photos of each Flickr user per day by the time sequence during their itinerary. The sets of sorted photos will reflect the visit order each user has gone through. The geographical data of the arranged photos can be converted into the topological travel routes. Finally, the routes will be mapped onto the base map using the proposed method. The collected data from the Flickr platform provide a wide range of information from static content to movement flow. These data are crucial in studying tourist activities and their related space.

KEYWORDS

Tourist Flow, Flickr, Segment Map, Photography

1. INTRODUCTION

Travel and tourism are essential economic activities in most countries around the world. As well as its direct economic impact, the tourism industry has significant indirect and induced impacts. The UN Statistics Division-approved Tourism Satellite Accounting methodology (*TSA:RMF 2008*) quantifies only the direct contribution of Travel and Tourism. However, WTTC recognises that Travel & Tourism's total contribution is much higher and aims to capture its indirect and induced impacts through its annual research (*Turner. 2017*).

The direct contribution of global travel and tourism to the Global GDP in 2016 was GBP1261.25bn (3.1% of GDP). This contribution is forecast to rise by 3.8% to GDP 1310.9bn in 2017. (Turner. 2017). It primarily reflects the economic activity generated by industries such as hotels, travel agents, airlines and other passenger transportation services (excluding commuter services). At the same time, it also includes the activities of the restaurant and leisure industries that are directly supported by tourists. The percentage of the contribution to the global city, such as London, is much more significant than the general data. London tourism sector employs 700,000 people – one in seven of the capital's jobs –and contributing 11.6 percent of its GDP, (*Deloitte LLP, 2013*) which hasn't counted for other indirect relevant industries like culture and art.

Monitoring and analysing the behaviour of the tourists become an increasingly important demand for the daily city maintenance routine. The tourist behaviour covers vary of activities conducted during the trip. For all the activities, none of them is static. The dynamic status of the activities raises the need for the tourist flow study. The flow in this paper indicates a series of collective movements from multiple origins to destinations. As Kevin Lynch insists, the flow is as important as the stationary physical parts while studying the city(*Lynch, 1959*). To study tourism and its related urban space, this paper investigates both the tourist flow data and other stationary urban characters. Flow analysis enhances the understanding of how multi-dimensional spatial characters in the city influent the spatial dynamics of tourism. It also contributes to understanding how the spatial structure of the city influence tourist behaviour and how the movement pattern is generated.

Before new technologies simplified the data collecting procedure, social scientists manage to retrieve the movement and stationary data through the on-site observation and survey. Space Syntax Limited has published a tourist movement report in 2005 that analysed the relations between the tourist flows and other spatial properties in London.(*Karimi, 2005*) The data was collected through the traditional methods by conducting on-site surveys of pedestrian flow at 18 important tourist areas per month between January 2004 and February 2005. Some findings have a great contribution to the London tourism study. Firstly, it revealed the diversity of movement flow from place to place. The results were presented in a detail movement pattern map that illustrates how the visiting intensity of the attraction spatially distributed. Secondly, the report shows the intensity curve of the tourist flow from morning to evening and from spring to winter. This help researchers understand how tourists manage their daily basis tour plan and how season and weather factor influence the tourist activities. Thirdly, the report revealed how the main attractions perform and relate to their neighbour attractions. The result indicates that the attractions within the urban space aren't separated from each other. The ones that close to each other have grouped and formed a cluster with the hierarchical network. This tourist movement study provides a solid foundation for the study of the tourist related spatial structure and character.

The tourist flow report has collected a large volume of valuable tourist flow data and has a significant contribution to the London tourist study. However, even with the huge human and time resource input, the available tourist movement data is still limited to 18 pre-selected sites and only covers the lunch period within a year. After more than 13 years, the collected dataset may already outdated. There are other alternative movements collecting methods, such as GPS or Wi-Fi tracking, but they all have the limitations when implementing it in city scale. The movement study based on GPS tracking requires the user group to record and upload the GPS data(*Zheng, 2019*). Tourists normally wouldn't upload their trip information to the internet in the GPS format. Therefore, the method can't collect sufficient data from the source without collaboration from the tourists. The movement study based on Wi-Fi tracking is one of the most popular techniques in modern spatial behaviour study. However, it requires

a huge investment to cover all the study area with expensive devices; hence, won't be considered in this research.

To collect the required movement data from the tourists, this research starts with investigating the data of the tourist activities, in the hope of retrieving the movement flow from these activities. There are five main activities that are commonly conducted by tourists. (*Karimi, 2005*)

The first main activity is sightseeing. This is the main purpose of the trip. Most of the tourists just literally walk around and observing the place they visit. This can be summarised as a process of searching and receiving expected new information during a trip. The targets could be the historical attractions, cultural shows, musical performance, art exhibitions etc. The second main activity during the trip is eating, this takes from 20% to 40% of all activities throughout the trip. It is hard to deny that, eating is the essential activities of all times in human life no matter if they are travelling, working or studying. The good experience of eating will enhance the overall experience of the trip, but it is the supplementary need, not the primary one.

The third important activity is photography. The invention of the camera provides the possibility for people to record the scenes of the moment with the realistic photographs. The falling of the price and increasing of the portability makes photography a popular activity in everyday life. Photography and travel are linked intrinsically. It also tightly relates to the sightseeing activities of the trip. In Modern tourism, photography no longer requires tourists to take an expensive and heavy camera with them, and the cost of taking and saving photos almost drop to zero. Increasing number of tourists tend to use photos to document their travel experience and share them with others(*Larsen, 2008*). Research conducted by Lo. I. S reveals the very high percentage of photo taking activities (89%), and high photo sharing activities from the tourists (41%). (*Lo, 2015*) These findings and papers from other researchers had proved that photographs by tourists could be a good material to study the activities of the tourists.

