Discovering tourist attractions of cities using Flickr and OpenStreetMap data

Farzad Vaziri[1]✉, Mirco Nanni[2], Stan Matwin[3,4], and Dino Pedreschi[1]

[1] Department of Computer Science, University of Pisa, Pisa, Italy
surname@di.unipi.it,
[2] KDDLab, ISTI-CNR, Pisa, Italy
name.surname@isti.cnr.it,
[3] Institute for Big Data Analytics, Halifax, Canada
[4] Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland
name@cs.dal.ca

Abstract. Tourism is a growing industry which needs accurate management and planning. Photography and tourism are inseparable; Photographs play the role of tourists' footprints during their visit to a touristic city. Nowadays, the large deployment of mobile devices and digital cameras has led to a massive increase in the volume of records of where people have been and when they were there. In this paper, we introduce a new method to automatically discover the touristic attractions of every single city with the use of two open-source platforms, Flickr and OpenStreetMap. We applied techniques to convert raw metadata of geotagged photos downloaded from Flickr to information about popular Points of Interest with the help of additional information retrieved from OpenStreetMap.

Keywords: Flickr · OpenStreetMap · Point of Interest · Tourist attraction · Geotagged photo.

1 Introduction

Photo-sharing websites such as Flickr contain millions of geotagged photos taken and shared by people from all over the world. These sites accommodate billions of photos, which are publicly available and annotated with different kinds of useful metadata: users, tags, titles and Spatio-temporal information. These crowd-sourced metadata pose new challenges in the domain of Spatio-temporal analysis to investigate and derive high-level human behavior information. Identifying touristic attractions in a city can be used for different purposes, e.g., for city management, urban planners, traffic engineers, and tourism authorities. It can help in planning easier access for tourists and locals, and also produce additional benefits, e.g., by commercial advertising, and providing diverse services. It can also be useful for tourists by recommending the best places to visit in their visiting time. This task can be easy in more popular touristic cities according to different available resources such as tourist information points and online websites, but for low-touristic cities, we would definitely have limited information available presentable to visitors. In this paper, we describe our approach to collect raw metadata from two different platforms and merging them by using novel techniques to discover tourist attractions from publicly available digital footprints. Geotagged photos metadata is increasingly used in tourism studies and urban planning and management [1, 9, 19, 20]. However, a good way of distinguishing visitors and local has not been investigated in previous works. Different techniques have been applied, but they have their own drawbacks that we wanted to address in this paper and solve by our approach.

Finding the tourist attractions for famous touristic cities is not a difficult task considering all the online information resources such as weblogs, websites and any other similar online touristic information providers. Naturally we can find plenty of lists of tourist attractions for cities like London, Paris, Rome, Florence, etc. But if we want this list for cities that are not being visited as frequently as those cities, even if we find some information, it will be very limited. Imagine we want to find out the touristic spots of a city for an application without having any empirical data or information from any other reliable resources. Our work has two main objectives; we want to use two open-source platforms to first, find a list of popular Points of Interest (POIs) for each city and the second, making this process automatic. We want to come up with a technique to discover top touristic spots of a city based on the real data gathered from the tourists. The output of our technique for highly-visited touristic cities will be compared with TripAdvisor to check their similarity, and if they are similar enough, we can generalize the technique for any other city. Since the information on websites like TripAdvisor is generated by users, it means that their lists can be incomplete. After all, we do believe in the wisdom of the crowd, and if we find out that a POI has been photographed by so many tourists but it does not exist in the attraction list provided by TripAdvisor, we can use our method to extend its results.

2 Related Work

Insight into the most attractive tourist places in a city is crucial for the municipality of a city or business planners engaged in strategic planning and decision-making to create a sustainable tourism industry. Finding these attractions will not be a challenging task in well-known touristic cities based on the census data or plenty of other online resources, but it will not be an easy task in smaller and not very popular cities.

