ENRICHING WALKING ROUTES WITH TOURISM ATTRACTIONS RETRIEVED FROM CROWDSOURCED USER GENERATED DATA

M. Mor*, S. Dalyot

Mapping and Geoinformation Engineering, The Technion, Haifa, Israel
matan.mor@campus.technion.ac.il, dalyot@technion.ac.il

Commission IV, WG IV/5

KEY WORDS: Social Media, Trajectory Reconstruction, Crowdsource Geotagged Photos, Tourism Context

ABSTRACT:

It is always a tourism challenge - and aspiration - to discover scenery routes and tourism attractions in unfamiliar areas. Tourism information is getting more extensive, comprehensive and complex, so first-time tourists have to manage and mine large volumes of data to better plan their trip. Nowadays, geotagged photos are uploaded by users to social media photo-sharing online websites, which become more popular and commonly used by travelers to share their tourism experiences. Handling, mining and interpreting these user-generated 'digital footprints' can be used to reconstruct travel trajectories of users to recover their activity and knowledge. In this research, we showcase Flickr geotagged crowdsource photo database as a source for mining users' trajectories to effectively compute walking tourism routes. Our methodology mines tourism context by conceptualizing a set of adaptive spatiotemporal descriptors to identify photographers that show tourism activity of first-time visitors. By implementing spatial clustering, we find popular locations that are traversed by these tourism-oriented photographers' trajectories. To analyze our approach, we develop a greedy route computation algorithm that seeks the most popular traversed locations between origin and destination points defined by the user. Results for two cities are presented, proving the robust mining and retrieving of valuable tourism context and information from social media photos. We evaluate and validate our results by comparing the computed walking routes to recognized tourism information. The computed walking routes are scenery and pass through the main popular tourism sights and landmarks in the city, including additional attractive places that are frequently visited by tourism-photographers.

1. INTRODUCTION

Trip planning is a task that requires skilled interaction with different sources (Choudury et al., 2010): one may rely on tourist guides and books, while others will use internet search engines and ask for recommendations from family and friends. Consequently, smart tourism solutions are developing fast, aspiring to keep track of the growing diverse travelers' demands and needs. Gradually, current solutions rely on the cumulative volumes of user-generated online repositories that did not exist until recently. These are constantly updating, becoming more accessible and straightforward to use. Thus, paper maps and tour guides are being replaced with internet and mobile guides, internet search engines and blogs.

Crowdsource geotagged user-generated data and information sources are increasing dramatically; social media websites, such as Twitter and Facebook, are becoming more commonly used to document and share peoples' daily activities. Similarly, Flickr allows photographers to depict and share their everyday activities and events, including tourism activity, relying mostly on geotagged photos. Including scenic and attractive places documented in social media activities, i.e., digital footprints, can significantly add to tourism experience. Social media crowdsource trajectories can provide up-to-date landmark-related routes that reflect current public interest (Basiri et al., 2017). Still, understating and extracting the human activities and patterns, which are based on location data only without knowing their context and labels, i.e., tourism in our case, is a complex challenge, mainly since only a share of these photos are related to travel and tourism experience.

In this paper, we investigate three main algorithmic challenges associated with online heterogenous crowdsource user-generated geotagged-data to compute tourism routes, where each relies on the output of the previous one: first, understanding which user (photographer in our case) traversing the explored space can be considered as a valid first-time visitor (tourist), since not all the data uploaded by photographers have tourism context; second, analyzing the space traveled and traversed by these tourism-oriented photographers, to discover popular tourism places (geolocations), as well as the tourism activities (patterns) that can be used as tourism information elements; third, computing walking routes, which use the retrieved information while maximizing the tourism popularity.

In this study, we focus on the development of algorithms to automatically compute tourism-oriented walking routes from Flickr, interpreted from the travel trajectories and activities of tourism-photographers. Instead of relying on and analyzing numerous photos that might not have tourism context, we aspire to identify users (photographers) that can be considered first-time visitor tourists to better retrieve tourism context. Accordingly, we rely on the experience of users who are non-locals, thus minimizing the potential of relying on the global experiences of many general photographers in the area that do not explicitly contribute to tourism information. These rely on the knowledge that local photographers normally take photos all over the city, while tourism-oriented photographers show distinct activities that relate to tourism landmarks, attractive and scenic places (Kadar, 2014). To this end, we analyze the metadata attributed to the geotagged photos from the Flickr database, automatically reconstructing travel trajectories while considering an array of spatiotemporal adaptive key descriptors to classify tourism-photographers. A cell-grid approach is then implemented to find popular cells traversed by their trajectories, where cells that show distinctive activities in terms of users and photos are indexed. A

* Corresponding author

minimum distance constrained heuristic greedy route computation approach is developed with the objective of maximizing the tourism popularity score, thus automatically seeking and computing the most comprehensive and attractive tourism walking route. Experimental results for two tourism areas - Tel-Aviv, Israel, and Manhattan, New-York, USA - are presented and discussed, analyzing the performance of the developed algorithms, qualitatively evaluating their tourism value using known familiar tourism information. Results show that the algorithms can generate tourism routes that pass through the main attractions and landmarks, while discovering additional popular places that although less known still attract photographers. The results are promising and can serve as a complementary source for computing tourism walking routes, focusing on the alternative free and easy-to-access source of tourism information.

