

# Event-based analysis of people's activities and behavior using Flickr and Panoramio geotagged photo collections

Slava Kisilevich, Milos Krstajic, Daniel Keim, Natalia Andrienko, Gennady Andrienko  
 University of Konstanz, slaks@dbvis.inf.uni-konstanz.de  
 Fraunhofer Institute IAIS, gennady.andrienko@iais.fraunhofer.de

## Abstract

*Photo-sharing websites such as Flickr and Panoramio contain millions of geotagged images contributed by people from all over the world. Characteristics of these data pose new challenges in the domain of spatio-temporal analysis. In this paper, we define several different tasks related to analysis of attractive places, points of interest and comparison of behavioral patterns of different user communities on geotagged photo data. We perform analysis and comparison of temporal events, rankings of sightseeing places in a city, and study mobility of people using geotagged photos. We take a systematic approach to accomplish these tasks by applying scalable computational techniques, using statistical and data mining algorithms, combined with interactive geo-visualization. We provide exploratory visual analysis environment, which allows the analyst to detect spatial and temporal patterns and extract additional knowledge from large geotagged photo collections. We demonstrate our approach by applying the methods to several regions in the world.*

**Keywords**—Geo visual analytics, geotagged images, spatio-temporal analysis, movement data, clustering

## 1 Introduction

In the last years, the amount of information created in the digital universe is growing rapidly. Some forecasts predict that the information added annually will increase more than tenfold, from 161 exabytes to nearly 1800 exabytes between 2006 and 2011 and images comprise its largest component [9]. Part of these information is distributed over Internet on photo-sharing websites, such as Flickr<sup>1</sup> and Panoramio<sup>2</sup>. These sites contain billions of photos, which are publicly available and annotated with different kinds of useful metadata: image size, tags, titles and spatio-temporal information - where and when the photo was taken.

In our work, we are interested in several different aspects of geospatial analysis of large photo collections re-

lated to individual and social activity, and movement of people. However, analysis of movement data is usually performed on data captured by GPS devices, since these devices are the most widely available source of this kind of data. The most important difference between GPS trajectories and user-generated photo datasets is in the way the data records are generated: while raw GPS data lack semantic meaning and background information, geotagged photos are hand-picked by humans. Each record from the photo dataset carries underlying information about the surrounding area and the owner of the photo. This data can be regarded as a set of trajectories of multiple users or as independent spatio-temporal events and can be used for analysis of attractive areas, places of interest, landmarks or user travel preferences. Besides, new insights about local individual and group dynamics could lead to recommendations and customized services in tourism, improvements in local administration and targeted product offerings.

In the exploration of geospatial image data, we could be interested in users' trajectories [2] that consist of sequences of photos taken during their photo sessions. Images taken by a single user, that are adjacent in time within a certain region can be seen as a trajectory taken by this user, while photo session can be regarded as a time window of a certain size in which the photos were taken.

Another question is *what* are the places where the activity of people is high. The places where a lot of people take photos present attractive areas that grab photographers' attention because of location specific attributes. Within a specific area there are different places, which could be ranked according to their attractiveness. Usually, these are region landmarks, but they could also be areas of temporal interest, because of an event happening in a specific time period. Thus, temporal attribute of the data is of much importance in the analysis of attractive areas and presents a task of its own. Besides analysis of specific areas, our goal is to perform comparison between different communities, their behavior and mobility and different attractive areas

<sup>1</sup><http://www.flickr.com>

<sup>2</sup><http://www.panoramio.com>

and points of interest in different time periods.

In this paper, we take a systematic visual analytics approach to accomplish different tasks in event-based analysis of geotagged image data. By considering the characteristics of the data, we define seven possible tasks for analysis of attractive places, points of interest, behavior, and mobility of people based on geotagged photos. Obviously, these huge collections require scalable computational techniques in order to perform efficient analysis. At the same time, we employ visualization and geovisualization techniques to represent the data and allow the analyst to interact with it in order to provide insight, detect spatial and temporal patterns, and extract meaning and additional knowledge during exploration of these multivariate datasets.

