Visual Analytics of Mobile Data

Benjamin Höferlin[*]
Institute for Visualization and Interactive Systems
University of Stuttgart, Germany

Markus Höferlin[†]
Visualisation Research Center Stuttgart
University of Stuttgart, Germany

Jürgen Räuchle[‡]
Institute for Visualization and Interactive Systems
University of Stuttgart, Germany

ABSTRACT

Analysis of mobile data without a predefined task or known search target requires an explorative analysis approach. We present a flexible and scalable visual analytics system, which enables the human analyst to efficiently discover unknown patterns in movement data. Our visual analytics system integrates different coordinated and interactive visualizations that all provide complementing views on the data. Scalability of the analysis process with respect to the amount of data is facilitated by different filtering and aggregation techniques. Two usage examples show the broad analysis capabilities of our approach in the context of Nokia's Mobile Data Challenge.

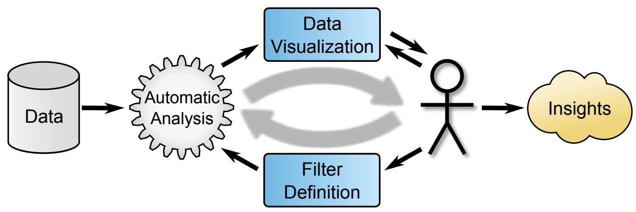

Figure 1: Structure of the presented visual analytics system for mobile data analysis.

1. INTRODUCTION

When analyzing time-dependent data without exact knowledge of the search target, usually patterns of common behavior or abnormalities are of interest. In such situations, in which the search target is vaguely defined or completely unknown, purely automatic data mining approaches are only applicable to a limited extent. One solution is to closely involve the users in the analysis process, for example by visual data analysis. However, such approaches often show limited scalability in complex data domains with Big Data.

Visual analytics addresses this issue by combining the tremendous processing power of automatic approaches with the excellent recognition capabilities of human analysts. The link between both worlds is established by visualization techniques and human computer interaction. The general structure of the developed visual analytics system is illustrated in Figure 1. This way pattern discovery can be iteratively steered by the human analysts integrating their domain knowledge. Visual analytics can be summarized as "the science of analytical reasoning facilitated by interactive visual interfaces" [15]. Yet it is more than just visualization. As Keim *et al.* [9] outline, it can be regarded "as an integral approach combining visualization, human factors and data analysis".

The visual analytics process follows the visual information seeking mantra formulated by Shneiderman [14]: "Overview first, zoom and filter, then details-on-demand". Therefore, we provide different views on the data by four complementing visualizations. These views include different levels of aggregation to empower the users selecting the desired level of detail. By automatic aggregation methods, the amount of information analyzed by the users remains manageable.

To address the open nature of questions posed to the system, two fundamental methods of information retrieval are provided by the visual analytics system: query by filter definition and visual exploration. A first overview of the patterns the data contains can be gained by visually exploring the different facets of the data. After a more precise information need has developed and can be expressed by a query to the system, filters can be defined to retrieve the required information. These filters are not only useful to reduce the amount of data presented, but also to verify hypotheses on potential patterns, which are developed during the analysis process.

[*]benjamin.hoeferlin@vis.uni-stuttgart.de
[†]markus.hoeferlin@visus.uni-stuttgart.de
[‡]raeuchjn@studi.informatik.uni-tuttgart.de

This material was prepared for the Mobile Data Challenge 2012 (by Nokia) Workshop; June 18-19, 2012; Newcastle, UK.