2. RELATED WORK

Since numerous sources of tourism information exist, e.g., internet websites, travel agencies and travel guides, having diverse opinions and sometimes contradicting information, travel planning can be a complicated task to achieve (Li et al., 2015). Sharing tourism information as in 'digital footprints' via different social online platforms (e.g., Facebook, Twitter, Instagram and Flickr) is becoming increasingly popular, mainly since nowadays the use of cameras and smartphones equipped with embedded GNSS sensors makes it easy to document, upload and share geotagged photos (Farzanyar and Cercon, 2015; Lim, 2015).

Crowdsourcing user-generated geotagged photos location can be ambiguous and misleading, or inaccurate. Hauff (2013), for example, showed that the location accuracy of Flickr in urban areas is better than 15 meters, whereas in rural areas it can reach hundreds of meters. Yang and Tang (2016) showed that location accuracy analyzed based on the photographers' trajectories, can be very low, even in urban areas, mostly due to reduced accuracy related to the GNSS measurement factors (e.g., multi-path). Other problems of shared data can occur, such as: duplicated data (e.g., Paldino et al., 2015), misplaced data when photos are taken from a distant position and geotagged to that location (Choudury et al., 2010), missing tourism information (e.g., landmarks) due to the prohibition of authorities to take photos (Salas-Olmedo et al., 2017), or biased data, since not all tourists make use of online social networks to share their photos (Garci-Palomarees et al., 2015). Hence, aggregation of location data is required, together with filtering and interpretation, to extract information that is relevant, consistent and homogenous.

In addition to filtering erroneous location data, methods for removing photos that do not contribute information relevant to tourism are developed. This is achieved mainly by filtering photos taken by locals, which could be considered irrelevant since they do not always present tourism context and activity - as opposed to photos taken by tourists. Accordingly, understanding and defining tourism activity is a mandatory preliminary stage for pre-processing geotagged photos designed to retrieve tourism patterns, context, and knowledge (Girardin et al., 2008). Salas-Olmedo et al. (2017), for example, presented a spatial autocorrelation comparison analysis of photographers' geotagged photos distribution in cities among different social sharing platforms (Twitter, Foursquare and Panoramio). The analysis used data collected over a limited time period of one week, under the assumption that this is the average visit time of tourists in cities. De Choudhury et al. (2010) set a limited duration of 21 days of travel itineraries, while Girardin et al. (2008) and Garci-Palomares et al. (2015) determined that the temporal criterion of visit duration of tourists will not exceed one month. The latter showed that there exists a differentiation in the geospatial distribution of tourists' photos, which are concentrated around the city's main sightseeing spots, as opposed to the locals' photos, which are more dispersed throughout the city. Kadar (2014) supports these findings, showing that tourists who have limited time usually consume tourism in known popular and tourism places, and accordingly will document these places with photos, whereas locals take photos in most city sites. Additionally, according to Kadar & Gede (2013), tourists uploading photos to Flickr vary, originating from diverse social groups involved in tourism consumption, hence it is possible to differentiate these tourism groups according to their unique activity.

To discover popular and attractive tourism places, statistical cluster processes are mostly used. Faraznyar and Cerocone (2015) and Subramaniyaswamy et al. (2015), for example, use the Mean Shift algorithm for arranging big data of geotagged photos to find group centroids. Other studies propose the use of DBSCAN and its variants (e.g., Sun et al., 2015, Korakakis et al., 2016, and Zhang et al., 2018). Becker et al. (2015) and Ali et al. (2013) choose to divide the geographic framework to equal cell sizes, searching for "popular" (cluster) cells, while Doytsher et al., (2017) developed a partitioned space to 3D equal size grid that copes also with the temporal dimension. Hio et al. (2013) develop a grid-based Region of Interest mining algorithm that considers density changes with neighboring cells to discover interesting regions. Temporal methods are also used for extracting valuable information about tourism phenomena in Places-of-Interest (POI) locations. Hsieh et al. (2012), for example, use temporal information to recommend time-sensitive trip routes by investigating the actual visit time and visit duration at a specific location. Lim et al. (2016) use timestamps to calculate the popularity of POIs based on the time a tourist spends at that location. Basiri et al. (2017) proposed an approach that checks travel speed between consecutive segments to find the average speed to classify the travel mode and discover stay points that can imply meaningful POIs along the routes. Aggregation of information retrieved by combing spatial and temporal parameters, as suggested in our research, can extend the above with information on tourists' characteristics, common tourism activities and places of visits.

