Implementing Visual Analytics Methods for Massive Collections of Movement Data

G. Andrienko, N. Andrienko

Fraunhofer Institute Intelligent Analysis and Information Systems
Schloss Birlinghoven, D-53754 Sankt Augustin, Germany
Telephone: +49 (0)2241 142486
Fax: +49 (0)2241 142072
Email: gennady.andrienko@iais.fraunhofer.de

1. Introduction

Exploration and analysis of large spatio-temporal datasets requires a synergetic work of humans and computers where the power of computational techniques is complemented with and directed by the human's background knowledge, flexible thinking, imagination, and capacity for insight. Visualisation and interactive visual interfaces are essential for a proper involvement of humans in problem solving. Hence, there is a need in visual analytics (Thomas and Cook 2005) environments integrating visualisation with database technologies, computerised data processing, and computational techniques.

In a EU-funded project GeoPKDD - Geographic Privacy-aware Knowledge Discovery and Delivery (IST-6FP-014915; see http://www.geopkdd.eu) - we develop visual analytics methods for analysis of massive collections of data about multiple discrete entities changing their spatial positions over time while preserving their integrity and identity (i.e. the entities do not split or merge). Such data will be henceforth called "movement data". Recently, we have developed a theoretical basis for the creation of methods for analysis of movement data (Andrienko and Andrienko 2007). In particular, we have defined the possible types of patterns that can be detected inside movement data and between movement data and data about other phenomena. Next, we have envisaged the kinds of data transformations, computations, and visualisations that could enable a human analyst to detect these pattern types in truly massive data, possibly, not fitting in a computer's memory. On the basis of the previous works (Tobler 1987; Dykes and Mountain 2003; Laube, Imfeld, and Weibel 2005; and others), we have suggested a set of techniques where a key role belongs to aggregation and summarisation of the data by means of database operations and/or computational techniques.

For a practical verification of this choice of techniques resulting from a theoretical analysis, we started a prototype implementation of a visual analytics toolkit for movement data. We present here some of the recently developed interactive tools combining visualisation with out-of-memory data processing, which includes spatial, temporal, and spatio-temporal aggregation. The presentation is preceded with a brief definition of movement data and related concepts.

2. Movement Data

Movement data can be conceptualised as a continuous mathematical function matching pairs entity + time moment with positions in space. This is an abstraction from real data, which are always discrete, i.e. contain positions only for some of the possible pairs. From positions of entities at different time moments other movement characteristics can be derived: speed, direction, acceleration (change of the speed), turn (change of the direction), etc.

The changes of the position and the other movement characteristics of an entity over time form the *individual movement behaviour* (IMB) of this entity, where *behaviour* is a synoptic concept differing from just a sequence of values of the characteristics attained at all time moments (see Andrienko and Andrienko 2006). An IMB has its own characteristics such as the path, or trajectory, travelled by the entity in the space, the travelled distance, the movement vector (direction from the initial to the final position), and the variation of speed and direction along the path.

A unity of the movement characteristics of a set of entities at some single time moment can be called *momentary collective behaviour* (MCB) of this set of entities. A MCB has such synoptic characteristics as the distribution of the entities in the space, the spatial variation of the derivative movement characteristics, and the statistical distribution of the derivative characteristics over the set of entities.

The concept corresponding to a holistic view of the movement characteristics of multiple entities over a time period (i.e. multiple time moments) can be called *dynamic collective behaviour* (DCB). We assume that the goal of the analysis of movement data is to describe, in a parsimonious way, the DCB of all entities during the whole time period the data refer to and to and relate them to properties of space, properties of time, properties and activities of the moving entities, and relevant external phenomena. For a valid and comprehensive analysis, a DCB should be considered from two different perspectives:

- as a construct formed from the IMBs of all entities, i.e. as the variation of IMB over the set of entities;
- as a construct formed from the MCBs at all time moments, i.e. the variation of MCB over time.