One of the significant changes in modern social activities is that people tend to conduct more social interactions with their friends through the internet. Among all the internet social interactions, searching, posting and sharing dominant among all other activities. Social media platforms have been widely used to maintain their social networks nowadays(*Hasnat ,2018*). Increasingly amount of everyday data has been shared by internet users through the platforms. They intend to use this to keep themselves active and draw more attention. Social media services, like Twitter, Facebook, Instagram and Flickr have been a fundamental base for the formation of the digital social networks. The promoting and sharing information and expressing opinions through social media by both users and organisations become a global standard. This shared information is not only the priceless memories of the users themselves but also valuable data for researchers to conduct their social science studies. The significant contributions these platforms bring to the world not only through their social services but also from the incredible valuable tools available for users and platform extension developers, namely the API (application programming Interface). The API of the social platforms bridges the massive data in the server to the outside world, allowing the users and developers to communicate and process the enormous database from these platforms(*Stieglitz, 2018*).

There are all types of platforms for different users' social needs. Some of them specifically target the users that enjoying the photo sharing and social interaction. Instagram and Flickr are two of the most popular platforms among them. The massive amount of travel photographs has been uploaded to these platforms. This research will explore the method of using the tourist's photographs from the platform to retrieve the tourist flow data within the urban network.

2. DATASETS AND METHODS

Flickr, as one of the most widely used photos sharing social platforms, has its advantages over other photo sharing platforms. Different from the daily activities-based sharing platform such as Instagram and Facebook, Flickr platform attracts more serious photography users. There are much fewer selfies and indoor social event photos than other platforms. Besides this, most of the photos been uploaded to the platform are about buildings, streets, attractions and other urban structures. This makes the data content from Flickr mostly reflecting the sightseeing activities conducted by the tourists during their trip.

The Verge report revealed that Flickr had a total of 87 million registered members and more than 3.5 million new images uploaded daily during 2013. (*Verge, 2013*) Besides its large user groups and data volume, it is also well known for its API high-level data accessibility. Photos and videos can be accessed from the Flickr platform with advanced detail searching requirement. The advanced users can batch upload or download photo datasets using the Flickr API framework. For example, the user can download a group of photos within a given geo-boundary by using the method "flickr.photos.search" with arguments of latitude, longitude, and radius.

Flickr's API is the server-side web API. It provides a programming interface that consists of multiple publicly accessible endpoints which response to the user's requests. The returned packages are normally in JSON or XML format. Both formats are widely used for many web applications, and most commonly, in the HTTP-based web server.

The content of the data packages that Flickr return to the users depends on what they have requested through API. To collect a complex dataset, the user should de-construct the complex requirements into several simple requests and send them to the API correspondingly. Since this process will receive multiple datasets as the result, Further process to cleaning, organising and combining data into one final set is required. The cleaned data will be used to generate the tourist movement flow and use it to study the relation to the urban space. To properly visualise spatial data, researchers can either take advantage of some existing visualisation programming libraries or like in this research, make use of the existing CAD/GIS software platform to visualise spatial data. There are several applications have the programming interface that can allow users to write plugins within the software, to take advantage of the application. These include Rhino Grasshopper 3D, QGIS, ArcGIS, Blender, Maya etc. The custom program in Grasshopper 3D platform has some unique functions that can't be found in other application. The platform supports all types of programming language including C++, C#, VB, Python and most importantly its visual programming environment. Like the Lego or the module for prefabrication house, There is the module like visual programming language blocks for an easy prototyping experience. It allows you to create or use the existing functions like laying bricks for building, which massively increase the process efficiency of the experiments. Therefore, this research chooses this platform as the development environment for the research.