Researchers aim to discover comprehensive tourism routes from user-generated information by using the calculated POI geographic information to guide photographers to locations that demonstrate tourism features (e.g., Lim (2015), Li et al., (2015), Choudury et al., (2010), and Yahi et al., (2015)). Yahi et al. (2015), Brilhante et al., (2015), and Mor and Dalyot (2018), for example, use Google Maps API to compute routes between POIs, while Sun et al. (2015) use the Dijkstra algorithm with weighted popularity road matrix to compute the routes. Chen et al., (2017) enriched the road network with scenic score considering density and dominate direction of geotagged photos and check-ins near each segment. Moreover, research focusing on walking route planning mainly studies the urban trails to retrieve tourism sights in a city. Becker et al. (2015) extracted photo trails for a single photographer, regardless of the time span, by comparing the extracted trail to a weighted POI geographic layer to explore tourists' patterns. Ali et al. (2013) and Choudhury et al. (2010) analyze photographers' pattern by a comprehensive study of their photo sequence to evaluate tourism routes, while Brilhante et al. (2015) used Wikipedia POIs information to analyze crowdsourced tourism trajectories. To extend these approaches, we aim to generate comprehensive tourism routes by investigating and interpreting the accumulated trajectory patterns of the tourism-photographers, without relying on external POIs information.

3. METHODOLOGY

We aim to crowdsource geolocations of Flickr photos to retrieve tourism context supporting the computation of tourism routes. User-generated photos can show data ambiguities and noise, together with the fact that only a share is taken during a tourism activity; thus, not all downloaded photos will have a tourism context. Accordingly, we rely on a higher data level that is the photographers (users), analyzing and interpreting their trajectories, dramatically reducing the discrete data volume needed for investigation and analysis to extract and interpret valuable continuous tourism context.

3.1. Geotagged Photos and Metadata

Flickr stores explicit metadata information (EXIF) about the uploaded photo and the user who took it. Each photo stores the geographic location in (X, Y) values, its URL, a Time Stamp, and a User ID.

3.2 Defining the Tourism User

We aspire to primarily rely on geotagged photos that have tourism context of first-time visitors, i.e., filter photos that were taken by locals or casual photographers passing through the area. Accordingly, we identify photographers that show tourism activity and tourism descriptors by processing spatiotemporal information retrieved from the trail of the user's geotagged photos. In addition, geotagged photos of a specific photographer that store identical spatial and/or temporal metadata are filtered, assuming redundant or erroneous data.

Using a rule of thumb temporal threshold related to the presumed visit duration of first-time tourists to the area, e.g., one-week or three-weeks, is not optimal. This is since different urban structures and users will most evidently show different activity patterns that affect tourism descriptors (Liu et al., 2018). Accordingly, we rely on the photographers' collective activity to find attractive tourism locations in the analyzed urban area. Since tourists mostly show similar tourism consumption for a specific area (e.g., Kadar, 2014, and Liu et al., 2018), we employ a set of adaptive tourism descriptors to compute the average values that describe the common activity of first-time visitors in a given area; values are calculated for each descriptor as follows:

1. Travel time (temporal): the time interval between two consecutive photos should be at least one second, and no more than a defined travel time threshold. A single trip travel time (including multi-day trips) threshold is calculated by analyzing the changing volume of photographers per day, and searching for the most shared visit duration among first-time visitors under the assumption that tourists usually stay for a short period (Yang et al., 2017). The average visit duration time (t_{avg}) (Equation 1) among (n) photographers is defined as the photographer's (i) travel time itinerary in the analyzed city between the first (t_s) and the last (t_e) geotagged photo timestamp.

$$t_{avg} = \frac{\sum_{i=1}^{n} t_e^i - t_s^i}{n} \quad (1)$$

2. Number of photos (volume): a tourism photographer will most evidently take several photos during a single trip (travel) while sightseeing and observing new and attractive sights. Implementing a threshold of at least three photos per user (e.g., De Choudhury et al. (2010)), is a mandatory step for retrieving a valuable trajectory of a tourism photographer.
3. Travel distance (spatial): the photographer is exploring the area by moving between distinct urban landmarks that are well distributed and spread in the area (Becker et al., 2015). Evaluation of a travel adaptive threshold is conducted by retrieving the average accumulated Euclidean distance (D_{avg}) (Equation 2) between ($j=1:m$) photos for each photographer ($i=1:n$).

$$D_{avg} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} \sqrt{(X_{j+1} - X_j)^2 + (Y_{j+1} - Y_j)^2}}{n} \quad (2)$$

4. Travel speed (spatiotemporal): the average travel speed V_{avg}, including multi-day trips, is calculated (Equation 3) according to the accumulated traveled distance (D_i) divided by the time interval between the last (t_e) and first (t_s) photo timestamp. Outliers of photographer's speed higher than 10 km/hr are excluded to ensure walking activity only, and excluding different mobility patterns (e.g., car, cycling).

$$V_{avg} = \frac{\sum_{i=1}^{n} \frac{D_i}{t_e^i - t_s^i}}{n} \quad (3)$$

Figure 1 depicts the pseudo-code for identifying tourism photographers. For each user in the *Users* array, we check the set of the aforementioned four adaptive descriptors. In case that the user validates all, we label him/her as a tourist, and store his/her information with the geotagged photos' metadata in a data structure formulated to analyze each photographer trajectory to investigate his/her activity patterns.

```
for each User in Users do
    if (User took more than N photos) And
      (0 > TimeResolution(betweenPhotos) < M_days) And
      (Distance(Route) < Y_Kilometers) And (Speed(Route) < S_speed)
    then
        Save (User);
        for each Photo in Photos do
            if (photo's user id = user) then
                Save (photo);
            end
        end
    end
end
```

Figure 1. Pseudo-code for identifying tourism photographers.