Different tasks have been investigated for the analysis of people's activities and behavior using Flickr and Panoramio geotagged photos. Most of the works done upon Flickr and geotagged photos, analyze the GPS information of the photo itself, and they normally do not have information about the location which was the subject of the photographer. Finding interesting and attractive locations was a task that in work [12] they studied by density-based clustering algorithms, such as DBSCAN. Flickr data with Mean-Shift Clustering method has been used to define landmarks as a uniquely represented specific location within the city, such as a sightseeing spot, a store, a building, a bridge and so on [13], but they will not have any idea about the place itself. It can be any kind of listed spots. Flickr geotagged photos have been used in different scale of spatial analysis, e.g., for analyzing travel diaries and patterns to identify preferred destinations and future travel intentions in country and continent scale [21]. In the work shown in [20], the travel behavior of inbound tourists to Hong Kong using geotagged photos of Flickr has been studied. The P-DBSCAN clustering algorithm has been used to identify the location in which tourists are most interested. They found a total of seven clusters, indicating there are seven areas of interest in Hong Kong inbound tourists. Seven clusters definitely do not give sufficiently detailed information about the exact place that tourists have been visited in the area. In [22] researchers used a density-based approach to discover the Regions of Interest (ROIs). For comparison purposes, they also applied mean-shift clustering to identify Regins of Interest, which is the approach used in [13] and [15]. In [2] a bayesian approach has been applied to assess different explanation of how trails are produced across different cities by analyzing sequences of geotagged photos uploaded to Flickr. For the informed specification of hypotheses they utilized additional data resources such as DBpedia [14] and YAGO [16] to add information about POIs. They tried to track the trails of users in a cell-based grid that were made out of POIs location retrieved from those resources. Apart from cell-based zones, researchers tried to recommend the most popular travel path within a region [18]. They developed algorithms to first identify ROIs and then used group trajectories to find the most popular paths to visit these ROIs. Flickr data has been utilized to build TripBuilder framework using Wikipedia to find POIs in a geographical area. The drawback of this work was its dependency on Wikipedia [4]. Even if we know it is a reliable resource, but we also know that it can not be always a complete and absolute solution.

3 Data collection, Cleaning and Fusion

The dataset was used in these experiments were collected by downloading photo metadata from Flickr using the site's public API. We firstly search for photos by providing a bounding box for the city or area that we want to discover the tourist attractions there. The coordinates of the bounding box are in a decimal degree form, which can be determined either from the Flickr itself or by using OpenStreetMap (OSM) or Google Maps. For each city in this paper, we retrieve photos taken from the first of November 2015 until the first of November 2016 that are geotagged to make sure that we have coordinates to further analysis. Using the Flickr API methods we extract metadata tags that contain Photo ID, Owner ID, Hometown, Current City, Country of the owner, Taken time, and GPS information that contains Latitude and Longitude. Table 1 shows the extracted metadata of some sample photos taken in London by different users. These pieces of information have been used in several studies mostly for behavior analysis and finding ROI since the subject of the photos is not provided by Flickr metadata. For overcoming this problem and finding the popular and top touristic attractions of a city we needed additional information that can help us find out what are the subject of the photos and if there are taken from a POI or just random subjects. OpenStreetMap is an open-source and highly used resource in other academic studies, therefore we applied some techniques to first find the photos taken from POIs and then assign the name of the most possible place to the photos. The interestingness and popularity of an attraction is defined with respect to the users own understanding (to take a photo or not), however, if a substantial number of people like to take photos from the same POI, it can suggest that the place is attractive.

We downloaded photos in three European cities: London, Paris, and Rome. And the reason why they have been chosen is that they are among the highly visited cities in Europe, and we have plenty of data for them on Flickr. Figure 1 shows the distribution and dispersal of photos in London city. Since we are interested in studying every single separate POI and not just a zone or region in a city, we needed to collect: (1) metadata of photos to analyze behavior of tourists (2) the information of the points to be able to assign the most accurate name to a photo in case of being a photo of POI.

Table 1. Metadata of photographs extracted from Flickr in a defined bounding box.

