

Mining Tourist Information from User-Supplied Collections

Adrian Popescu
TELECOM Bretagne France
adrian.popescu@enst-
bretagne.fr

Gregory Grefenstette
EXALEAD France
ggrefens@exalead.com

Pierre-Alain Moëllic
CEA, LIST
pierre-alain.moellic@cea.fr

ABSTRACT

Tourist photographs constitute a large part of the images uploaded to photo sharing platforms. But filtering methods are needed before one can extract useful knowledge from noisy user-supplied metadata. Here we show how to extract clean trip related information (what people visit, for how long, panoramic spots) from Flickr metadata. We illustrate our technique on a sample of metadata and images covering 183 cities of different size and from different parts of the world.

Categories and Subject Descriptors

H.2.8 [Database Management]: Database applications – Data mining, Image databases; Spatial databases and GIS.

General Terms

Algorithms, Experimentation, Human Factors.

Keywords

Tourist sites, visit times, geographical gazetteer, k-Nearest Neighbors, image classification, bags of visual words, Flickr.

1. INTRODUCTION

E-tourism sites allow users to interactively plan trips. For example, HomeAndAbroad.com (H&A hereafter) allows users to compose thematic trips (such as Family Time, Local Culture or History Buff) in 100 cities. Schmap.com, an e-tourism platform with 200 destinations, mashes-up interactive maps, text and photographs to provide a rich description of tourist sites. Both sites, detailed and generally accurate, describe a limited number of destinations because manual compilation of tourist information is costly. Automating the acquisition of tourist information from “tourist memory” data in web-based user-supplied photographic collections, e.g. Flickr, is an appealing alternative to manual work. We describe how to (i) discover and characterize a large number of trips from Flickr metadata and recombine their elements to propose new trips, (ii) discover visit duration times at tourist sites by analyzing textual metadata and classifying images into interior/exterior views and (iii) find locations that offer panoramic views via binary image classification. We model trips as graphs where tourist sites are nodes (labeled with place names) and paths between these sites are edges.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CIKM'09, November 2–6, 2009, Hong Kong, China.

Copyright 2009 ACM 978-1-60558-512-3/09/11...\$10.00.

2. RELATED WORK

Rattenbury et al. [13] was an early attempt to discover both event and place names from Flickr geolocated textual metadata, resulting in an application [1] for geographic image retrieval, with representative and popular tags overlaid on a scalable map. Quack et al. [12] downloaded 200,000 georeferenced Flickr images from nine urban areas and clustered them using local image descriptors to discover place names and events, linking some places to their Wikipedia articles. In contrast to [12], [3] and [8], we do not limit ourselves to geolocated photographs since many pictures of places do not have associated location information. For instance, there are less than 1000 Flickr photos geotagged in the *Timișoara* but 31,000 images tagged with this city’s name. Elsewhere [10], we detailed methods for combining information from heterogeneous sources to extract place names, place coordinates, place types and popularity values. Here we exploit these same results but shift our focus towards discovering additional trip related information. Zheng et al. [14] recorded GPS tracks of 107 users during a year to determine the interestingness of tourist sites. The authors of [5] also focus on mining similar traveling sequences from multiple users’ GPS logs while the authors of [9] retrieve maximum periodic patterns from spatio-temporal metadata. Girardin et al. [4] analyzed the tourist flows in the Province of Florence, Italy, based on a corpus of georeferenced Flickr photos and their results contribute to understanding how people travel. None of these approaches mines image content to produce knowledge about places, a mining process which is a central piece of our approach. Whereas [14] or [5] rely on accurate GPS traces for small scale regions and obtained from a relatively reduced number of users, we exploit noisy Flickr data but which covers most interesting tourist regions of the world. As a result, we are able to propose itineraries in any region of the world that is sufficiently covered by Flickr data.

Visit duration is an important characteristic of trips and it is classically estimated by domain experts [7]. The automatic extraction of visit duration from Flickr metadata was only recently explored [11] but no separation between sightseeing and sightseeing + interior visits was proposed. Building on this latter work, we use visual image classification to separate these two types of visits and to calculate typical visit times for each case.

3. DATASET

We collected textual metadata and images with Flickr’s public API. Since many Flickr tourism photographs are related to cities [6], we selected 183 regions or cities, both popular (*Paris, London, New York, Tokyo*) and less so (*Timișoara* or *Brașov* in *Romania, Szeged* in *Hungary, Segovia* in *Spain*). As of May 2009, we retrieved over 71 million metadata associated with Flickr

photos: all with titles, tags, time stamps and user IDs, plus a small fraction with geotags.