2. RELATED WORK

Visual Analysis of movement data is traditionally covered by research in the fields of geoVisualization and geospatial visual analytics. These concepts relate the visualization of geographical data with the interactive manipulation of its graphical representation. Furthermore, geographic data mining and information retrieval are principal components to geospatial visual analytics. In the context of these re-

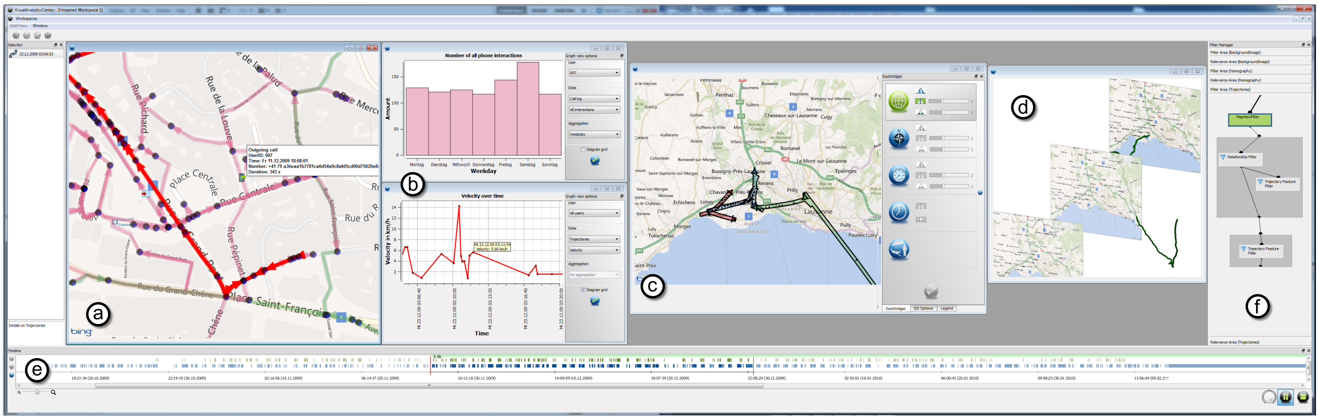

Figure 2: Screenshot of the visual analytics system with its different components: a) map view, b) chart views, c) Interactive Schematic Summaries, d) VideoPerpetuoGram, e) timeline, and f) filter graph.

search areas, different methods and tools have emerged (see Andrienko *et al.* [4, 3, 2] for an overview) that are related to the presented visual analytics system.

Besides analysis of common movement paths and normal behavior of a group of participants, we additionally demonstrate applicability of visual analytics to the analysis of individuals' mobile data. Tools for visual analysis of geographical data that focus on individuals' movement data are most related to our approach. There are two systems that match these criteria.

A tool called spatial history explorer was introduced by Mountain [11] for the visual analysis of individual mobile data to develop geographic filters that represent the geographic context of mobile individuals in different scenarios. The spatial history explorer utilizes three linked views to address the spatial, temporal, and attribute distribution of the movement data, respectively. Further, interactive spatio-temporal query definition allows to select the data relevant for analysis.

Shen and Ma introduced MobiVis [13], a graph-based visualization that integrates timevarying spatial and social information of mobile data. Their visual analytics system supports temporal filtering via an interactive time chart and semantic filtering via an ontology graph. Further, comparison of individual and group behavior patterns is facilitated by a compact radial representation they term *behaviour rings*.

Both systems cover parts of the presented visual analytics system. Our system, however, offers a more thorough set of coordinated views on the data, which are carefully selected to support revelation of spatial, temporal, spatio-temporal, and attribute patterns. Filtering and aggregation techniques allow for scalable human-centered information retrieval by visual data exploration and interactive query definition in the precense of Big Data. Besides movement data, supplementary mobile data is included in the analysis, too. Finally, our system allows to seamlessly break down analysis from behavioral patterns of large groups to interaction of few individuals to the actions of a single person.

3. DATA PREPROCESSING

In this paper, we use mobile data provided by the *Nokia Mobile Data Challenge* (MDC) [10]. To capture prominent patterns of the participants' behavior, we focus on their movement trajectories and enrich this data with additional events recorded by their mobiles such as call logs, system messages, Bluetooth connections, etc.. Please note the term participant refers to the subjects involved in data recording. Trajectories are generated within a preprocessing step. Therefore, we route the GPS samples by *Bing Maps REST* services. Based on time information and subject's estimated movement speed, we split the samples in distinct traces and decide whether to use road or footway routing option. Several plausibility checks (e.g., distance between the original trajectory and the routed has to be within a certain ratio) were applied to maintain high quality of the generated trajectories. GPS samples with adequate positional confidence, but large distance to the next street were directly transferred to the dataset to preserve hiking, skiing, and boat trips. Though the high quality of trajectories generated, the users may fall back to raw GPS samples to capture the data in detail, if necessary.

