Mining Travel Patterns from GPS-Tagged Photos

Yan-Tao Zheng[1], Yiqun Li[1], Zheng-Jun Zha[2], and Tat-Seng Chua[2]

[1] Institute for Infocomm Research, Singapore
[2] Department of Computer Science, National University of Singapore, Singapore
{yzheng,yqli}@i2r.a-star.edu.sg, {zhazj,chuats}@comp.nus.edu.sg

Abstract. The phenomenal advances of photo-sharing services, such as Flickr^{TM}, have led to voluminous community-contributed photos with socially generated textual, temporal and geographical metadata on the Internet. The photos, together with their time- and geo-references, implicitly document the photographers' spatiotemporal movement paths. This study aims to leverage the wealth of these enriched online photos to analyze the people's travel pattern at the local level of a tour destination. First, from a noisy pool of GPS-tagged photos downloaded from Internet, we build a statistically reliable database of travel paths, and mine a list of regions of attraction (RoA). We then investigate the tourist traffic flow among different RoAs, by exploiting Markov chain model. Testings on four major cities demonstrate promising results of the proposed system.

Keywords: Geo-mining, human mobility analysis.

1 Introduction

The prevalence of photo capturing devices, together with the advent of media-sharing services like Flickr^{TM}, have led to voluminous digital photos with text tags, timestamp and geographical references on the Internet. Different from other community-contributed multimedia data, these photos connect geography, time and visual information together and provide a unique data source to discover patterns and knowledge of our human society. In this study, our focus is to discover people's travel patterns within a local tour destination, by exploiting the geographically calibrated photos on photo-sharing websites, like Flickr [1]. The rational is that the time-referenced and GPS-tagged photos implicitly document the spaitio-temporal movements of their photographers. A large volume of such GPS-tagged photos can give rise to a statistical data source of people travel trails, as shown in Figure 1.

Studies on people mobility and travel behavior within a local tour destination have always been important topics to mobile applications and location based services. In general, there exist two types of methods to acquire detailed travel

[1] http://flickr.com

K.-T. Lee et al. (Eds.): MMM 2011, Part I, LNCS 6523, pp. 262–272, 2011.

(a) GPS-tagged photos in San Francisco

(b) Tourist movement trajectories in San Francisco

Fig. 1. GPS-tagged photos implicitly document the spaitio-temporal movements of their photographers. The movement trajectories of photographers shown in (b) can be generated from GPS-tagged photos shown in (a). (For better viewing, please see the original color pdf file.)

data: (1) a survey with questionnaire on people's location histories [12]; and (2) location-acquistion devices for people to wear, such as GPS, cellular phone, etc, [19,12]. The issue with the first method is its expensive and time-consuming manual process, while the second method gives rise to unavoidable privacy issue that makes most people reluctant to participate in the study. In this study, we propose an Internet-driven approach to acquire people's travel information from GPS-tagged photos on the Internet. The advantage of this approach is that tourist mobility analysis can readily scale up onto a multitude of tour destinations. Such an automated travel pattern analytic approach can be tremendously useful to many geo-spatial applications. For example, the travel sequence analysis can reveal the crowd's choice of popular tour routes and help to monitor the traffic patterns of tourists.

To perform travel pattern analysis, we need first to build a statistically reliable database of people's travel paths. To do so, we discriminate tourist[2] photos v.s. non-tourist ones, based on the *mobility entropy* of a photo sequence pertaining to one photo uploader. The rational is simple: the mobile nature of sightseeing renders the photos of a true tourist to be spread over a large spatial extend within the tour destination. A Z-test is then applied on the movement trajectories generated from these tourist photos to ensure that the resulting photo trails are statistically reliable. Though outliers and noise might still exist, the travel patterns are expected to be statistically significant, when the number of trail samples is large. Figure 1 (b) depicts ~2000 mined photo trails in San Francisco. As shown, such large number of photo trail trajectories reveal conspicuous travel patterns. As our interest is to model the tourist mobility among different regions of attractions, we mine a list of *regions of attractions* (RoA) within a tour destination, by borrowing the approach in our previous work [18]. The premise is: a dense cluster of photos from different people indicate a region of frequent

