Analysis of community-contributed space- and time-referenced data by example of Panoramio photos

Abstract

Space- and time-referenced data published on the Web by general people can be viewed in a dual way: as independent spatio-temporal events and as trajectories of people in the geographical space. These two views suppose different approaches to the analysis, which can yield different kinds of valuable knowledge about places and about people. We present several analysis methods corresponding to these two views. The methods are suited to the large amounts of the data.

1 Introduction

In the age of Web 2.0 more and more people publish various kinds of contents on the Web. Some kinds of contents have spatial and temporal references, for example, the photos in Panoramio [9] and flickr [4] linked to the places where they were taken and supplied with the dates and times of the shots. Such data can serve as a source of knowledge about the places and about the interests, behaviors, and mobility of the people. The knowledge may be valuable for local administrations, tourist services, advertising agencies, and other organizations. However, the data are not easy to analyze. One of the problems is the huge number of entries, which calls for scalable computational techniques. At the same time, the involvement of a human analyst, who perceives spatial and temporal patterns and gives them meaning, is essential.

The community-contributed Web entries having spatial and temporal references can be viewed, on the one hand, as independent spatio-temporal events. On the other hand, the entries made by the same person can be considered as a trajectory of this person in the geographical space, which tells something about the movement and behavior of this person. The whole dataset can be viewed as a set of trajectories of multiple people. These two views suppose different approaches to the analysis, which can yield different kinds of knowledge. Another example of data that can be viewed in such a dual way is data about mobile phone calls. We call this class of data *event-based movement data*.

Analysis of event-based movement data is a relatively new research topic. A series of papers have been published by Girardin and co-authors (e.g. [5] and [6]). They analyze concentrations and movements of tourists at the scales of a city (e.g. Rome) and a geographical region (e.g. central Italy) using Flickr photos and, in some studies, mobile phone calls made by the tourists in the same time periods. Concentrations are shown on heat maps produced by dividing the area of interest by a rectangular grid and counting the number of photos in every grid cell. Interpolation techniques are used in order to smooth the visualization. The movements are visualized in an aggregated form by means of flow maps, where predefined places are connected by lines with the widths proportional to the numbers of tourists that moved between the places. In [3], another research team uses geo-annotated Flickr photos to find concentrations of activity and most popular places on Earth. For that they use a non-parametric MeanShift clustering algorithm. The authors have also used the temporal information available from the photos to generate a few example maps showing the movements of photographers. The trajectories are represented by lines drawn on the maps in a semi-transparent mode; however, no further analysis is made.

2 Our Approach and Methods

We take a systematic approach to the analysis of event-based movement data. We define possible types of analysis tasks related to the views of the data as events and as trajectories. We distinguish space-centered tasks, where the data are used to study the properties of the space and places, from

agent-centered tasks targeting at the properties and behaviors of the people (in general, moving agents). For the tasks defined, we try to find or develop appropriate methods. Some of the methods are briefly presented here.
The example dataset we use for the presentation consists of about 590 000 photos made on the territory of Germany during the period from January 1, 2005 till March 30, 2009. The methods we present do not take into account the contents of the photos but only their spatial and temporal references.

2.1 Spatio-temporal aggregation of events

Aggregation helps an analyst to cope with large amounts of data. Spatial and temporal aggregation of movement data viewed as independent events can be done by means of database queries [1]. For the spatial aggregation, the territory is divided into suitable compartments, for example, using a regular grid. For the temporal aggregation, the time may be treated as linear or as cyclic and, respectively, divided into consecutive intervals or into intervals of the daily, weekly, or yearly cycle. The results of the aggregation are time series of counts related to the spatial compartments (number of photos, number of different people) and, possibly, other statistics such as the mean number of photos per person.

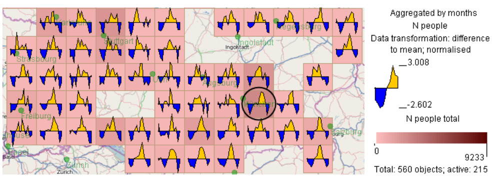

Figure 1: A possible map display of aggregated data.