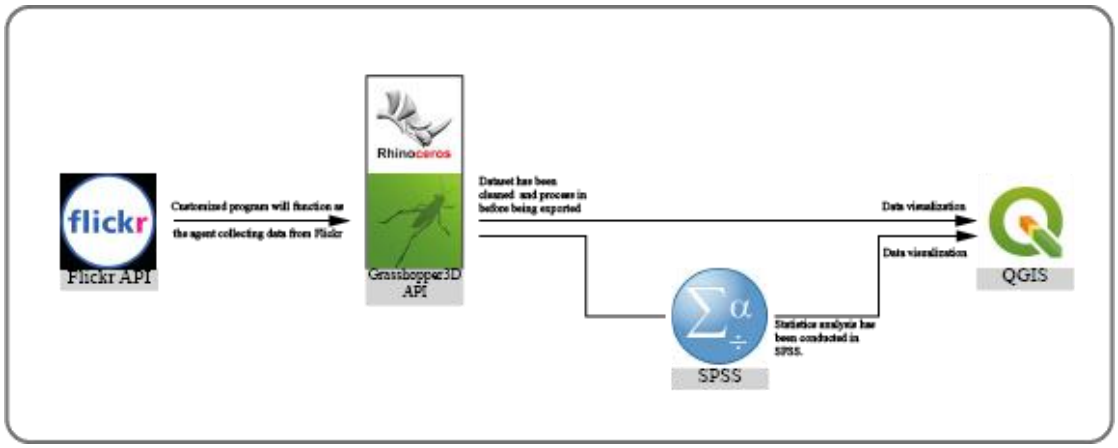

Figure 2A. The Overview of the Data Collecting, Analysing and Visualizing

To utilise the data from the Flickr server, studying the contents within the Flickr dataset is necessary. Flickr as the photo-sharing platform organises all its server data in the photo-oriented basis. When a photo is uploaded, other data relates to the photo will be packed and uploaded together with it. Among these data, some of them are compulsory, some of them are optional. Information like user ID, photo-taking date and time are always required. The system will generate the ID for each photo for the index purpose. There are other optional data that not always exist in the data package, but is important for the research. For instance, the geolocation is necessary for generating the tourist movement traces. Without the geotags, the proposed method will not function. Fortunately, since the geotags help users to categorise and organise their photo archive, most of the users prefer to provide the ger-tags together with the photos in Flickr. Other data such as Photo titles and content tags are both optional but important. It helps researchers to categorise and study the photo package without doing image recognition.

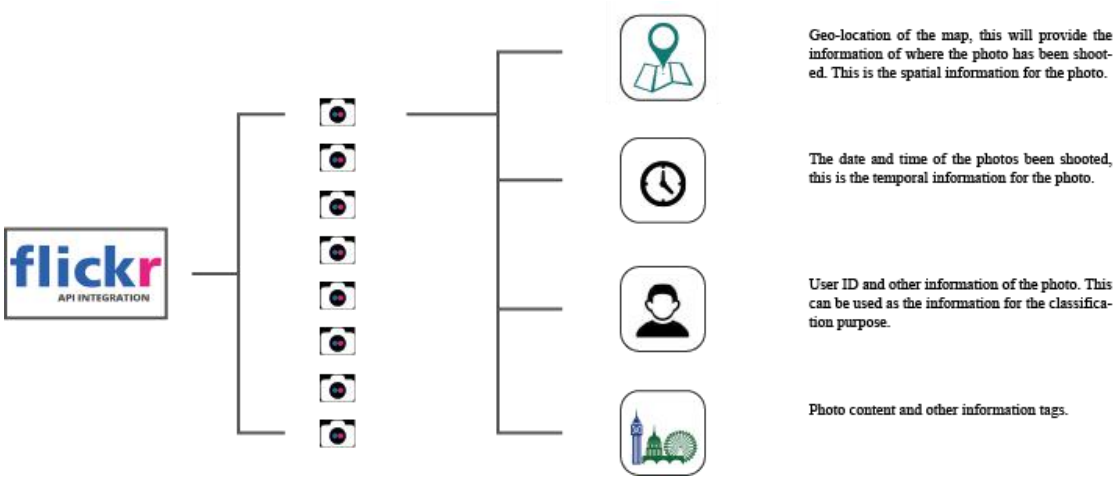

Figure 2B. Flickr Data Package Contents

In this research, a unique solution for collecting, generating and analysing the tourist movement flow through Flickr data has been proposed. In this chapter, the process will be explained and discussed in detail. This will start from the introduction of data collecting process, then move to the discussion of organising dataset and generating the tourists' topological movement. Finally, a discussion about how topological movement could be converted and mapped onto the spatial network will be discussed.

To generate the movement pattern from 2013 to 2017, the first step is to collect the data from the Flickr platform within the 5 years. The Flickr data covers the most recent tourist photo-taking activities and reveals how Flickr users travelling and recording their moment during their trips. There are two boundaries have been proposed and implemented in this research. The first boundary is a circle with its centre at the location (51.513122, -0.089786) in the City of London and with a radius of 10 km. This boundary covers most of the area within London M25, where most of the tourist attractions and activities locate. The Second circle boundary is the narrowed down boundary of the first one. It focuses on the main activities in the central London area and includes more than 90% of the tourist activities in London. Most of the hottest tourist attractions of London located within it. The second boundary only be used at the final phase to focus more on the main tourist activities and save computation resources for tourist flow calculating process.

Figure 2C. Flickr Photos Distribution, Greater London (2013-2017)

Besides the proposed boundary filter for the data collecting, there are two other filters been applied to the requests. One is the user identity filter. Comparing with other photo sharing platform, Flickr has its advantage in its targeting user group, there are very few selfies and indoor daily life image been uploaded to the site, which is good for the research purpose. Despite this, it is still important to filter out the photos that are taken by local people. The simplest method to locate them is to filter the users whose major photography location is London or has London in their tourist history. By removing this group of the user, the dataset will better reflect the tourist activities. Another filter is the movement validity filter. Among all the photos in the dataset, there are a certain number of photos that are the only photo taken by the user that day. The single photo will not be able to use for generating movement, therefore, should be removed.

Comparing with other social media platforms, the Flickr platform provides the most advanced API that can return highly customised datasets. However, it doesn't mean the data packages can be used for this research straight away after been downloaded. The Flickr server, like other social media platform, has the strict rules to prevent the abuse of data collecting through the API service and to protect Flickr users privacy. These rules increase the difficulty for researchers to collect the required dataset. There are two main limits that affect the efficiency of the data collecting process. Firstly, it sets the upper limit of 250 photos for each request. If the number of photos exceeds the limit, it will separate the data into multiple pages, and will only show the 250 photos by default. Secondly, if the return data can't be filled into a single page, the return data will in the disrupted order, and duplicated data will be randomly inserted into each page. This means even data from all pages have been collected, they still requires efforts to process this raw data into a clean and organised order.