Photo ID	User ID	Hometown	City	Country	Time	Latitude	Longitude
40286364662	71393709@N06	Bexleyheath	Bexleyheath	United Kingdom	2016-01-09 11:38:08	51.495283	-0.139898
36941816673	41802762@N03	Ruislip, Middlesex	Ruislip, Middlesex	United Kingdom	2016-01-09 14:23:56	51.496714	-0.144807
37365836965	34427470616@N01	NaN	London	United Kingdom	2016-01-09 12:43:10	51.469116	-0.116745
36553200763	34427470616@N01	NaN	London	United Kingdom	2016-01-09 12:43:23	51.469116	-0.116745
35922940183	41802762@N03	Ruislip, Middlesex	Ruislip, Middlesex	United Kingdom	2016-01-09 16:49:44	51.402257	-0.194545

OpenStreetMap is an open-source platform that provides data about *Nodes*, *Ways* and *Relations* for every geographical area in ***pbf*** file format [5]. For every element, there is a variety of tags. These tags are part of the metadata of the OSM maps, and they provide information about the location such as Name, Latitude, longitude, and OpenStreetMap Features tag which contains information about the category and the type of the location. Considering this work is all about tourists and points of interest, among the *Nodes*, *Ways*, and *Relations*, we filter *Ways* and *Relations* out and keep the specific *Nodes* by specifying the type and the category of the *Nodes* that we want. OpenStreetMap represents physical features on the ground (e.g., roads or buildings) using tags attached to its basic data structures (its *Nodes*, *Ways*, and *Relations*). Each tag describes a geographic attribute of the feature being shown by that specific *Node*, *Way* or *Relation* [6]. OpenStreetMap's free tagging system allows the map to include an unlimited number of attributes describing each feature. The community agrees on a certain key and value combinations for the most commonly used tags, which

[5] `https://wiki.openstreetmap.org/wiki/PBF_Format`
[6] `https://wiki.openstreetmap.org/wiki/Map_Features`

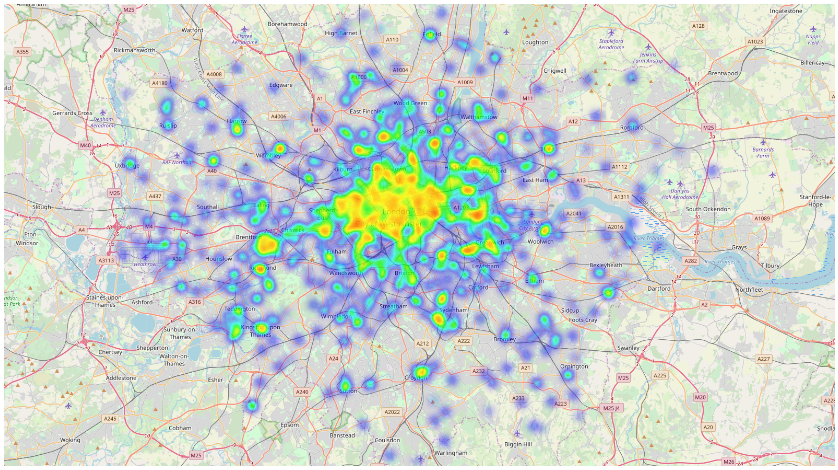

Fig. 1. Heat Map of distribution and dispersal of photos in London city.

act as informal standards. However, users can create new tags to improve the style of the map or to support analyses that rely on previously unmapped attributes of the features. Later on, we collect these features as a description of the category of the POIs. The values of the keys can specify the category and the type of the *Node*.

The data we are dealing with is encumbered by a number of issues that need to be solved before its use of analysis and visualization. In our work, we use open-source tools related to OpenStreetMap, i.e., OSMOSIS and OSMCONVERT. Osmosis is a Java command-line application for processing OSM data. The tool consists of pluggable components that can be chained to perform a larger operation. For example, it has components for reading/writing databases and files, deriving/applying changes to data sources, and sorting data, etc [7].