For each city, we sorted metadata to extract potential one day trips. We retain the set of pictures for a given user on a given day, if (i) there are $N \geq 15$ pictures in the picture set, (ii) the time stamp between the first image and the N th image is more than one hour, and (iii) the unique sort of the concatenated annotations of each of the N pictures yielded more than $N/3$ different textual annotations (in order to eliminate bulk uploaded tags). With these three criteria we extract information from photo sessions which contain a variety of pictures over an extended period on a given day (the likely behavior of a tourist), and which were individually tagged. At this stage, there remained 723303 photos taken during one day visits, with their tags and timestamps (1% of the original set). We downloaded these images and processed them on a 100 processors cluster using a texture-color descriptor [2] in order to be able to classify them further into interior/exterior images and panoramic views.

4. DISCOVERING TRIPS

4.1 Extracting Place Names from Wikipedia

To extract trips, we need to start from list of interesting sites in each examined region. Since the reported extraction accuracy in research into extracting tourist sites from georeferenced collections (85% in [13], 90% in [10]) is not sufficient for our current purposes, we decided to use only Wikipedia to mine possible points of interest, assuming that many touristy interest points are already geolocated there. We processed Wikipedia dumps available in the 10 most frequent languages to increase coverage, using a list of geographic patterns to select only geolocated articles and considered their titles to be place names. To associate types to these place names we built a multilingual vocabulary of geographic concepts containing 158 terms and covering most types of landmarks likely to be found in urban areas. Place names were categorized using a similar procedure to the one in [10], matching geographic vocabulary concepts against article categories and against the first sentence in the article. Tourist site types are determined by performing a frequency count of matching elements in the article body, retaining the most frequent. For place names that appear in several languages, we aggregate results from all language versions and retain the most frequent category. Results in [10] show this simple categorization procedure is precise. At this point, each place name is characterized by the triplet (name, GPS coordinates, type). The number of landmarks per region yields high values for major cities (*London*, 3451; *Paris*, 3075; *Tokyo*, 600) and small values for other cities (*Segovia*, 40; *Timișoara*, 50).

4.2 Characterizing trips

A trip is defined as a one day visit which includes one or several tourist sites. To find trips, we compare the list of place names extracted from Wikipedia to image sets associated with a region and retain only those sets that match at least a place name in that region. Following [11], we consider that a Flickr user was “visiting” a site if (i) the interval between the earliest and latest timestamp is at least ten minutes inclusive, and (ii) there are at least 5 images manually tagged with that site name within that interval, and (iii) the interval between any two images is not longer than 20 minutes or 5 times the average time between photos tagged with the name (whichever smaller). These

conditions mimic the activity of someone present at a site for while, taking a number of pictures later tagged with the site name (i.e., “visiting” the tourist site), and actively taking pictures of tourist attractions during the visit.

In figure 1, we present the distribution of the number of trips discovered for the 183 cities in the evaluation dataset. We grouped cities using 100 items intervals in order to have a succinct but informative representation. A majority of cities (112) has up to 100 associated trips whereas 10 cities have over 1000 associated trips (all in major cities and tourist destinations such as *London* (2578), *New York* (2412) or *Paris* (2213)). Some cities with a small number of associated trips: *Timișoara, Romania* (11), *Segovia, Spain* (52) and *Quimper, France* (2) elicit a limitation of our work but also of existing work on user-supplied collections [3] or [8] which only present results for major cities. Flickr representation of smaller cities is likely to improve over time but it is currently rather limited.

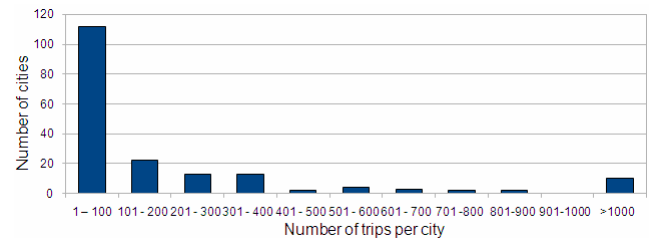


Figure 1. Distribution of the discovered trips.