4. VISUAL ANALYTICS SYSTEM

In this section, we briefly introduce the visual analytics system developed for the MDC. A screenshot of the complete system is depicted in Figure 2. The system provides various coordinated views that each highlights different aspects of the data. This means that the different visualizations are tightly coupled to support the users in understanding the data and its facets.

A fundamental view on the data is given by the map view (Figure 2(a)), which illustrates the trajectories in their natural domain. This view is very important, since it can be used to gain an overview over the distribution of the participants' movements and to retrieve detail information on a single trajectory. It consists of a map superimposed with trajectories, which are linearly interpolated between the samples. The trajectories of the participants are drawn semi-transparently to alleviate the effect of overdrawing in the distinct colors of the subjects. At levels of higher detail (i.e. high zoom factor is used), we draw additionally to the trajectories its samples and glyphs that represent additional cell phone information such as battery state, incoming or outgoing calls, etc.. By hovering over these glyphs, the users can access the raw data via tooltip. For the map itself, we rely on different web services provided by Bing, Google, and Open Street Map. Thus, several features that are natively available in these services, are accessible by the users such as road view, satellite view, bird's eye view, panoramio image layers, and Google's *What's here* service, which provides local information in the web browser. With the help of these web services, the users can use typical navigation in digital maps such as pan, rotation, and zoom.

To enable efficient exploration of movement data, we utilize an approach called *Interactive Schematic Summaries* (ISS) that was introduced in the context of video surveillance [7]. By this approach users can explore the movement data by scatter/gather browsing trajectory clusters. Therefore, trajectories are aggregated into clusters according to their similarity with respect to various aspects among them position, movement direction, movement speed, and temporal properties. Hence, the trajectories are scattered into clusters that are visually represented by a couple of schematic arrows simplified by a trajectory bundling technique to account for visual clutter (Figure 2(c)). After gathering one or more clusters by clicking them, a filter is applied to the dataset and all trajectories selected are re-scattered into new clusters. Hence, this view allows to quickly gaining overview on the main movement paths and recurring movement in the dataset.

The chart view (Figure 2(b)) depicts time series of features derived from trajectories (e.g., velocity, accuracy) and distributions of additional information such as call logs or system status messages. Different temporal aggregation intervals can be selected by the users such as minutes, hours, days, or weekdays. Aggregation allows besides the selection of the desired level of detail also identifying general behavior of the participants and recurring patterns. Knowing these patterns, deviation from general behavior could easily be identified (cf. with the first usage example in section 5). Additional trajectory information is presented in the color of the participant. Interaction with the chart by zooming and panning facilitates to focus on relevant parts of the data.

In contrast to the other views that aggregate either spatial or temporal dimensions of the data, the *VideoPerpetuoGram* (VPG) provides a detail on the spatio-temporal behavior of the trajectories. The VPG was originally designed for video data [5, 8] and principally related to space-time cubes known from geographic information systems. In the mobile data context, the VPG displays several instances of the map to provide context for the trajectories crossing the 3d space-time volume in real-time (Figure 2(d)). The VPG can be considered as interactive dynamic stream representation of the movement data. To cope with long movement durations, fast-forward playback of the movement data is also possible and can be steered by the users.

The timeline (Figure 2(e)) provides overview of the available data intervals (green and blue blocks), the different data sources loaded into our system (each row of blocks represents a dataset), and the current playback position when the VPG is used (red vertical line). Further, the users can navigate in time by scrolling and zooming the timeline and by constraining the time interval to be analyzed. Data outside the selected temporal range is depicted by faded colors.

Analysis is supported by linking the data representations of different views. The objects selected in one view (e.g., a group of trajectories) are highlighted in the other views, too. This way, coordinated views (cf. [12]) help to establish visual correspondence between the different data representations.