Figure 1 gives an example of how aggregation results can be visualized and explored. The data have been aggregated spatially by grid cells and temporally by months, irrespective of the years. The map fragment shows the south of Germany. The shading of the cells portrays the total number of different people who made their photos in these cells. Dynamic filtering has been applied so that only the cells visited by at least 250 photographers are visible. The diagrams represent the yearly variation of the number of photographers. The counts resulting from the aggregation have been transformed into the differences from the local mean values normalized by the standard deviations. The horizontal axis of a diagram represents 12 months from January to December. The vertical dimension represents the deviation from the local mean value. The color filling serves to improve the visibility of the diagrams and the differentiation between the positive and negative values (yellow and blue, respectively). The common and distinct features of the seasonal variations in different places are easily seen. For instance, a peculiarity can be noted in the area of Munich (marked in the figure): the number of photographers is very high from August till October, which may be related to the famous Oktoberfest (beer festival).

2.2 Spatial clustering of events

Density map is a good technique to visualize concentrations of events. A disadvantage of the grid-based approach described in [5][6] is that grids dont reflect the natural data distribution. Therefore, the resulting map may inaccurately show the places of concentration. Appropriate clustering methods are needed to find the natural areas of high concentration without imposing any artificial division of the territory. We propose to use density-based clustering algorithms (DBCA). The benefits are the following: (1) regions with density below some threshold are counted as noise and wont be reflected in visualization. (2) The resulting clusters are not restricted to a specific shape or size and properly reflect the natural distribution of the data.
A frequently used approach to visualizing densities on a map is color coding of the counts of points inside areas, as in [5][6] and in Figure 1 (background shading). A drawback is that the information about the locations of the individual photos is lost. We propose to use the notion of influence function introduced in [7]. For every photo in a cluster, the cumulative sum of the influences of the other photos is calculated using the Gaussian. Thus, every photo gets a weight based on the proximity of other photos around. When a density map is built, every photo gets its normalized color value. An example is given in Figure 2.

The map fragment portrays the central part of Berlin. The densities of the photos are represented using the color scale shown in the upper right corner of the image. The idea is to express the popularity of the place (i.e. the density of photos in it) as

Figure 2: A density map showing the popular places in Berlin.

the degree of heat. The right side of the color scale with red-hot and white-hot shades corresponds to the most popular places. Thus, the white spots in Figure 2 highlight the areas around the Bundestag, Brandenburg Gate, Potsdam Square, on the Museum Island, and around the Alexanderplatz.

2.3 Spatiotemporal analysis of clusters of events

The Growth Ring Maps technique supports the exploration of the frequencies and temporal patterns of events occurring in the same places. For the Panoramio dataset, we defined the significant places on the basis of the density-based spatial clustering of the events. For each place, we represent the photos made in it by pixels placed around the central point in an orbital layout [8] according to the times of taking the photos: the earlier the photo was made, the closer the pixel is to the central point. Color-hue is used to map semantic properties of the events or places. Thus, seasonal differences in visiting the places may be investigated by mapping the seasons to four distinct colors (winter-white, spring-green, summer-red, and autumn-orange). The resulting Growth Ring Maps show simultaneously the intensity of taking photos at different locations and the seasonal differences. Figure 3 presents the results for Berlin and Konstanz. While Berlin is visited through the whole year, Konstanz appears as a popular place in the warm season, as indicated by the Growth Rings with very little amounts of white color.

Figure 3: Growth Ring Maps showing the seasonal differences of taking photos for Berlin (top) and Konstanz (bottom). The small amount of white color in Konstanz indicates that warm months are more popular at this region, whereas Berlin is visited all over the year.

2.4 Analysis of flows

For this analysis, we build flow maps showing aggregated moves between places, i.e. how many people have moved from one place to another. There are two major differences from what is described in [5] and [6]. First, our flow maps show not only the amounts but also the directions of the movement by special half-arrow symbols. It is easy to see whether the movements between two locations are one- or two-directional and, in the latter case, whether one of the directions prevails over the other. Second, we do not use any predefined places but divide the territory into appropriate compartments on the basis of clustering of the positions from the trajectories. By varying the clustering parameter (specifically, the desired cluster radius), we can do the analysis at different spatial scales. An example is presented in Figure 4. We have taken the sequences of photos made in Berlin and surrounding area and divided them into subsequences, or sessions, assuming that a time interval of 8 hours or more between two photos means the beginning of a new session. The sessions have been treated as trajectories. First, we tessellated the territory into bigger areas using clustering with the desired radius 5km. The upper map fragment shows the flows between the areas. By means of interactive filtering, we have hidden the flows corresponding to less than 10 trajectories. We can see three disjoint regions of major movement: the central part of Berlin (on the east), Potsdam (on the southwest), and suburbs on the west of the central Berlin. This means that the photo sessions are restricted in space. The time intervals between the photos taken in different regions are usually longer than 8 hours.

