Computers, Environment and Urban Systems xxx (2013) xxx–xxx

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/compenvurbsys

Road-based travel recommendation using geo-tagged images

Yeran Sun *, Hongchao Fan, Mohamed Bakillah, Alexander Zipf

GIScience Research Group, Institute of Geography, University of Heidelberg, Berliner Straße 48, Heidelberg 69120, Germany

A R T I C L E I N F O

Article history:
Available online xxxx

Keywords:
Volunteered geographic information
Travel recommendation
Routing planning
Spatial data mining
Semantic routing

A B S T R A C T

Geotagged photos on social media like Flickr explicitly indicate the trajectories of tourists. They can be employed to reveal the tourists' preference on landmarks and routings of tourism. Most of existing works on routing searches are based on the trajectories of GPS-enabled devices' users. From a distinct point of view, we attempt to propose a novel approach in which the basic unit of routing is separate road segment instead of GPS trajectory segment. In this paper, we build a recommendation system that provides users with the most popular landmarks as well as the best travel routings between the landmarks. By using Flickr geotaggged photos, the top ranking travel destinations in a city can be identified and then the best travel routes between the popular travel destinations are recommended. We apply a spatial clustering method to identify the main travel landmarks and subsequently rank these landmarks. Using machine learning method, we calculate the tourism popularity of the road in terms of relevant parameters, e.g., the number of users and the number of Point-of-Interests. These popularity assessments are integrated into the routing recommendation system. The routing recommendation system takes into consideration both the popularity assessment and the length of the road. The best route recommended to the user minimizes the distance while including maximal tourism popularity. Experiments were conducted in two different scenarios. The empirical results show that the recommendation system is able to provide the user good travel planning including both top ranking landmarks and suitable routings in a city. Besides, the system offers user-generated semantic information for the recommended routes.

1. Introduction

Web 2.0 technologies enable users of social media to make contributions or to communicate with each other. Among the various types of information contributed and shared by users on social media, the geographic one is called Volunteered Geographic Information (VGI, Goodchild, 2007). The most common providers of VGI are Flickr, OpenStreetMap, Twitter, Facebook, YouTube, Wikimapia, Foursquare, etc. VGI has been used in research on tourism, disaster and crisis management, transportation, etc. As a typical photo-sharing provider, Flickr provides a platform where users can share their photos with metadata (e.g., size, time when photo was taken, location, camera type, etc.) and add textual information on the photos. Such textual information includes title, tag, etc. Georeferenced photos in Flickr are associated with locational, temporal and textual information, which reflect the behaviours and activities of users, particularly tourists. For instance, spatial distribution of Flickr images can reveal the underlying process of tourists' footprints. With a combination of spatial and the temporal dimensions, Flickr images can be used to uncover the trajectories and movements of tourists. From this point of view, Flickr can be employed to provide users with travel recommendations based on the identification of tourists' spatial and temporal patterns from images. In addition to the georeferenced photos, georeferenced posts in Twitter, geolocated check-ins in Foursquare, user-generated GPS trajectories, etc., have potentials of being leveraged for travel recommendations (Zheng, 2012).

From an insight into location-based social network (LBSN), Zheng and Zhou (2011) classify travel recommendations into two types: generic and personalized. Basically, a generic travel recommendation system should provide users the top ranked landmarks, travel sequences, and travel experts in a specified region (e.g., a city). In the contrast, a personalized system is able to provide an individual user with a distinctive recommendation based on the travel preferences and histories of the users (e.g., Cheng, Chen, Huang, Hsu, & Liao, 2011; Majid et al., 2012). Apart from the tourism attraction or landmark, routing is another important aspect of travel recommendation. GPS history data is often used to track the footprint of the user, which becomes an indicator of the user's trajectory and movement (e.g., Zheng, Zheng, Xie, & Yang, 2010). Representative or typical routings of tourists can be mined from huge amounts of tourists' trajectories and thus are recommended to the user (e.g., Lu, Wang, Yang, Pang, & Zhang, 2010). Some researchers take advantage of the trajectory similarity calculation to offer the

* Corresponding author. Address: Room 101, Berliner Straße 48, D-69120 Heidelberg, Germany. Tel.: +49 6221 54 5528; fax: +49 6221 54 4529.
E-mail address: yeran.sun@geog.uni-heidelberg.de (Y. Sun).

0198-9715/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compenvurbsys.2013.07.006

individual user the most "similar" routing, which is acquired from the tracking histories of the other users, as the personalized routing (e.g., Cheng et al., 2011). To enhance the real-time usability of the travel recommendation application, time constraint is also taken into account in some research works (e.g., Lu et al., 2010).

In general, a good travel recommendation system should provide the user with the most interesting (popular) landmarks including routing paths as the most people recommended. Additionally, for tourists, an optimal routing should take into account not only the distance but also the tourism popularity. The research result of Popescu and Grefenstette (2009) shows that tourists normally do not take the shortest path between landmarks. Tourists usually like popular streets where they can glimpse more touristic attractions (e.g., churches, squares, statues, memorials, etc.) or satisfy some personalized needs (e.g., eating, shopping, mailing post card, etc.) before reaching their destinations. Thus, to measure the popularity of a road, not only the images taken of the road itself, but also the images taken of the Points-of-Interest (POIs) on or near the road should be taken into consideration. Moreover, in addition to the popularity of the road itself, POIs on or near the road should be taken into account to recommend the best routing. However, the previous research works generate recommended routings from the historic trajectories (or trajectory segments) of users directly. Unlike these previous research works, we have an alternative perspective on the routing generation, according to which a recommended routing is composed by a set of separate roads (or road segments) which are connected in road network. Therefore, the advantage of our approach is that we are able to consider the tourism popularity of road as a new feature of road's popularity in a routing planning problem, and based on the values of features including road popularity as well as other attributes (e.g., POIs), we are able to find out the best road set constituting the best overall routing.

In this paper, we propose a road-based travel recommendation paradigm combining the landmarks and the routing. The travel landmark can be identified by the georeferenced images with a spatial clustering method. The best routings between the travel landmarks are recommended to users in terms of calculating the total value of the defined recommendation index. This defined recommendation index is constituted by the popularity and distance of the road as well as number of the POIs. Section 2 reviews related work on travel recommendation using georeferenced photos. Section 3 introduces the proposed approach to recommending travel. Section 4 presents the experimental results and analysis. At last, the paper makes a conclusion and presents future work.

2. Related works

In tourism research, the leverage of geotagged images and GPS history data mostly focuses on hotspot and landmark identification, trajectory and movement mining as well as trip recommendation. The related research works on these aspects are introduced in the following sub-sections.

2.1. Hotspot and landmark identification

Through geotagged images, Kennedy, Naaman, Ahern, Nair, and Rattenbury (2007) firstly used spatial clustering method to recognize the landmarks, and then employed a location-driven approach to generate representative tags for these landmarks. Rattenbury, Good, and Maaman (2007) proposed an approach to extract place semantics from Flickr tags automatically. Comparative results showed that the proposed scale-structure identification (SCI) outperforms the baseline methods. Following this, Rattenbury and Naaman (2009) developed a new approach called TagMaps to extract semantic information of place, and compared it with the previous methods. The results showed that TagMaps method outperformed the existing ones. Using Flickr, Girardin, Dal Fiore, Ratti, and Blat (2007) identified local regions of tourist concentration in terms of the point density of tourists. Subsequently, based on the most active regions obtained by spatial data clustering, Girardin, Calabrese, Dal Fiore, Ratti, and Blat (2008) presented the spatial and temporal distribution of tourists during their trips to Rome. In addition, Girardin, Vaccari, Gerber, and Ratti (2009) measured the attractiveness of POI in terms of the presence of photographer. Crandall, Backstrom, Huttenlocher, and Kleinberg (2009) employed a non-parametric clustering method named *mean shift* to discover the significant landmarks at a global level and evaluated their method within several specific cities.

To mine interesting locations and classical travel sequences in a geospatial region, Zheng, Xie, and Ma (2009) proposed a HITS (Hypertext Induced Topic Search)-based model to infer a user's travel experiences and the interest of a location considering the relationship between user and location. The results showed that there existed a mutual reinforcement relationship between location interest and user travel experiences. Meanwhile, employing both users' travel experiences and location interests, they achieved the best performance of detecting the classical travel sequences Jankowski, Andrienko, Andrienko, and Kisilevich (2010) presented a geovisual analytic approach to discover people's preferenceson landmarks from Flickr photos. Their results could help to distinguish between sites that are occasionally popular among the tourists and sites already known as city landmarks. Moreover, Ji, Gao, Zhong, Yao, and Tian (2011) applied a spectral clustering method to identify landmark region and mined representative photos.

In this work, we employ DBSCAN method to identify the landmarks. It is a classic and popular density-based clustering method, which has several generalized forms, e.g. GDBSCAN and ST-GDBSCAN. Compared to the most of clustering methods we mentioned above, DBSCAN can identify the clusters of arbitrary shape without the specification of cluster count *a priori*. However, two parameters are required in DBSCAN procedure which might influence the clustering result. Unfortunately, there is not yet perfectly general method to search the optimal parameters which is adaptive to different applications. Therefore, we firstly generate different results of clustering at different values of the parameters. We subsequently compare these results and finally find out an appropriate pair of parameters. Among the clusters generated with this pair of parameters, we have three steps to distinguish the clusters resulted from landmarks with the ones resulted from some events or other things which are unrelated to tourism. Consequently, the identification of landmark will be enhanced.