3.3. POI Identification

To investigate and extract the main popular and frequently traversed locations (POIs) by the photographers for a defined area, and to analyze the activity pattern according to the photographers' trajectories, we use a cell-grid structure approach. This allows for fast and intuitive analysis of the photographers' activity patterns in space by analyzing the cells traversed by their trajectories. Becker et al. (2015) defined a cell size of 200X200 meters for Berlin, whereas Eli et al. (2013) defined a cell size of 125X125 meters for Amsterdam. Additionally, Brilhante et al. (2015) showed that a value of 200 meters represents the approximated distance among POIs in Wikipedia. Accordingly, we have decided to use a cell size of 250X250 meters, also taking into consideration GNSS localization errors in urban areas (inaccurate geo-tagging values of the photos). Also, since we analyze walking routes that are mostly at least several hundreds of meters, the cell-based approach should ensure that each cell will have a unique POI, overcoming alternative clustering processes that use non-adaptive Kernel values (e.g., density-based) that can lead to numerous dispersed photos clustered to a single POI – instead of several ones. For each cell, the most attractive geographic location of the physical POI location relies on a centroid calculation that is derived from the geographic

location of all geotagged photos that fall in the cell's extent. According to this approach, instead of analyzing photos per cell to measure popularity, we analyze and find the cumulative photographers' visits. Our hypothesis is that this will provide a better indication regarding the popularity score of each cell: cells with a higher volume of visiting photographers, which are defined as tourists (according to Section 3.2), are considered more attractive, and hence frequently visited by tourists. Ranking of cells according to their tourism popularity measure enables us to better understand the actual tourism consumption for a given area, in terms of patterns and activities.

3.4 Route Computation

For seeking the most popular cells between the route's origin and destination points, we develop a heuristic greedy route optimization algorithm that maximizes the popularity score of the computed route based on a defined hierarchal ranking model. This approach ensures implementation according to the minimum distance constraint between the origin and destination points to avoid self-intersection, while enriching the tourism experience by traversing the popular cells. The model defines a set of three criteria for choosing popular cells in the greedy approach, according to this descending significance level:

1. The number of tourism photographers: the adjacent cell that has the highest number of tourism photographers, validating an evident tourism activity.
2. The number of geotagged photos: the adjacent cell that has the highest number of geotagged photos, to guarantee diversity of tourism activity.
3. Euclidean distance: the closest cell to the destination point.

A routable graph $S \rightarrow T$ between the origin point (S) and the destination point *(T)* is defined as a directed graph $G = (V, E)$, where V is a set of vertices and E is a set of edges. Each vertex V_i represents the centroid location derived from the geotagged photo locations in a cell. Each edge E_i indicates the transition between two adjacent cells according to the popularity transition defined from the hierarchal ranking model. The graph, depicted in Figure 2, is defined by routing from the origin point to the destination point (forward) - and vice versa (backward), whereas both routes might not geometrically coincide. Geometric conditions are required to avoid route self-intersection while advancing towards the destination point of the route. To handle these, we employ the shortest distance to the destination point constraint, together with removing the visited cells from the data structure. Using heuristic greedy approach will not ascertain a global solution ensuring the crossing of all popular cells and finding the optimal popular route; however, it will ensure crossing the ones that are the most popular along with the shortest route constraint. Accordingly, a popularity score test is conducted on the forward and backward calculated routes, and the route with the highest popularity score between the two is chosen as the chosen tourism route.

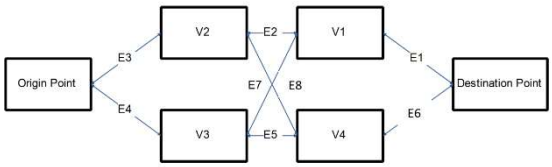

Figure 2. The directional graph between origin and destination points.

The pseudo-code of the developed greedy heuristic route computation approach is depicted in Figure 3. The algorithm uses the function *FindUsersCell* to search the most popular neighboring cell having the largest number of users closest to the current cell (e.g., origin cell), under the constraint that the distance to the destination (cell) should get smaller. In case the closest cells do not fall under the popularity ranking of the first criterion, the analysis is conducted based on the second popularity measure that ranks cells according to the number of existing geotagged photos via the function *FindPhotosCell* (similarly to *FindUsersCell*). In case the algorithm finds a cell that it already passed, it will search for another cell, while updating the current cell in the calculated route using the function *ADD*. Cells that were already added to the route are deleted from the database using the function *DELETE*, so the algorithm will not consider them, ensuring that the calculated route is not self-intersecting. In case there are no popular (tourism) cells nearby, the route is calculated based on the cell that is the closest to the destination point (third criterion) existing in the neighboring cells of *CurrentCell*. The algorithm will continue until the route reaches the destination point (cell).

```
CurrentCell = StartCell (origin);
while CurrentCell Is Not EndCell (destination) do
    if (neighboring cells are Not empty with users) then
        //find cell with maximum accumulated users - NN method;
        NextCell=FindUsersCell(CurrentCell) ;
    else
        //find cell with maximum volume of photos;
        NextCell=FindPhotosCell(CurrentCell) ;
    end
    Add NextCell to route;
    Delete CurrentCell from database;
    CurrentCell=NextCell;
end
```

Figure 3. Pseudo-code of the heuristic greedy route optimization algorithm.