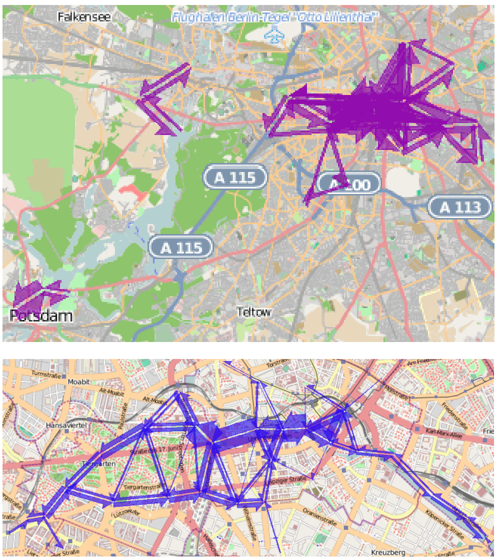

Figure 4: The flows of the photographers in Berlin at two different spatial scales.

The lower map fragment represents the movements in the central part of Berlin aggregated at a smaller spatial scale (the cluster radius was 1km). By filtering, we have removed the flows corresponding to less than 40 trajectories. We see that major movements occur along the street Unter den Linden. It is possible to see asymmetric two-directional movements. Thus, more photographers move along Unter den Linden in the direction from east to west towards Brandenburg Gate than from west to east. An opposite tendency can be seen on the southwest and south of this street, where the eastward movement Zoo - Potsdam Square - Checkpoint Charlie clearly prevails over the westward movement.

2.5 Analysis of interactions in space and time

One of the possible analysis tasks is detection and investigation of collective movement patterns and social behaviors of the photographers. We support this task by tools for interaction analysis [2]. We use the term interaction for the kind of event when two or more people (in general, moving agents) are located in the same place at the same time. We have developed a computational method for extracting interactions from movement data. For each pair of trajectories, it tries to find respective positions such that the spatial and temporal distances between them are within the given thresholds. In case of detecting such positions, the following positions of the trajectories are checked. Spatial and temporal indexing of trajectory fragments is used for the sake of efficiency.
Of primary interest are interactions with long duration (i.e. photographers spend significant time together) and repeated interactions between the same photographers. In order to detect these, we applied the tool to the 9884 trajectories having at least 10 positions (i.e. photos made). We used the spatial threshold of 1km and the time threshold of 2 hours and got 3032 interactions that lasted minimum one hour. From the 9884 photographers, 2587 had some interactions with others; however, only 1145 photographers had two or more interactions with others and, among them, only 224 photographers had repeated interactions with some of the other photographers. The maximum number of interactions per photographer is 100; the number of repeated interactions between a pair of photographers ranges from 2 to 36.

Figure 5 demonstrates an example of multiple interactions between two photographers. 14 interactions took place in Berlin and surroundings during the period from February 24, 2008 till July 19, 2008. The interactions are represented on a map (top), time line display (middle), and in a space-time cube together with the trajectories of these two people (bottom). The trajectories are shown as blue

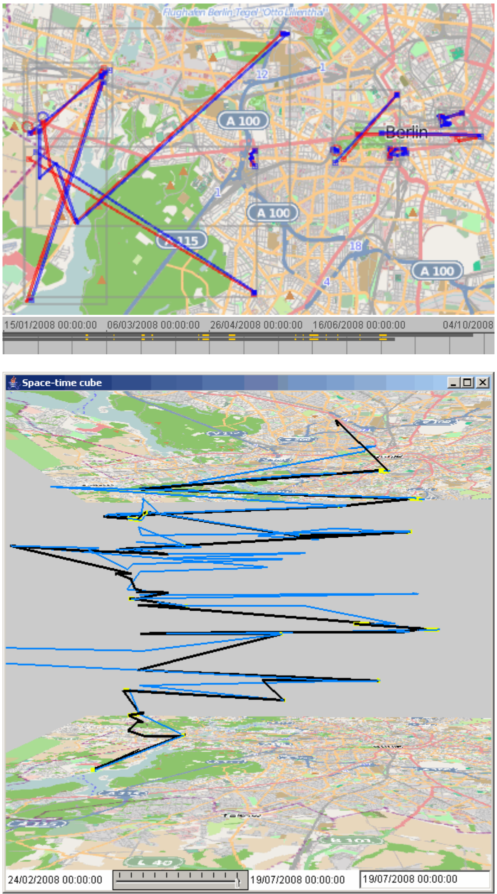

Figure 5: Multiple interactions of two photographers.