2.2. Trajectory and movement mining

For pedestrian navigation based on mobile devices, Hile et al. (2008) proposed a system for the automatic generation of navigational instructions from collections of geotagged photos. The instructions consist of a sequence of landmark images augmented with directional instructions. Using geotagged photos, Andrienko, Andrienko, Bak, Kisilevich, and Keim (2009) built a flow map showing aggregated moves of photographers between different places. Defining some criteria to calculate the duration of the tourist's visit from Flickr photos, Popescu and Grefenstette (2009) exploited temporal information associated with touristic sites. Also, to extract trip-related information, Popescu, Grefenstette, and Moëllic (2009) characterized the discovered trips and computed the visit times of the sites. Aiming at solving the problem of automatic travel route planning, Lu et al. (2010) provided a customized trip plan enabling tourists to specify personal preference such as visiting location, visiting time/season, time duration of travel, and destination style in an interactive manner to guide the system. Similarly, Arase, Xie, Hara, and Nishio (2010) detected people's frequent trip patterns, i.e., typical sequences of visited cities and stay durations as well as descriptive tags that characterize the trip patterns. With respect to the trip theme, the trips were categorized using tags and titles of photos as well as visited cites as features. The proposed method mined frequent trip patterns for each trip theme category. Jankowski et al. (2010) presented a geovisual analytics approach to discovering people's movement patterns from Flickr photos. The results show that most of the tourists' itineraries are short and highly local. Choudhury et al. (2010) extracted photo streams of individual users on Flickr and aggregated them into a POI graph. Subject to the constraints of user's time and destination, travel itineraries were then automatically constructed from the graph based on the popularity of the POI. To explore the representative trajectory patterns, Yin, Cao, Han, Luo, and Huang (2011) generated many trajectories and ranked the trajectory patterns by quantitatively utilizing an exemplar-based algorithm. To discover moving flock patterns among pedestrians, Wachowicz, Ong, Renso, and Nanni (2011) proposed an extraction algorithm based on the notion of collective coherence. Consequently, the base trajectories of the moving flock patterns were mined from GPS history data. Using GPS traces, Zheng and Xie (2011) performed a travel recommendation devoted to offering a user with top interesting locations and travel routines in a given geospatial region in consideration of both the number of users and their knowledge. Aiming at low-sampling-rate GPS data, Lou et al. (2009) proposed a novel global map-matching algorithm called ST-Matching to match the GPS trajectories of users with the digital map. Using Markov chain model, Zheng, Zha, and Chua (2012) investigated the tourist movement patterns in relation to the regions of attractions from geotagged photos. Subsequently, the topological characteristics of travel routes made by different tourists were analyzed. Wei, Zheng, and Peng (2012) proposed a Route Inference framework based on Collective Knowledge (RICK) to construct the popular routes from uncertain historic trajectories. Foursquare check-ins and taxi trajectories were chosen as two experiment cases.

2.3. Travel recommendation

Based on user's location history from GPS, Takeuchi and Sugimoto (2006) proposed a shop recommendation system for real-world shopping. Considering a "convenience" metric in making recommendation, Horozov, Narasimhan, and Vasudevan (2006) proposed an enhanced collaborative filtering method to recommend restaurants to mobile users. To provide personalized travel recommendations for users owning a preference for a specific type of landmark in a city, Clements, Serdyukov, Vries, and Reinders (2010) proposed a method to predict a user's favorite locations from Flickr photos. They reranked the popular locations for individual user in terms of calculating the similarities between the geotagged image distributions of different users. Using a probabilistic Bayesian learning method, Cheng et al. (2011) proposed a personalized recommendation considering specific user profiles or attributes (e.g. gender, age, race). Zheng et al. (2010) integrated the users' comments into the travel recommendation system based on the GPS history data. The system can discover interesting locations together with possible activities that can be performed there. Considering the location correlation, Zheng and Xie (2010) performed a personalized location recommendation system based on the user-generated GPS trajectories, by taking into account a user's travel experiences and knowledge. Majid et al. (2012) presented a context-aware personalized travel recommendation system considering the temporal and weather context of tourist. The system acquired users' specific travel preferences by building the travel similarities between users from their travel histories. Subsequently, it recommended tourist the personalized travel locations in unvisited regions. To enhance travel planning in unfamiliar regions, Yoon, Zheng, Xie, and Woo (2012) proposed a social itinerary recommendation system by learning from multiple user-generated GPS trails.

The above works on travel recommendation can offer users the prominent travel landmarks and further the ordering of landmarks, which compose travel routings. Besides, personalized preference is considered in the recommendation. Moreover, some works present a trip routing based on user trajectory from GPS history data. Each historic trajectory of the user can be considered as a historic individual routing. The recommended routing is always represented as the average of the historic routings (trajectories) or the one with the highest ranking. Nevertheless, such recommended trip routings sometimes lack convenience and effectiveness for users to position and navigate on the digital map, since they are based on GPS trajectory instead of separate road in the road network. Specifically, users might conduct their travelling by walking, biking, car, bus, or hybrid way. It might be inconvenient and ineffective to recommend a user, who wants to walk in a city, a routing based on historic trajectories of users driving cars. Moreover, road-based approach can consider more the hidden spatial associations of roads in different directions than trajectory-based approach. Compared to common tourists, tourists of GPS trajectories, especially the ones who are driving or bicycling, tend to choose a straightforward line. On the other hand, few tourists tend to use GPS device to record their trajectories when travelling. Thus, coverage and representativeness of GPS trajectory data is relatively low compared to Flickr images which are generated by a larger variety of tourists. Besides, the routing recommenders in state-of-the-art neglect the tourism attractions and POIs on the road, to which tourists always pay more attention than the road itself.

Contrary to the previous research works, through leveraging Flickr geotagged images, this paper aims to generate recommended routing from separate roads instead of historic trajectories of users. In other words, the atomic unit of the routing is road segment rather than trajectory segment. Moreover, to recommend the suitable routing, in addition to the popularity of the road, POIs on the road as well as the length of the road are considered in the proposed recommendation system.

3. Approach

3.1. Framework of travel recommendation system

We define the travel recommendation for a user u as $\{L(1), Rou(1, 2), L(2), \ldots, L(n-1), Rou(n-1, n), L(n)\}$. $L(i)$ is the recommended landmark for the user in a city. $Rou(i-1, i)$ is the recommended travel routing between starting landmark $L(i-1)$ and ending landmark $L(i)$. $Rou(i-1, i)$ is constituted by a sequence of separate roads represented as $(r_1, r_2, \ldots, r_k)$.

As shown in Fig. 1, the outputs of the travel recommendation system are landmarks and routings. For this purpose, we need to identify and rank the landmarks using geotagged images of Flickr firstly. Then, we calculate the tourism popularity of the road by assigning geotagged images to the appropriate roads.

3.2. Noise elimination

Some of images related to residents' activities (e.g. wedding, birthday party, house decoration, etc.) or some temporal events unrelated to tourism (weather change, accident, etc.) do not have any positive contribution to travel recommendation. In other words, such images are noise and thus should be pruned for travel recommendation. Fortunately, the vast majority of such images are contributed by residents and thus can be removed by our following

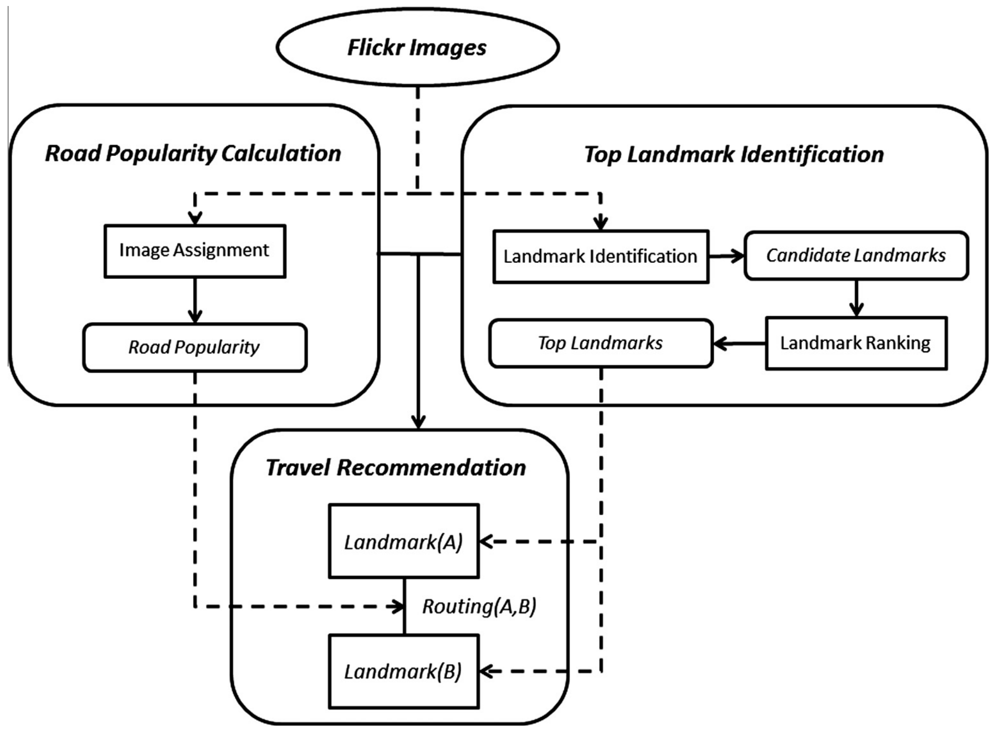


Fig. 1. Prototype of the proposed travel recommendation system.

method. We distinguish residents from the users of the geotagged image set. Based on this, we discard noise images that are taken by residents. Usually, in a city, a tourist stays short time (e.g., one week) in a whole year, thus the taken dates of her images should be usually within the same month or at most two successive months. While a resident who stays in the city most of the year and thus he can take images in almost all months. Compared to tourist, resident can take image in many months of a year. Therefore, an entropy filtering method is applied here to distinguish residents from the users of the image set. The method is described as:

$$E(u) = -\sum_{i}^{Mon(u)} P_i(u) \log P_i(u) \tag{1}$$

$$P_i(u) = \frac{D_i(u)}{\sum_{i}^{Mon(u)} D_i(u)} \tag{2}$$

$D_i(u)$ is the number of days on which the user u had taken images within Month i in the targeted region. $Mon(u)$ is the number of the months in which the user u had taken images in the targeted region. For example, a user uploaded 5 images on Flickr. Among these 5 images, two were taken on 2010/12/30, one is taken on 2011/01/01, and other two are taken on 2011/03/03. According to Eqs. (1) and (2), the $E(u) = -(2/5)\log(2/5) - (1/5)\log(1/5) - (2/5)\log(2/5) = 1.522$.