The visual analytics techniques enabling these complementary views are introduced in sections 4 and 5, respectively. Prior to that, we briefly describe how the original movement data are pre-processed in the database to become suitable for the analysis.

3. Data Preparation

Originally, movement data come as a set of records with the structure <*e_id, time, lat, lon*>, where *e_id* is the identifier of an entity, *time* is the time moment of the measurement, *lat* and *lon* are the latitude and longitude, respectively, of the geographical position of the entity measured at the specified time moment. The data are stored in a relational database. First, we use database functions to link the records of each entity into a temporally ordered sequence. Practically, this means that new fields are added to each record: the time of the next record referring to the same entity, the distance (difference) in time between the two measurements, and the spatial distance between the respective positions.

A temporally ordered sequence of all positions of an entity is not a meaningful object for analysis since the entity does not necessarily moves all the time. It is reasonable to divide the sequence into trips or movement episodes, which describe the movements from one stop to another. Since the original data do not contain explicit information about stops, stops can be defined on the basis of the temporal and spatial distances between the measurements: whenever two measurements are close in space but distant in time, it may be assumed that there is a stop between them. However, there are lots of possibilities for defining what is "close in space" and what is "distant in time". Therefore, we have implemented an interactive tool for splitting movement data into episodes where the user can specify different thresholds and divide the same data in different ways (Figure 1). Thus, when we have data about daily movements of people, choosing the threshold of 2 hours for the temporal distance will yield trips from home to work and back and round trips (e.g. for shopping) starting from home and ending at home. The threshold of 5 minutes will give us the trips to shops, post offices, etc., and the threshold of 30 seconds will result in getting small-scale moves between road crossings. Hence, the user gets an opportunity to analyse movement data at different scales. The user may also apply some other kinds of division rules; for instance, the data can be divided into moves between pre-specified regions of interest.

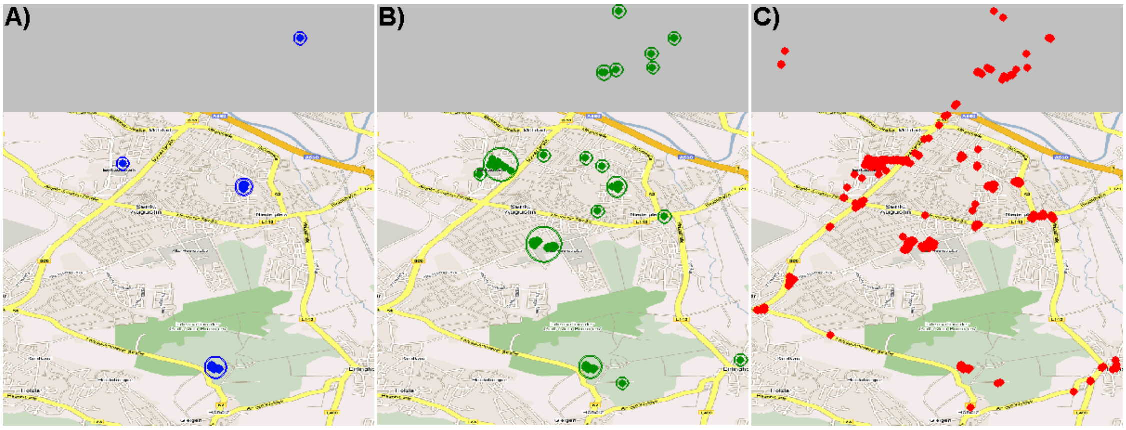

Figure 1. Stops retrieved from data describing long-term movement of a car. The set of stops depends on the choice of the temporal threshold: 2 hours (A), 5 minutes (B), and 30 seconds (C). In A) and B), the circles built around the stop positions indicate the probable regions of interest of the car driver.