and black lines; the interactions are marked in yellow. It can be seen that the photographers do not always move and make photos together but meet from time to time. Among the detected interactions, there are cases when two photographers, apparently, traveled together from one city or area to another, e.g. from Leipzig to Hamburg. Specifically, there are 27 cases when the maximum distance between the places of taking photos in one interaction is 20km or more. Some of the long-distance interactions are demonstrated in Figure 6.

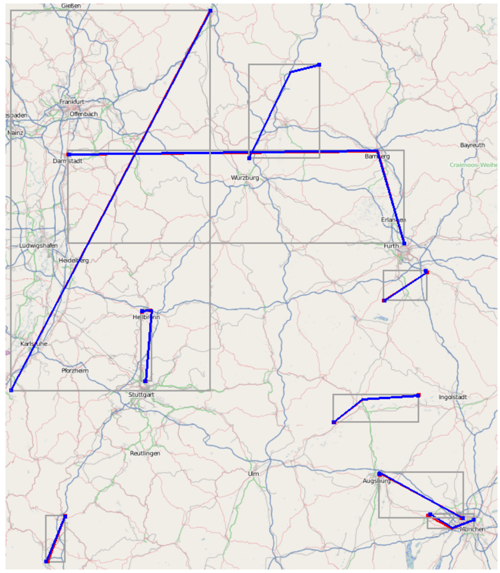

Figure 6: Multiple interactions of two photographers.

3 Conclusions

In most of the existing approaches to analysis of movement data it is assumed (in most cases implicitly) that the available position records represent continuous trajectories and, hence, intermediate positions can be obtained by means of interpolation between known positions. However, there are many cases when the position records are temporally sparse and irregular, which means that the data cannot be handled in this way. This kind of data can be called event-based movement data. We suggest a set of visual analytics methods combining computational techniques with interactive visual displays to support the analysis of such data. The methods are oriented to different classes of analysis tasks and enable the analyst to discover different kinds of patterns in the data, as is demonstrated by example of Panoramio photos.

4 Acknowledgments

The work has been done within the research project ViAMoD Visual Spatiotemporal Pattern Analysis of Movement and Event Data, which is funded by

DFG Deutsche Forschungsgemeinschaft (German Research Foundation) within the Priority Research Programme Scalable Visual Analytics (SPP 1335).

References

[1] Andrienko, G., and Andrienko, N. Spatio-temporal aggregation for visual analysis of movements. In *Proceedings of IEEE Symposium on Visual Analytics Science and Technology (VAST 2008), IEEE Computer Society Press,* 51-58, 2008.

[2] Andrienko, N., Andrienko, G., Wachowicz, M., and Orellana, D. Uncovering Interactions between Moving Objects. In *Cova, T.J., Miller, H.J., Beard, K., Frank, A.U., Goodchild, M.F. (Eds.): GIScience, 5th international conference, Proceedings,* 16-26, 2008.

[3] Crandall, D., Backstrom, L., Huttenlocher, D., and Kleinberg, J. Mapping the Worlds Photos. *International World Wide Web Conference.* 2009.

[4] Flickr. http://flickr.com

[5] Girardin, F., Calabrese, F., Dal Fioro, F., Ratti, C., Blat, J. Digital Footprinting: Uncovering Tourists with User-Generated Content. *Pervasive Computing, IEEE,* 7, 36-43, 2008.

[6] Girardin, F., Dal Fioro, F., Ratti, C., Blat, J. Understanding of Tourist Dynamics from Explicitly Disclosed Location Information. In *4th International Symposium on LBS and Telecartography, Hong-Kong, China,* 2007.

[7] Hinneburg, E. and Keim, D. An efficient approach to clustering in large multimedia databases with noise. *AAAI Press.* 58-65, 1998.

[8] Janetzko, H. Mansmann, F., Bak, P., and Keim, D.A Spatiotemporal analysis of Sensor Logs using Growth Ring Maps. *IEEE Transactions of Computer Graphics and Visualization*,2009 (accepted).

[9] Panoramio. http://www.panoramio.com