When the value of $E(u)$ is higher, the user is more likely to be a resident. We use a threshold $E_{threshold}$ to distinguish residents from tourists. If $E(u) > E_{threshold}$, the user u is considered as a resident; otherwise, he is considered as a tourist.

Furthermore, we need to further eliminate noise images unrelated to tourism from the images of tourists. Although taken by tourists, some images have tags, such as "sarah", "jump", "pose" and so forth, which reflect some persons (e.g., one of users' friends) or persons' gestures, do not have any positive contribution to the popularity of POI or road. Meanwhile, these tags unrelated to tourism are most likely to be used by single user instead of other users (Chaudhry & Mackaness, 2012). Therefore, like Chaudhry and Mackaness (2012), we discarded the image whose tag has only one user.

3.3. Hotspot detection

3.3.1. Landmark identification

To discover the prominent travel locations in a city, a spatial clustering method is applied for the geotagged images of tourists to identify the hotspots of tourism. As a density-based clustering method, DBSCAN (Ester, Kriegel, Sander, & Xu, 1996) is able to identify the clusters of arbitrary shape with no need to specify the number of clusters *a priori*. Two parameters are required in DBSCAN procedure: the radius (*Eps*) and the minimum number of points required to form a cluster (*MinPts*). These two parameters control the size and the number of the clusters.

Furthermore, in order to distinguish the detected clusters resulting from landmarks from the ones related to events or other things, we analyzed the tag usage, spatial distribution and the user pattern of the images within each cluster. Three steps were conducted:

(1) Select the clusters whose images were totally contributed by over ten users.
(2) Choose the landmarks located within the boundaries of the clusters as the candidate landmarks from the POI dataset.
(3) Match the most frequent tags of each cluster with the names of candidate landmarks.

In addition to distinguishing landmark with event or other issues, elimination of some noise and errors can be conducted by these three steps. Step 1 aims to eliminate the clusters resulted from tourist's duplicate photographing. For instance, a restaurant (e.g., McDonald's) might lead to a local cluster consisting of 30 images taken by only one or two customers who did not have anything else to do apart from taking picture casually during the dinner. In another example, there is a square reflected by totally 30 images as well, which are contributed by 15 persons. Compared to the square, the restaurant cannot be considered as landmark because landmark should have relatively large numbers of images and users simultaneously. Here we choose ten as the minimum threshold of user count. Step 2 aims to eliminate noise and error of user's tagging. Tourists sometimes take images of a high landmark (e.g. a tower) from a long distance and still use the name of this landmark as the tag of the images. These images could be considered as indicators of the landmarks. However, with the DBSCAN method based on distance these images are actually contained by a local cluster which represents another landmark duo to shorter distance. Thus, we only take into account the landmarks within the spatial region of the cluster as the candidate ones. Otherwise, there would be some errors occurring in case that the majority of the images composing a detected local cluster representing a nearby landmark belong to such kind of images actually indicating a landmark far away. Besides, some users might make a mistake when tagging images recording same types of objects. For instance, a tourist who does not have knowledge of church' architectural style, might confuse Catholic churches with Protestant churches and thus tag them with wrong names. Therefore step 2 reduces the occurrence of such mistakes as well through the restriction of spatial distance. Step 3 is dedicated to finding out popular landmarks.

After the three steps above, the top landmarks must be identified.

3.3.2. Landmark ranking

To rank the landmarks detected using the method in Section 3.3.1, we need to quantify the tourism popularity of the landmark based on the geotagged images. To measure the popularity of a place (e.g., landmark, POI, street, etc.) using user-generate images, in general the difference of contribution among images should be considered. For user-generated images, there are basically two kinds of differences. One is the difference among tourists with different levels of travel knowledge and experiences. For instance, based on Zheng et al. (2009), tourists of rich knowledge and travel experiences would provide more reliable and valuable data for travel recommendation than novel tourists. An experienced tourist might be more interested in the highly ranked landmarks than a novel tourist. The experienced tourist thus is more likely to upload the images reflecting the landmark of more popularities since such images are more worthy of attention. Moreover, she can tag the image with more accurate information to indicate the name of the landmark which the image reflects. For instance, to tag an image recording the *Old Pinakothek*, a user of rich knowledge would like to use "oldpinakothek" while a novel tourist might use "pinakothek". In this case, the user of rich knowledge provides data with more accurate textual information than the novel tourist because "pinakothek" is ambiguous in consideration of the existence of another *Pinakothek* nearby called *New Pinakothek* in Munich. Therefore, the contributions of experienced tourists should be of more importance and values than novel tourists.

The other is the difference among images taken by the same tourist. As Schlieder and Matyas (2009) substantiated already, the landmark popularity contributed by user-generated image follows a power law distribution. Specifically, the first image of a user increases the popularity of the landmark, while subsequent images of that same user add less and less popularity. For instance, when arriving at the *Avenue des Champs-Elysee*, a tourist takes a couple of images. When uploading these images, she is most likely to tag the first one with "Champselysee" since this image is recording the *Avenue des Champs-Elysee* itself instead of other POIs at this famous avenue. Then she, in turn, tags the other images with e.g., "louisvuitton", "hôteldelapaïva", etc., which record some popular POIs at this avenue. In this case, the first image should contribute more to the popularity of the *Avenue des Champs-Elysee* than the other images.

Here we name the first and the second kind of difference as the inter- and intra-tourist difference of image contribution respectively. It is difficult to measure the inter-tourist difference of contribution currently duo to the limitation of obtaining user' travel experiences and knowledge from Flickr images. Because the majority of Flickr users' profiles are incomplete, the level of user' knowledge and experiences is always unavailable. At the same time, there exists a potential of measuring the intra-tourist difference through exploring spatial-temporal patterns of user' footprints generated by Flickr images (e.g., Zheng et al., 2009). However, in regard of obtaining the experiences and knowledge of the user, the data quality or fitness-for-use of Flickr image is not so good as that of the GPS data and the user-generated Check-in data (e.g. Foursquare).For example, the data quality of Flickr data, such as positional, temporal and textual accuracies, is not that good. Besides, the spatial and temporal distribution of Flickr image is considerably heterogeneous. For a landmark, s tourist only takes one image although she visits it four times totally in a year; while another tourist takes more than ten images during the first visit but never takes any more for the later visiting. More details about the disadvantages of Flickr data will be presented in Section 3.4.2. Thus, we need to propose an innovative approach in the future, or at least modify some existing methods (e.g., HITS-based model proposed by Zheng et al. (2009)) originally applied to GPS data and consequently make them adaptive to user-generated geotagged images.

Therefore, in this research we currently only take into account intra-tourist difference of contribution based on an assumption that there is no inter-tourist difference of contribution; in other words, the contribution to the popularity of landmark is the same among tourists.

To measure the popularity of landmark, we take advantage of a measurement method proposed by Schlieder and Matyas (2009) based on a fact that tag frequency of social media follows the power law distribution (Guy & Tonkin, 2006). Consider the images assigned to landmark L were contributed by m different users. $U_L = \{u_1, \ldots, u_m\}$ is the user set of the images assigned to landmark L. The number of images taken by user u in the region of landmark L is $n_L(u)$. The popularity of the landmark is computed as follows:

$$Pop(L) = \sum_{u \in U_L} (1 + \log n_L(u)) = m + \sum_{u \in U_L} \log n_L(u) \tag{3}$$

With this measure, the first image of a user increases the popularity of the landmark by 1, while subsequent images of that same user add less and less popularity: the second image raises popularity by log 2 and the n-th image by log $(n/n-1)$.

3.4. Road popularity calculation

3.4.1. Indicator of road popularity

Users take images on street or inside the buildings representing POIs (e.g., restaurants) located on the street. Images owning the tags which contain road names act as the indicators of road popularity. Unfortunately, the vast majority of the images are not able to indicate the street or the road where they were taken directly, because the tags do not contain names of streets explicitly. It would lead to data sparsity if we only consider the geotagged images which explicitly indicate road names. Actually, images whose tags do not contain road names explicitly still have some contributions to road popularity since they record the POIs on roads. When standing on a street, a tourist might take photos of some buildings along the street. Then she uploads the image with a tag indicating a building name rather than the street name since she is more interested in the building representing a POI than the street itself. For instance, there are a large number of images about the *Old Pinakothek* in Munich, they are normally tagged with the name of "old pinakothek" or "pinakothek". Although the *Old Pinakothek*is located on the *Arcis Street*, these photos are rarely tagged with "arcisstreet" or "arcisstr.". But such images can indicate that some tourists had been on this street before, because the POIs which they reflect are geo-located on this street.

Despite in the absence of tag indicating road name, such images still have the potential of being the indicators of road popularity by recording the POIs on these streets. Therefore, to address the data sparsity issue, we also consider these images as popularity indicators.

3.4.2. Image assignment

Based on above discussion, in a densely populated urban area we can take into account almost all the tourism-related geotagged images as the indicators of the road popularity, since these images, more or less, record or reflect roads themselves or POIs located on the roads. The problem is thus to determine, for an image, of which street (road) it should act as a popularity indicator. Essentially, this problem is how to assign these images to the correct roads.