4. Variation of Individual Movement Behaviour

In (Andrienko and Andrienko 2007), we argue that clustering techniques should be applied to explore the variation of IMB over the set of entities when the set of movement data is large. Clustering is applied to movement episodes obtained as described in the previous section. To visualise the results of the clustering, it may be inappropriate to represent the original movement episodes by graphical symbols (such as lines on a map) coloured according to the clusters the episodes belong to since the number of such symbols and the number of intersections and overlaps between them may be very large. Therefore, generalisation and aggregation has to be applied to results of clustering. It is also meaningful to apply generalisation and aggregation to the whole set of movement episodes. The basic idea is that, firstly, a finite set of "places" is defined, secondly, the episodes are represented as sequences of moves between these "places", thirdly, the moves with the same start and end "places" are grouped together, and, finally, various aggregate characteristics are computed for the groups: number of moves, minimum, maximum, and median speed, duration, etc. The "places" may be user-specified regions of interest and/or areas around stops (such as in Figure 1A and B). The results of the generalisation and aggregation may be represented as vectors (arrows) on an animated map (Figures 2 and 3) or in a space-time cube where the visual attributes of the vectors (thickness, colour, transparency) encode some of the aggregate characteristics. Another useful visualisation is a movement matrix, which may also be animated (Figure 3), where the rows and columns correspond to the "places" and graphical symbols in the cells portray selected aggregate characteristics.

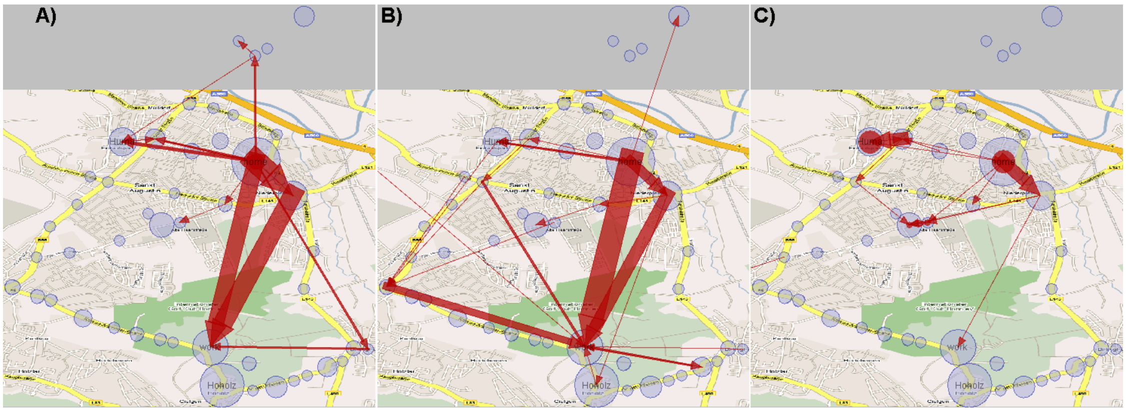

Figure 2. An animated map portrays aggregated moves between user-specified places as arrow symbols connecting the places. The thickness of an arrow is proportional to the number of moves that occurred between the respective places during the time interval selected by means of the time slider Figure 3). The screenshots A, B, and C correspond to three successive 1-hour intervals.

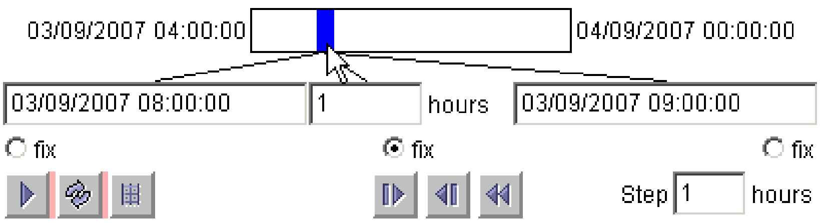

Figure 3. A time slider is used to select time intervals for time-related data displays, such as the map in Figure 2 and movement matrix in Figure 4, and to run display animation.