[2] Tourist here is defined in a board sense to comprise both foreign and local people traveling for pleasure.

visits and popular appeal and is highly probable to be a region of attractions. After mining the list of RoAs, we represent tourist movement as a visit sequence of RoAs and exploit the Markov chain model [3] to analyze the tourist traffic statistics between different RoAs. The Markov chain model is widely used in various disciplines to analyze the trend of spatio-temporal movement and outcomes of sequential events [14]. Based on the first-order dependence in Markov chain, we estimate the statistics of visitors traveling from one region to another. Such tourist traffic analysis helps to indicate centric regions of attractions (RoA), which have influx of tourists from many other RoAs.

Overall, this study aims to exploit GPS-tagged photos on the Internet to analyze the travel patterns in a local destination. To the best of our knowledge, this is the first approach that leverages GPS-tagged photos for tourist traffic analysis. We demonstrate the proposed approach on four major cities in the world, i.e., Paris, London, San Francisco and New York City; and experiments show that the proposed approach can deliver promising results.

2 Related Work

In recent years, the advent of media-sharing services, such as Flickr^{TM} and Youtube^{TM}, has led to voluminous community-contributed photos and videos available on the Internet. Together with socially generated textual and spatiotemporal metadata, these enriched multimedia data have spurred much research on discovering knowledge and patterns of our human society. Kennedy *et. al* proposed to discover aggregate knowledge of a geographical area, by analyzing spatiotemporal patterns of tags of Flickr photos in the area [8]. Similarly, Rattenbury *et. al* [13] and Yanai *et. al* [16] analyzed the spatiotemporal distribution of photo tags to reveal the inter-relation between word concepts (namely photo tags), geographical locations and events. Li *et. al* [10] and Zheng *et. al* [18] learned the geographical and visual appearance knowledge of tourist landmarks from community contributed photos on the Internet. The commonality between the aforementioned work and this study is that they all aim to extract some knowledge and patterns from photos with textual and spatiotemporal metadata, while the difference is that this study focuses on mining traveling patterns of tourists.

The study on tourist travel pattern within a tour destination has been a popular geographic research topic. Mckercher and Lau [12] attempted to identify the movement patterns and styles of tourists within an urban destination. Asakura and Iryo [1] investigated the topological characteristics of tourist behavior in a clustering approach. Lewa and McKerchera [9] explored the urban transportation and tourist behavior modeling to identify explanatory factors that influence tourist movements. Compared to the work above, this study differs mainly in two aspects. First, the travel information in the previous work is mainly acquired via a manual survey with a limited number of tourist respondents. Consequently, the studies [12,1] covered only one or two tour destinations. In contrast, the proposed approach mines the travel information from Internet photos, which renders the data acquisition highly efficient, and thus, allows the travel analysis to easily scale up to a multitude of destination. Second, constrating to existing approaches [12,1], this study analyzes the travel traffic by modeling it as sequence data via Markov chain model.

3 Approach

The overall framework consists of two major modules, i.e., building the travel path database and analyzing the travel traffic patterns.

3.1 Building the Travel Path Database

Given a set of GPS-tagged photos $\mathbb{P} = \{p\}$ within a tour destination, we build the database of travel paths. A photo p is a tuple $(\theta_p, \wp_p, t_p, u_p, \varrho_p)$, containing the unique photo ID θ_p, tagged GPS coordinates $\wp_p$ in terms of latitude and longitude, time stamp t_p when photo was taken, photographer/uploader ID u_p and tagged text ϱ_p. Here, the tourist travel movement is modeled at a daily basis. According to photographer ID u_p, we organize photos of each photographer in a day in a chronical sequence $< p_0, ..., p_k >$, which is defined as below.