Some researchers used map-matching algorithms to assign the points of GPS trajectories to the appropriate roads (e.g., Lou et al., 2009). Map-matching is the process of aligning a sequence of observed user positions with the road network on a digital map. It is applied to many trajectory-based applications, e.g., moving object management, traffic flow analysis and driving directions (Lou et al., 2009). While some map-matching methods were already proposed (e.g., Lou et al., 2009) to deal with low-sampling rate GPS trajectories (e.g., one point every 2 min), they are not applicable to Flickr images, since a large amount of these geotagged images are not useful to generate a trajectory. More specifically, if we consider a series of images taken by the same user on the same day, if all of them were taken at the same place, they cannot constitute a trajectory in the absence of neighbouring locations and directional information. Particularly, the majority of the images in our case study are such ones; this will be explained with more details and supported by statistical data in Section 4.3. Thus, the existing map-matching algorithms cannot perform well on Flickr geotagged images. At the same time, map-matching algorithms are proposed initially to assign GPS points of a trajectory to the right road. However, those approaches are mostly useful only for vehicles. For example, Lou et al. (2009) suggested that the speed of the vehicle can be used to determine if a part of a trajectory that falls between a highway and a nearby parallel road should be assigned to the highway or the road. In our context, we cannot make these types of assumptions, as we have no information on the transportation means of people taking pictures. This means that map-matching algorithms probably are not that applicable to Flickr images even in the case that the majority of the images could constitute trajectories. Map-matching algorithms take into consideration the positional context of the vehicle and the network's topological information. Although their approach does not apply to our case since, like in the previous example, the assumptions are only suitable for the vehicles, we adopt a similar approach here by taking into consideration the spatial and topological characteristics of the environment where the pictures are taken.

Therefore, in this research, we have developed another image assignment method, explained below.

Before that, we clarify the idea behind our proposed assignment method. Firstly, we introduce two types of images in two scenarios respectively.

Scenario 1: Firstly, for the images with tags indicating road names and the ones located on roads exactly (e.g., a point representing an image is located within a polygon representing a road in digital map) or images extremely close to roads (e.g. the distance between a geo-tagged image and its closest road is less than 3 m), should be assigned to the closest roads (road segments) directly (e.g., point A and B in Fig. 2).

Scenario 2: Apart from the kind of images in Scenario 1, the vast majority of the images are the ones recording or reflecting the POIs. Among such images, some are located within the buildings representing POIs. For a building representing a POI, almost every tourist would go into it from the main entrance instead of using other doors. Meanwhile, usually the entrance of POI is located on or at least closest to the POI's nearest road in space (see Fig. 2). In such case, the tourist absolutely had walked on the road where the entrance is located. For building representing POI, therefore, images located within it can be assigned to the closest road of the POI regardless of the distances of these images to other roads (see point C, D, E and F in Fig. 2).Besides, for most of these images, the roads to which they are assigned via the location of POI entrance are actually the roads closest to themselves in space (see point C, D and E in Fig. 2).

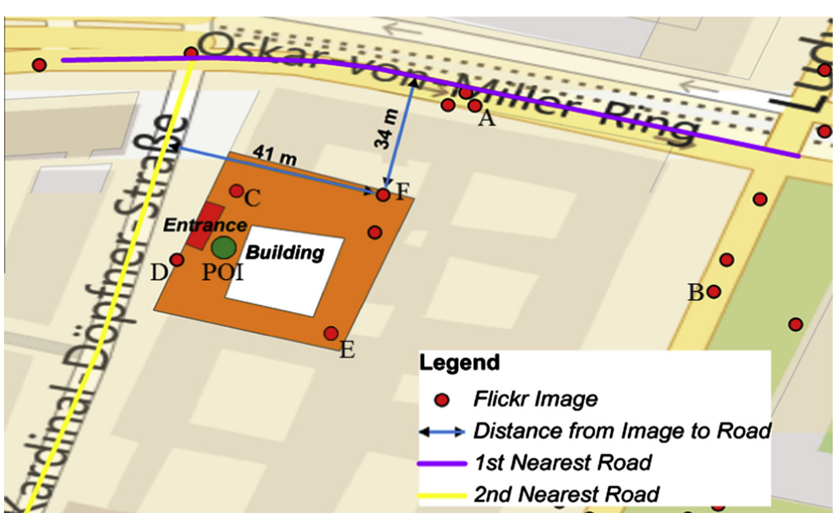


Fig. 2. The problem of assigning image in the road corner region.

Please cite this article in press as: Sun, Y., et al. Road-based travel recommendation using geo-tagged images. *Computers, Environment and Urban Systems* (2013), http://dx.doi.org/10.1016/j.compenvurbsys.2013.07.006

Here we call the types of images described in Scenario 1 and Scenario 2 ***Type 1*** and ***Type 2*** respectively. For images of ***Type 1*** and ***Type 2***, the roads to which they are most likely assigned are their 1st closest roads in space. Meanwhile in densely populated urban area where the road is densely distributed, like Munich in this research, images of ***Type 1*** and ***Type 2*** are able to represent the vast majority of images used for travel recommendation. Therefore, we can assume in general that almost all the images used for travel recommendation could be most likely assigned to the 1st closest roads. In this case, however, there would be potential false assignments when we deal with images taken on the corner region. For instance, for an image displayed as Point F in Fig. 2, if the POI entrance is considered, it should be assigned to the 2nd nearest road (yellow colour); otherwise, it should be assigned to the first 1st road (violet colour). Although the possibilities of the images assigned to other closest roads instead of the 1st roadsare relatively low, we still need to consider them in image assignment.

For this kind of problem, we built a model by which images can be assigned to not only1st, but also 2nd, ..., or *k*th nearest roads. In order to reduce the complexity of modelling, for each image currently we only take into account the 1st nearest road or 2nd nearest road as the candidate roads of assignment. In other words, we assume that all of the tourists' images should belong to their 1st nearest roads or 2nd nearest roads. Although only two nearest roads are considered, the model would still be rather reasonable. For an image of ***Type 1***, two candidate roads are definitely enough for it to be assigned to (e.g., Point A and B in Fig. 2). For an image of ***Type 2***, taking advantage of two candidate roads is rather reliable as well, considering the entrances of the buildings representing POIs as we discuss above. Specifically, for almost all buildings representing POIs, the entrances of them are exclusively located on their two nearest roads which are always the two nearest roads of the images recording or reflecting the POIs simultaneously. In Fig. 2, it is clear that for the images represented as Point C, D, E and F separately, the two nearest roads are the same with that of the corresponding POI entrances.

Finally, for the images of ***Type 1***and ***Type 2***, this assignment becomes a binary classification problem. The classification category set is {1,0}, where 1 means that the image is assigned to 1st nearest road, and 0 means that it is assigned to 2nd nearest road. Through the noise elimination of Section 3.2, we can acquire the image set among which the image belongs to either ***Type 1***or ***Type 2*** as the experimental data. Therefore, we propose a classification-based approach to assign these images to the roads in this research. The classification features as well as the detailed classification algorithms are presented as follows.

3.4.3. Classification features

We can expect that the distances of image to the 1st and 2nd nearest roads have powerful influences on the classification (assignment) result. Basically, we select these two distances as the classification features. In addition to the distances, the number of POIs is also regarded as the classification feature, since tourists are sometimes attracted by one road duo to its POIs offering a variety of services. We need to assign the POIs to the correct roads as well. If a POI has address information, we assign it to the appropriate roads by matching the street name extracted from its address with the name of the road; otherwise, we assign it to its closest road. Concerning in general the different importance of various POIs from the perspective of a tourist, here we use three categories of POIs, which are visiting attractions, eating, and other tourism activities (e.g., lodging). According to the statistical result of the Foursquare POIs in Munich, the number of users who checked-in at these three categories of POIs are 13%, 14% and 34% of the total number of users who checked-in at all POIs respectively. Thus, these three categories of POIs are considered as important POIs to tourist. Among these 3 categories of POIs *tourism attractions* rank first, followed by *eating & drinking* sites, and *others* (e.g. hotel, bank, etc.) rank last. This ranking is supported by a statistical result of the Foursquare POIs as well. The average number of users who check-in at *tourism attraction*, *eating & drinking* sites and *others* is 105, 82 and 67 respectively, which represents 1.4, 1.1 and 0.9 times the average number of users who checked-in at all POIs respectively. If one category of POI has a higher average number of users who checked-in at, we consider that it is more attractive for tourists.

The subcategories of these three categories of POIs are presented as following

Tourism Attraction (***Category 1***): museum, palace, church, gallery, memorial, monument, square, university, etc.
Eating & Drinking (***Category 2***): restaurant, cafe, canteen, vending machine, etc.
Others (***Category3***): hotel, post office, bank, ATM, Info centre, toilet, etc.

In addition, according to the problem shown in Fig. 2, the location of POI entrance could be integrated into classification model. In this case, we could generate some new feature variables, e.g. the distance between image and its 1st and 2nd nearest POI entrances, which would be very likely to enhance the assignment of image. However, in this research we are not able to leverage the location of POI entrance due to the limitation of data availability.

Here, therefore the classification feature set is represented as

$$(Dis_1, Dis_2, Num_1^1, Num_2^1, Num_1^2, Num_2^2, Num_1^3, Num_2^3) \tag{4}$$

where Dis_1 is the distance of the image to the 1st nearest road; Dis_2 is the distance of the image to the 2nd nearest road; Num_1^1 is the number of POIs of ***Category1*** on the 1st nearest road; Num_2^1 is the number of POIs of ***Category1*** on the 2nd nearest road; Num_1^2 is the number of POIs of ***Category2*** on the 1st nearest road; Num_2^2 is the number of POIs of ***Category2*** on the 2nd nearest road; Num_1^3 is the number of POIs of ***Category3*** on the 1st nearest road; Num_2^3 is the number of POIs of ***Category3*** on the 2nd nearest road.