## 2 Related Work

In recent years, various general techniques for analysis of movement data and human activities in particular were proposed. Different techniques for 3D geo-visualization of space-time patterns of people’s travel experience and mobility is presented in [16]. An approach for discovering different types of motion patterns is proposed in [17] and applied on football players to find meaningful group motion patterns. Two types of algorithms for mining interesting patterns from trajectories acquired by GPS-enabled devices are proposed in [10]. In the first type, the trajectories are converted into a sequence of stops or important parts (regions in which an object stayed more than a predefined time interval) before the algorithm for mining interesting patterns is applied. In the second type, the identification of important parts in a trajectory is part of the algorithm for mining patterns. Progressive clustering of trajectories of moving objects is presented in [18]. The authors combined clustering with visual interaction to let the analyst apply different distance functions based on the particular characteristics of trajectories under investigation. Visualization techniques (aggregations, ringmaps) of daily repeating activities like travel, work, shopping are presented in [19]. An algorithm for finding interesting places and mining travel sequences from GPS trajectories is proposed in [20]. The algorithm detects frequent sequences on different scales, taking into account the interestingness of the visited place and the experience of a user.

Research on movement data is usually done on trajectories acquired by GPS-enabled devices. However, large-scale GPS datasets, which would allow us to perform qualitative analysis on the level of a city or country, are still not available. On the other hand, geotagged photo collections could be obtained on the world scale, which makes them a valuable resource for the analysis of people’s activities. Importance of analysis of tourist activity, their behavior and interests using geotagged photos has been recently addressed in series of publications of Girardin and

co-authors in the framework of urban and tourism studies. Concentration and movement of tourists at the scale of a city is analyzed in [11] using Flickr geotagged photos. For this, the authors identified tourists in the city of Rome using user profiles and built heatmaps to visualize regions of high tourist concentration. The heatmaps were created by dividing a region into cells, counting the number of people who took photos in every cell and smoothing the visualization by interpolating between values of every cell. However, no detailed analysis of the method, its advantages and disadvantages was provided. In addition, flow maps were used to visualize tourist movement between visited places. These places were connected by lines whose widths were proportional to the number of tourists. Mean-shift, a non-parametric clustering algorithm, was used in [6] to find the most attractive places on Earth on a local and city scales using Flickr photos. The authors presented examples of maps with movements of people. However, no detailed analysis of the movement was presented.

## 3 Analytical Framework

We take a systematic approach suggested in [3] to defining possible types of analysis tasks related to the data about geotagged photos. We consider these data as a specific instance of the generic data type describing events, i.e. objects having positions in space and time. This data type can be represented by the formal model

$$O \rightarrow S \times T \times A_1 \dots \times A_n, \quad (1)$$

where  $O$  is the set of objects (events),  $S$  is the space (set of places),  $T$  is the time (set of moments), and  $A_1 \dots A_n$  are additional attributes of the events, if available. Table 1 lists the major analysis tasks defined generically on the basis of the data structure and instantiated for the case of geotagged photos.

To perform these tasks, an analyst needs a visual representation of the data and the way to interact with it. The visual environment in which the analyst could interactively perform data exploration and analysis should implement many of the general visual data exploration approaches such as brushing, focusing, multiple views, linking [4] and support geo-visualization approaches such as direct depiction, visualization of abstract data summaries, and extraction and visualization of computationally extracted patterns to allow the analyst to quickly generate and test his hypothesis.

Taking into consideration basic requirements of a geo-visualization environment and requirements needed for analysis of the tasks, we developed a framework in which Google Earth plays a primary role of geo-browser [13]. The integration of Google Earth into a custom application allowed us to compensate on the lack of direct support of geo-processing by implementing the geo-processing

Table 1: Generic tasks in analysis of event data and their specification for Panoramio/Flickr photo data

Task Category	Component in focus	Generic	Specific
Pattern identification	Spatial position	Find patterns in the spatial distribution of the events (A) For the whole time period (B) For selected time intervals	Detection and analysis of attractive areas sections 3.3.1, 3.3.2
Pattern identification	Temporal position	Find patterns in the temporal distribution (frequency) of the events (A) On the whole territory (large area) (B) In selected places	Analysis of temporal distributions in selected places: section 3.3.3
Pattern comparison	Spatial position	Compare the spatial distribution patterns (A) In different time intervals (B) For different subsets of events	Comparisons between time intervals: section 3.3.4 Comparisons between communities of people: section 3.3.6, 3.3.7
Pattern comparison	Temporal position	Compare temporal distribution patterns (A) In different places (B) For different subsets of events	Comparison of temporal distributions in different places: section 3.3.3

as part of the framework and enrich the functionality of Google Earth by such methods which are not available in its stand-alone version like multiple views, linking, and region selection using a drawing tool.