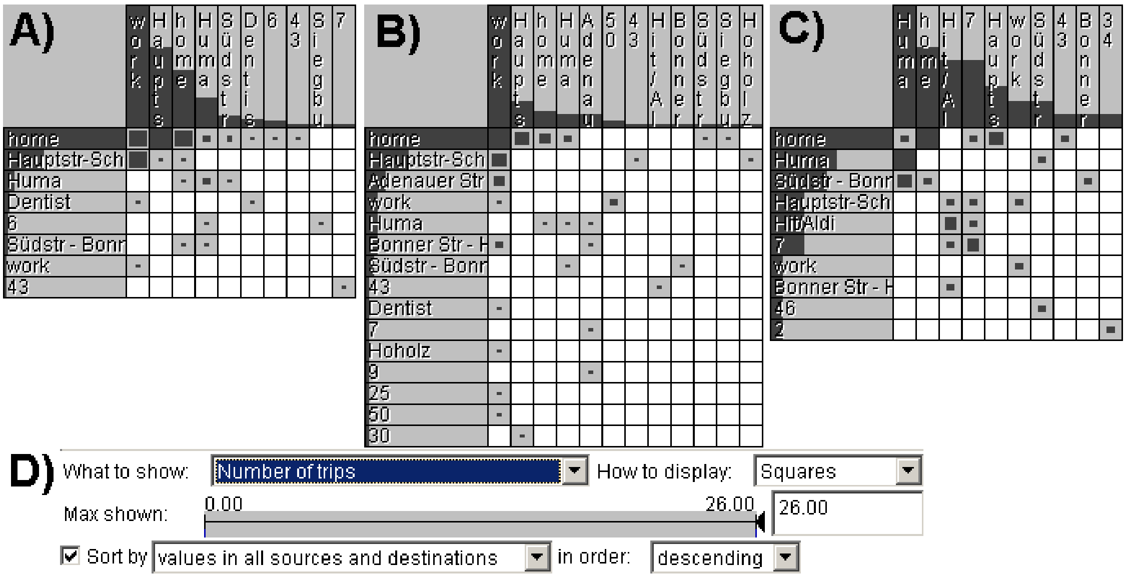

Figure 4. An animated movement matrix represents the same data as the map in Figure 2 and is controlled by the same time slider (Figure 3). The section D presents the interactive controls of the matrix.

5. Variation of Momentary Collective Behaviour

We suggest a set of interactive tools that supports the following scenario of the exploration of the variation of MCB over time:

1. Discretise the time by dividing it into appropriate intervals; the user can try out various divisions.
2. Discretise the space by building a regular rectangular grid; the user can try out grids with various origins and cell sizes.
3. For each grid cell and time interval, compute aggregated characteristics: number of data records, number of different entities, statistics of the speeds, statistics of the time per entity, etc.
4. Select the aggregate characteristics one by one and visualise them on an animated map by colouring or shading of the grid cells or by graduated symbols drawn inside the cells (Figure 5).

The computations are done in the database.