Before using OSMOSIS, we use OSMCONVERT to convert and process OSM files [8]. With it, we extract *OSM ID* which is unique for every single *Node*, *Latitude*, *Longitude*, *Name* of the *Node* and the *Keys* that have been picked as tourism features of *Nodes*. They become key features we are going to use for distinguishing the categories of the POI. In the first step, OSMOSIS is used to filter a *pbf* file. The bounding box of the city or the area that we downloaded from the Flickr API is filtered and then from the Map Features, keys that are related to tourist attraction are picked. Such keyword-based selection allows us to extract touristic *Nodes* from others. The result is, therefore, a map containing the metadata of the *Nodes* that are related to tourism in a defined bounding box. Table 2 shows the information about POIs extracted based on these tolls and techniques from OSM.

[7] https://wiki.openstreetmap.org/wiki/Osmosis
[8] https://wiki.openstreetmap.org/wiki/Osmconvert

Table 2. Information extracted from the OpenStreetMap about the POIs.

OSM ID	**Latitude**	**Longitude**	**Name**	**Category**
36969174	51.3601417	-0.2166396	*Cheam Library*	library
257810726	51.6665265	-0.1425318	*George Grey's Obelisk*	monument
267470837	51.4956267	-0.2548747	*The Tabard Theatre*	theatre
268039103	51.596843	-0.1951207	*Stephens Collection*	museum
268988407	51.4890059	-0.3667031	*The Ramblers*	artwork
370187938	51.587506	-0.131498	*Hornsey Moravian Church*	place of worship

We use a simple strategy to deal with duplicated location names. Duplication happens in two circumstances: (1) *Nodes* are in the same geographical area and simply have different components e.g., in a park every single statue can have the name of the park on them but with different ID and latitude and longitude (2) *Nodes* that are completely different but just accidentally happened to have an identical name. For the first type of duplicated names, we try to aggregate those that are close to each other by defining a radius and assign the same name for them and calculating a new latitude and longitude by calculating coordinates in between them. And for the second type, we simply define a new name for each of them. At the end of this process, we will have photos that are assigned to POIs with unique information.

4 Methods and experiments

So far we have two separate datasets that by their own can not provide any information about attractions and tourists: one provides the photos of the user, but we do not know if they are tourist or no. Moreover, we have no idea if these photos have been taken from tourist attractions or no. The other supplies *Nodes* that are likely related to tourism based on the initial filter we applied on the features of the *Nodes*. But we do not know if they are a POI, and how popular are there among visitors. We are facing two main problems: (1) what are the subject of photographs (2) if the user who has taken the photo is a local or a tourist. For solving the first problem, our heuristic for assigning a name to a photo is that a tourist will to take a picture of POI from a certain bounded distance. We bound this distance with a threshold based on the scale of the city. If a photo has been taken inside of a predefined radius of distance of a *Node*, we can assume that the subject of the photo is that specific *Node*. We are then confronted with the problem of calculating the distances between *Nodes* and photos. Even after applying some filtering and selecting tourism-related *Nodes* from the OSM data we still have a massive set of distances to compute. Calculating the distance of every single photo from every single *Node* is computationally too expensive for the resources at our disposal. Since we have coordinates of photos and *Nodes*, we use **KD-tree** spatial indexing methods to find the nearest neighbor for each photo. For every photo and *Node* we have Latitude and Longitude therefore we store data organized by xy-coordinates. The general idea behind KD-tree is that it is a binary tree in which every leaf node is a k-dimensional point. Every non-leaf node can be thought of as implicitly generating a splitting hyperplane that divides the space into two parts, known as half-spaces. Points to the left of this hyperplane are represented by the left subtree of that node and points to the right of the hyperplane are represented by the right subtree. The hyperplane direction is chosen in the following way: every node in the tree is associated with one of the k dimensions, with the hyperplane perpendicular to that dimension's axis. So, for example, if for a particular split the "x" axis is chosen, all points in the subtree with a smaller "x" value than the node will appear in the left subtree and all points with larger "x" value will be in the right subtree. In such a case, the hyperplane would be set by the x-value of the point, and its normal would be the unit x-axis [3]. Using this algorithm we find the closest *Node* to every single photo [17], if the distance between them is less than the predefined threshold, we assign the name and other information of that *Node* to the photo. Photos that do not have any nearby *Node* within the predefined distance we assume that they are just random photos without any subject related to a POI, and we simply remove them from the dataset. Now, we can say that the remaining photos in the dataset are taken from places that most probably are touristic attractions, but we still can not be sure since locals can take a photo of a random place which is not from city attractions. In this phase, we deal with the users, and we need to distinguish tourists from locals. Photos taken by tourists can be more related to city attractions than local users' photos.