#### 4 Data collection and preprocessing

The dataset was collected by downloading photo metadata from Flickr and Panoramio sites, using publicly available APIs. The downloading of metadata from Flickr was performed similar as described in [6]: an initial user id was used to download his photo metadata. Then, we downloaded all the user’s contacts. To speed up the process of retrieving heterogeneous users, we retrieved all groups to which the user belongs, and using group information we were able to retrieve all the people who belong to these groups. This process was applied again on other users. We collected 86, 314, 466 entries from 4, 137, 248 users to the time of writing this paper, which allows us to apply spatiotemporal analysis virtually on every region in the world.

Panoramio data was downloaded by providing boundaries of a specific region of interest. We collected 9, 296, 104 entries from 738, 341 users for Western Europe and 3, 374, 535 entries from 391, 595 users for North and South America. In both datasets a number of attributes were obtained, including those that are most important for our tasks: owner id, photo id, photo URL, geographical coordinates, and timestamp.

During the data collection process we converted geographical coordinates expressed in degrees to Universal Transverse Mercator (UTM) such that analysis tasks that require Euclidian distances can directly use UTM coordinates.

Geo-referenced photo data may introduce a level of uncertainty. In some cases, coordinates could refer to the position of the photographer, while in others they refer to the

location of the object being photographed. This justifies our use of visual analytics methods, which require human involvement in addition to automated analysis in the exploration of the data.

Occasionally, the temporal information was not correct. Thus, in the flickr dataset 6, 229 photos have wrong dates and 50, 076 photos have dates after February 1, 2010 while the data have been collected before February 2010. Moreover, 9, 943, 161 combinations of user id and time stamp occurred several times, 30, 377, 849 records in total. We have to disregard entries in which timestamps were not provided or were incorrect for the tasks, where temporal aspect has a crucial role. 72, 454 photographers stored just a single photo, therefore these entries are not suitable for some of the tasks.

The data are being collected since the beginning of June, 2009 and, as to the end of January, takes up around 100 GB of our server’s disk space.

#### 5 Tasks and visual analytics components

In this section we consider seven possible tasks for analysis of movement data based on geotagged photos.

##### 5.1 Analysis of attractive areas

Finding interesting or attractive locations is usually of great importance in analysis of people’s mobility. While interestingness or attractiveness depends on the context of the problem, the analysis is usually done by finding places where trajectories of many people intersect often, or where people’s activity is more frequent than in other places. In case of photo data, people travel and take photos of interesting places. The interestingness is defined with respect to the user’s own understanding (to take a photo or not), however, if a substantial number of people like to take photos in the same region, it can suggest that the place is attractive.

One of the approaches to find attractive places is to divide an area into cells and count the number of people taking photos in every cell. This approach has several disadvantages. The division into cells is done irrespective of the data distribution, thus the results are sensitive to the size of the cell. Therefore, another approach, which has to capture the essence of the data, is required.

Density based clustering algorithms, such as DBSCAN [7] or OPTICS [5], are good candidates for analysis of attractive areas. Based on the notion of density connectivity between points using distance and density thresholds, these algorithms are able to find clusters of different shapes, sizes, filter outliers (regions where density is less than a predefined threshold) and have regions of different densities within a cluster. These algorithms require minimum two parameters: the radius threshold around a point and the minimum number of points (density threshold) within the radius.

In this task we used DBSCAN in order to analyze attractive regions in the city of Munich during the month of March 2008. 399 flickr users from total 20,200 were retrieved for that month.

