Conference Paper

Computing Touristic Walking Routes using Geotagged Photographs from Flickr

Author(s):
Mor, Matan; Dalyot, Sagi

Publication Date:
2018-01-15

Permanent Link:
https://doi.org/10.3929/ethz-b-000225591

Rights / License:
Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International
Computing Touristic Walking Routes using Geotagged Photographs from Flickr

Matan Mor, Sagi Dalyot

Mapping and Geoinformation Engineering, The Technion, Technion City 3200003, Haifa matan.mor@campus.technion.ac.il, dalyot@technion.ac.il

Abstract. Tourism information is getting extensive, comprehensive and complex, thus tourists have to manage and mine large volumes of data and information to better plan their trip. Geotagged photographs uploaded by users to social photo-sharing online websites are today frequently used by tourists to describe their tourism experience, sometimes even replacing textual description. We focus on Flickr geotagged photograph database to automatically compute touristic walking routes. Instead of simply clustering photographs to extract places that might not be associated with tourism, we suggest using a set of spatio-temporal descriptors that are associated with the activities of touristic photographers, to mine and interpret photographs that have a tourism-context. Cell-based clustering is used on the retrieved photographs to find popular regions and places of interest traversed by trajectories made by the photographers that show tourism characteristics. A bi-directional constrained pathfinder Nearest Neighbor route calculation algorithm is developed to compute routes that visit the most popular touristic locations among photographers. Preliminary results for Manhattan are presented, proving reliable interpretation, mining and retrieving of valuable tourism information from social media Flickr geotagged photographs, computing interesting and realistic touristic walking routes.

Keywords. Geotagged Big Data, Activity Interpretation, Route Calculation

1. Introduction

Smart tourism solutions are developing fast, aspiring to keep pace with the growing number of dynamic volumes of big data and information sources that did not exist until recently, continuously updating and becoming more accessible and straightforward to use. As a result, paper maps and tour guides are becoming obsolete, being replaced with internet and mobile
guides, internet search engines, apps, and blogs. Dynamic crowdsourced user-generated geotagged data and information sources are increasing dramatically, and social media websites, such as Flickr, Twitter, and Facebook, are more commonly used to share travel and tourism experience. Since these dynamic social media websites serve as collective up-to-date knowledge, mining and extracting relevant and updated photograph trails of photographers can serve as a smart solution to touristic walking route planning, mainly in unfamiliar areas (e.g., De Choudhury et al. 2010, Chareyron et al. 2014). This requires two main challenges to deal with: 1) identify and understand the underlying tourism context by implementing data mining processes to interpret and discover popular places that should be included in the route; 2) compute an optimized walking route that should serve as the most touristic one. Current research aims to discover tourism routes extracted from user-generated geotagged photograph and Places of Interest (POI) databases. Li et al. (2015), for example, calculate POI geographic information to guide to locations resembling tourism features, while Sun et al. (2015) use the Dijkstra algorithm with weighted popularity road matrix to compute touristic routes. Becker et al. (2015) retrieve photograph trails for a single user regardless of time span by comparing the extracted trail to a weighted known POI geographic layer to explore tourists’ patterns. Wang et al. (2016) analyze photographers’ pattern by analyzing their photograph sequence to evaluate touristic routes.

In this study, instead of relying and analyzing numerous discrete geotagged photographs that exist on Flickr that might not have a tourism context, we aspire to identify users (photographers) that can be considered as tourists based on their activity and analyze their trajectory patterns; thus, we filter less relevant (and ‘noisy’) photograph data. Relying on these photographers’ activities we can interpret and discover popular touristic places visited by them according to the accumulated photographs that have a tourism-context. To this end, we have devised an array of spatio-temporal key indicators to classify touristic photographers and analyze their accumulated travel trajectories (touristic trail). We interpret popular cells traversed by their trajectories. These cells are used as input in a constrained bi-directional Nearest Neighbor (NN) route calculation algorithm to automatically seek and calculate the most touristic and comprehensive walking route, without having to rely on an external POI database.

2. Methodology

Our methodology, depicted in Figure 1, relies on three consecutive stages with the aim of mine and interpret touristic photographs (stage 1), cluster and rank popular touristic places and POIs (stage 2), and construct the touristic route (stage 3), as follows:
2.1 Classifying Touristic Photographers
To retrieve photographs that have a tourism context, we identify photographers (users) that show touristic activity and descriptors, aspiring to filter photographs uploaded by local residents or random photographers. This is achieved by implementing a set of spatio-temporal parameters and identifiers that characterizes a touristic activity, namely: the number of photographs taken on a trip; the time interval between consecutive photographs; the duration of a trip travel; the distance traveled; and, the traveling speed. By linking all geotagged photographs taken by these photographers, we can analyze the touristic setting of the area.