The number of tourist sites to which our algorithm associated visit times is smaller than 10 for 78 cities out of 183; is between 11 and 40 tourist for 81 cities and there are also 4 cities with over 100 tourist sites with associated visit times. Cities in the first case include *Timișoara* (4 sites with associated visit times) or *Quimper, France* (1). Cities with the largest number of associated tourist sites are *London* (322), *Paris* (155), and *Berlin* (134). The number of trips and the number of tourist sites with associated visit times are usually well correlated.

After this step, for each set of images taken during a one day visit, we know the sites visited and in which order. Since we also know the GPS coordinates of each site from Wikipedia, trips can be situated in space and used for guiding other users. Approximate paths followed between two consecutive visited sites can be produced from Flickr metadata using a sampling method. We calculate the paths associated to a trip by going through the GPS coordinates associated to images in a trip and retaining a new image whenever there is a difference of at least 100 meters between this photo and the last retained photo.

4.3 Computing visit times

Beside spatial information, temporal information is also central for describing trips. At H&A, experts provide visit time estimations as intervals, with minimum estimations around 30 minutes or one hour and maximum going up to a full day. From time stamps associated to Flickr, we compute visit times by subtracting the time stamp of the first photo tagged with a place name from that of the last photo. We can estimate usual behavior, by averaging individual visit times, and compute average, maximum and minimum visit times and then add these visit durations to the tourist sites descriptions, a technique easily applicable to any tourist site found in Flickr.

To evaluate our automatically extracted estimations against H&A manual estimates, we retained 464 attractions from 25 cities. Given the high number of evaluated sites, we present the distributions of average results compared to the maximum estimations from H&A, showing a high correlation of our automatically extracted visit times and H&A values. When compared to maximum manual estimations, the highest concentration of automatically extracted visit times is around 0.4 – 0.5 of the manual estimations. This finding supports initial results in [11] showing that people often tend to visit tourist sites more rapidly than expert estimations.

Visit times are obtained by averaging durations for external sightseeing and sightseeing+interior visits, performing interior/exterior image classification using a k-NN algorithm applied to texture-color image features [2] in order to separate the two types of visits. 1000 exterior and 1000 interior diversified images constitute our ground truth. An image can be classified as interior or exterior using a threshold (smaller or equal to $k=20$). For a threshold value of 15, precision reaches 0.89 and recall 0.65, meaning that one image out of 10 is wrongly classified, when 2/3 of the images are classified. For a visit of a site to be considered as combined, at least three images have to be detected as interior. With this condition, the probability for a trip to be wrongly classified as sightseeing+interior is very small.

Visit times for the external sightseeing and sightseeing+interior visits were computed on 271 tourist sites which had at least two associated visits of each type and their distributions are presented in figures 2 and 3. Visit times associated to sightseeing tend to concentrate below the minimum manual estimation whereas times for interior visits are concentrated above the same estimation. As expected, interior visits are often longer than exterior visits.

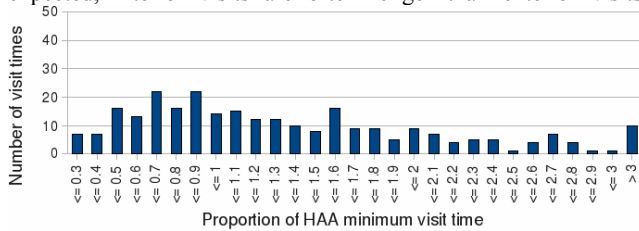


Figure 2. Distribution of automatically extracted visit times for sightseeing compared to H&A minimum estimations.

The separation of visits into interior and exterior is interesting for determining how people visit a particular site. For instance, for museums, such as *Musée d'Orsay* a majority of visits is interior (76 out of 107) whereas the same is true only for 37 visits out of 133 associated with *Arc de Triomphe*. A high proportion of interior visits indicates that the interior of a particular site was seen by many tourists and it is worth entering that attraction. Whereas human experts consider that each tourist enters, our results demonstrate when Flickr users limit themselves to external sightseeing.

In section 4.1, we presented a method for associating types to tourist sites and arrange them in a hierarchical structure. This structure can be exploited to compute average visit times per type (*cathedral, palace or bridge*) and to propose estimations for other tourist sites of the same types which do not have associated visit times. As a result, the coverage of our approach is improved and we can estimate visit times for any tourist site which is typed with one of the 158 types in the geographic vocabulary.