In this research, an integrated solution for generating the tourist movement pattern from the collected Flickr data will be proposed. This includes the dataset combining and cleaning process, the topological movement generating process, and the movement pattern converting process. This solution aims to generate the ready-to-use movement dataset for the tourist-related urban spatial study.

To achieve proper coverage of the Flickr data, the first step is to divide the 5-year time into monthly basis 60 sub-ranges. Then for each of the sub range, multiple requests will be sent to collect the data. The number of requests depending on the data size of each subrange. This method provides acceptable

data resolution and offers the monthly pre-ordered dataset for the monthly and yearly statistical analysis.

Another challenge waiting to be tackled is the structure of the received dataset, as mentioned previously, the Flickr server has applied several algorithms to obfuscate the datasets before sending it back to the client. The received data packages thus have duplicate and disorder information. To make each photo and its data unique for analysis, Duplicate data need to be removed from the set. Since all data in the Flickr server are photo-oriented, each photo has its unique ID. Therefore, all the Photo ID need to be checked to find out any non-unique photos, if there are any photo ID appears more than once, Duplicate photos and their related data need to be removed from the dataset.

By executing the above process, a refined dataset has been prepared for further analysis. This dataset included 125,315 Flickr Photos taken by 4367 users from 2013 to 2017, which included 864,417 tags.

The photo tags together with the photo location information, form the big image of the tourists co-present and distribution pattern. However, It is clear that tourist behaviour is not only about the static co-present pattern, but also about the movement flow. Besides the traditional pedestrian movement survey, the use of Flickr data to generate movement patterns could be a low-cost alternative or auxiliary method for the tourist activity study. The photography is one of the main activities that reflect what tourist have experienced during the trip. If the tourist has taken a sufficient amount of photos during his trip, the photos and the information come with the photos may reflect more information about the trip than people expected.

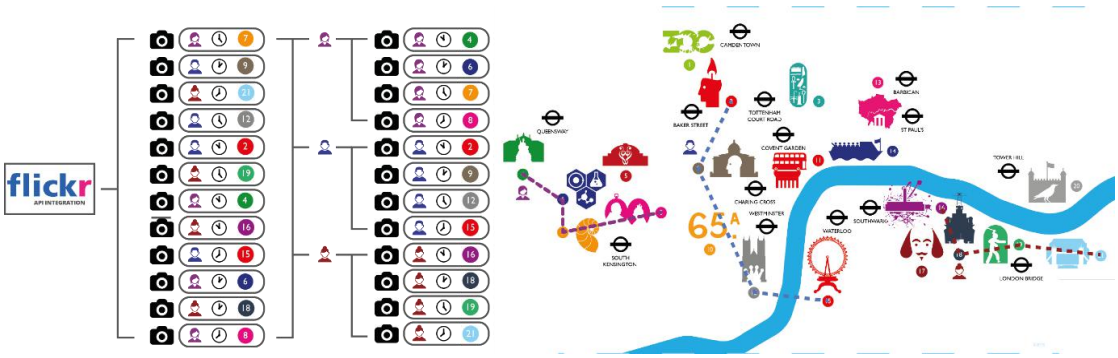

Figure. 2D Movement Pattern Construction Diagram (Base map from Mapway)

To convert the Flickr data into the movement traces, the research objective needs to be refocused from the stationary photo patterns to the user activities in the daily timeline. Flickr photos and its related data would thus need to be grouped by the user ID and date, then be rearranged by time. These processes organised all the photos into the list of photos for each user per day. Each list of data then will be the storytellers narrating the novels of the trips.

Figure 2Ea shows Flickr photos taken by user Z on the 28th of March 2014, from 10 am to 6 pm. These photos, with their geo-tags and time stamps, indicate Z's photo shooting activities of the day. It further indicates the time he stayed at the various locations and his interests at the time.

Movement traces are the geographical storyline of the trip. These can be constructed by connecting the geolocation tags within the previous data lists. Each of the generated traces reflects the journey and the places the user has visited during the day. For instance, user Z arrived at Westminster tube station in the morning; he took the first few photos of the Westminster Church and Palace of Westminster, then he walked along the Thames river and took several photos during the walk. After that, he visited Covent Garden, Smithfield Market, Holborn and China Town in the time sequence. Once all these connected photo base on the timeline, the topological route will be generated. (Figure 2Ea, Figure 2Eb) Not all the users take a large volume of photos from the beginning to the end of their trip; some user only took a small number of photos. The generated routes from these photos only reflect part of their trips. However, when all five years data was overlaid onto the map. The influence of the partial routes has been neutralised.

As Figure 3A shows, most parts of the tourist movement pattern were located in the centre area of London. To reduce the computing cost for the final pattern conversion and to narrow down the focus, the second boundary used for further research is the smaller area within the first boundary. It has the location in Soho as the centre point and a 5km radius. This boundary has also been used to collect other spatial properties in central London for comparative study.

Figure 2Eb shows one of the movement routes generated by the Flickr data. However, this pattern is composed of routes being generated from Flickr photos. The routes are thus the topological representation of the movement from one photo location to another. To make the topological pattern comparable with other spatial patterns on the segment map, the routes needed to be converted into the proximate movement routes based on the context map. In this research, most of the other spatial properties have been mapped onto the segment map. Therefore, the topological movement routes need to be converted into the segment network.

