Exploration of Massive Movement Data: a Visual Analytics Approach

Gennady Andrienko, Natalia Andrienko
Fraunhofer Institute IAIS (Intelligent Analysis and Information Systems)
Schloss Birlinghoven; 53754 Sankt Augustin, Germany

ABSTRACT

To make sense from large amounts of movement data (sequences of positions of moving objects), a human analyst needs interactive visual displays enhanced with database operations and methods of computational analysis. We present a toolkit for analysis of movement data that enables a synergistic use of the three types of techniques.

INTRODUCTION

Thanks to the recent advent of inexpensive positioning technologies, data about movement of various mobile objects are collected in rapidly growing amounts. Potentially, these data are a source of valuable knowledge about behavioral and mobility patterns. To gain an understanding of these patterns, an analyst needs a visual representation of the data, which is the most effective way to support human perception, cognition, and reasoning. However, purely visual methods of analysis (e.g. Hägerstrand 1970), even being enhanced with interactive techniques (Andrienko et al. 2000, Kraak 2003, Kapler and Wright 2005), are not scalable to large datasets. Such methods need to be combined with database operations and computational analysis techniques helping to handle large amounts of data. Some approaches have been suggested recently. Forer and Huisman (2000) and Dykes and Mountain (2003) summarize movement data into surfaces, but this is not suitable for analyzing routes. Buliung and Kanaroglou (2004) envelop bunches of trajectories and compute the central tendency, which works well for similar and close trajectories. Laube et al. (2000) combine visualization with data mining methods oriented to specific types of patterns.

In our earlier work (Andrienko and Andrienko 2007), we suggested a theoretical model of movement data and, on this basis, defined the possible types of movement patterns and the techniques that could enable the detection of these patterns. In (Andrienko et al. 2007) we have presented a framework and a toolkit for analysis of movement data based on a synergy of visualization, database operations and computations. Here, we focus on the visual and interactive components of the toolkit.

MAKING SENSE FROM POSITION SEQUENCES

Movement data acquired by position tracking usually lack any semantics. The records basically consist of time stamps and coordinates. In particular, there are no explicitly defined trips with specified origins and destinations and no semantically identifiable places. To understand the data, an analyst should be able to link them to his/her prior knowledge and interpretable information from other sources. Visualization is essential for this purpose.

Finding significant places

One important task in analysis of movement data is to extract and interpret the places of stops. Our toolkit supports this task in the following way. First, the positions of stops with user-specified minimum duration are extracted from the database. Second, a spatial clustering tool is applied to find groups of spatially close positions, which indicate repeatedly visited places. Third, the results are shown on a map where the positions are marked by colored point symbols (each cluster receives a unique color). The map provides the spatial context and thereby helps the analyst to interpret the places. Additional help may come from temporal histograms showing the distribution of the stops within temporal cycles (daily, weekly, etc.). Thus, the histograms in Figure 1 show the frequencies of stops of a personal car for minimum 3 hours by days of the week (A) and by hours of the day (B). The colored bar segments represent the results of the spatial clustering of the stops. It is vividly seen that the stops of the "blue" cluster occur only on the working days and mostly in the morning times. The stops of the "red" cluster occur all days and mostly in the evenings. A plausible conclusion is that the "blue" cluster of positions is situated near the place where the person works and the "red" cluster is near person's home.