Definition 1. ***Photo sequence*** *of a photographer u_p is a chronological sequence $P =< p_0, ..., p_k >$ of photos pertaining to the photographer u_p, where k is the number of photos and t_i is time stamp of photo i with $t_i < t_{i+1}$.*

By representing the geographical calibration $\wp_p$ of photo p in ordinary Cartesian coordinates (x_p, y_p), we define the movement of a photographer, in the notation of [5], as follows:

Definition 2. *The* ***photo trail*** *of a traveler corresponds to a spatio-temporal sequence (ST-sequence) $S =< (x_0, y_0, t_0)\ , ..., (x_k, y_k, t_k) >$ drawn from chronologically ordered photo sequence P in a one-to-one corresponding manner, where $(x_i, y_i) \in \mathbf{R}^2$.*

Based on Definitions 1 and 2, we construct the ST-sequence of movement trajectory of a photographer/uploader, by concatenating photos in the order of their time-stamp in a daily basis. We then classify these spatio-temporal sequences to tourist and non-tourist trails. The premise for classification is that the mobile nature of sightseeing renders the photos of a true tourist to be spread over a large spatial extend within the tour, as shown in Figure 2.

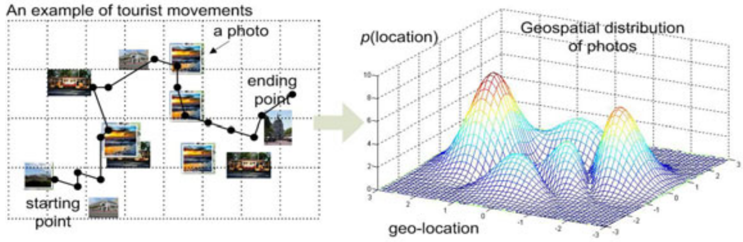

Fig. 2. Toy example of a tourist movement trajectory

Entropy based Mobility Measure. In a probabilistic perspective, the mobility complexity leads to a geospatial distribution of photos with reasonably high entropy. We, therefore, exploit this **mobility entropy** to discriminate the tourist and non-tourist movement trajectories, by utilizing the concept of Shannon entropy in Information Theory. Let $p(x, y)$ denote the geospatial density of photos with geospatial coordiantes (x, y) pertaining to the photographer/uploader. The mobility entropy $H_{mob}(S)$ of a movement trajectory $S =< (x_0, y_0, t_0)$ $, ..., (x_k, y_k, t_k) >$ is computed as follows.

$$H_{mob}(S) = -\sum_{i}^{n}\sum_{j}^{m} p_{ij}(x, y) \log p_{ij}(x, y), \tag{1}$$

where $p_{ij}(x, y)$ is a discrete geospatial distribution of photos in grid (i, j). By partitioning the tour destination into $n \times m$ grids, $p_{ij}(x, y)$ is estimated by the counts of photos in grid (i, j). To discriminate photo trails, we empirically set an mobility entropy threshold ε_{mob}. The photo trail S is then classified as a tourist one, if $H_{mob}(S) \geq \varepsilon_{mob}$.

Statistical Significance of Travel Paths. We need to ensure that the resulting travel path database are statistically reliable. To do so, we perform statistical significance test on the resulting photo trails. Here, we characterize the photo trajectory of a photographer/uploader with the number k of his/her visiting places. To some extent, the number of visit places indicates the mobility complexity of tourist. The number of visiting places is determined by the number of photos with unique geospatial coordinates. (Geospatial coordinates within a small distance will be considered as one.)

As the number of photo trajectory samples is relatively large, we approximate the tourist itinerary in terms of number of places k with normal distribution $\mathcal{N}(\mu_k, \sigma_k)$, based on *central limit theorem* in probability theory [2]. The mean μ_k and standard deviation σ_k are estimated from the trajectory samples. This normal model of tourist itinerary implicitly assumes individual tour itinerary samples are independent from each other. This independence assumption is reasonable, in the way that a tour itinerary depends on many factors, including tourist personal preference, his/her prior visits, total tour duration, tour destination demography, etc.

With the normal model, we apply Z-test to evaluate the statistical significance of a photo trajectory k. The null hypothesis H_0 in the test is defined as follows.

$$H_0 : k \sim \mathcal{N}(\mu_k, \sigma_k), \tag{2}$$

where H_0 states that the itinerary k generated from GPS-tagged photos is statically significant, rather than generated from noisy photos by chance. The z-score z is then computed as below:

$$z = \frac{(k - \mu_k)}{\sigma_k/\sqrt{n}}, \tag{3}$$

where n is the number of photo itinerary samples. By looking up the z-score in a table of the standard normal distribution, the corresponding p-value can be obtained. A lower p-value indicates a lower probability that the null hypothesis H_0 holds [17]. If p-value is less than a threshold τ, the null hypothesis H_0 is rejected and the photo itinerary is deemed to be statistically insignificant or unreliable and discarded subsequently.