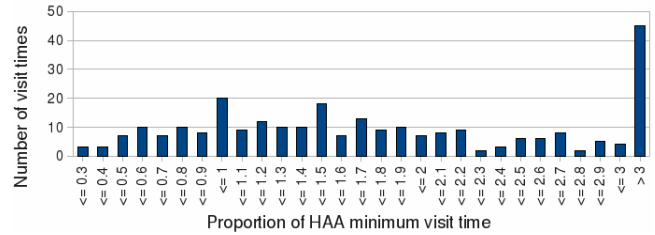


Figure 3. Distribution of automatically extracted visit times for interior visits compared to H&A minimum estimations.

4.4 Trip Examples

The four trips presented in Figure 4 convey some interesting information about our trip extraction method. Their duration is variable, from under five hours (red trip) to over 9 hours (yellow trip). Landmark visits cover only a fraction of the total visit times and transitions between sites (rectangles on the timelines) might depend on activities not involving picture taking (transport).

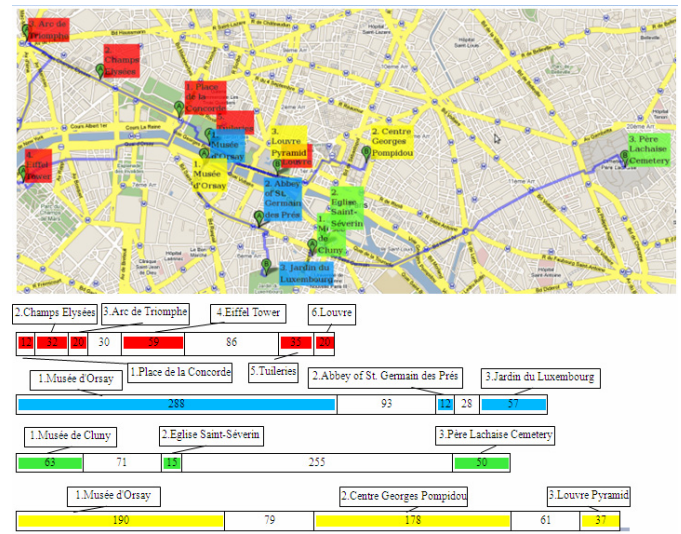


Figure 4. Four different trips (red, blue, green, yellow) in Paris overlaid on Google Maps and their timelines. Visited hotspots for each trip are presented in the order of visiting. On the timelines, transitions between tourist sites are depicted by rectangles and visit times are given in minutes.

An automated system proposing feasible trips must include realistic pauses. For the red trip, there are only two significant transitions (30 min between *Arc de Triomphe* and *Eiffel Tower* and 86 min between *Eiffel Tower* and *Tuileries*). The first three visited tourist sites (*Place de la Concorde*, *Champs Elysées* and *Arc de Triomphe*) and the last two (*Tuileries* and *Louvre*) are adjacent and the tourist made no pause between them. From just these four plotted trips, we can see there is a high concentration of visited landmarks in Central Paris. Tourist heat maps could be obtained by plotting all trips discovered for a region. Visit times are highly variable: two visits of *Musée d'Orsay* (blue and yellow) correspond to long visits (nearly 5 and over 3 hours) seeing the art collection in this museum. In contrast, the interval tagged with *Louvre* in the red trip (20 min) corresponds to sightseeing. Likewise for *Eglise Saint-Séverin* (green trip, 15 min) and *Abbey of St. Germain des Prés* (blue trip, 12 min). Interior/exterior

image classification allows us to separate the two types of visits and propose different duration estimations for each case.

Visit order for the red trip illustrates an interesting fact: tourists don't necessarily take the shortest path between landmarks. A shorter trip would have started at the *Louvre* and finished at the *Eiffel Tower*. Some factors influencing visiting order such as personal preferences or opening hours must be taken into account by an automated trip advisor when proposing visits.

Elements of known trips (sites and paths between them) can be recombined to create a large number of novel trips. Recombining trip elements could be especially useful if user preferences are known. Mining user preferences falls outside of the scope of this paper but represents a promising future research direction.

5. PANORAMIC VIEWS DETECTION

Panoramic spots are of high interest for tourists. With the automated detection of panoramic spots, tourists can preview generic city views before going there. We performed binary image classification to identify panoramic views by using 1000 varied panoramic Flickr images and 1000 Flickr images representing diversified scenes as an external class (outliers). Classification was performed using a kNN with k=10.