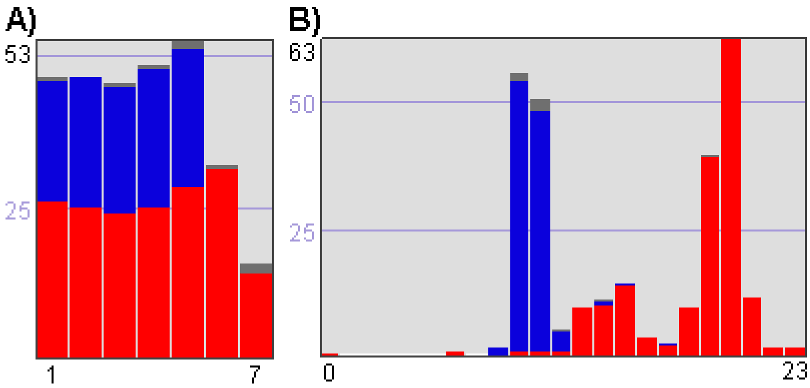

Figure 1: temporal histograms show results of clustering of stop positions

Extracting trips and exploring the routes

The sequence of position records representing the movement of an entity needs to be partitioned into subsequences corresponding to trips. The notion of trip may be application- and goal-dependent. Our toolkit allows the users to divide data in several ways: by stops, by spatial gaps, by temporal cycles, and by places of interest. The division is done by means of database operations. After that, the data may be loaded into the visual analysis system where each subsequence forms a single object further called trip or trajectory. The trips may be directly displayed on a map (as lines with specially marked starts and ends) but intersections and overlaps between the trips make the display illegible. Hence, this way of presenting trips can only be used in combination with interactive filtering tools, in particular, time filter (Figure 2). The user selects a time interval, and all displays in the system show only data from this interval. When a trip does not fully fit in the selected interval, the appropriate part is shown. The user may choose a convenient temporal granularity, which may range from seconds to years depending on the time span of the data. The time filter can also be used as a device for display animation: the user can drag the slider (blue bar) representing the current time interval, or use the buttons. Besides the time filter, there is an interactive attribute filter, which selects trip subsets on the basis of values of various attributes: trip duration, moving entity, start and/or end time, day of the week, hour of the day, etc.

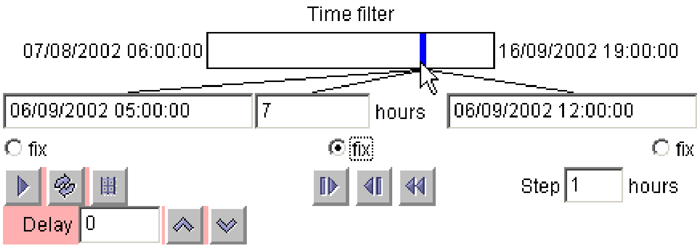

Figure 2: the interface of the interactive time filter

The display of individual trips combined with interactive filtering does not support an overall view of the whole set of trips and comparison between subsets of trips. A suitable approach is based on clustering, i.e. grouping trips by similarity and consideration of the groups. Trips may be similar in different respects: they may fully or partly coincide in space, or just have similar shapes, or have common starts and/or ends; they may be fully or partly synchronous or disjoint in time but have similar dynamic behaviors. It depends on the application and goals of analysis which of these respects are relevant. Therefore, we have implemented a clustering tool that allows the analyst to choose an appropriate similarity measure (also called distance function) from a number of alternatives. Our toolkit provides a range of distance functions and allows extension with new functions. An important property of the clustering tool is that it builds clusters from the data subset satisfying current filters rather than the whole set. This property can be utilized, in particular, for the kind of analysis that may be called "progressive clustering": the analyst selects one or a few clusters and refines them by re-applying the clustering tool with a different distance function or different parameter settings. For instance, the trips of a private car may be first clustered according to the closeness of their starts and ends. Then, the cluster of trips from work to home can be selected and further clustered according to the route similarity in order to detect typical routes from work to home (Figure 3) as well as atypical routes ("noise", which is not shown in Figure 3).

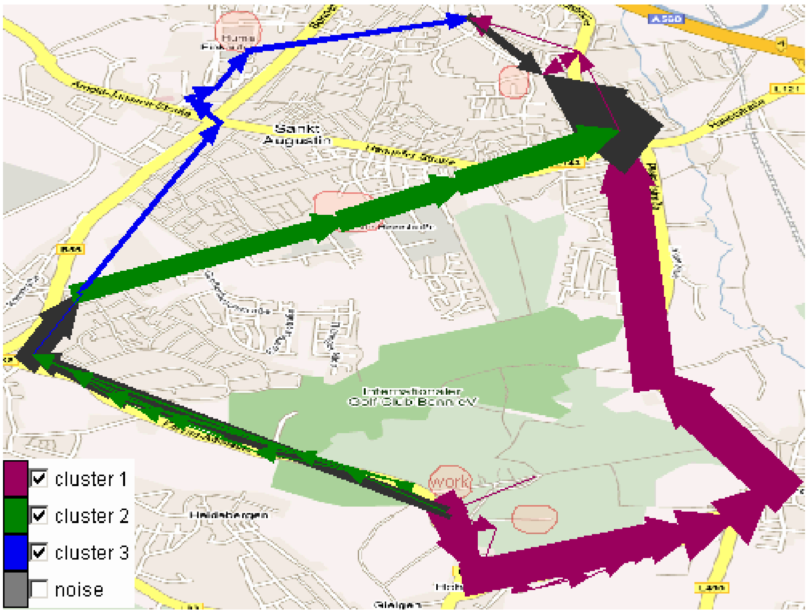

Figure 3: three clusters of trips are represented in a summarized form on a map display