We propose to visualize the results of the clustering by providing boundaries of clusters using convex hull algorithm 1(a) which is the fastest way to determine an area based on a set of point observations (photo points in our case). The analyst has a visual feeling where people concentrate in the city and which parts of the city attract more people. The visualization can be combined with filtering options, such that only clusters in which the number of people is more than a predefined threshold will be visualized. In our example, the largest area is situated in the center of the city where 71 people took photos.

Several approaches can be proposed to compare obtained clusters: (1) Comparing boundaries of clusters (see left part of Fig.1(a)); (2) Applying color to a cluster where color is mapped according to a specific attribute. The right part of 1(a) demonstrates clusters with colors mapped to the number of people who took photos in clusters (Color Brewer and Log scale were used to map colors); (3) Obtaining statistics for an individual cluster by clicking on the boundary of the cluster (see Fig 1(a)); (4) Producing histogram of clusters using some dependent variable. Fig. 1(b) shows a histogram of clusters and number of people who took photos in those clusters. The statistical information about every cluster and histogram are presented using Google Earth balloon feature.

## 5.2 Visualization of attractive areas using density maps

As was discussed in Section 5.1, the common and fast approach to analysis of attractive areas is to split the region into grid cells and count the number of people who took

photos in every cell. The simple way to interpret these results is to build heatmaps [12] using interpolation between values of every cell. While this approach can provide quick understanding of the level of attractiveness, it is a closed-box solution, which does not allow any further analysis to be carried out on top of heatmaps. We propose to use density based clustering as described in Section 5.1 to build density maps [14].

Density maps are used to highlight areas of people's activity within a cluster. Areas with different concentration are visualized by using color scale. This gives the analyst a possibility to differentiate between areas with lower concentration and areas with high concentration. The density maps are built using different resolution levels which allows inspection of all places together (left part of Fig1(c)) or by close zooming as presented in the right part of Fig.1(c). The density map is built as follows: (1) Clusters are obtained using a density based clustering algorithm as described in Sect 5.1 (2) In a cluster, the weight of every geotagged photo is calculated using density function based on its relative position to photos of other users in a cluster. Therefore, the weight of a photo will be high if a lot of photos of different users are located near that photo. The weight of a photo can indicate high activity or importance of a region around that photo and can be used as a representative image to highlight this activity. (3) Every geotagged photo location is mapped to a color using the weight calculated in the previous step. In contrast to heatmaps, density maps don't produce false points due to interpolation. Only existing photos, which belong to a cluster, will be visualized. This approach allows applying other techniques on top of density maps such as splitting dense clusters into local clusters and finding representative images.

The obtained clusters may span over large territory due to a popularity of an area (and parameters chosen for density based clustering algorithm) and have different densities. Such clusters can be further split into a number of local clusters with a given radius by utilizing the weight of a photo as a split heuristic. Since a weight can be regarded as a value of local maximum, the splitting is performed by selecting the most weighted photo as a centroid of a local cluster and assigning all photos that are located within a predefined radius to this cluster. This is repeated for each most weighted photo that has not been assigned to any local cluster yet. Local clusters can be used for locating representative images, which reflect the main attractions in a cluster. Since every geotagged location has a photo behind it, it can be visualized and treated as a representative photo of a cluster. Using weight of every photo, the representative image of the cluster can be defined as a photo that has the maximum weight in the cluster (see Fig.1(d)). For the illustration of density maps we selected a region of St.

Martin Island. Its main attractive areas can be seen on the left part of Fig 1(c), where a yellow color is used to depict the most concentrated areas. We focus on one such region, Maho Beach, which is one of the most attractive areas on the island, due to its vicinity to the airport landing strip (the right part of Fig.1(c)), where planes fly very low before they land.

Maho Beach spans over large area (people take photos of landing planes from nearby places such as Hotels and pubs) and constitutes one dense cluster, therefore we split it into series of local clusters having the same radius (100 meters). Using weights of photos, we retrieve the location of the most weighted photo in every cluster and display the image that belongs to that photo location (Fig.1(d)).