In its current stage, the algorithm is considering urban structure and environmental constraints using Google Maps Direction API[1] applying walking travel mode route planning via the extracted waypoints (POIs) represented by the cells' centroid.

4. EXPERIMENTAL RESULTS

Two different cities are analyzed as case studies to test and validate the methodology: Tel-Aviv, Israel, and Manhattan, New-York, USA. Flickr photos were downloaded for a period of 6 years: 2010 to 2016. The differences in the urban characteristics of both cities, together with the differences of their respective Flickr photo datasets (volume), will help to assess and analyze the robustness and effectiveness of the developed algorithms. Table 1 depicts the main statistics for the Flickr photo data used for both cities, showing that although the area chosen in Manhattan is only double the area chosen in Tel-Aviv, the number of users is 10 times higher, and the number of photos is 5 times higher.

City	Number of Photographers	Size in Km	Photo Volume
Tel-Aviv	2168	10X11	63831
Manhattan	22665	14X15	358691

Table 1. Main Flickr photo data statistics for Tel-Aviv and Manhattan.

[1]https://developers.google.com/maps/documentation/directions/

4.1. Tourism Photographers Identification

Preliminary parameters were defined to exclude erroneous data and outliers. Outliers of a minimum route distance value of 1 m, and a minimum time interval of 1 second of the photographer trajectory were defined to filter irrelevant data. Using these two thresholds, the number of photographers has decreased by more than 50%: 9026 in Manhattan, and 1085 in Tel-Aviv.

The adaptive descriptors are computed for each city to retrieve the shared tourism activity related to a first-time tourism visitor. By iteratively changing the threshold value of the maximum visiting days, the values for New-York of the four descriptors are calculated, depicted in Table 2 and in Figure 4. Analysing both, it is apparent that for certain descriptors there is a sharp value change between 10 to 20 days (e.g., photographers volume), while for others there is no significant value change (e.g., travelled distance and photo volume). Tel-Aviv yielded similar results. Based on these results, the duration travel time of 10 days is defined as a first-time tourism activity for both cities, whereas the travel time threshold was defined as the calculated average value plus one standard deviation. Accordingly, the travel time value of 5.3 days was calculated for Manhattan, and 5.7 days for Tel-Aviv. To validate the implementation of the adaptive descriptors, we have cross-referenced them with official surveys. For Tel-Aviv, for example, a survey made in 2017 by the Israeli Ministry of Tourism[2] found that the average number of days tourists stay in Tel-Aviv is 5.1 days. This value supports the results, validating the holistic approach of working with adaptive values.

Area	Number of days	Travel Time (days)	Number of Photos	Travel Distance [km]	Travel Speed [Km\Hr]
Tel Aviv	20	3.67	25.00	10.00	0.48
	10	2.50	25.00	9.00	0.50
	5	1.58	22.00	8.59	0.63
	3	0.99	24.00	6.55	0.73
Manhattan	20	3.10	16.67	6.95	0.45
	10	2.41	16.70	6.76	0.48
	5	1.70	16.06	5.89	0.54
	3	1.12	14.97	4.94	0.66

Table 2. Descriptor values computation for changing duration visit.

Analyzing the tourism photographers' trajectory distribution reveals that most are clustered close to the city centers - a pattern that is not associated with locals (also corresponding with the findings of Garci-Palomares et al. (2015)). The use of the speed threshold (> 10 km/hr) helped excluding photographers traveling long distances over a short time, which should resemble a travel mode other than walking. As a result, approximately 20% of all the users (photographers) in both cities were identified as showing tourism characteristics of first-time visitors.

4.2. Cell-Grid Trajectory Analysis

Figure 5 depicts a Natural Jenks classification heat map based on the number of photos per 250X250 m cell resolution for the area of Manhattan, New-York; the darker the color, the more photos exist. The left image depicts the developed approach, in which only photos of tourism photographers are accumulated, while the right image depicts all Flickr photos that exist for that area. In the left image, a clear distinction exists among different neighboring cells having varying values, making it easy to identify more tourism popularly and visited locations. Dissimilar, the right image shows relatively identical values for neighboring cells, distributed uniformly in the central area of Manhattan, which makes it hard to identify the tourism popular locations. This demonstrates that using tourism photographers, instead of photos, helps in more clearly revealing the popularity and attractiveness of different areas that are derived from tourism photographers' activities and patterns.