Discovering Regions-of-Attractions (RoA). As our focus is on tourist mobility analysis at macro-level, a comprehensive list of regions of attractions within a tour destination is needed. Here, we define the region of attraction (RoA) as follows.

Definition 3. *A* region of attraction r *is a spatial extent in the geographical feature space of Cartesian coordinates* (x, y)*, where a considerable number of tourist movement trails pass through. RoA can be modeled as a spatial neighborhood function* $F(x_i, y_i) : \mathbf{R}^2 \rightarrow \{0, 1\}$*.*

In the spirits of our previous work [18], we develop a density-based model to discover regions of attractions, by analyzing the geospatial distribution of GPS-tagged photos. As stated in Definition 3, a region-of-attractions is a communal and interpretable spatial concept shared by a multitude of people. In other words, a RoA corresponds to a spatial extent, where many tourists visit and photograph. Clustering on GPS-tagged photos then become an intuitive solution to discover the list of regions of attraction.

Here, we adopt DBSCAN algorithm [4] to perform geospatial clusting on GPS-tagged photos for the following reasons. First, DBSCAN is a density-based clustering algorithm. Intuitively, it tends to identify regions of dense data points as clusters. This density driven approach just fits our task well, as the high density of photos implicates the popular appeal of the region. Second, DBSCAN algorithm supports clusters with arbitrary shape. This is critical to our task, as shapes of RoA can be spherical, linear, elongated etc. Third, DBSCAN is demonstrated to have good efficiency on large-scale data. (cf. [4] for more details of DBSCAN.)

After obtaining clusters of photos, we then determine the name and spatial extent of RoA, by examining the GPS coordinates and text title of component photos. We compute the frequency of n-grams of all photos text titles in each cluster. The name of RoA is determined to be the photo title with highest frequency. The geospatial extent of RoA is the area defined by the GPS coordinates of its member photos. Similar to [18], the resulting RoA is validated by the number of unique photographers/uploaders. This is to further ensure the popular appeal of RoA.

3.2 Transition Traffic between RoAs

Based on the concept of RoA, we define the *transition statistics between RoAs* as below.

Definition 4. *The transition statistics between RoAs depicts how tourist traffic flows from one RoA to another. It is defined as transition probabilities among different RoAs.*

By defining the tourist travel as a sequence of RoA, we investigate how tourists move from one RoA to another using the Markov chain model, in the spirit of [15].

In a statistical perspective, we model the movement of a tourist as an independent stochastic random process. The state space of the stocastic process is the set of RoA $\{r\}$ in the tour destination. Let $T = \{0, 1, 2, ...\}$ denote the time index of the moves of a stochastic process. The stochastic process representing tourist movement $\{R_t\}_{t \in T}$ is referred to as a *Markov chain* (MC), if the value of next state does not depends on any previous states, given the value of current state, as defined below.

$$\begin{aligned} P(R_{t+1} = r_{t+1} | R_t = r_t, R_{t-1} = r_{t-1}, ..., R_0 = r_0) \\ = P(R_{t+1} = r_{t+1} | R_t = r_t), \end{aligned} \tag{4}$$

where R_t is the random variable of RoA, r_t is a value of R_t and $r_t \in \{r\}$. In Markov chain model, each move in the state space $\{r\}$ is called a *step*. As each movement occurs after one unit time step, the stochastic process of tourist movement is modeled by a stationary discrete Markov chain. The transition probability $P(r_j \mid r_i)$ from RoA r_i to r_j can then be estimated by counting the tourists moving from RoA r_i to r_j. Accordingly, the RoA transition can be represented by a directed graph $G = (V, E)$, in which vertex V corresponds to RoA and edge E represents the transition statistics.