3.4.4. Classification methods

In this paper, we applied machine learning methods for the classification. For the purpose of classification, machine learning is always superior to other classification methods (e.g., statistical methods), particularly when the relationships between aims and factors are unclear or insignificant. In our classification task, the relationships (associations) between classification category and feature variables are not clearly correlated. Thus, two classic machine learning methods, namely Binary Logistic Regression and Support Vector Machines, which are widely used in remote sensing, land use, data mining, etc., are chosen for the classification in this research. These two methods are deployed for the same selected training and test data set separately. The results of these two methods are then compared in terms of validation, and at last the method with a higher overall accuracy is chosen to assign all the rest images. The two methods are introduced as following:

3.4.4.1. Binary Logistic Regression. Binary Logistic Regression (BLR) is able to measure the relationship between the binary dependent variable and one or several independent variable(s), by computing the probabilities of the outputs for the different category. BLR allows assessing the contribution of individual variables to the classification and subsequently choosing the important variables to

Table 1
Filtering accuracy of user type with different threshold.

Month count (k)	3	4	5	6	7
E(u)	$3*(-1/3)\log_2(1/3)$	$4*(-1/4)\log_2(1/4)$	$5*(-1/5)\log_2(1/5)$	$6*(-1/6)\log_2(1/6)$	$7*(-1/7)\log_2(1/7)$
Threshold value	1.6	2.0	2.3	2.6	2.8
Filtering accuracy	87.5%	88.2%	89.0%	87.5%	87.1%

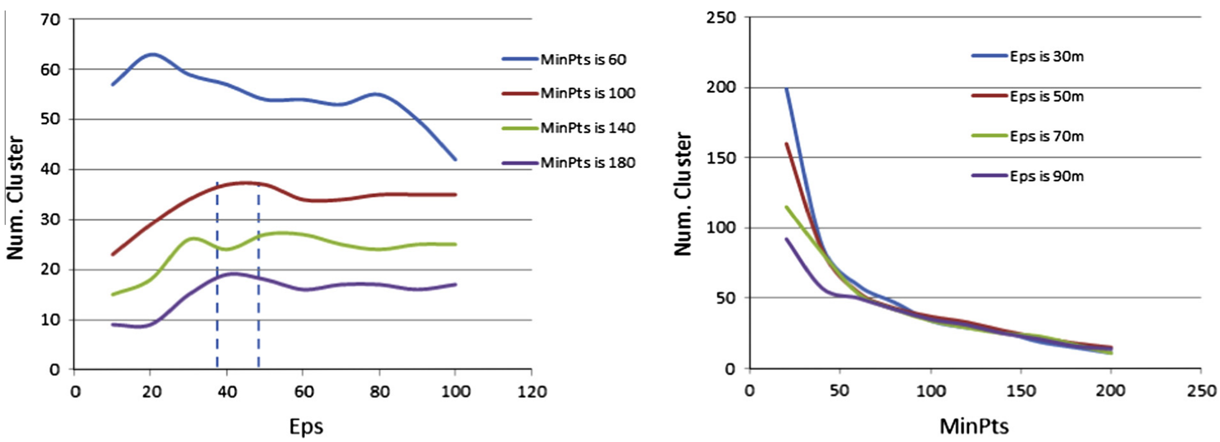


Fig. 3. The number of clusters detected with different value of parameters.

build a new model, leading to an enhancement of the classification. The BLR model is represented as

$$P(y) = \frac{1}{1+e^{-(a_0+a_1x_1+a_2x_2+a_3x_3+a_4x_4+a_5x_5+a_6x_6+a_7x_7+a_8x_8)}} \tag{5}$$

where $y \in \{1, 0\}$ is the dependent variable, and (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8) are the independent variables a_0 is the intercept, and a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8 are the coefficients of the independent variables ($x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$). In this classification problem, the targeted classification category (1 or 0) and the eight classification features in Eq. (4) correspond to the dependent variable Y and eight independent variables in Eq. (5) respectively.

3.4.4.2. Support Vector Machines. Support Vector Machines (SVM) is firstly proposed by Vapnik (1995) to solve the classification problem. Particularly, it can avoid the over-fitting problem in data training. Using kernel function, SVM can efficiently perform non-linear classification. It is widely used to address classification, prediction and regression problem in many fields (e.g. image processing, remote sensing, etc.).

The formula can be briefly presented as follows:

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} y_i y_j \alpha_i \alpha_j k(x_i, x_j) - \sum_{j=1}^{l}\alpha_j \\ \text{s.t.} \quad & \sum_{i=1}^{l} y_i \alpha_i = 0 \\ & 0 \leqslant \alpha_i \leqslant C, i = 1, \cdots l \end{aligned}$$

where $y \in \{1, -1\}$ is the classification variable, and x is the feature variable. Similarly, the eight features defined in classification feature set are all chosen as the feature variables, which correspond to independent variables x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8 in BLR model separately. $K(x_i, x_j)$ is the kernel function which maps the input features into high-dimensional feature space.

3.4.5. Popularity of road

Like the popularity of landmark, the popularity of road should also take into account the inter- and intra-tourist difference of contribution. Duo to the availability of data as well as limitation of algorithm that we discuss in Section 3.3.2, here we consider only the intra-tourist difference of contribution with an assumption that the contribution of image to the popularity of road is the same among tourists. In order to calculate the popularity of individual

Table 2
The top-10 ranked landmarks identified by analysing geotagged images.

Ranking	Landmark	Popularity	Reviews in TripAdvisor	Related tags
1	Marienplatz	535.8	2013	Marienplatz, square
2	English Garden	184.8	1449	English garden, garden
3	BMW World & BMW Museum	162.8	1040	Auto, bmw, car, museum, welt
4	Hofbräuhaus	151.6	557	Hofbräuhaus, brewery,beer
5	Church of Our Lady	120.1	94	Frauenkirche, church, lady
6	New & Old Pinakothek	86.2	551	Pinakothek, art, museum
7	Olympic Park	56.2	511	Olympiapark, olympic
8	Deutsches Museum	52.3	707	Deutsches museum
9	Nymphenburg Palace	47.3	553	Nymphenburg, palace, schloss
10	TierparkHellabrunn	10.3	247	Tierpark,zoo, hellabrunn

Table 3
Classification accuracy of BLR and SVM methods.

Prediction							
BLR				SVM			
Observation							
Class	1	0	Accuracy	Class	1	0	Accuracy
1	81	0	100%	1	81	0	100%
0	23	3	11.5%	0	21	5	19.2%
Total			78.5%	Total			80.4%

Table 4
Selected variables in the BLR model.

Variables	B	S.E.	Wald	df	Sig.	Exp (B)
x_1	3.235	1.740	3.456	1	0.063	25.400
x_3	20.125	54382.509	0.000	1	1.000	5.495
x_4	.234	36633.066	0.000	1	1.000	1.264
x_5	–.274	49222.006	0.000	1	1.000	0.760
x_6	−22.577	23143.635	0.000	1	0.999	0.000
x_7	−2.380	1.178	4.083	1	0.043	0.093
Intercept (a_0)	2.464	18323.206	0.000	1	1.000	11.755

road, we applied the method used in landmark ranking as well. Similarly, consider the images assigned to road r were contributed by m different users. $U_r = \{u_1, \ldots, u_m\}$ is the user set of the images assigned to road r. The number of images taken on road r by user u is $n_r(u)$. The popularity of the road can be represented as follows:

$$Pop(r) = m + \sum_{u \in U_r} \log n_r(u) \tag{7}$$

3.5. Routing recommendation

3.5.1. Recommendation index

In addition to the number of the images, the number of POIs and the length of road are considered in the routing recommendation. Here we define a recommendation index named $Rec\_Index(r)$ to quantify the overall popularity of the road r. The $Rec\_Index(r)$ of the road mainly consists of two parts: tourism popularity and POI usability.

Regarding the different categories of POI mentioned in Section 3.5, POI usability of the road r is described as

$$Poi(r) = w_1 Num^1_{POI}(r) + w_2 Num^2_{POI}(r) + w_3 Num^3_{POI}(r) \tag{8}$$

$Num^1_{POI}(r)$, $Num^2_{POI}(r)$ and $Num^3_{POI}(r)$ are number of POIs of ***Category 1***, ***2*** and ***3*** separately. w_1, w_2 and w_3 are the weights of $N^1_{POI}(r)$, $N^2_{POI}(r)$ and $N^3_{POI}(r)$, and $w_1 + w_2 + w_3 = 1$.

Therefore, $Rec\_Index(r)$ is computed as follow

$$Rec\_Index(r) = \alpha_1 Pop(r) + \alpha_2 Poi(r) - \beta\, Leng(r) \tag{9}$$

$Pop(r)$ is the popularity of road r in Eq. (7). $Leng(r)$ is the length of the road r, which is used as a penalty to avoid generating unpractical routing whose length is too long for user. α_1, α_2 and β are the weights of $Pop(r)$, $Poi(r)$ and $Leng(r)$ separately.

3.5.2. Best routing generation

Considering two landmarks L_p and L_q, in terms of the $Rec\_Index(r)$ in Eq. (9) the best routing can be generated by

$$Rou^*(L_p, L_q) = argmax_{Rou(L_p,L_q)} \sum_{r \in Rou(L_p,L_q)} Rec\_Index(r) \tag{10}$$

$Rou(L_p, L_q)$ is a routing between L_p and L_q, and r is a separate road in $Rou(L_p, L_q)$. $Rou^*(L_p, L_q)$ is the one which is of the maximum value of total $Rec\_Index(r)$.