### 5.3 Analysis of temporal distributions of events

Temporal distributions of events can reveal some interesting expected or unexpected behavioral patterns. For example, the activity of people can increase during a national holiday or local event such as football match. Fig.1(e) presents graphs of the daily and monthly activity of people as number of persons who took photos in a specific area as a function of time. After inspecting the graphs the analyst can narrow down the region of interest or/and the time frame to get more detailed results. Also, this approach can be used to perform comparison of multiple temporal distribution patterns. The analyst can find similarities and differences between events occurring in different places at specific time intervals by comparing their graphs in multiple views or using overlays in a single view.

### 5.4 Comparison of spatial distributions in different times

Different places may attract different number of people in different periods of time. It can be influenced by weather conditions or other events like sport or holiday. Revealing such differences may be important to local administrations or tourist services. Fig.1(f) presents cluster boundaries in Munich (depicted in yellow) for two different time periods. The left part shows the places where people took photos during the October Festival 2008; the right image shows the places during month of March, 2008. It can be seen, that during the festival people are mostly concentrated within the area of the event (large cluster in the south-west part of the map) and less interested in other attractive places, while in March people mostly visit places located in the center of the city.

The analyst may also be interested in splitting large dense clusters into local ones (as described in Section 5.1) filtering those local clusters where the number of people is below some predefined threshold. Orange clusters in Fig.1(f) correspond to local clusters having more than 9

people.

### 5.5 Detailed analysis of clusters

As was already mentioned, the cluster boundaries are obtained by applying convex hull algorithm on the photo locations belonging to a cluster. While cluster boundaries provide a general feel of the extents of a cluster, they don't provide any information about particular locations of taken photos. Finding out where the photos were taken in a cluster and which photos are assigned to a cluster can be important for a detailed analysis of attractive areas. Fig. 1(g) demonstrates this approach by visualizing photos belonging to a cluster as red circled placemarks. Using the interaction environment of our framework, different information can be obtained, such as owner id, photo id, title and the time the photo was taken.

### 5.6 Ranking of sightseeings in a city and validating findings by comparison of ranks from different communities

Very often, when we plan our trip, we want to find out which locations are more interesting to visit than the others. In other words, we would like to be able to order the places by some criteria. In general, user-defined ranking of places can be used in such scenarios as tour planning or recommendation. There are web sites that provide rankings of points of interest (POIs) in the world based on explicit or implicit scores provided by users, such as reviews, opinions, comments or ratings. One of these websites is TripAdvisor (<http://www.tripadvisor.com>)

Another possibility to rate the places is to use photo data, by counting the number of people who took photos in the POI vicinity. A simple automatic approach would be to take every photo and to assign it to a nearest POI. This approach is simple but inaccurate for two reasons. First, we need to obtain the POI database for a particular region, which can be a problematic task. Second, POIs can be of different sizes. For example, a zoo may span several square kilometers, while a statue covers several square meters. However, the POI location is usually described by a single coordinate, so it can happen that a photo taken within the natural boundaries of POI won't be assigned to it, because it is located far away from the center coordinate defined for a POI.

Analysis of title or tags may introduce other inaccuracies due to missing information, differences in naming or different languages. This is the case when the ranking based on manual outline of a POI region may be more appropriate and provide more control to the analyst. Google Earth is very handy because it allows the user to easily locate any POI and navigate to it, so there is no need to have an external POI database. The manual outline of a region is added as an extension in our framework. For this, we've

defined 50 most interesting sightseeings in Berlin<sup>3</sup> and outlined them as shown in Fig.1(h). We counted the number of people in every region using Flickr (47859 photos and 4567 users) and Panoramio (39625 photos and 6630 users) datasets. The results of ranking are shown in Fig.1(h) using different perspectives where the numbers represent the ranks starting from 1 (top rank). The left part of Fig.1(h) shows the results using 3D perspective where the size of vertical lines are inversely related to the rank of a POI, while the right part of Fig.1(h) shows the ranked places from above. The navigation to a specific POI and acquisition of POI ranks can be done by clicking on the POI region. We don't impose any visualization style in our example, but colors or symbol sizes can be used to allow the user to visually differentiate between ranks.