To represent massive movements, as in Figure 3, we apply computational summarization. The algorithm is described in (Andrienko et al. 2007). It comprises three major steps: extraction of characteristic points of the trajectories, i.e. starts, ends, stops, and turns; generalization from points to areas by building circles around the points; and collecting moves (fragments of the trajectories) between pairs of circles into aggregate moves, from which a new map layer is built. The aggregate moves are represented by arrows (vectors) showing the movement directions; the thickness is proportional to the number of moves. Each aggregate move keeps references to the original moves it unites and dynamically reacts to any filter applied to the trajectories: time filter, attribute filter, or cluster filter (bottom left of Figure 3). When none of the original moves satisfies current filter conditions, the aggregate move does not appear on the screen; otherwise, the thickness of the vector is adjusted to the number of the filter-compliant original moves. If all these moves belong to trajectories of the same cluster, the vector is colored in the color of the cluster; otherwise, it is shown in gray.

Exploring movement dynamics

3D views where two dimensions represent space and one represents time (Hägerstrand 1970) are good for exploring the speed of movement and its variation over time. This approach can be effective even for examining multiple trajectories if they do not intersect (in particular, if they follow the same route). One of the distance functions in our toolkit groups trajectories by similarity of the routes and similar dynamics of the movement. Selected clusters can be explored and compared in a 3D view as shown in Figure 4, which is quite legible despite the number of trajectories displayed.

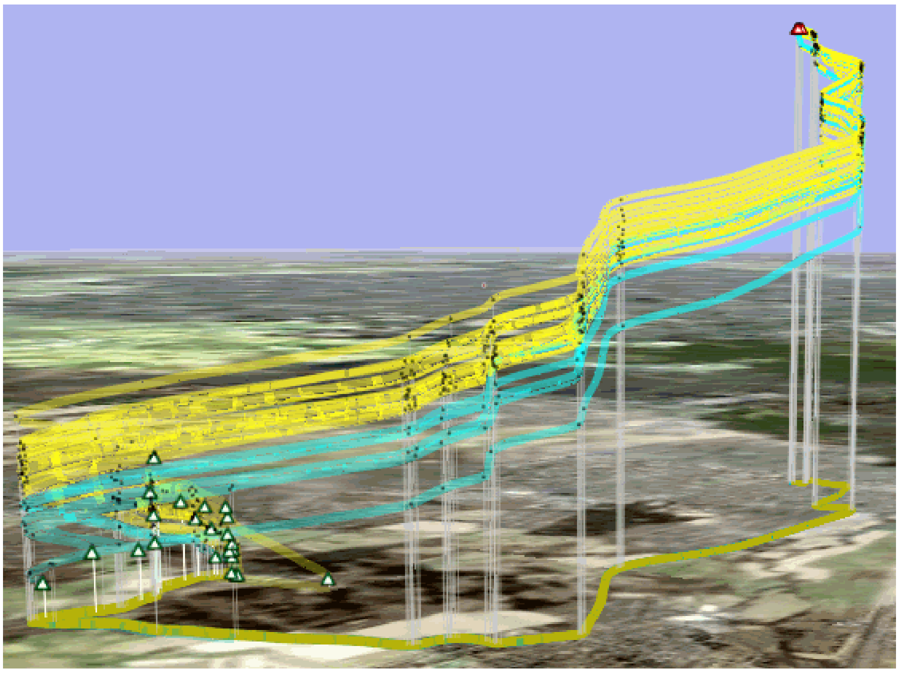

Figure 4: two clusters of trips follow the same route but differ in the dynamics

Here, the vertical dimension represents the time relative to the ends of the trips: the ends of the trips have the same vertical position and the remaining points of each line are shifted down in relation to this position proportionally to the temporal distance to the end of the trip. The lines in yellow and cyan represent the trips of two clusters. On the left and in the centre of the image, the vertical positions of the yellow lines are higher than those of the cyan-blue lines. This means that the trips of the first cluster were shorter in time than the trips of the second cluster. The cyan-blue lines are steeper than the yellow ones, which signifies that the speeds in the second cluster of trips were lower than in the first cluster.

Figure 5 demonstrates another way of looking at the speeds of movement: the heights of the points above the surface are proportional to the speeds of movement in these points. Again, the legibility of the display owes to previous clustering and selection of one cluster of trips following the same route.