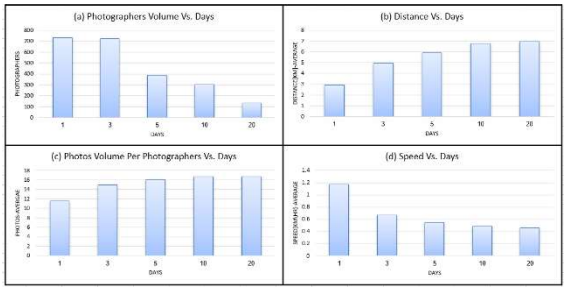

Figure 4. Tourism descriptors calculated for the photos downloaded from Flickr for the area of Manhattan, New-York : (a) Number of photographers vs. traveled number of days; (b) Accumulated traveled average distance vs. traveled number of days; (c) Average photo volume per photographer vs. traveled number of days; (d) Average speed vs. traveled number of days.

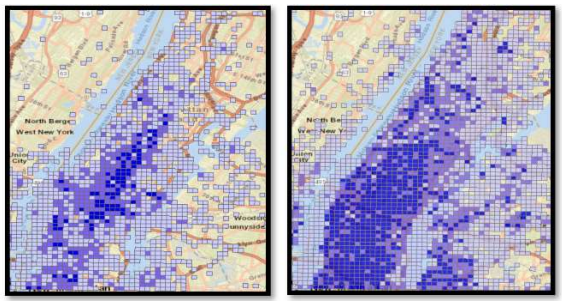

Figure 5. Heat map comparison of Flickr photos in Manhattan (based on Natural Jenks classification): tourism photographers based (left) and photo-based (right): bright blue - lower values, dark blue - higher values (background: OpenStreetMap).

4.3. Route Computation

4.3.1 Tel-Aviv

The origin point is defined in the Israel Defence Forces History Museum, and the destination point in the Old Jaffa Port (Tel-Aviv South); the walking route is approximately 2 km in length. Two routes are computed, whereas the algorithm chose the backward route, considered more tourism due to its higher popularity rate: the forward route relied on 174 accumulated tourism photographers, while the backward route on 355. The resulting yellow route, depicted in Figure 6, is generated by Google Maps Direction API using as waypoint input the retrieved POIs; the shortest walking route, computed via Google Maps, is depicted in blue. To evaluate the route tourism context, the POIs traversed by the computed route are compared to main landmarks and attractions existing in the area, which are published in TripAdvisor - a recognized and established online crowdsource tourism platform. The shortest route traverses one attraction in the area - the Old Jaffa Clock Tower (C). The computed tourism route, however, is much more tourism rich, passing through more main attractions in the area, such as Mitcham-Ha'Tachana (A), Kikar Kedumim - Old Jaffa (the main square in the area) (F), and the Ilana Goor Museum (G) – all stored in TripAdvisor and are highly ranked as tourism attractions. Additional attractions are also traversed in the computed route – owing their popularity

[2] https://motwebmediastg01.blob.core.windows.net/nop attachment/8862_english5.pdf (page 6)

score, which do not appear in TripAdvisor, such as the coastline (B and D) and Habrasha Park (E). Examining these locations, it is evident that they serve as attractions in the area that photographers enjoy visiting in their trip.

Figure 6. Tel-Aviv walking routes: tourism (yellow) and shortest (via Google Maps) (blue). Main traversed landmarks and attractions (with TripAdvisor ranking where exists): Mitcham-Ha'Tachana (A), the beach area (B), the Old Jaffa Clock Tower (C), a viewpoint from Old Jaffa (D), Habrasha Park (E), Kikar Kedumim (F), and the Ilana Goor Museum (G) (background: Google Maps).

4.3.2 Manhattan

The origin point is defined in the Grand Central Terminal, and the destination point in the Manhattan Cruise Terminal. The forward route is considered more tourism due to its higher popularity rate: 568 accumulated tourism photographers, with respect to 282 in the backward route, validating that the optimal tourism route passes through more popular cells (maximum popularity score). Figure 7 depicts the resulting walking route computed by the developed algorithm (yellow) and the shortest one generated by Google Maps (blue) - approximately 5.2 km and 3.0 km in length, respectively. Figure 7 shows that the shortest route passes through Time Square (F). Many main attractions and landmarks are traversed by the tourism route, validated when compared to TripAdvisor. Attraction G does not appear in TripAdvisor; checking its popularity measure shows that this location is very popular among tourists (Hell's Kitchen neighborhood and a photographing opportunity of the area), attracting many photographers, such that it contributes to the tourism experience when walking the route.