Figure 2Ea shows the photos taken by user Z on 28th of March 2014. Figure 2Eb is the converted topological map mentioned in the previous section. From Figure 2Eb to Figure 2Ec is the process of this section going to discuss.

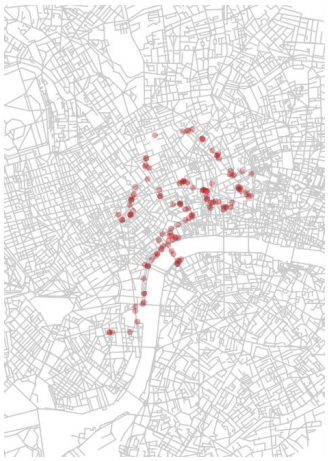

Figure 2Ea Locations

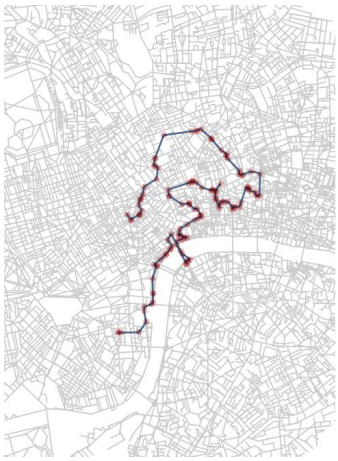

Figure 2Eb Topological route

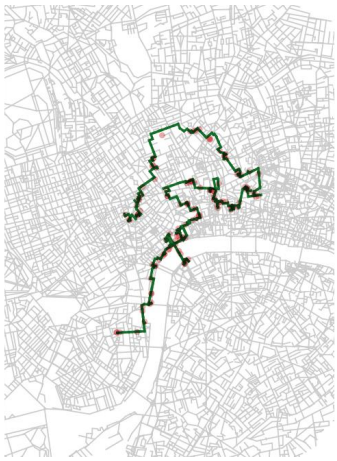

Figure 2Ec Converted route

The converting process requires a base map, in this research, it is the segment map. Firstly, the segment map needs to be converted into the dual-graph map. Then each topological route needs to be deconstructed into multiple origin-destination. Each of the origins and destinations will be projected to the closest graph node in the dual graph. Dijkstra Algorithm will be used to select the paths between each origin and destinations. After this, paths will be combined back into single routes that composed of dual graph connections. Finally, the dual graph routes will be converted back into the segment map as shown in Figure *2Ec.*

The method this research applied has revealed some similarities to the process of creating the Angular Choice Map. However, the goal and outcome were significantly different. In the natural movement theory, the potential origins and destinations cover every piece of the segment roads. Therefore, the choice value of the natural movement counts the shortest routes generated from all road segments to others. For the tourist flow, it is not the case. Since the topological movement routes from Flickr data could be converted to groups of origins and destinations, the movement intensity value of the road segment will be the sum of all the pass-through count of the shortest routes being generated by the given origins to the destinations. Therefore, the photo-taking frequency directly influences the accuracy of the generated routes. In this research, if the distance between two photo locations is larger than 2km, the generated route may not reflect what it should be in the actual situation. Therefore, that extra-long shortest path of the route will not be generated between the origin and destination.

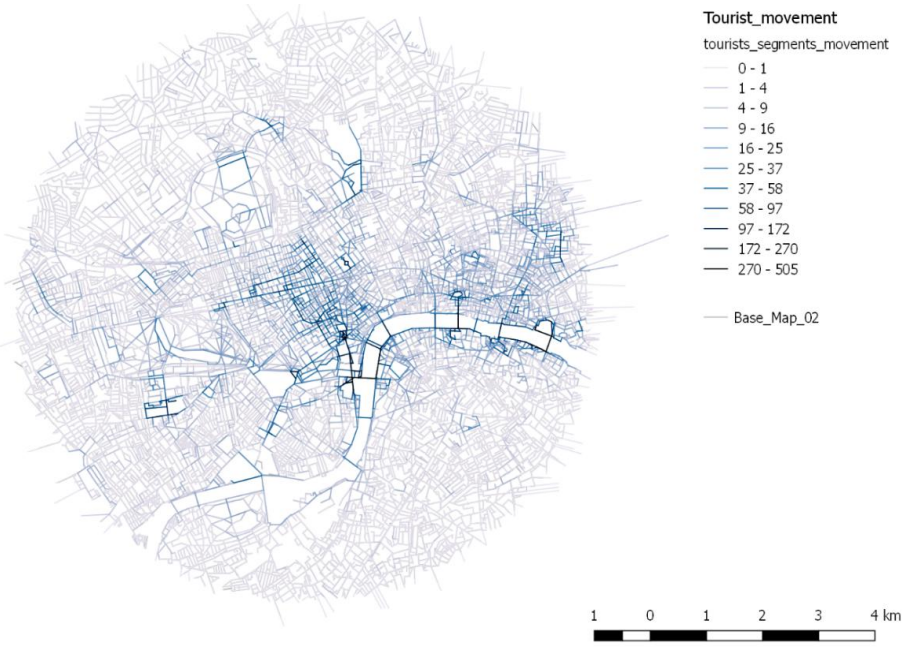

Figure 2F The Converted Tourist Movement Flow

3. RESULTS

Three data maps have been produced in this research. All of them provide valuable information for the tourist study. In this section, each of the dataset will be discussed, then be compared with the traditional on-site survey data collected by the Space Syntax Limited from 2004 to 2005.