The ranking also allows us to compare and validate behavioral patterns of different groups. For example, we would like to compare how close the ranking results are between Flickr, Panoramio and TripAdvisor communities. Many approaches to achieve this exist: we could use parallel coordinates, statistical measures or tabular comparison. We decided to use Andrews Plot, which is proven to be a good graphic method for detecting outliers [1] in multivariate data. The plotting is performed by converting  $n$ -dimensional data into a function in a two-dimensional space of the form:

$$f_x(t) = x_1\sqrt{2} + x_2\sin(t) + x_3\cos(t) + x_4\sin(2t) + \dots, \quad (2)$$

where  $-\pi \leq t \leq \pi$  and coefficients of  $x_1, x_2, x_3, \dots$  are terms in Fourier series.

The variables listed first give more weight to the function and it is suggested [8] that the original data is provided in the order of importance. Fig.1(i) presents the variations between ranks given by members of 3 communities. The X-axis represents the 23 most ranked places sorted in descending order using Flickr ranked places as a reference point. Y-axis is a frequency variation described by Equation 2. It can be seen, that Panoramio has smaller fluctuations around Flickr, which suggests that Panoramio ranking is more similar to Flickr than ranks from TripAdvisor.

## 5.7 Comparison of attractive area patterns of different communities

The comparison of rankings is not the only way to compare behavioral patterns. The analyst may select the region of interest and apply clustering algorithm using data from different sources. The obtained clusters may reveal some interesting patterns and differences between communities. 1(j) shows the results of finding highly visited re-

gions in London during the year 2008 using a variant of DBSCAN algorithm with adaptive densities[15], applied on Panoramio (5266 photos from 1082 users) and Flickr (94765 photos from 12062 users). Since the number of users in Flickr is almost 11 times larger, we used 40 users as a *MinPts* threshold parameter [7] for Panoramio data and 410 users for Flickr.

The algorithm produced 5 dense clusters using Panoramio and 4 dense clusters using Flickr data. The observation reveals that two places in London are highly visited by people from both communities: Victoria Memorial and the area of Piccadilly Circus. However, the Greater London Urban Area is not among the highly visited places in case of Panoramio. Likewise, the London Eye is not among the highly visited places in case of Flickr.

## 6 Discussion and future work

In this paper, we analyzed the structure of the event-based movement data to define systematically seven tasks for event-based analysis of people's travel activities, behavior and mobility using geotagged photo data, collected and shared by people from all over the world. We showed that attractiveness of places can be obtained from photo data and presented computational and visual techniques to analyze these places with and without temporal dimension. We presented a way to assess attractiveness of places based on ranking and compared differences between three communities (Flickr, Panoramio, TripAdvisor). We suggested a set of visual analytics methods combining computational techniques with interactive visual displays to support the analysis of such data.

In the future, we shall address in more detail the analysis tasks for the trajectory-oriented view. Likewise, we will concentrate on definition of further tasks related to analysis of photo data and develop new computational and visual analytics methods.

## Acknowledgements

This work was partially funded by the German Research Society (DFG) under grant GK-1042 (Research Training Group "Explorative Analysis and Visualization of Large Information Spaces"), and by the Priority Program (SPP) 1335 ("Visual Spatio-temporal Pattern Analysis of Movement and Event Data").

The authors wish to thank two anonymous reviewers for their helpful comments.

## References

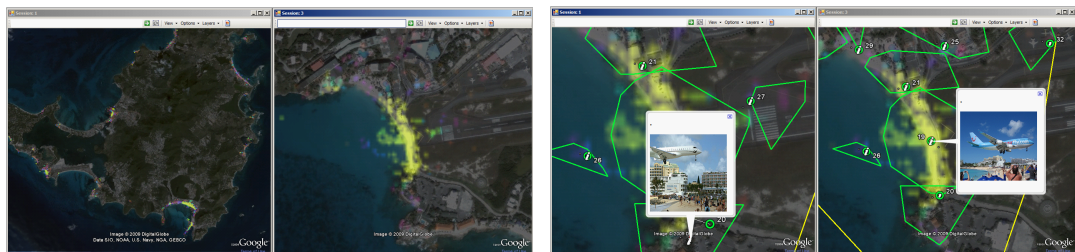
- [1] DF Andrews. Plots of high-dimensional data. *Biometrics*, 28(1):125–136, 1972.