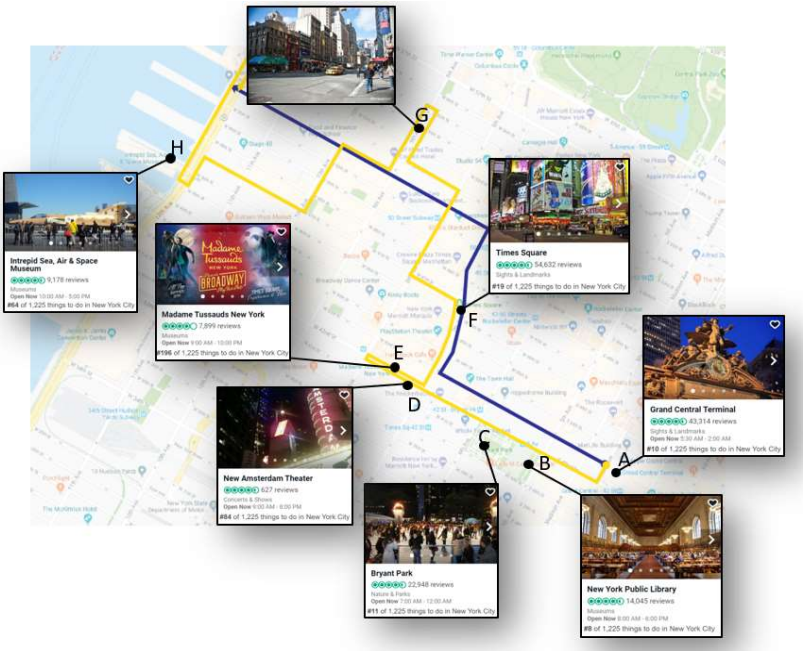

Figure 7. Manhattan walking routes: tourism (yellow) and shortest (via Google Maps) (blue). Identified landmarks and attractions along the tourism route (with TripAdvisor ranking where exists): Grand Central Terminal (A), New York Public Library (B), Bryant Park (C), New Amsterdam Theatre (D), Madame Tussauds New York (E), Times Square (F), Hell's kitchen (G), Intrepid Sea, Air & Space Museum (G) (background: Google Maps).

5. CONCLUSIONS AND FUTURE WORKS

This research explores and analyses the trajectories of photographers exploring urban areas by mining and interpreting crowdsource user-generated geotagged photo data in social media, with the motivation of enriching walking routes with tourism context. Instead of analyzing large volumes of discrete photo locations, which can generate data noise and outliers in terms of tourism context, we have devised a set of spatiotemporal descriptors to automatically identify and retrieve photographers' trajectories that show tourism activity and behavior. Reconstructing the trajectories of these tourism photographers and analyzing their urban activities, the automatic identification of popular local areas that are visited by them is made possible. To asses this, a hierarchal heuristic greedy route computation algorithm that relies on the accumulated popularity information derived from the photographers' trajectories was developed. Experiments showed promising results, where the computed routes for two case studies validated the tourism optimization criteria, validating that the user has a tourism experience while walking in an unfamiliar area. The generated routes were evaluated qualitatively by cross-referencing them with external tourism databases, showing that the developed approach is reliable in terms of the tourism perspective and the user experience. Moreover, the algorithms have the potential to identify popular and attractive places that might be missing from authoritative tourism databases, although frequently visited by tourists in the area, proving the capacity to retrieve additional and up-to-date data and information.

In future work, we plan to modify hardcoded values, such as the grid size and cell ranking, with adaptive values that will be calculated based on the photo data based on the scale of the area. As part of the knowledge recovery, we will consider using the textual interest to understand the context of different users and their diversities, generating more tuned and context-aware recommendations (e.g., local vs. global attractions). In addition, machine learning algorithms will be investigated to improve the identification of tourism activity of photographers and analysis of heterogeneity of the retrieved data. Furthermore, holistic approaches of route computation algorithms will be investigated, such as the orienteering problem and the salesman problem, to develop compatible and optimize routes in terms of maximizing the tourism experience and take into consideration time and space constraints. More experiments with longer routes and in more cities, as well as in rural areas, are planned, together with the use of supplementary crowdsource user-generated data.

The results of this study can contribute to tourism consumption, mainly since contributed user experience in online social media is becoming customary, storing information that might not exist elsewhere, together with the fact that the update-rate is very high. Tourists will be able to navigate and tour in unfamiliar destinations without a-priory knowledge, while up-to-date and reliable tourism data stored in online crowdsource social media will contribute and support to developed context-aware tourism routes attuned to their interest and criteria.

ACKNOWLEDGEMENT

The authors gratefully acknowledge Flickr open public data that is under the Creative Commons (CC) license.

REFERENCES

Basiri, A., Amirian, P., Winstanley, A., & Moore, T. (2017). Making tourist guidance systems more intelligent, adaptive and personalised using crowd sourced movement data. Journal of Ambient Intelligence and Humanized Computing, 1-15.

Becker, M., Singer, P., Lemmerich, F., Hotho, A., Helic, D., & Strohmaier, M. (2015, December). Photowalking the city: Comparing hypotheses about urban photo trails on Flickr. In International Conference on Social Informatics (pp. 227-244). Springer, Cham.

Brilhante, I., Macedo, J. A., Nardini, F. M., Perego, R., & Renso, C. (2015). Planning sightseeing tours using crowdsensed trajectories. SIGSPATIAL Special, 7(1), 59-66.

Chen, C., Chen, X., Wang, Z., Wang, Y., & Zhang, D. (2017). ScenicPlanner: planning scenic travel routes leveraging heterogeneous user-generated digital footprints. Frontiers of Computer Science, 11(1), 61-74.