The first one is the map of stationary data. It shows the spatial distribution of the photography activities through geo-tags and time-tags and reveals other activity contents through photo-tags. The map covers information in the 10km radius boundary between 2013 and 2017. The number of Flickr photos taken by users varied periodically every month; it reaches its peak in summer around June, then the number starts to drop gradually in the following month. Finally, it reaches to its valley between January and February. When comparing the site survey tourist flow graphs with the Flickr generated tourist flow graph, it seems that the shape of the Flickr generated graph is closer to the max flow graph instead of the average flow. Both the graph from Flickr and survey max flow indicate the tourist flow reach to its peak in June, which is different from the survey average flow graph. In the Space Syntax tourist report, it indicates that the max flow from the busiest gates should be more representative then averaging all gates together at each area. The solution avoids the problem of bias towards the fewer gates area.

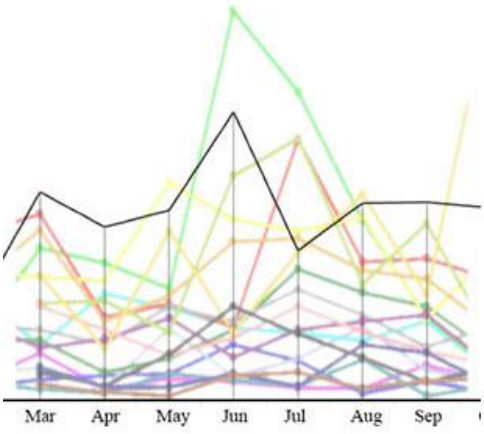

Figure 3Aa Comparison: Flickr/Survey Max Flow

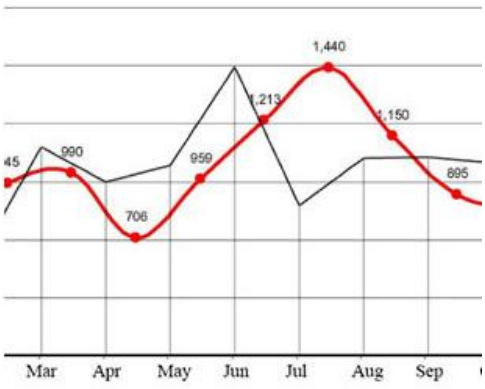

Figure 3Ab Comparison: Flickr/Survey average Flow

Besides the monthly tourist intensity changes, the tag data also provide the image of the hourly tourist intensity status. The graph clearly indicates the activity intensity reached its peak at around 2 pm. Additionally, there is the second intensity peak appeared in the evening during the summer time. The site survey hourly flow graphs are placed next to the Flickr graph. The results also are similar to site survey data.

The tag data also could be used to identify the geographical distribution of the tourist intensity. The data indicates that there are four areas have the highest intensity in London. They are the City of London, Westminster, Camden and Shoreditch. This information together with the intensity pattern plays an important role in defining the boundaries for the meso and microscale research. Besides this statistical result, the photo tags also indicate the most favourite building types by tags. They are bridges, towers, Thames River, parks, squares and museums. Also, there is a tag been frequently mentioned in many photos, it is the word "art".

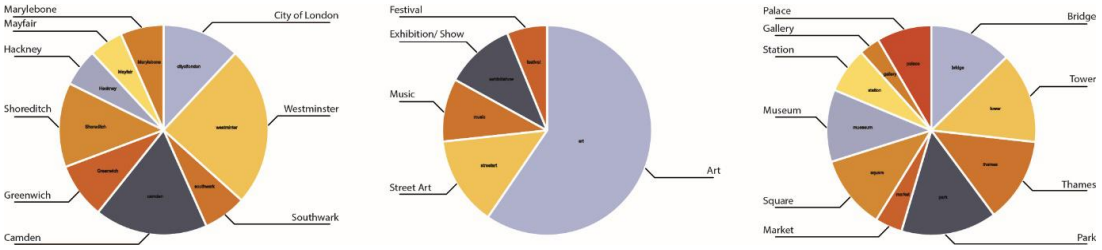

Figure 3B Tag Frequency (2013 to 2017)

The discussion of the first dataset is mainly about the photo content tag analysis. For the second generated dataset, the discussion will be about the movement pattern spatial distribution. The topological movement pattern generated by the Flickr data covered the large area of Greater London. The data map has revealed several findings of the tourist movement. Firstly, the tourist movement pattern didn't distribute the same as the natural movement. The movement traces highly concentrated on some areas of London, most of which are located in the city centre. Secondly, two types of movement flow can be identified through the shape of the clusters. One of the cluster shapes is the point-based circle, while another is linear based ribbon. For the point based high-intensity clusters, the areas attract tourist from all over the city. No specific direction has significant tourist movements flow into the area. For instance, the V&A area in South Kensington, Buckingham Palace, Greenwich, British Museum, Covent Garden is this type of cluster. For the ribbon shape cluster, such as the riverside of Thames, Oxford Street, Whitehall Street, Millennium Bridge, Buckingham Palace Road. These clusters show very strong directed movement. This type of clusters reflects the strong flow along the area, which is usually generated by the shopping activities or strong tourist attractions along the direction. Thirdly, most of the bridges on the River Thame were highly used. This indicates that there are many tourists has the travel plan included the attractions on both sides of the river. Three bridges appeared to have the most significant movement flows. These were the Westminster Bridge, the Millennium Bridge, and the Tower Bridge. Fourthly, ribbon shape clusters were mainly found in central London, and the clusters in suburban areas were mostly point based. This indicated the movement flows between their neighbourhood and tourist attractions in London neighbourhood are weak.