One of the simplest ways that already has been used in previous studies is checking the duration of stay. Researchers calculated the difference between the time-stamps of the users' first and last photographs taken in the specific area. If the difference is smaller than a predefined number of days, the user can be considered a visitor; otherwise, he/she is a local [5–8]. In these studies, they used a threshold of 30 days in the region of Florence which is appropriate on the regional level. This technique followed in another research [11] but since their study was in a smaller scale and in urban tourism, they decided to have a shorter period as thresholds that correlates better with the duration of stay of typical visitors in touristic cities. They analyzed and visualized tourist movement in Budapest and an average of 3 days in the case of the city selected for their study was their threshold. In another study, the same technique for distinguishing tourists and locals in some European touristic historic cities used to measure activities in those cities [10]. This technique is an efficient way to determine whether a photo is taken by a local or by a tourist, but we know that different cities according to their geographical size and also the number of attraction points that they can have will need variant stay time to visit main POIs. Duration of stay can vary city by city and will not be a trustworthy way to determine locals and tourists in overall. Flickr API has a method that can provide us some information about the users' profile such as the "Country", "Hometown" and the "Current City" of a user. This additional information can give us a much more precise idea about if a user is tourist or local than just depending on the duration of stay. This solution, however, requires additional API method call for every single photo. This is going to happen once in the data collection process and further ahead, they will not add any extra cost in our complexity. Considering that not all the users provide this information in their profile, we should have another strategy to be still able to distinguish them. Hence, we still need the duration of stay technique in case of missing information about the "Hometown" or "Current City" of the user. Our final approach to determine locals and visitors uses both techniques, but conditionally. If information about a user is available, it will rely on that, otherwise, the duration of the stay will be applied.

We initially check if these extra data about users are available, if yes, then we check if the "Country" and one of the "Current" or "Hometown" cities and desired place are identical. The user is determined as a local if they match. For example, if someone's either hometown or current city is London, he is not counted as a tourist. We also check the country, because there are cities in different countries with the same name, e.g., there is London in England, the United States and Canada. In case of lacking this information, we go back to the "duration of stay" technique. We calculate the time difference between the first and last photo taken by a user in the city, and if this time interval is bigger than a predefined threshold, he/she is not a tourist. This stay time can be defined based on the size of the city, how touristic is it and how long does normally take to be visited by a tourist.

Now, we have made a dataset, containing the photos of POIs taken by tourists who have been visited the city. For finding out what POIs are the real tourist attraction, the frequency of appearance of a POI in the dataset can be an evident measurement to extract top attraction, but we need to keep in mind that, it can be affected by users interest. Imagine a user like a place and takes several photos of that place, it does not mean that the place is really popular compared to another place that has fewer pictures. A more reasonable parameter can be the number of users that have taken pictures of a POI. For example, if POI A has been photographed by 10 users, it is more popular than a POI that has been photographed by 5 users even if in total place B has more photos in the dataset. In the end, for having a better and more reliable output, we filter out the places that have been photographed by less than a certain number of users. We apply the same method for the users that have photographed less than a certain number of places. The information coming out of tourists that take pictures regularly during their stay in different places would be much more informative than those with a few photographs. After filtering these POIs and users out, we list them based on their popularity. Table 3 shows top attractions for London city, derived from our method and TripAdvisor and Wikipedia websites. Most of these POIs presented by these two online resources, either showed up with same or similar names in our list or with different names but pointing to the same geographical location. They might have been tagged based on different subject. For example a place can be tagged on a name of sculpture instead of the location that the statue has been located there.

Table 3. Tourstic attractions provided by TripAdvisor and our method for London city.