De Choudhury, M., Feldman, M., Amer-Yahia, S., Golbandi, N., Lempel, R., & Yu, C. (2010, June). Automatic construction of travel itineraries using social breadcrumbs. In Proceedings of the 21st ACM conference on Hypertext and hypermedia (pp. 35-44). ACM.

Doytsher, Y., Galon, B., & Kanza, Y. (2017, November). Emotion maps based on geotagged posts in the social media. In Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities (pp. 39-46).

El Ali, A., Van Sas, S. N., & Nack, F. (2013, February). Photographer paths: sequence alignment of geotagged photos for exploration-based route planning. In Proceedings of the 2013 conference on Computer supported cooperative work (pp. 985-994). ACM.

Farzanyar, X., & Cercone, N. (2015). Trip pattern mining using large scale geo-tagged photos. In Proceedings of the International Conference on Computer and Information Science and Technology (p. 113).

García-Palomares, J. C., Gutiérrez, J., & Mínguez, C. (2015). Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS. Applied Geography, 63, 408-417.

Girardin, F., Fiore, F. D., Ratti, C., & Blat, J. (2008). Leveraging explicitly disclosed location information to understand tourist dynamics: a case study. Journal of Location Based Services, 2(1), 41-56.

Hauff, C. (2013, July). A study on the accuracy of Flickr's geotag data. In Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval (pp. 1037-1040). ACM.

Hio, C., Bermingham, L., Cai, G., Lee, K., & Lee, I. (2013, December). A hybrid grid-based method for mining arbitrary regions-of-interest from trajectories. In Proceedings of Workshop on Machine Learning for Sensory Data Analysis (p. 5). ACM.

Hsieh, H. P., Li, C. T., & Lin, S. D. (2012, August). Exploiting large-scale check-in data to recommend time-sensitive routes. In Proceedings of the ACM SIGKDD International Workshop on Urban Computing (pp. 55-62). ACM.

Kádár, B. (2014). Measuring tourist activities in cities using geotagged photography. Tourism Geographies, 16(1), 88-104.

Kádár, B., & Gede, M. (2013). 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.

Korakakis, M., Mylonas, P., & Spyrou, E. (2016, October). Xenia: A context aware tour recommendation system based on social network metadata information. In Semantic and Social Media Adaptation and Personalization (SMAP), 2016 11th International Workshop on (pp. 59-64). IEEE.

Li, J., Yang, Y., & Liu, W. (2015). Exploring personalized travel route using POIs. International Journal of Computer Theory and Engineering, 7(2), 126.

Lim, K. H. (2015, May). Recommending tours and places-of-interest based on user interests from geo-tagged photos. In Proceedings of the 2015 ACM SIGMOD on PhD Symposium (pp. 33-38). ACM.

Lim, K. H., Chan, J., Leckie, C., & Karunasekera, S. (2016, March). Towards Next Generation Touring: Personalized Group Tours. In ICAPS (pp. 412-420).

Liu, Q., Wang, Z., & Ye, X. (2018). Comparing mobility patterns between residents and visitors using geo-tagged social media data. Transactions in GIS, 22(6), 1372-1389.

Mor, M., & Dalyot, S. (2018). Computing touristic walking routes using geotagged photographs from Flickr. In Adjunct Proceedings of the 14th International Conference on Location Based Services (pp. 63-68). ETH Zurich.

Paldino, S., Bojic, I., Sobolevsky, S., Ratti, C., & González, M. C. (2015). Urban magnetism through the lens of geo-tagged photography. EPJ Data Science, 4(1), 5.

Salas-Olmedo, M. H., Moya-Gómez, B., García-Palomares, J. C., & Gutiérrez, J. (2018). Tourists' digital footprint in cities: Comparing Big Data sources. Tourism Management, 66, 13-25.

Subramaniyaswamy, V., Vijayakumar, V., Logesh, R., & Indragandhi, V. (2015). Intelligent travel recommendation system by mining attributes from community contributed photos. Procedia Computer Science, 50, 447-455.

Sun, Y., Fan, H., Bakillah, M., & Zipf, A. (2015). Road-based travel recommendation using geo-tagged images. Computers, Environment and Urban Systems, 53, 110-122.

Yahi, A., Chassang, A., Raynaud, L., Duthil, H., & Chau, D. H. P. (2015, March). Aurigo: an interactive tour planner for personalized itineraries. In Proceedings of the 20th international conference on intelligent user interfaces (pp. 275-285). ACM.

Yang, X., & Tang, L. (2016). Crowdsourcing Big Trace Data Filtering: A partition-and-Filter Model. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, 41.

Yang, L., Wu, L., Liu, Y., & Kang, C. (2017). Quantifying tourist behavior patterns by travel motifs and geo-tagged photos from Flickr. ISPRS international journal of geo-information, 6(11), 345.

Zhang, W., Tan, G., Lei, M., Guo, X., & Sun, C. (2018). Detecting tourist attractions using geo-tagged photo clustering. Chinese Sociological Dialogue, 3(1), 3-1.