The comparison between the generated movement pattern and the on-site survey also shows some interesting findings. The coloured circles stand for the average tourist flow collected by on-site survey, while the dark purples lines overlaid at the background reflect the movement flow converted from Flickr data. There is an on-site survey gate located in the Richmond town centre. It indicates the low level of movement flow. However, high movement flow could be found in the Royal Botanic Garden by the Flickr data. This means the location of the pre-defined gate in this area might not be proper. Beside this, there are some new emerged tourist host spots were not included in the 2005 tourist movement report. They are Wembley Park, Olympic Park and Camden Town. They are all well-known and trendy in recent year. This indicates that the tourist intensity of the city is dynamically changing every year. Constant monitoring of the changes is crucial for the research.

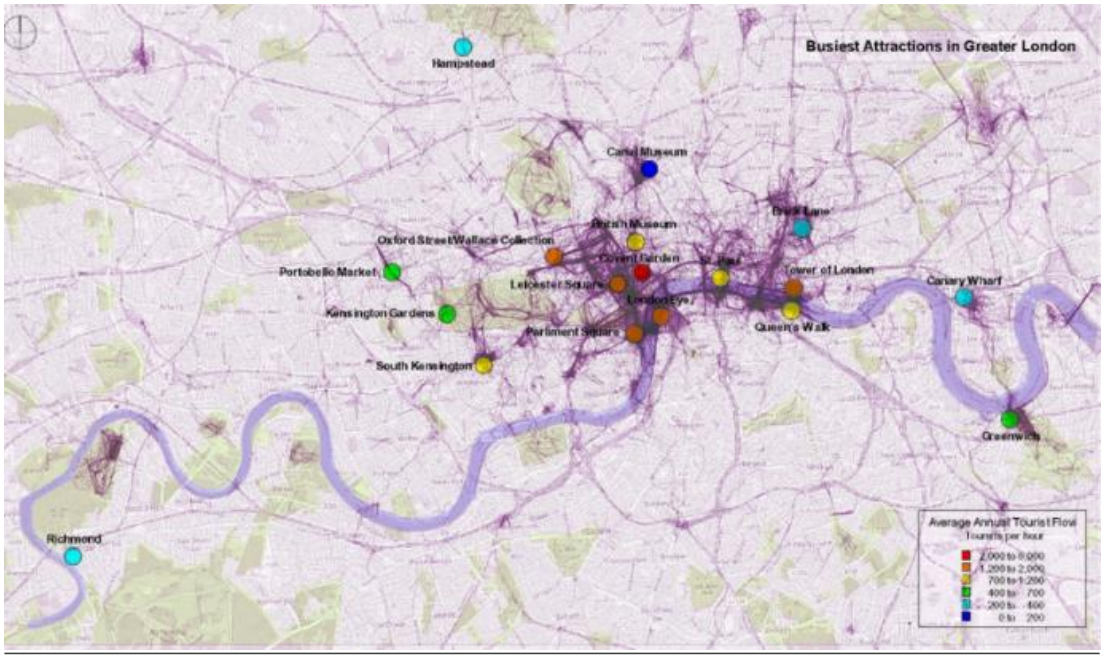

Figure 3C Comparison: Flickr: Topological Flow/ Survey: Average Tourist Flow

The third dataset is the segments map of the tourist movement flow. It covers the converted tourist flow within the centre of London. (Figure 2F) In this data map, the tourist movement pattern has been mapped onto the road segment. In space syntax study, many of the spatial analyses are based on the segment map. Since both the movement data and spatial characters became the properties of the road segments, it will be easier for the researchers to study the relations among them. The map *Figure 3D* contains both the on-site survey data and Flickr converted data. The arrows with colours from red to blue represent the movement data collected by the traditional method. The segment roads with varies of colour and thickness from grey to black reflect the converted movement data from Flickr. The busiest areas and quietest areas from both data source area identical. High movement flow along both side of the river bank, on bridges, Trafalgar square and around Covent Garden.