Besides aggregation, filters are the fundamental technique to reduce the amount of data displayed in our visual analytics system. This way, scalability with respect of Big Data is maintained. Furthermore, filters are also utilized to verify the hypotheses developed during the analysis process. In the context of MDC, we use four types of filters. A whitelist type of filter is automatically created in the gather step of the ISS view. Trajectories not within the selected clusters are excluded by this filter. In a similar way, users can generate a region of interest filter via the map view, which excludes all trajectories from further processing that are not crossing the currently chosen map section. A query by sketch filter allows finding trajectories that are similar to a prototypical trajectory drawn on the map by the users. The measure of similarity and the properties to be considered (e.g., position, direction, velocity) are defined by the users. Interactions between participants of the MDC could be detected by the relationship filter, which takes the similarity of multiple trajectories into account that overlap in time. Using this filter, meetings of the participants could easily be discovered. Filters applied to the dataset are presented in the filter graph that is depicted in Figure 2(f). The filter graph provides overview on the applied filters and serves as analysis history. Users are able to jump back and forth in history to evaluate alternative analysis paths.

5. USAGE EXAMPLE

In this section we give two examples of mobile data analysis with our visual analytics system. The first example is dedicated to discover abnormal patterns of a single MDC participant, whereas in the second example we show how relations between multiple participants can be analyzed. Please note that due to space limitations, only brief overview can be provided. For more details we refer to the supplementary material.

In the first example we analyze the mobile data of one participant. Figure 3(a) gives a first impression of the paths the subject took in the area of Lausanne. As first analysis step, we determine the main routes taken by the subject. The ISS view in Figure 3(b) exhibits the three major characteristics of the subject's movement. The red cluster shows mainly trajectories that have their origin or destination within the dashed circle. The blue cluster includes movement between Saint-Sulpice and a region near the Lausanne train station. Trajectories that move towards the harbor (red circle) or come from the harbor are bundled within the green cluster. Please note that we only depict a representative excerpt of the trajectories within the clusters to reduce visual clutter and overdrawing. After switching the similarity measure from mean position to positional variation, we re-cluster the whole dataset to get another impression of the participant's movements. The dataset splits into clusters of different movement radius (Figure 3(c)). We discover that the red cluster, which includes trajectories with small movement radius (the main spots are highlighted with red circles, since the screenshot does not convey the interactive highlighting of our system), also covers trajectories from the harbor area. Since we found in the last step, the harbor area to be one of the main destinations of the participant, we decide to have a closer look into the movement related with this region. The subject's movement profile exhibits that almost