<sup>3</sup>[http://www.visitberlin.de/english/sightseeing/e\\_si\\_sehenswuerdigkeiten.php](http://www.visitberlin.de/english/sightseeing/e_si_sehenswuerdigkeiten.php), January 2010

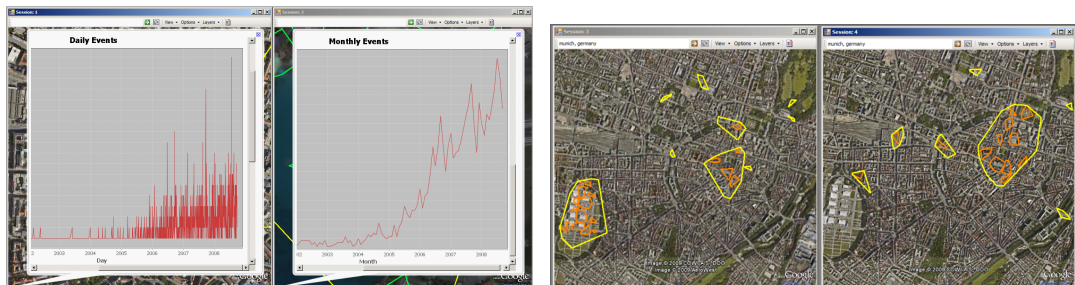
- [2] G. Andrienko, N. Andrienko, P. Bak, S. Kisilevich, and D. Keim. Analysis of community-contributed space-and time-referenced data (example of flickr and panoramio photos). In *Proceedings of IEEE Visual Analytics Science and Technology (VAST 2009)*, pages 213–214, 2009.
- [3] N. Andrienko and G. Andrienko. *Exploratory analysis of spatial and temporal data: a systematic approach*. Springer Verlag, 2006.
- [4] N. Andrienko, G. Andrienko, and P. Gatalaky. Exploratory spatio-temporal visualization: an analytical review. *Journal of Visual Languages and Computing*, 14(6):503–541, 2003.
- [5] M. Ankerst, M.M. Breunig, H.P. Kriegel, and J. Sander. OPTICS: Ordering points to identify the clustering structure. *ACM SIGMOD Record*, 28(2):49–60, 1999.
- [6] D.J. Crandall, L. Backstrom, D. Huttenlocher, and J. Kleinberg. Mapping the world’s photos. In *Proceedings of the 18th international conference on World wide web*, pages 761–770. ACM, 2009.
- [7] M. Ester, H.P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proc. KDD*, volume 96, pages 226–231, 1996.
- [8] B.S. Everitt and G. Dunn. *Applied multivariate data analysis*. 2001.
- [9] J. Gantz, C. Chute, A. Manfrediz, S. Minton, D. Reinsel, W. Schlichting, and A. Toncheva. An updated forecast of worldwide information growth through 2011. 2008. <http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf>.
- [10] F. Giannotti, M. Nanni, F. Pinelli, and D. Pedreschi. Trajectory pattern mining. In *13th ACM SIGKDD international conference on Knowledge discovery and data mining*, page 339, 2007.
- [11] F. Girardin, F.D. Fiore, C. Ratti, and J. Blat. Leveraging explicitly disclosed location information to understand tourist dynamics: a case study. *Journal of Location Based Services*, 2(1):41–56, 2008.
- [12] F. Kessler, R. McMaster, H. Howard, and T. Slocum. *Thematic Cartography and Geovisualization*. Prentice Hall, 2008.
- [13] S. Kisilevich, D. Keim, and L. Rokach. A generic google earth-based framework for analyzing and exploring spatio-temporal data. In *12th International Conference on Enterprise Information Systems*, 2010.
- [14] S. Kisilevich, F. Mansmann, A. Bak, P. Tchaikin, and D. Keim. Where would you go on your next vacation? - a framework for visual exploration of attractive places. *Proceedings of the GeoProcessing 2010*, 2010.
- [15] S. Kisilevich, F. Mansmann, and D. Keim. P-dbscan: A density based clustering algorithm for exploration and analysis of attractive areas using collections of geo-tagged photos. In *1st International Conference on Computing for Geospatial Research & Application*, 2010.
- [16] M.P. Kwan and J. Lee. Geovisualization of human activity patterns using 3D GIS: a time-geographic approach. *Spatially integrated social science*, 27, 2004.
- [17] P. Laube, S. Imfeld, and R. Weibel. Discovering relative motion patterns in groups of moving point objects. *International Journal of Geographical Information Science*, 19(6):639–668, 2005.
- [18] S. Rinzivillo, D. Pedreschi, M. Nanni, F. Giannotti, N. Andrienko, and G. Andrienko. Visually driven analysis of movement data by progressive clustering. *Information Visualization*, 7, 3(4):225–239, 2008.
- [19] J. Zhao, P. Forer, and A.S. Harvey. Activities, ringmaps and geovisualization of large human movement fields. *Information Visualization*, 7(3-4):198–209, 2008.
- [20] Y. Zheng, L. Zhang, X. Xie, and W.Y. Ma. Mining interesting locations and travel sequences from GPS trajectories. In *Proceedings of the 18th international conference on World wide web*, pages 791–800. ACM New York, NY, USA, 2009.