4. Experimental results

4.1. Preprocessing data

The proposed algorithm is implemented and tested with Flickr photos in Munich, Germany. All the images were taken during the year 2010 and 2011. Totally, 45,950 images were acquired via Flickr API (http://www.flickr.com/services/api/). From these images, we need to extract the images taken by tourists using the entropy filtering method described in Section 3.3. Firstly, we have selected a sample of images whose user location is available

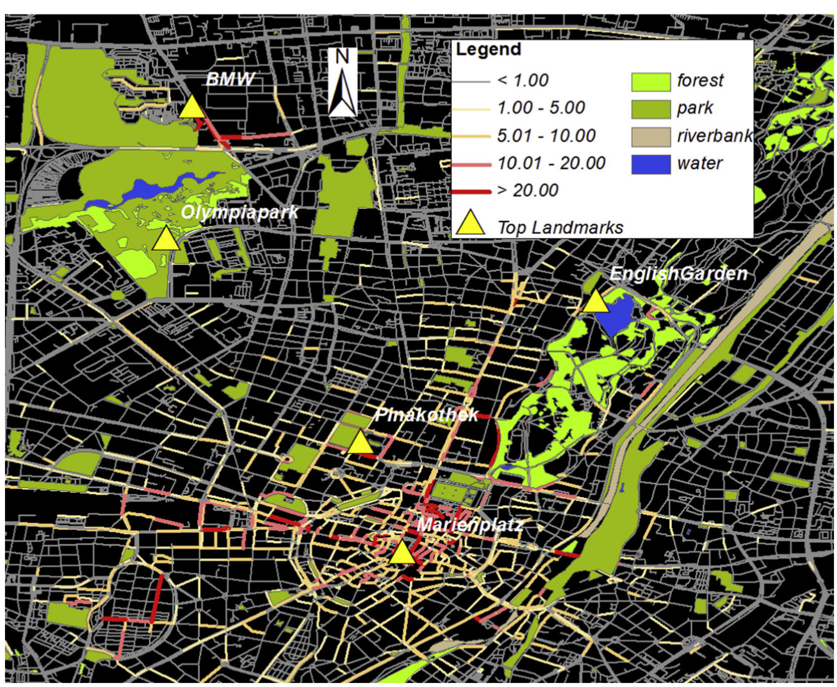


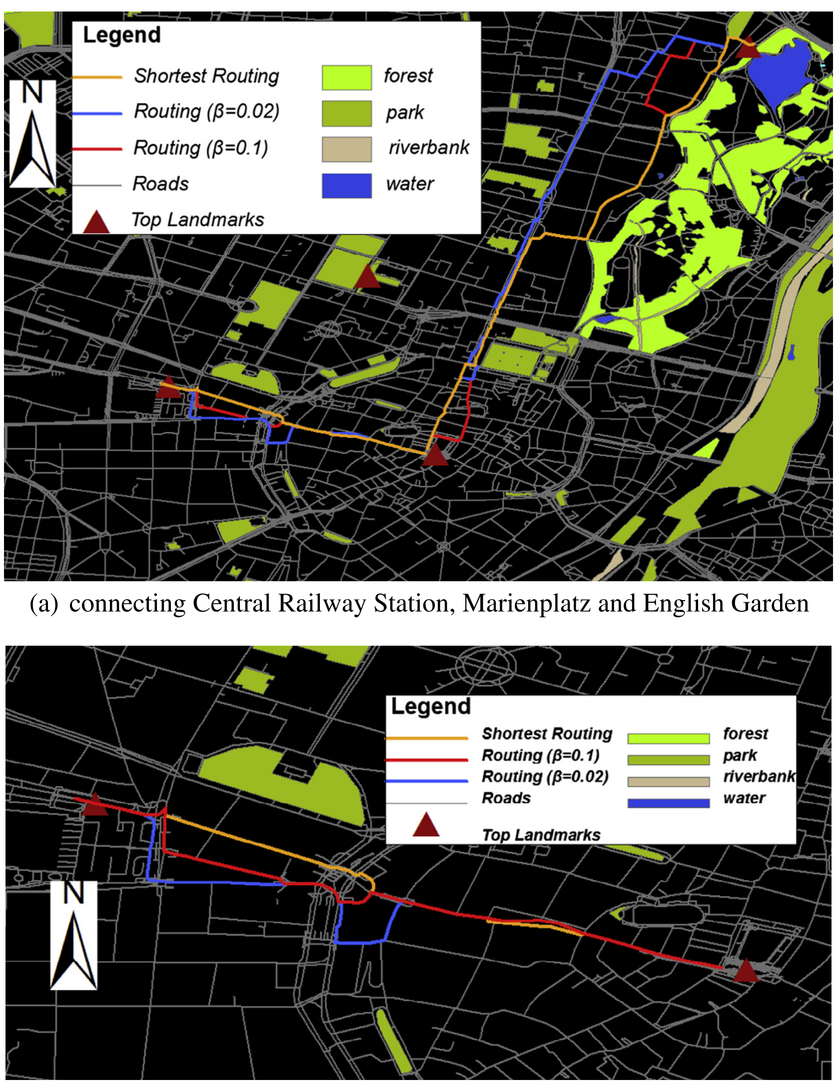
Fig. 4. The tourism popularity of the road in Munich.

in terms of the user profile to acquire an appropriate threshold for distinguishing tourists from residents among image contributors. Simply, we suppose that the images of the user in one year are equally distributed among k months. According to Eq. (1), we in turn calculated the $E(u)$ at different k. For instance, there is a user who took 12 pictures in one year. If these pictures are equally distributed among three months, in other words, 3 pictures per month, the $E(u)$ is $3 * (-1/3)\log_2(1/3)$ which equals to 1.6. If these pictures are equally distributed among four months, the $E(u)$ is $4 * (-1/4)\log_2(1/4)$ which equals to 2. We attempted to select the appropriate one as the best threshold among these values of $E(u)$ in terms of taking them as the threshold value for the sampling data set separately. Table 1 shows the filtering accuracy of the user type of the same sampling images with different $E(u)$ as the threshold value. Thus, 2.3 is the best threshold with a filtering accuracy of 89% and it was chosen as the filtering threshold. With this filtering threshold, a total of 32,558 images taken by tourists were distinguished from the 45,950 images. Subsequently, among these 32,558 images, we discarded 1223 images whose tags are not indicating the popularity of any POI or road but reflecting some persons, persons' gestures and other contents unrelated to tourism. Each one among these discarded images is the one whose tag has only one user and thus was easily selected. Consequently, 31,335 geotagged images taken by 2422 users were finally chosen as the study data in this paper. In addition, the road network and POI data were downloaded from OpenStreetMap (http://www.openstreetmap.org/).

4.2. Detecting top landmarks

The DBSCAN method needs two parameters. We generated different results of clustering at different values of the parameters. Fig. 3 demonstrates the distribution of cluster numbers with different *Eps* and *MinPts* separately. Apart from the curve with a *MinPts* of 60, the other curves are similarly changing. When the value of *Eps* is between 40 m and 50 m, the cluster count reaches a maximum value. The number of detected clusters is always decreasing with the increase of *MinPts*. The decreasing becomes slow when *MinPts* is up 50. Here we made *Eps* = 50 m and *MinPts* = 100, and consequently found out 37 clusters.

The dictionary of landmarks in this paper was from *TripAdvisor* (http://www.tripadvisor.com/) which recommends the important landmarks in terms of tourists' reviews and comments. With the three steps in 3.3.1, we identified 15 landmarks. Table 2 shows the top-10 ones. Except *Frauenkirche* and *Deutsches Museum*, landmarks of higher popularity always own more reviews in



(a) connecting Central Railway Station, Marienplatz and English Garden

(b) between Central Railway Station and Marienplatz

Fig. 5. The recommended routings with $\alpha_1 = 1$ and $\alpha_2 = 5$.

Please cite this article in press as: Sun, Y., et al. Road-based travel recommendation using geo-tagged images. *Computers, Environment and Urban Systems* (2013), http://dx.doi.org/10.1016/j.compenvurbsys.2013.07.006

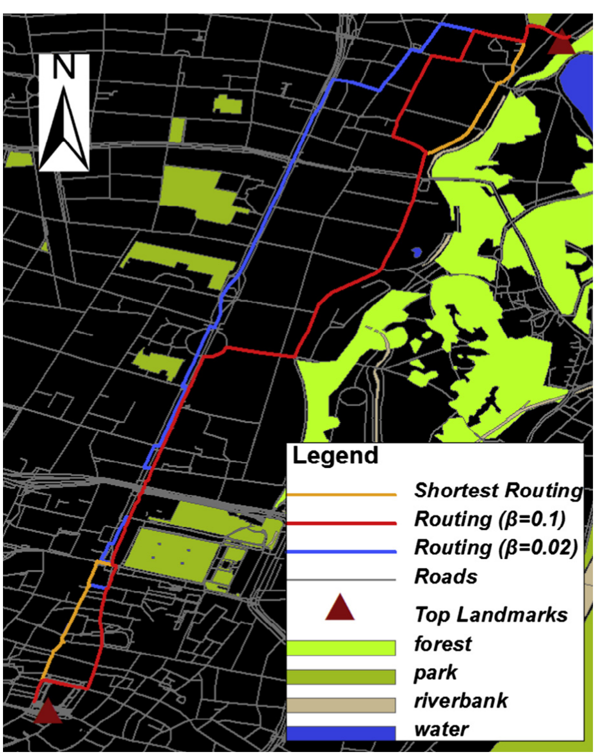
TripAdvisor. Moreover, the top three landmarks with a popularity of up 100 had been reviewed over 1000 times in *TripAdvisor*.

4.3. Calculating tourism popularity of road

Firstly, we analyzed the data. We assume that a Flickr user's trajectory should consist of images (1) taken by this user on the same day; and (2) associated with more than one geo-location. Among the 31,335 images part of our experimental data, only 40% meet these requirements. Therefore, for the task of image assignment, we did not apply the existing map-matching method (which is based on trajectory), since the majority of images are not useful to generate trajectories. Instead, we used our classification-based method that exploits spatial attributes of images, as stated in Section 3.4.

Before applying the BLR and SVM methods for the classification, building model from the training data is necessary. At the same time, we need to evaluate the two methods by using a test data set, and subsequently choose the one with better performance to classify the rest data.

In the data set, there are 452 images bearing the tags able to indicate road name. From these 452 images, 320 images (nearly 1% of the total data set) are randomly selected to build and evaluate the classification model. Among them, 213 images are chosen as the training data and the other 107 ones as test data. Table 3 shows the validation results of classification using the two methods. Among the test data, 81 images belong to *class 1* and 26 ones to *class 0*. *Class 1* means image is actually assigned to its 1st nearest road and *Class 0* means image is assigned to its 2nd nearest road. The total validation accuracies of BLR and SVM are 78.5% and 80.4% respectively. These two methods perform almost same for our purpose. Both accuracies approximately 80% demonstrate that our proposed methods are fine in this study case. For both methods, the false predictions are almost resulted from that some images actually belonging to 2nd nearest roads (*Class 0*) are assigned to 1st nearest roads (*Class 1*) (see Table 3). In the presented work, we use SVM which is slightly higher as the classification model to predict the assignment categories of the rest 32,238 images.