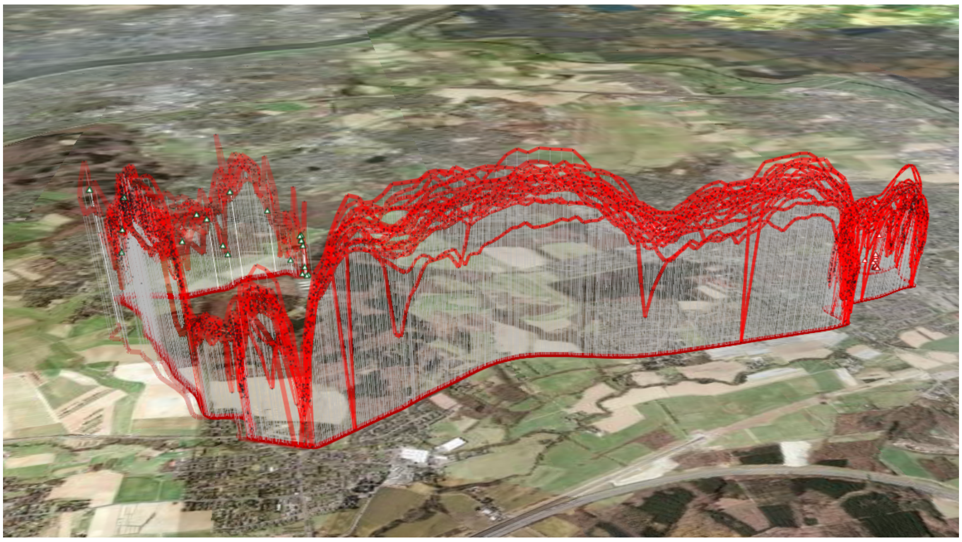

Figure 5: the vertical positions of the points of the lines represent the speeds

CONCLUSION

Interactive visual displays play the key role in supporting sense-making from movement data but are insufficient when the data are large. Our framework combines visualization with database operations and computations so that the techniques are complementary and mutually reinforcing. The generic database techniques enable handling large datasets and are used for basic data processing and extraction of relevant objects and features. The computational techniques, which are specially devised for movement data, aggregate and summarize these objects and features and thereby enable the visualization of large amounts of information. The visualization enables human cognition and reasoning, which, in turn, direct and control the further analysis by means of the database, computational, and visual techniques.

The reported work has been partly funded by EU in the project GeoPKDD - Geographic Privacy-aware Knowledge Discovery and Delivery (IST-6FP-014915; http://www.geopkdd.eu).

BIBLIOGRAPHY

Andrienko, N., and Andrienko, G. 2007. Designing visual analytics methods for massive collections of movement data. Cartographica, v.42 (2), 117-138

Andrienko, G., Andrienko, N., Wrobel, S. 2007. Visual Analytics Tools for Analysis of Movement Data. ACM SIGKDD Explorations, 9 (2), (in press)

Andrienko, N., Andrienko, G., Gatalsky, P. 2000. Supporting Visual Exploration of Object Movement. In Proc. Working Conf. Advanced Visual Interfaces AVI 2000 (Palermo, Italy, May 2000), ACM Press, 217-220, 315

Buliung, R.N., Kanaroglou, P.S. 2004. An Exploratory Data Analysis (ESDA) toolkit for the analysis of activity/travel data. Proceedings of ICCSA 2004, LNCS 3044, Springer, Berlin, 1016-1025

Dykes, J. A., Mountain, D. M. 2003. Seeking structure in records of spatio-temporal behaviour: visualization issues, efforts and applications, Computational Statistics and Data Analysis, 43, 581-603

Forer, P., Huisman, O. 2000. Space, Time and Sequencing: Substitution at the Physical/Virtual Interface. In Information, Place and Cyberspace: Issues in Accessibility (Eds: Janelle, D.G., Hodge, D.C.), Springer, Berlin, 73-90

Hägerstrand, T. 1970. What about people in regional science? Papers of the Regional Science Association, 24, 7-21

Kapler, T., Wright, W. 2005. GeoTime information visualization, Information Visualization, 4(2), 136-146

Kraak, M.-J. 2003. The space-time cube revisited from a geovisualization perspective. In Proc. 21st Int. Cartographic Conf. (Durban, South Africa, Aug. 2003), 1988-1995

Laube, P., Imfeld, S., Weibel, R. 2005. Discovering relative motion patterns in groups of moving point objects. Int. J. Geographical Information Science, 19(6), 639–6