Extracted from Flickr&OSM data	**Listed in TripAdvisor and Wikipedia**
Big Ben, Tower Bridge, International Brigades Memorial, George Washington, Winston Churchill, Buckingham Palace Millennium Bridge, Covent Garden Market, Westminster Scholars War Memorial, Traitor's Gate, Manufacture, British Optical Association Museum, Oxford Circus, Through Blue, Agatha Christie Memorial, Covent Garden Piazza, London Film Museum, Carnaby Street, Shakespeare, Victoria Museum, Betjemen Sculpture, Horatio Nelson, The Cenotaph, Minerva, War Memorial, Bridge Theatre, South Bank Book Market, St Jame's Park, Nelson Mandela, Brick Lane,London Bridge	Tower of London, The British Museum, Churchill War Rooms, Tower Bridge, National Gallery, Westminster Abbey, V&A - Victoria and Albert Museum, Borough Market, Natural History Museum, Hyde Park, Millennium Bridge St. Paul's Cathedral, Chelsea FC Stadium Tour & Museum, Houses of Parliament, Covent Garden, Camden Market, Shakespeare's Globe Theatre, Greenwich, Covent Garden Piazza, Buckingham Palace, Regent's Park, Imperial War Museums, Emirates Stadium Tour and Museum, Kensington Gardens, Wallace Collection, Museum of London, HMS Belfast, Highgate Cemetery

5 Conclusion

Different investigations used Flickr data to study human movements more specifically tourist behavior. Even if they applied different techniques to determine who is a tourist and who is a local, their methods do not seem to be very accurate. Since Flickr metadata does not provide any information neither about the subject of the photos nor type of user (local or tourist), therefore none of those techniques can be completely reliable for discovering single and separate POIs. For user type discernment, the duration of the stay is a technique which as been used in different studies. In this work, we added user information to have much more precise distinction between locals and tourists. Comparing the country, hometown and current city of the user with the city that photos have been taken can be more intuitive than just the duration of stay. Most of the works are focused on tourist behavior, activity, and task analysis but very few of them tried to extract top attractions out of these data. They either used clustering techniques to find ROI or tried to use additional information about POIs to make cellular-based layout and track trails of users. In this paper, we tried to provide extra information using OpenStreetMap to make the real POIs extraction possible by relying on the wisdom of the crowd. Our extracted list may not match completely with the list of attractions provided by websites like Trip-advisor or Wikipedia but most popular ones pop up in both lists. There are cases that might have a slightly different name but still the same places or even named differently. Accordingly, we can generalize the method and apply it on not highly touristic cities where we do not have many resources about touristic attractions. As we mentioned at the beginning cities like London has much available information but for other cities in the country, our technique can help to extend limited available resources.

Another achievement of this study is about helping other information providers. We all know new attractions have been made or removed in cities and all the websites need to be updated. Moreover, we know that websites like TripAdvisor and Wikipedia are open and users add information into their database and it can not be a complete and pure source of city attractions to rely just on it. The POIs that are popular in our provided list, that do not exist in this website can actually be a real attraction that needs to be added into these lists. If there are so many people taking pictures of a place, most probably there is an attraction that grabs the photographer's attention. And in addition there is a possibility that these places that have not been listed on those websites, they worth to visit whereas a lot of tourists have been visiting them.

The results of this work have great potential to be carried out in tourist movement analysis and predictions. After all, we can make a trajectory for every single tourist according to the photos timestamp and investigate predictability of his/her behaviour.

Acknowledgements

The work and results reported in this paper are part of broader research in tourist movement analysis that I did during my Ph.D. at the University of Pisa, Italy. I want to thank Prof. Matwin who is one of the co-authors of this paper because of all his support and help alongside my own supervisors (Dr. Nanni and Prof. Pedreschi) to finish my Ph.D. and finalize this research.