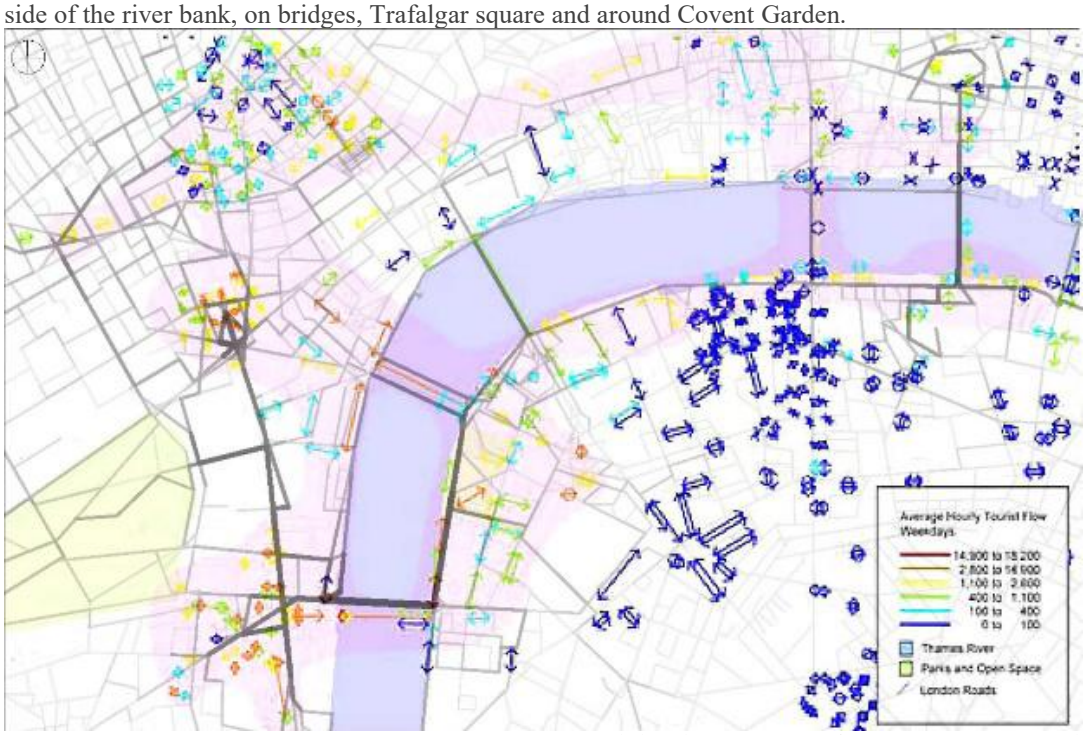

Figure 3D Comparison: Flickr: Flow Segment Map/ Survey: Average Tourist Flow

4. CONCLUSIONS

This research proposed a new data collecting method for tourist movement flow. The data collected through the method covers a wide range of important information for tourist behaviour and urban space study. This method shows the advantages of the traditional survey for data collecting. Firstly, it reduces the cost of the data collecting process. Comparing with the on-site survey, the new method neither requires a survey team nor requires the expensive equipment. All the raw data is already stored on the social media platform, what research needs to do is to process and convert the data. Secondly, it can theoretically cover the unlimited size of the research area. The constraints are the computation power of the machine. Thirdly, the movement data generated by this method covers a wider time range. It shows the data from the first day the social media platform started operating to the most recent day, from the early morning to the midnight, which is much more flexible than the traditional site survey methods.

As the comparison revealed, the movement pattern between traditional movement data collecting method and the proposed method are mostly identical. Despite some of the data from the method are different from the 2005 tourist report data, the differences are in the acceptable range causing by the urban changes. It is still the early experimental stage of the method; however, the positive result reveals a potential and bright future for the tourist movement and space study.

The collected movement pattern has been stored on the target spatial networks. When research aims to compare it with other spatial properties, such as land use, POI, space integration values, it is straightforward for the researchers to run the statistical analysis between these values.

There are some limitations to this method. The movement data accuracy of this method relies on the number of active users of the social media platform. Some of the most widely used platforms such as Facebook does not provide users with the request dataset while the one has the support has the issue of dropping active users. In the next step, the research will focus on developing a movement data retrieving method where the source data is collected from multiple platforms.

It also should be clear that the tourist movement pattern generated by this method is the movement pattern map, it did not reflect the actual value of movement flow. To calculate the actual movement flow, it requires this method to work together with the traditional on-site survey. The relative intensity value will function as the scale factor applying to the actual count of the on-site survey. This will help to transform the regional movement flow map into a city scale movement map.

REFERENCES

Deloitte LLP. Tourism: Jobs and Growth the Economic Contribution of the Tourism Economy in the UK. 2013. Oxford Economics.

https://www.visitbritain.org/sites/default/files/.../Tourism_Jobs_and_Growth_2013.pdf

Great London Authority. 2015. Take A Closer Look Cultural Tourism in London. A Cultural Tourism Vision for London 2015-2017.

https://www.london.gov.uk/what-we-do/arts-and-culture/cultural-places-and-creative-spaces/cultural-tourism-vision-london

Larsen, Jonas. 2008. Practices and Flows of Digital Photography: An Ethnographic Framework. Mobilities 3 (1): 141–60. https://doi.org/10.1080/17450100701797398.

Lo and McKercher B. 2015. Ideal Image in Process: Online Tourist Photography and Impression Management. *Annals of Tourism Research* 52 (May): 104–16. https://doi.org/10.1016/j.annals.2015.02.019.

Lynch K, 1959, The Image of The City, MIT Press.

Nikjoo, Adel, and Hamed Bakhshi. 2019. The Presence of Tourists and Residents in Shared Travel Photos. Tourism Management 70 (February): 89–98. https://doi.org/10.1016/j.tourman.2018.08.005.

Karimi, K. 2005. Tourist Flow in London, Space Syntax Limited.

Turner, R. 2017. The Economic Impact of Travel & Tourism. World Travel Tourism Council https://www.wttc.org/-/media/files/reports/economic-impact-research/regions-2017/world2017.pdf

Stieglitz, S. 2018. Social Media Analytics – Challenges in Topic Discovery, Data Collection, and Data Preparation. International Journal of Information Management 39 (April 2018): 156–68. https://doi.org/10.1016/j.ijinfomgt.2017.12.002.

United Nations, 2008. The Tourism Satellite Account: Recommended Methodological Framework 2008, UN Trade Statistics

Zheng et al, 2018, Understanding the Tourist Mobility Using GPS: How Similar Are the Tourists? Tourism Management 71 https://doi.org/10.1016/j.tourman.2018.09.019.