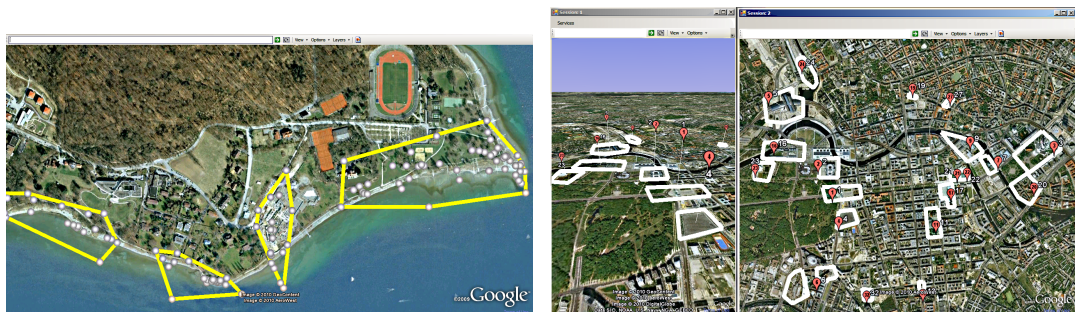
(a) Cluster visualization using convex hull (left and right) and conveying additional information using color (right) (b) Aggregated cluster statistics using histogram. X-Axis: cluster ids, Y-axis: number of people who took photos in a cluster



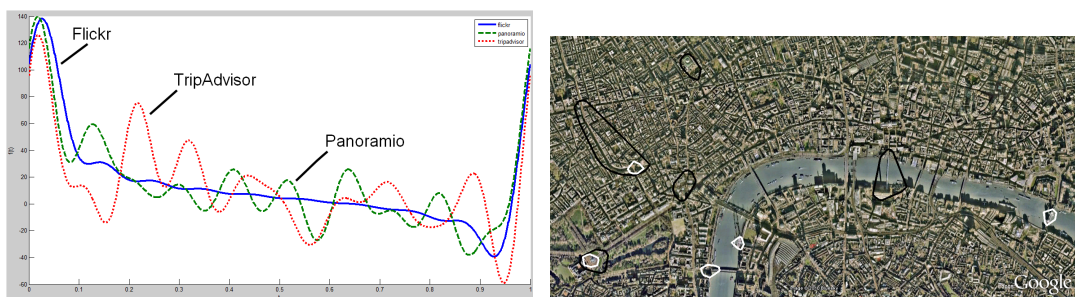
(c) Attractive areas using density maps with different level of resolutions (d) Representative images of local clusters



(e) Daily and monthly events (f) Comparison of spatial distributions in different time intervals



(g) Visualization of exact places where people took photos (h) Ranked POIs (Berlin) using 2D and 3D perspectives



(i) Comparison of ranking scores between Flickr, Panoramio and Tripadvisor communities (j) Comparison of behavioral patterns between Flickr (black clusters) and Panoramio (white clusters) communities

Figure 1: Illustration of visual analytics tasks