4 Experiments

GPS-tagged photos used in this study were downloaded from Flickr^{TM}, by using its publicly available API [7]. To download photos, the name of a tour destination, such as Paris, London, etc, is fed in as query to retrieve a set of seed photos. Then, the owner ids of these seed photos are retrieved. Based on the owner ids, we download the entire collection of user's shared photos to ensure the completeness of the generated photo trail.

4.1 Travel Path Database

We download photos in four major cities: London, Paris, New York City and San Francisco. In total, we collected ∼769k GPS-tagged photos from ∼23k Flickr users. Based on Definition 1, we concatenate photos of a photographer into photo sequences in a daily basis. Following Section 3.1, we build a local travel database consisting of 8047 person-day trips by 5010 people in total. In average, each city has ∼2000 person-day trips. This significantly outnumbers the manually collected tourist movement datasets of existing tourist mobility analysis works [12,11], not to mention that the database can be easily augmented by downloading more GPS-tagged photos. Figure 1 and 3 show the movement trajectories generated from GPS-tagged photos in New York City, San Francisco, Paris and London plotted on Google Earth.

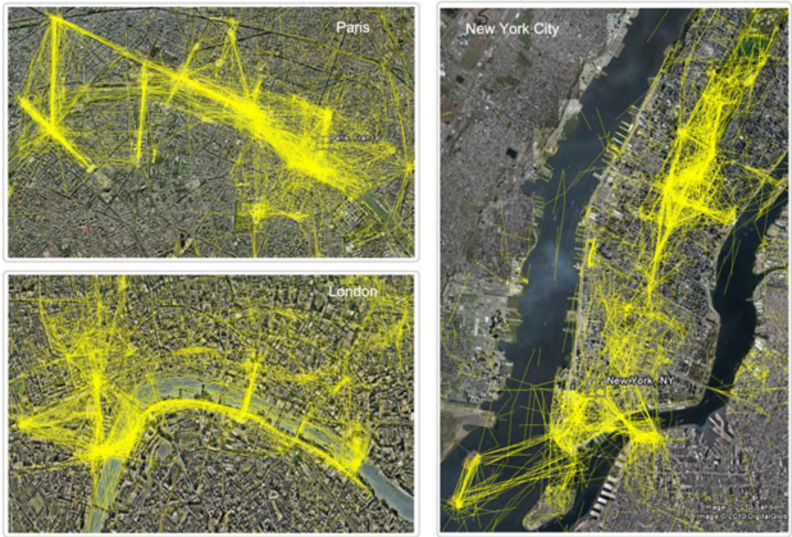

Fig. 3. Tourist travel trails generated from GPS-tagged photos in Paris and London

Regions of Attractions. By taking the photos in tourist movement trajectories as input, we discover the regions-of-attractions (RoAs) in a density-based approach, as presented in Section 3. In total, we discover 80 RoAs with 18 in London, 19 in Paris, 23 in New York City and 20 in San Francisco. Among them, only 1 out of 80 RoAs is false, which is "San Francisco Pride Parade". This event is misclassified as a RoA, as it gives rise to voluminous photos with strong geospatial pattern.

In the local travel database, the average cardinality of a daily trip routes is 3.5 RoA visits per day. This number is similar to the average visit of 3.7 RoAs per day in the tourism study [12].

4.2 Tourist Traffic Analysis

Popularity of RoA. The popularity of a RoA can be estimated by its tourist traffic volume, namely the number of people that have photographed in the region. Table 1 summarizes the top 3 most visited (most popular) RoAs in the four cities. For each RoA, the percentage of tourists that visit it is also computed. We compare Table 1 against the list of top 3 attractions in Yahoo!Travel [6] and found that two lists share 42% identical RoAs. The attraction popularity in Yahoo!Travel is estimated based on the feedback scores provided by Yahoo users. This overlap of popular RoAs suggests that despite of different background, people tend to agree on the most popular attractions to some extent.

Transition Traffic between RoAs. We utilize Markov chain model to estimate the transition probability $P(r_j|r_i)$. $P(r_j|r_i)$ indicates how tourist traffic

Table 1. Top three most visited RoAs and percentage of tourists in the four cities. SF: San Francisco.