City	Panorama points
Berlin	Spree, Brandenburg Gate, Berlin Cathedral, Frensehturm, Reichstag
London	Tower Bridge, London Eye, South Bank, Tate Modern, St. Paul's
Paris	Arc de Triomphe, Eiffel Tower, Sacré Coeur, Montmartre, Trocadero

Table 1. Top 5 sites offering panoramic views

The simplest way to discover panoramic points is to analyze linguistic patterns. Panoramic views of taken from a particular landmark L are often annotated with the pattern "... from L". We translated this pattern to French, Spanish, Italian and German and selected up to 200 images which matched the pattern for each object. However this analysis is unreliable because many photos annotated with this pattern are not panoramas, so text matching is complemented with image classification. We rank tourist sites offering panoramic views by counting the number of different users having uploaded panoramic views taken from a particular site. We illustrate results in table 1, where we present the top 5 points in 3 cities. Categories of places offering panoramic views include: tall structures, constructions on elevated points, elevated points and bridges. Whereas we expected a large proportion of elevated points, we found surprisingly many sites near bodies of water.

6. CONCLUSION

We presented a method for mining tourist information from a large scale user-supplied photographic collection, extracting a large number of trips for each location with associated spatial and temporal information, differentiating sightseeing and sightseeing + interior visits and finding tourist sites that offer panoramic views of surrounding regions. These trips can be used as such or recombined to generate new ones. We found that automatically extracted visit times differ from manual expert-based estimations in that they are often shorter. Also, many visits are limited to sightseeing, without people actually entering the sites. This

process is fully automatic and easy to scale. Although Flickr currently under-represents smaller cities, the size of the collection increases over time and better coverage is likely to be obtained in the future. We showed that it is possible to extrapolate information from sites which are well represented in Flickr to other tourist sites of the same type and propose visit times for the latter. In the future we will focus on individual behavior to extract tourist preferences for proposing personalized trips. Personalization may refer to the types of sites visited, to the trips length or to the visit rhythm.

7. ACKNOWLEDGMENTS

This research is part of Georama, a French research project funded by ANR.

8. REFERENCES

- [1] Ahern, S., Naaman, M., Nair, R. and Yang, J. 2007. World Explorer: Visualizing Aggregate Data from Unstructured Text in georeferenced Collections. In Proc. of JCDL 2007.
- [2] Cheng, Y.-C., Chen, S.-Y. Image classification using color, texture and regions. *Image Vision Computing*, 21(9), 2003.
- [3] Crandall, D., Backstrom, L., Huttenlocher, D., Kleinberg, J. Mapping the World's Photos. Proc. of WWW 2009.
- [4] Girardin, F., Dal Fiore, F., Blat, J., Ratti, C. Understanding of Tourist Dynamics from Explicitly Disclosed Location Information. In Proc. of the *4th International Symposium on LBS and Telecartography* (Hong-Kong, China, 2007).
- [5] Gonotti, F., et al. Trajectory pattern mining. In Proc. of KDD 2007 (San Jose USA, Aug. 2007), ACM Press, 330-339.
- [6] Grefenstette, G., Pitel, G. Image Specific Language Model: Comparing Language Models from Two Independent Distributions from Flickr and the Web. *J. Language Forum*, 34, 2008.
- [7] Home&Abroad – <http://homeandabroad.com>
- [8] Kennedy, L., Naaman, M. Generating diverse and representative image search results for landmarks. Proc. of WWW 2008.
- [9] Mamoulis, N. et al. Mining, Indexing and Querying Historical Spatiotemporal Data. In Proceedings of KDD'04 (Seattle USA, August 2004), ACM Press: 236-245.
- [10] Popescu, A., Grefenstette, G., Moëllic, P.-A. Gazetiki: automatic construction of a geographical gazetteer. In Proc. of JCDL 2008 (Pittsburgh, PA, June 2008).
- [11] Popescu, A., Grefenstette. Deducing Trip Related Information from Flickr. In Proc. of JCDL 2008.
- [12] Quack, T, Leibe, B., van Gool, L. World-Scale Mining of Objects and Events from Community Photo Collections, In Proc. of ACM CIVR'08.
- [13] Rattenbury, T., Good, N., Naaman, M. 2007. Towards Automatic Extraction of Event and Place Semantics from Flickr Tags. In Proc. of SIGIR 2007.
- [14] Zheng, I, Zhang, L., Xie, X., Ma, W.-Y. Mining Interesting Locations and Travel Sequences from GPS Trajectories. In Proc. of WWW 2009.