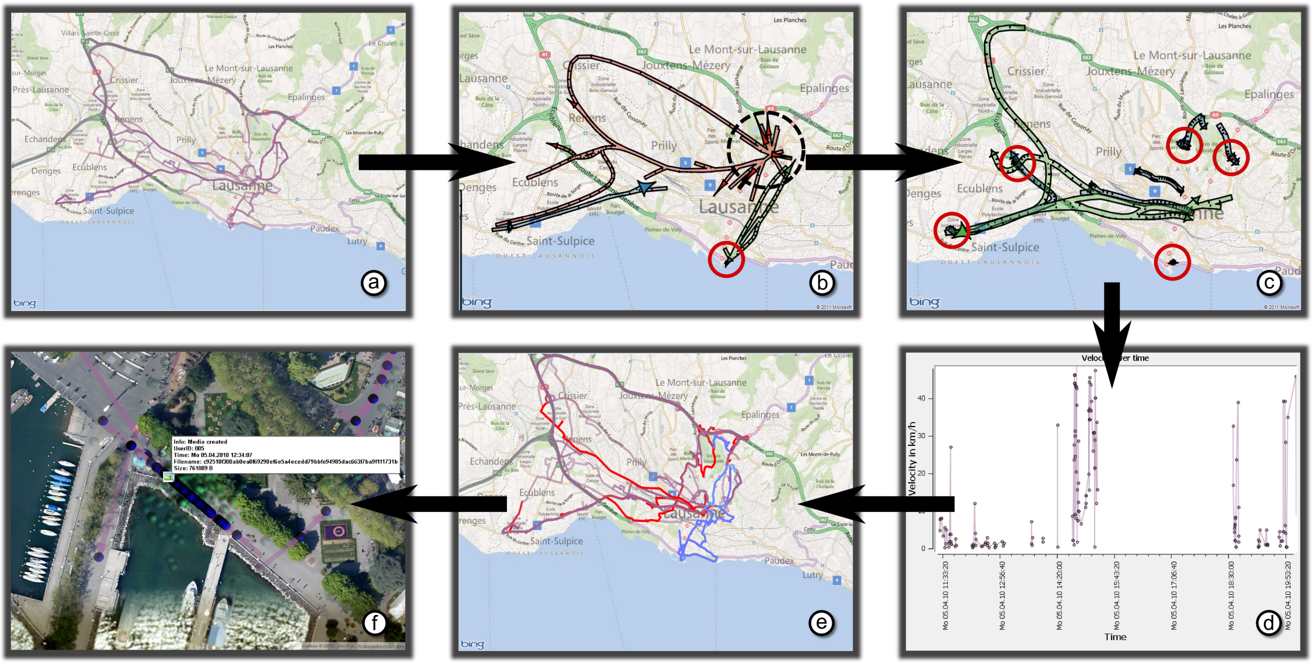

Figure 3: First usage example including the analysis of the mobile data of one MDC participant. Details are provided in the text.

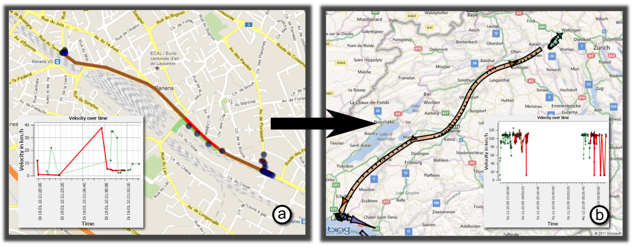

Figure 4: Interaction analysis of two MDC participants. Details are provided in the text.

all trajectories that access the harbor, stem from a single day (cf. Figure 3(d)). The velocity profile of the trajectories suggests that the subject went to the harbor several times by a motorized vehicle and had a walk at the harbor. Since this is untypical movement behavior of the subject (cf. Figure 3(e): typical movement at Mondays in red, the untypical movement at the mentioned day in blue), we look up the date and find out it is a public holiday in Lausanne. Detail information on the movement and mobile data enriched by raw GPS positions reveal that the person went to the Ouchy park and also took a photo there (see Figure 3(f)). Based on the information provided by the questionnaire and the place the photo was taken (next to the carousel) we assume the subject was accompanied by children.

To detect relations within the MDC participant group, we decide to filter interactions of the whole dataset. After we applied a relationship filter 26 meetings (i.e. subjects that are not farther than 15 meters away from each other, while their trajectories overlap in time) could be detected out of 13390 trajectories. We pick one of those related trajectories of two participants for further analysis (cf. Figure 4(a)). Based on the velocity profile of their movement and their common route, we assume that both walk together and may know each other. Further analysis shows that both subjects visited the *École Polytechnique Fédérale de Lausanne*. We assume the subjects study together and search for further indication of their relationship. Applying the ISS (clustering by position) on the dataset, it shows that the set of common trajectories of the subjects include a trip to a city near Zürich (see Figure 4(b)). Google's "What's here" service exhibits that their common destination is a concert of a foreign singer. However, there is a temporal shift between their particular arrival and departure of almost exactly one hour (cf. Figure 4(b)). A closer look on the GPS data reveals that the subject changed the time zone a day before both presumably drove together to the concert. Since this happened about two weeks before official clock change, it might be caused by accident, or in order to obfuscate the movement data.