	RoAs	Percentage (%)
SF	1. Golden Gate Bridge	27.6
	2. Pier 39	22.9
	3. Union Square	20.3
New York City	1. Times Square	35.6
	2. Rockefeller Center.	29.3
	3. Brooklyn Bridge	22.9
Paris	1. Notre Dame	38.7
	2. Eiffel Tower	31.0
	3. Arc de Triomphe	30.5
London	1. London Eye	43.6
	2. Trafalgar Square	34.5
	3. Tower Bridge	29.9

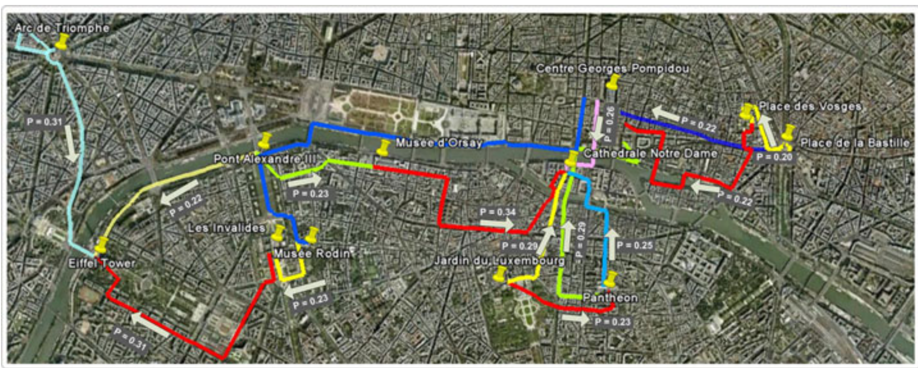

Fig. 4. Traffic transitions among RoAs in downtown, Paris, with transition probability $P(r_j|r_i) > 0.2$. For better viewing, please see the original color pdf.

moves from one RoA to another. A reasonably high value of $P(r_j|r_i)$ suggests that RoA r_j and r_i are coupled in the way that tourists tend to visit RoA r_j right after r_i. Figure 4 displays the RoA transitions with probability $P(r_j|r_i) > 0.2$ in downtown area of Paris. As shown, the coupled RoAs are usually geographically adjacent to each other. Moreover, it is also observed that people tend to prefer certain direction when visiting two coupled RoAs. For example, the transition probability P(Eiffel Tower |Arc de Triomphe) that tourists move from "Arc de Triomphe" to "Eiffel Tower" is 0.31, while the transition probability P(Arc de Triomphe | Eiffel Tower) in the opposite direction is only 0.12. This suggests that tourists might share similar preference in tour route planning. (For space limit reason, the RoA transition in the other three cities are not illustrated.)

Centric RoA. Tourist traffic tends to flow from several RoAs to a central one. We denote this central RoA as *centric RoA*. Specifically, we define centric

Table 2. Centric RoAs in the four cities

	Centric RoAs
San Francisco	Union Square, Chinatown
New York City	Time Square, Brooklyn Bridge
Paris	Eiffel Tower, Cathedrale Notre Dame
London	London Eye, Trafalgar Square

RoA as the one with transition probability $P(\text{centric RoA} \mid r_i) > 0.15$ for more than 3 RoA r_i. Table 2 summarizes the centric RoAs in the four cities. Figure 4 shows that "Eiffel Tower" and "Cathedrale Notre Dame" are centric RoAs in Paris, as they receive influx of tourists from several adjacent RoAs. The centric RoA might be determined by several factors, including popularity, geographical location, transportation convenience, etc. In a way, the centric RoA is the place where people congregate and meet each other.

5 Conclusion

Analysis on tourist mobility are important to tourism bureaucracy and industries. However, the cost of collecting detailed travel data is formidable. GPS-tagged photos available on the Internet implicitly provide spatio-temporal movement trajectories of their photographers. In this paper, we proposed to leverage these GPS-tagged photos to analyze the tourist travel behavior at the local level of a tour destination. We first built a statistically reliable tourist movement trajectory database from GPS-tagged photos, by utilizing an entropy-based mobility measure and Z-test. A list of regions of attraction (RoA) in a tour destination is then built, based on the frequency of tourist visits. We then investigated tourist traffic flow among different RoAs, by exploiting markov chain model to interpret tourists traffic transition. Finally, tourist travel patterns were analyzed by performing a sequence clustering on tour routes. Testing on four major cities, including San Francisco, New York City, Paris and London, demonstrated that the proposed approach can deliver promising results. One of our future works is to argument the tourist movement trajectory database and extend the travel analysis to a large scale of tour destination.