(c) between Marienplatz and English Garden

Fig. 5 (*continued*)

Besides, to observe the influences of the variables on the classification result, the selected variables as well as their coefficients in the last stage of the stepwise BLR are listed in Table 3. It indicates that x_1 and x_7 are the most influential variables with the significance value (Sig.) <0.1. It means the distance to 1st nearest road as well as the number of POIs of ***Category 1*** on 1st nearest road have significant effects on the prediction. However, the effects of the other variables on the result are not significant. Contrary to the BLR, SVM model are not allowed to see the coefficients or weights of variables because the variables were together transformed into a high-dimension space during the classification.

Moreover, although the total accuracy is fine, based on the classification result of the *Class 0* and the coefficients of the variables, the two methods somewhat do not perform pretty well. The main reason is probably the selection of the features (variables) instead of the methods themselves. In other words, some of the features (e.g. x_3 and x_4) are not that correlated with assignment result (dependence variable) according to significance values (Sig.) in Table 4.

In general, SVM and BLR can be considered as effective methods to deal with the task of image assignment in this research. Firstly, considering disadvantages of Flickr images over GPS data as well as the unclear relationships between classification categories and feature variables, machine learning is probably more applicable to this task than both map-matching algorithms and other classification methods (e.g., statistical methods). Secondly, the total test accuracies of these two methods are rather good (approximately 80%).

On the basis of the abovementioned classification model using SVM method, we computed the tourism popularity of road. Here we mapped the popularity of road in Fig. 4. We can see that for Flickr user the road of East–West direction is more popular than the one of North–South direction in Munich.

4.4. Generating travel routing

Selecting two landmarks, we need to find out the recommended routing between them. Simply, we applied the classic Dijkstra's algorithm (Dijkstra, 1959) to generate the best routing between two landmarks by using the proposed Recommendation Index in Eq. (9) instead of the road length. Although we in general introduce our proposed approach at a road level, this algorithm aiming at routing generation is actually applied at a level of road segment in this research, since road segment can be considered as "smaller" road in space.

We assume that a novice user uses our travel recommendation system when she is on the central railway station after arriving Munich. She selects two ones from the 10 top ranked landmarks generated by our proposed system. Concerning the weights in Eq. (9), we generated routing at different values of the weights. Because the influence of α_2 (the weight of POI number) is relatively weak on the final result, in Fig. 5 we showed only the generated routings with different value of β while value of α_1 and α_2 is

Table 5
Top tags of semantic routings.

Start and end locations	Routings	*Top-three* POI-related tags
Marienplatz–English Garden	Baseline	Theatiner church, odenosplatz, church of st.louis
	β = 0.02	Town hall, theatiner church, theater
	β = 0.1	Leopoldstrasse, opera, church of st.louis

Table 6
Comparison of the routings generated by proposed and baseline methods.

Routes	Start landmark	End landmark	Baseline method			Proposed method		
			Image Num	POI Num	Check-in Num	Image Num	POI Num	Check-in Num
1	Railway Station	Marienplatz	2005	55	333	2167	87	331
2	Marienplatz Square	English Garden	2029	35	228	2298	98	413
3	Marienplatz Square	Olympic Park	1336	45	162	1715	57	122
4	Marienplatz Square	Nymphenburg Palace	2109	88	440	2408	90	493
5	Marienplatz Square	BMW World	1869	95	397	3824	109	516
6	English Garden	Deutsches Museum	2785	74	361	2226	119	542
7	Nymphenburg Palace	Hofbräuhaus Beer garden	4739	122	893	4753	124	948
8	Pinakothek Art Gallery	BMW World	385	74	100	533	38	19
9	Olympic Park	Deutsches Museum	873	96	298	4515	140	629
10	Marienplatz Square	TierparkHellabrunn Zoo	388	36	80	4083	92	577

constant. These routings are connecting Central Railway Station, and two top landmarks (*Marienplatz* and *English Garden*). To make comparison, we also demonstrate the shortest routing as the baseline routing exclusively considering road length. The comparison indicates that the generated routing is more similar to the baseline routing when β (the weight of road length) is higher. In addition, the total lengths of the recommended routings are all less than 2 times of that of the corresponding shortest routings. It shows the feasibility of the routing recommended to the user.

In addition, we display a simple semantic routing example. Table 5 shows the most frequent POI-related tags of images assigned to the roads composing the generated routings in Fig. 5. With such tags, the system can offer some user-generated semantic information for the routing.

4.5. Routing quality

The quality of the generated routings cannot be evaluated directly, since we do not have access to user-generated GPS trajectory data or tourist survey data. But here, we still evaluate the generated routings by comparing them with the shortest routings generated by the baseline method. We used Flickr image, OSM POI, as well as Gowalla Check-in located on or close to (with a distance <50 m) the roads of routing. Like Foursquare, Gowalla is a LBSN service and the information of its user-generated georeferenced check-in, such as user ID, longitude, latitude and time, is available (Cho, Myers, & Leskovec, 2011). The Flickr images are the ones that were already preprocessed in Section 4.1. Similarly, we applied the entropy filtering method (see Section 3.2) to check-in data to eliminate the ones generated by local residents with the same filtering threshold (i.e., 2.3) that we used for Flickr images.

Ten routings generated by our proposed method with $\alpha_1 = 1$, $\alpha_2 = 5$and $\beta = 0.1$ are compared with 10corresponding routings generated by baseline method. Table 6 displays the start and end landmarks of the routings as well as comparison measurements. With *Wilcoxon test*, the image count, POI count and check-in count of the routings generated by proposed method are all statistically higher than that generated by baseline method. The *p*-values are 0.032, 0.044 and 0.049 respectively, all of which are <0.05. It indicated that the routings generated by proposed method are significantly of more user activities than that by baseline method. Compared to the shortest routing, the generated routing by our method should be somehow good, at least with more user activities.

5. Conclusions and future work

This paper presented a new travel recommendation approach integrating landmark and routing. The routing is based on road instead of the GPS trajectory used in the researches of the state-of-the-art. Using spatial clustering method, we generated some clusters of images. Subsequently, we identified important landmarks from these clusters and ranked these landmarks in terms of tourism popularity. To recommend the routings between landmarks, we calculated the tourism popularity of road. Before that, we assigned geotagged image to the correct road in terms of employing data mining method. Finally, appropriate routing between two landmarks was generated using a recommendation index proposed in the paper. An empirical evaluation of our proposed approach was conducted on the Flickr image set. The experimental results demonstrate that the approach is able to recommend the user suitable routings considering image, POI and road length.

Duo to the data quality (e.g., incompleteness, tagging error, etc.) of the Flick data, to enhance the travel recommendation, there are some aspects should be taken into account in the future research. Firstly, other data sources (e.g. highly accurate GPS data and other VGI) should be considered to be integrated into our system to enhance the recommendation. The identification and ranking of landmarks should take into consideration the experiences and knowledge of tourist who have visited these landmarks. The challenge here mainly relies on how to gather information that can help to accurately assess the experiences and knowledge of tourist. More importantly, to enhance the assignment of images to appropriate roads, we in addition will attempt to integrate the location of POI entrance into the image assignment. In that case, it would generate new feature variables (e.g. the spatial distance between image and POI entrance) which could probably enhance the performance of our proposed method. We also could employ the spatial distribution of POI entrance to improve the image assignment by mining some spatial association rules between POI and road in the road network. To obtain the location of POI entrance, we will seek to acquire the official data of urban buildings or extract such information from some existing map services (e.g., OSM, Google Map, etc.) using some methods (e.g., geocoding). In addition to the location of POI entrance, the boundary of the building representing POI is probably helpful to the image assignment and thus should be integrated into the recommendation system, considering the topological relationship between image (point) and building (polygon) representing POI. On the other hand, we will attempt to take advantage of more advanced methods to assign the images to the correct road. For instance, instead of only two candidate roads, more roads should be taken into consideration as the candidate roads for the image to be assigned to, and thus a more complicated model is probably needed to be built. We also plan to use other classification methods (e.g., other machine learning methods, statistical methods, etc.) to improve the assignment accuracy. At the same time, the recommendation system should pay more attentions to the personalized need of user. Moreover, a context-aware recommendation would be more helpful to users by integrating context information (e.g., weather, user task, user preference, time, season, etc.) into the image assignment as well as the popularity calculation. Apart from the visit of landmark, other activities (e.g. eating & drinking, shopping, etc.) that might influence the final travel decision of the user should be integrated into the personalized recommendation system. Subsequently, the result of routing recommendation should be validated in terms of the test conducted by volunteers or tourists. At last, a real-time recommendation is vital to users, particularly the smart phone owners. Thus the time needed to spend on landmarks as well as the routings become important constraints and thus should be considered in a time-constrained travel recommendation application.

Acknowledgements

This research is supported by CSC (China Scholarship Council). We would like to express our gratitude to the three anonymous referees for their comments and thank Christian Sengstock (Database Systems Research Group, Institute of Computer Science, University of Heidelberg) for sharing the research data.

References

Andrienko, G., Andrienko, N., Bak, P., Kisilevich, S. & Keim, D. (2009). Analysis of community-contributed space-and time-referenced data (example of flickr and Panoramiophotos). In *IEEE symposium on visual analytics sciences and technology*, October 12–13, 2009, New Jersey, USA (pp. 213–214).

Arase, Y., Xie, X., Hara, T. & Nishio, S. (2010). Mining people's trips from large scale geo-tagged photos. In *Proceedings of the international conference on Multimedia*, October 25–29, 2010. Firenze, Italy (pp. 133–142). New York: ACM.