References

1. Barchiesi, D., Preis, T., Bishop, S., Moat, H.S.: Modelling human mobility patterns using photographic data shared online. Royal Society open science **2**(8), 150046 (2015)
2. Becker, M., Singer, P., Lemmerich, F., Hotho, A., Helic, D., Strohmaier, M.: Photowalking the city: Comparing hypotheses about urban photo trails on flickr. In: International Conference on Social Informatics. pp. 227–244. Springer (2015)
3. Bentley, J.L.: Multidimensional binary search trees used for associative searching. Communications of the ACM **18**(9), 509–517 (1975)
4. Brilhante, I., Macedo, J.A., Nardini, F.M., Perego, R., Renso, C.: Where shall we go today?: planning touristic tours with tripbuilder. In: Proceedings of the 22nd ACM international conference on Information & Knowledge Management. pp. 757–762. ACM (2013)
5. Girardin, F., Calabrese, F., Dal Fiore, F., Ratti, C., Blat, J.: Digital footprinting: Uncovering tourists with user-generated content. IEEE Pervasive computing **7**(4), 36–43 (2008)
6. Girardin, F., Calabrese, F., Dal Fiorre, F., Biderman, A., Ratti, C., Blat, J.: Uncovering the presence and movements of tourists from user-generated content. In: Intnl Forum on Tourism Statistics. Citeseer (2008)

7. Girardin, F., Fiore, F.D., Ratti, C., Blat, J.: Leveraging explicitly disclosed location information to understand tourist dynamics: a case study. Journal of Location Based Services **2**(1), 41–56 (2008)
8. Girardin, F., Vaccari, A., Gerber, A., Biderman, A., Ratti, C.: Quantifying urban attractiveness from the distribution and density of digital footprints (2009)
9. Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W., Prasad, S.: Extracting and understanding urban areas of interest using geotagged photos. Computers, Environment and Urban Systems **54**, 240–254 (2015)
10. Kádár, B.: Measuring tourist activities in cities using geotagged photography. Tourism Geographies **16**(1), 88–104 (2014)
11. Kádár, B., Gede, M.: Where do tourists go? visualizing and analysing the spatial distribution of geotagged photography. Cartographica: The International Journal for Geographic Information and Geovisualization **48**(2), 78–88 (2013)
12. Kisilevich, S., Krstajic, M., Keim, D., Andrienko, N., Andrienko, G.: Event-based analysis of people's activities and behavior using flickr and panoramio geotagged photo collections. In: 2010 14th International Conference Information Visualisation. pp. 289–296. IEEE (2010)
13. Kurashima, T., Iwata, T., Irie, G., Fujimura, K.: 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. ACM (2010)
14. Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P.N., Hellmann, S., Morsey, M., Van Kleef, P., Auer, S., et al.: Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia. Semantic Web **6**(2), 167–195 (2015)
15. Lu, X., Wang, C., Yang, J.M., Pang, Y., Zhang, L.: Photo2trip: generating travel routes from geo-tagged photos for trip planning. In: Proceedings of the 18th ACM international conference on Multimedia. pp. 143–152. ACM (2010)
16. Mahdisoltani, F., Biega, J., Suchanek, F.M.: Yago3: A knowledge base from multilingual wikipedias (2013)
17. Maneewongvatana, S., Mount, D.M.: Its okay to be skinny, if your friends are fat. In: Center for Geometric Computing 4th Annual Workshop on Computational Geometry. vol. 2, pp. 1–8 (1999)
18. Shafique, S., Ali, M.E.: Recommending most popular travel path within a region of interest from historical trajectory data. In: Proceedings of the 5th ACM SIGSPATIAL International Workshop on Mobile Geographic Information Systems. pp. 2–11. ACM (2016)
19. Vu, H.Q., Leung, R., Rong, J., Miao, Y.: Exploring park visitors activities in hong kong using geotagged photos. In: Information and Communication Technologies in Tourism 2016, pp. 183–196. Springer (2016)
20. Vu, H.Q., Li, G., Law, R., Ye, B.H.: Exploring the travel behaviors of inbound tourists to hong kong using geotagged photos. Tourism Management **46**, 222–232 (2015)
21. Vu, H.Q., Li, G., Law, R., Zhang, Y.: Travel diaries analysis by sequential rule mining. Journal of travel research **57**(3), 399–413 (2018)
22. Zheng, Y.T., Zha, Z.J., Chua, T.S.: Mining travel patterns from geotagged photos. ACM Transactions on Intelligent Systems and Technology (TIST) **3**(3), 56 (2012)