2.2 Cell-Grid Popularity Analysis
The cell-based approach is implemented to geographically cluster the accumulated photographers’ activity patterns by analyzing their traversed trajectories to identify popular touristic places: cells with a higher number of visiting touristic photographers are considered to be more touristically popular and frequently traveled. The most attracting location for a cell relies on a centroid calculation of all geotagged photographs that fall in its extent, which presumably coincides with the position of the touristic and popular attraction (e.g., landmark, site), i.e., POI.

2.3 Route Calculation
Searching popular cells is carried out via a constrained bi-directional NN pathfinder method. Using NN will not ascertain crossing all popular cells’ POIs, but will assure crossing the ones that are the most popular along the shortest route constraint. In our implementation, two routes are calculated, forward (origin-destination) and backward (destination-origin). Both routes
will not always pass through the same popular cells, such that a popularity test is implemented to quantify and compare both, recommending the one that is the most popular. To construct routes that are logical in relation to the existing road network, our preliminary route calculation algorithm relies solely on the road network arrangement without considering specific environmental constraints, such as the orienteering problem in networks. We use Google Maps Direction API by applying walking travel mode via the extracted waypoints (POIs) as input to compute a logical walking route.

3. Preliminary Results

Manhattan, New-York, USA, with an approximated area of 210 sq. km. is chosen as a case study. Flickr database was downloaded in July 2016, having 22665 users, and a total of 358691 geotagged photographs. Preliminary parameters were defined for excluding erroneous data and outliers, while together with the touristic photographer descriptors, only 1846 users (photographers) were identified as having touristic characteristics - less than 10%. Statistics revealed that photographers’ visit time duration is below four days, traveling 13 Kilometers (values for CEP90), validating the tourism travel behavior indicators used for filtering.

Origin point was defined at Grand Central Terminal, and destination point at Manhattan Cruise Terminal. Computing walking route via Google Maps API generated two options, where the optimal shortest route is approximately 2.7 km in length, depicted in grey in Figure 2 (left). Figure 2 (right) depicts the bi-directional routes computed by our algorithm: red (forward) and green (backward), where the forward one was considered as more touristic due to its higher popularity rate: 568 accumulated photographers, as opposed to 282. The resulting yellow route was computed by Google Maps Direction API using the retrieved waypoints (POIs) as input. The final touristic route, depicted in Figure 3, which is slightly longer than the shortest route, passes through main touristic landmarks and attractions, depicted as A to E, which were automatically retrieved without relying on an external POI database. When TripAdvisor’s landmark recommendations and rankings of the area are compared to the retrieved POI locations, all are represented. Hell’s kitchen, depicted as POI D in Figure 3, which is popular among tourists and photographers, is not listed as POI in TripAdvisor’s database. This proves that our algorithms are capable of retrieving local popular attractions, which are not always listed in external POI databases, commonly visited by tourists and are attractive among photographers, and thus most obviously can be included in the planned touristic walking route. Moreover, the number of POIs retrieved is realistic: 5 landmarks for a walking distance of approximately 3.5 km, ascertaining not to communicate excessive information to the tourist.
Figure 2. Google Maps routes (left), and the automatically computed touristic walking routes in Manhattan (right), superimposed on the grid-cell heat map with popularity colors: forward route (red), backward route (green), and Google Maps Direction API route (yellow) that uses the retrieved waypoints of the forward route as input (background: Google Maps).

Figure 3. Automatically computed touristic route in Manhattan passing through main landmarks and attractions (with TripAdvisor ranking, where available): Bryant Park (A), New Amsterdam Theater (B), Time Square (C), Hell’s kitchen (view of Hearst Tower from 9th Avenue) (D), Intrepid Sea, Air & Space Museum (E) (background: Google Maps).
4. Conclusions and Future Work

This research presents a methodology that relies on spatio-temporal descriptors, which are associated with the activities of touristic photographers, to mine and interpret photographs that have a tourism-context to retrieve touristic popular attractions. By doing so, we filter with high certainty photographs that have no touristic significance, which is commonly the case when simply clustering all existing photographs. Preliminary experiments show very promising results, where the automatically computed touristic routes validate the premise of passing through touristic sights and landmarks. Moreover, routes do not significantly deviate from the general direction and distance between the origin and destination points, serving as interesting and realistic touristic walking routes, validating the optimization process suggested here. Small deviation segments exist, which are the result of relying on the centroid locations as waypoints in the routing; this is planned to be handled in future work, tuning and developing a more complex route calculation algorithm. Future work will also include the use of an adaptable clustering method to replace the cell-based one and the replacement of the hard-coded thresholds in the touristic photographers’ classification stage. Results show the potential that local, alternative and less known attractions - yet ones that attract photographers and tourists - are successfully identified. This proves that our algorithms retrieve important and updated (perhaps even temporal) attractions that do not exist in other external databases. These ensure the user has a touristic experience while walking in an unfamiliar area, proving the capacity to retrieve dynamic and up-to-date data and information from geotagged big data sources to computing comprehensive touristic routes.

References


Chareyron, G., Da-Rugna, J., & Rainbault, T. (2014). Big data: A new challenge for tourism. In Big data (Big data), 2014 IEEE international conference on (pp. 5-7). IEEE.