Chaudhry, O. & Mackaness, W. (2012). Automated extraction and geographical structuring of Flickr Tags. In *Proceedings of the 4th international conference on advanced geographic information systems, applications, and services*, January 30 – February 04, 2012, Valencia, Spain (pp. 134–139).

Cheng, A., Chen, Y., Huang, Y., Hsu, W. H. & Liao, H. M. (2011). Personalized travel recommendation by mining people attributes from community-contributed photos. In *Proceedings of the 19th ACM international conference on Multimedia*, November 28–December 01, 2011, Scottsdale, Arizona, USA (pp. 83–92). New York: ACM.

Cho, E., Myers, S. A., & Leskovec, J. (2011). Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1082–1090). New York, NY, USA: ACM.

Choudhury, M.D., Feldman, M., Amer-Yahia, S., Golbandi, N., Lempel, R. & Yu, C. (2010). Automatic construction of travel itineraries using social breadcrumbs. In *Proceedings of the 21st ACM conference on Hypertext and hypermedia*, June 13–16, 2010, Toronto, Ontario, Canada (pp. 35–44). New York: ACM.

Clements, M., Serdyukov, P., Vries, A. & Reinders, M. (2010). Using flickrgeotags to predict user travel behaviour. In F. Crestani, S. Marchand-Maillet, H. H. Chen, E. N. Efthimiadis and J. Savoy (Eds.), *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, 19–23 July 2010, SIGIR '10, Geneva (pp. 851–852). New York: ACM.

Crandall, D., Backstrom, L., Huttenlocher, D. & Kleinberg, J. (2009). Mapping the World's Photos. In *Proceedings of the 18th international conference on World wide web*. April 20–24, 2009, Madrid, Spain (pp. 761–770). New York: ACM.

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik, 1*, 269–271.

Ester, M., Kriegel, H. P., Sander, J. & Xu, X. (1996). A density based algorithm for discovering clusters in large spatial databases with noise. In Simoudis, E., Han, J. W., & Fayyad, U. M. (Eds.) *Proceedings of the 2nd international conference on knowledge discovery and data mining*. August 02–04, 1996, Portland, Oregon, USA (pp. 226–231). Menlo Park, California: AAAI Press.

Girardin, F., Calabrese, F., Dal Fiore, F., Ratti, C., & Blat, J. (2008). Digital footprinting: Uncovering tourists with user-generated content. *IEEE Pervasive Computing, 7*(4), 36–43.

Girardin, F., Dal Fiore, F., Ratti, C., & Blat, J. (2007). Leveraging explicitly disclosed location information to understand tourist dynamics: a case study. *Journal of Location Based Services, 2*(1), 41–56.

Girardin, F., Vaccari, A., Gerber, A., & Ratti, C. (2009). Quantifying urban attractiveness from the distribution and density of digital footprints. *Journal of Spatial Data Infrastructure Research, 4*, 175–200.

Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal, 69*(4), 211–221.

Guy, M., & Tonkin, E. (2006). Folksonomies: Tidying up Tags? In: D-Lib Magazine, 12(1).

Hile, H., Vedantham, R., Cuellar, G., Liu, A., Gelfand, N., Grzeszczuk, R. & Borriello G. (2008). Landmark-based pedestrian navigation from collections of geotagged photos. In *Proceedings of the 7th international conference on mobile and ubiquitous multimedia*, December 03–05, 2008, Umeå, Sweden (pp. 145–152). New York: ACM.

Horozov, T., Narasimhan, N. & Vasudevan, V. (2006). Using location for personalized POI recommendations in mobile environments. In *Proceedings of the international symposium onapplications on Internet applications and the Internet*, 23–27 January 2006. Phoenix, AZ, USA (pp. 124–129). Washington, DC: IEEE Computer Society.

Jankowski, P., Andrienko, N., Andrienko, G., & Kisilevich, S. (2010). Discovering landmark preferences and movement patterns from photo Postings. *Transactions in GIS, 14*(6), 833–852.

Ji, R., Gao, Y., Zhong, B., Yao, H., & Tian, Q. (2011). Mining flickr landmarks by modeling reconstruction sparsity. *ACM Transactions on Multimedia Computing, Communications, and Applications, 7*(Supplement), 31.

Kennedy, L., Naaman, M., Ahern, S., Nair, R., & Rattenbury, T. (2007). How flickr helps us make sense of the world: Context and content in community-contributed media collections. In R. Lienhart, A. R. Prasad, A. Hanjalic, S. Choi, B. P. Bailey and N. Sebe (Eds.), *Proceedings of the 15th international conference on Multimedia*, 24–27 September 2007, MULTIMEDIA '07, Augsburg (pp. 631–640). New York: ACM.

Lou, Y., Zhang, C., Zheng, Y., Xie, X., Wang, W. & Huang, Y. (2009). Map-matching for Low-sampling-rate GPS Trajectories. In *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems*, 04–06 November, 2009, Seattle, Washington, USA (pp. 352–361). New York: ACM.

Lu, X., Wang, C., Yang, J., Pang, Y. & Zhang, L. (2010). Photo2Trip: generating travel routes from geo-tagged photos for trip planning. In A. D. Bimbo, S. F. Chang and A. W. M. Smeulders (Eds.), *Proceedings of the internationalconference on Multimedia*, 25–29 October 2010, MM '10, Firenze (pp. 143–152). New York: ACM.

Majid, A., Chen, L., Chen, G., Mirza, H. T., Hussain, I., & Woodward, J. (2012). A context-aware personalized travel recommendation system based on geotagged social media data mining. *International Journal of Geographical Information Science, 27*(4), 662–684.

Popescu, A. & Grefenstette, G. (2009). Deducing trip related information from Flickr. In J. Quemada, G. León, Y. S. Maarek and W. Nejdl (Eds.), *Proceedings of the 18th internationalconference on World Wide Web*, 20–24 April 2009, WWW '09, Madrid (pp. 1183–1184). New York: ACM.

Popescu, A., Grefenstette, G. & Moëllic, P. A. (2009). Mining tourist information from user-supplied collections. In *Proceeding of the 18th ACM conference on Information and knowledge management*, November 02–06, 2009, Hong Kong, China (pp. 1713–1716). New York: ACM.

Rattenbury, T., Good, N. & Maaman, M. (2007). Towards automatic extraction of event and place semantics from flickr tags. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*. July 23–27, 2007, Amsterdam, The Netherlands (pp. 103–110). New York: ACM.

Rattenbury, T., & Naaman, M. (2009). Methods for extracting place semantics from Flickr tags. *ACM Transactions on the Web, 3*(1), 1–30.

Schlieder, C., & Matyas, C. (2009). Photographing a city: An analysis of place concepts based on spatial choices. *Spatial Cognition and Computation, 9*(3), 212–228.

Takeuchi, Y. & Sugimoto, M. (2006). CityVoyager: An outdoor recommendation system based on user location history. In: J. Ma, H. Jin, L. Yang, & J. Tsai (Eds.), *Ubiquitous intelligence and computing*, Vol. 4159 of Lecture Notes in Computer Science (pp. 625–636). Berlin: Springer.

Vapnik, V. (1995). *The nature of statistical learning theory*. New York: Springer-Verlag.

Wachowicz, M., Ong, R., Renso, C., & Nanni, M. (2011). Finding moving flock patterns among pedestrians through collective coherence. *International Journal of Geographical Information Science, 25*(11), 1849–1864.

Wei, L., Zheng, Y. & Peng, W. (2012). Constructing popular routes from uncertain trajectories. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, August 12–16, 2012, Beijing, China (pp. 195–203). New York: ACM.

Yin, Z., Cao, L., Han, J., Luo, J. & Huang, T. (2011). Diversified trajectory pattern ranking in geo-tagged social media. In *Proceedings of the eleventh SIAM international conference on data mining*, SDM 2011 (pp. 980-991), April 28–30, 2011, Mesa, Arizona, USA.

Yoon, H., Zheng, Y., Xie, X., & Woo, W. (2012). Social itinerary recommendation from user-generated digital trails. *Personal and Ubiquitous Computing, 16*(5), 469–484.

Zheng, Y. & Xie, X. (2010). Learning location correlation from GPS trajectories. In *Proceedings of 11th international conference on mobile data management*. 23–26 May, 2010, Kanas City, Missouri, USA (pp. 27–32). Washington, DC: IEEE Computer Society.

Zheng, Y. (2012) Tutorial on location-based social networks. In *Proceedings of the 21st international conference on world wide web*. April 16–20, 2012, Lyon, France.

Zheng, Y., Xie, X., & Ma, W. Y. (2009). Mining interesting locations and travel sequences from GPS trajectories. In *Proceedings of the 18th international conference on world wide web*. April 20–24, 2009, Madrid, Spain (pp. 791–780). New York: ACM.

Zheng, V. W., Zheng, Y., Xie, X. & Yang, Q. (2010). Collaborative location and activity recommendations with GPS history data. In J. Quemada, G. Leon, Y. S. Maarek and W. Nejdl (Eds.), *Proceedings of the 19th international conference on World Wide Web*, 20–24 April 2009, WWW '10, Raleigh, NC (pp. 1029–1038). New York: ACM.

Zheng, Y., & Xie, X. (2011). Learning travel recommendations from user-generated GPS traces. *ACM Transactions on Intelligent Systems and Technology, 2*(1), 1–29. Article 2.

Zheng, Y., Zha, Z., & Chua, T. (2012). Mining travel patterns from geotagged photos. *ACM Transactions on Intelligent Systems and Technology, 3*(3), 1–18.

Zheng, Y., & Zhou, X. (2011). *Computing with spatial trajectories*. Springer.

Please cite this article in press as: Sun, Y., et al. Road-based travel recommendation using geo-tagged images. *Computers, Environment and Urban Systems* (2013), http://dx.doi.org/10.1016/j.compenvurbsys.2013.07.006