6. CONCLUSION AND DISCUSSION

We presented a visual analytics framework for mobile data that enables its users to discover unknown patterns in a scalable way by combining automatic analysis with the recognition capabilities of human experts. With the help of two usage examples we were able to briefly demonstrate the analysis capabilities of our visual analytics approach. However, it becomes obvious from the two examples that visual analysis of individuals might be a threat to privacy. Although, there are efforts to investigate and establish privacy safeguards to location-based data analysis [6, 1, 2], such as the concept of k-anonymity or geographic masks, their applicability to user-centered visual analytics of individuals' movement data (in contrast to general movement patterns of groups) remains an open research question. Future work will focus on the question how automatic pattern recognition approaches that go beyond clustering can be incorporated into such a visual analytics system.

7. ACKNOWLEDGMENTS

This work was funded by German Research Foundation (DFG) as part of the Priority Program "Scalable Visual Analytics" (SPP 1335).

8. REFERENCES

[1] O. Abul, F. Bonchi, and M. Nanni. Never walk alone: Uncertainty for anonymity in moving objects databases. In *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on*, pages 376–385. IEEE, 2008.

[2] G. Andrienko, N. Andrienko, U. Demsar, D. Dransch, J. Dykes, S. Fabrikant, M. Jern, M. Kraak, H. Schumann, and C. Tominski. Space, time and visual analytics. *International Journal of Geographical Information Science*, 24(10):1577–1600, 2010.

[3] G. Andrienko, N. Andrienko, I. Kopanakis, A. Ligtenberg, and S. Wrobel. Visual analytics methods for movement data. In D. P. Fosca Giannotti, editor, *Mobility, Data Mining and Privacy*, Geographic Knowledge Discovery, chapter 13, pages 375–408. Springer-Verlag Berlin, Heidelberg, 2008.

[4] G. Andrienko, N. Andrienko, and S. Wrobel. Visual analytics tools for analysis of movement data. *ACM SIGKDD Explorations Newsletter*, 9(2):38–46, 2007.

[5] R. Botchen, S. Bachthaler, F. Schick, M. Chen, G. Mori, D. Weiskopf, and T. Ertl. Action-based multifield video visualization. *IEEE Trans. on Visualization and Computer Graphics*, 14(4):885–899, 2008.

[6] S. Bridwell. The dimensions of locational privacy. *Societies and cities in the age of instant access*, 88(5):209–225, 2007.

[7] M. Höferlin, B. Höferlin, D. Weiskopf, and G. Heidemann. Interactive schematic summaries for exploration of surveillance video. In *Proc. Int. Conf. on Multimedia Retrieval*, pages 9:1–9:8. ACM, 2011.

[8] M. Höferlin, B. Höferlin, D. Weiskopf, and G. Heidemann. Uncertainty-aware video visual analytics of tracked moving objects. *Journal of Spatial Information Science (JOSIS)*, 2:87–117, 2011.

[9] D. Keim, F. Mansmann, J. Schneidewind, J. Thomas, and H. Ziegler. Visual analytics: Scope and challenges. *Lecture Notes in Computer Science*, 4404:76–90, 2008.

[10] J. K. Laurila, D. Gatica-Perez, I. Aad, J. Blom, O. Bornet, T.-M.-T. Do, O. Dousse, J. Eberle, and M. Miettinen. The mobile data challenge: Big data for mobile computing research. In *Proc. of Mobile Data Challenge by Nokia Workshop, in conjunction with Int. Conf. on Pervasive Computing*, 2012.

[11] D. Mountain. Visualizing, querying and summarizing individual spatio-temporal behaviour. In M.-J. K. Jason Dykes, Alan M. MacEachren, editor, *Exploring Geovisualization*, chapter 9, pages 181–200. Elsevier, London, 2005.

[12] J. Roberts. State of the art: Coordinated & multiple views in exploratory visualization. In *Proc. of Int. Conf. on Coordinated and Multiple Views in Exploratory Visualization*, pages 61–71. IEEE, 2007.

[13] Z. Shen and K. Ma. Mobivis: A visualization system for exploring mobile data. In *Visualization Symposium, 2008. PacificVIS'08. IEEE Pacific*, pages 175–182. IEEE, 2008.

[14] B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. In *Proc. of Symposium on Visual Languages*, pages 336–343, 1996.

[15] J. Thomas and K. Cook. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society, 2005.