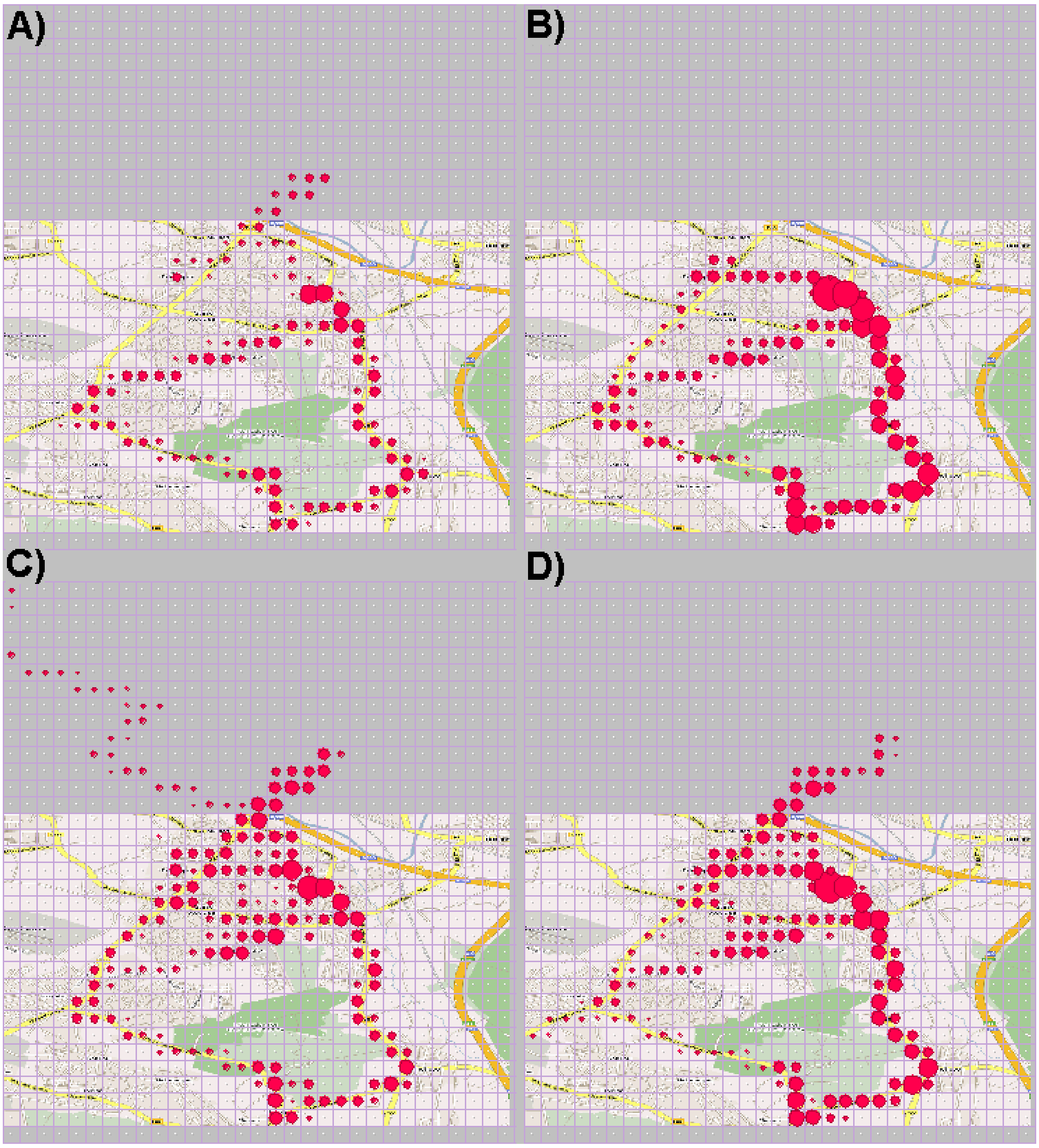

Figure 5. Car tracking data have been aggregated spatially by cells of a regular grid and temporally by user-specified time intervals (e.g. intervals of equal length). The screenshots A, B, C, and D have been produced from an animated map where the numbers of records fitting in the cell and in the corresponding intervals are represented by graduated symbols.

6. Conclusion

We have presented the status of an ongoing work where much is still to be done. The chosen approaches show promise; however, intensive experiments will be required when the conceived toolkit is ready.

7. References

Andrienko N and Andrienko G, 2006, *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*, Springer, Berlin, Germany.

Andrienko N and Andrienko G, 2007, Designing Visual Analytics Methods for Massive Collections of Movement Data, *Cartographica*, 42(2):117-138.

Dykes JA and Mountain DM, 2003, Seeking structure in records of spatio-temporal behaviour: visualization issues, efforts and applications, *Computational Statistics and Data Analysis*, 43 (Data Visualization II Special Edition):581-603

Laube P, Imfeld S, and Weibel R, 2005, Discovering relative motion patterns in groups of moving point objects, *International Journal of Geographical Information Science*, 19(6):639–668

Thomas JJ and Cook KA (eds), 2005, *Illuminating the Path. The Research and development Agenda for Visual Analytics*, IEEE Computer Society, USA

Tobler W, 1987, Experiments in migration mapping by computer, *The American Cartographer*, 14 (2): 155-163