References

1. Asakura, Y., Iryo, T.: Analysis of tourist behaviour based on the tracking data collected using a mobile communication instrument. Transportation Research Part A: Policy and Practice 41(7), 684–690 (2007)
2. Degroot, M.H., Schervish, M.J.: Probability and Statistics, 3rd edn. Addison Wesley, Reading (2001)
3. Diaconis, P.: The Markov chain Monte Carlo revolution. Bull. Am. Math. Soc., New Ser. 46(2), 179–205 (2009)

4. Ester, M., Kriegel, H.-P., Jörg, S., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of Conference on Knowledge Discovery and Data Mining, USA, pp. 226–231. ACM, New York (1996)
5. Giannotti, F., Nanni, M., Pinelli, F., Pedreschi, D.: Trajectory pattern mining. In: Proceedings of Conference on Knowledge Discovery and Data mining, pp. 330–339. ACM, New York (2007)
6. Yahoo!Travel, `http://travel.yahoo.com/`
7. Flickr API, `http://www.flickr.com/services/api/`
8. Kennedy, L., Naaman, M., Ahern, S., Nair, R., Rattenbury, T.: How flickr helps us make sense of the world: context and content in community-contributed media collections. In: Proceedings of conference on Multimedia, pp. 631–640. ACM, New York (2007)
9. Lewa, A., McKerchera, B.: odeling tourist movements: A local destination analysis. Annals of Tourism Research 33(2), 403–423 (2006)
10. Li, X., Wu, C., Zach, C., Lazebnik, S., Frahm, J.-M.: Modeling and recognition of landmark image collections using iconic scene graphs. In: Proceedngs of European Conference on Computer Vision, pp. 427–440 (2008)
11. McKercher, B., Lew, A.A.: Tourist flows and the spatial distribution of tourists. In: A Companion to Tourism, ch. 47, p. 36 (2004)
12. Mckercher, B., Lau, G.: Movement patterns of tourists within a destination. Tourism Geographies 10(3), 355–374 (2008)
13. Rattenbury, T., Good, N., Naaman, M.: Towards automatic extraction of event and place semantics from flickr tags. In: Proceedings of ACM SIGIR, pp. 103–110. ACM, New York (2007)
14. Upton, G.J.G., Fingleton, B.: Spatial Data Analysis Categorical and Directional Data, vol. 2. Wiley & Sons, Chichester (1989)
15. Xia, J.C., Zeephongsekul, P., Arrowsmith, C.: Modelling spatio-temporal movement of tourists using finite markov chains. Math. Comput. Simul. 79(5), 1544–1553 (2009)
16. Yanai, K., Kawakubo, H., Qiu, B.: A visual analysis of the relationship between word concepts and geographical locations. In: Proceeding of the ACM International Conference on Image and Video Retrieval, pp. 1–8. ACM, New York (2009)
17. Zha, Z.-J., Yang, L., Mei, T., Wang, M., Wang, Z.: Visual query suggestion. In: Proceedings of ACM International Conference on Multimedia, pp. 15–24 (2009)
18. Zheng, Y.-T., Zhao, M., Song, Y., Adam, H., Buddemeier, U., Bissacco, A., Brucher, F., Chua, T.-S., Neven, H.: Tour the world: building a web-scale landmark recognition engine. In: Proceedings of International Conference on Computer Vision and Pattern Recognition, Miami, Florida, U.S.A (June 2009)
19. Zheng, Y., Zhang, L., Xie, X., Ma, W.-Y.: Mining interesting locations and travel sequences from gps trajectories. In: Proceedings of the 18th International Conference on World Wide Web, WWW 2009, pp. 791–800. ACM, New York (2009)