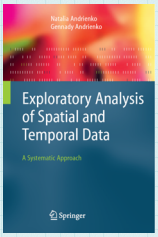
Scalable Visual Analytics Methods for Massive Collections of Movement Data

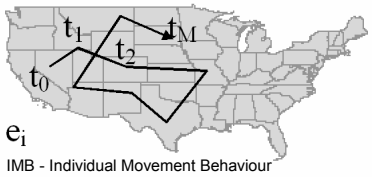
1. Theory



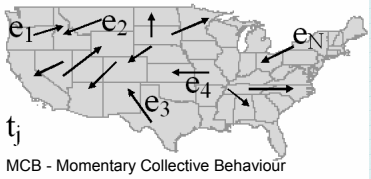
General theory:
Natalia Andrienko and Gennady Andrienko,
"Exploratory Analysis of Spatial and Temporal Data. A Systematic Approach", Springer, December 2005
(functional data model => definition of patterns => appropriate data analysis tools)

Specialization for movement data:
Natalia Andrienko and Gennady Andrienko,
"Designing visual analytics methods for massive collections of movement data", *Cartographica*, 2007, v.42 (2)
(types of patterns in movement data => appropriate visual analytics methods)

Goal of analysis: describe the Dynamic Collective Behaviour (DCB) by appropriate patterns.
A DCB has two aspects:
1) Individual movement behaviours of all entities; 2) Momentary collective behaviours in all time moments



IMB - Individual Movement Behaviour



MCB - Momentary Collective Behaviour

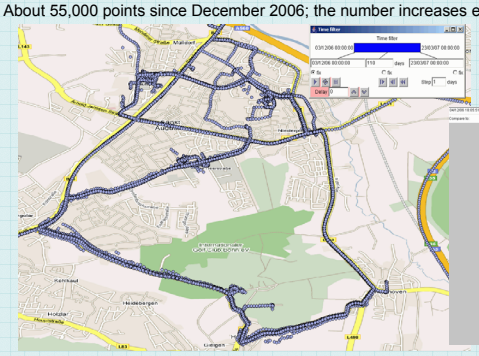
Factors that influence movement characteristics:
- Properties of space (characteristics of terrain, accessibility, objects and their functions, way of use)
- Properties of time (temporal cycles, daylight, typical activities)
- Properties and activities of the moving entities (individual properties, way and means of movement, purposes and causes, activities)
- Related spatial, temporal, and spatio-temporal phenomena

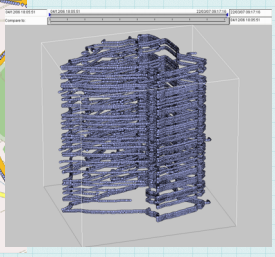
Types of patterns in movement data:
Descriptive patterns:
- Variation of IMB over the set of entities: similarity of characteristics, co-location in space, synchronization in time, co-incidence in space and time
- Variation of MCB over time: constancy, change, trend, fluctuation, pattern change, repetition, periodicity, symmetry

Connectional patterns:
- correlation (co-occurrence of any characteristics); influence; structure

2. Experiment: Data Collection & Processing

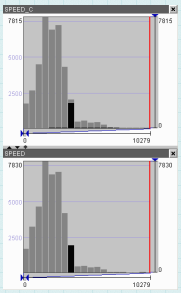
Collected data:
N;Time;Lat;Long;Height;Course;Speed;PDOP;State;NSat
…
13;22/03/07 08:52:09;50.777057;7.206522; 67.9;117.7;34.003;3.8;1808;4
14;22/03/07 08:52:12;50.776925;7.206858; 66.9;117.5;37.151;3.8;1808;4
15;22/03/07 08:52:15;50.776813;7.207263; 67.0;99.2;39.188;3.8;1808;4
…
About 55,000 points since December 2006; the number increases every day

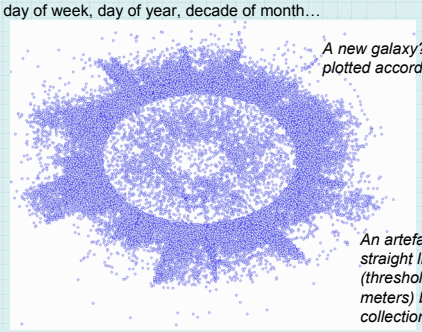




Cleaning and enrichment of the data: (computations in Oracle)
- Duplicates (id+time) are removed
- Points are connected into sequences
- Computed dX, dY, dT, distance
- Derived speed, course, acceleration and turn in each point
- Additional temporal components are easy to extract from the database: day of week, day of year, decade of month…

Measured speed Vs. computed:

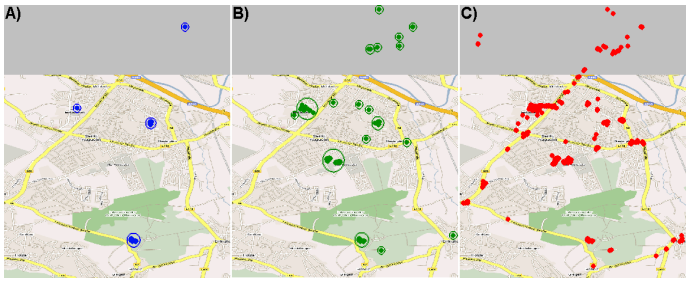



A new galaxy? (points are plotted according to dx & dy)

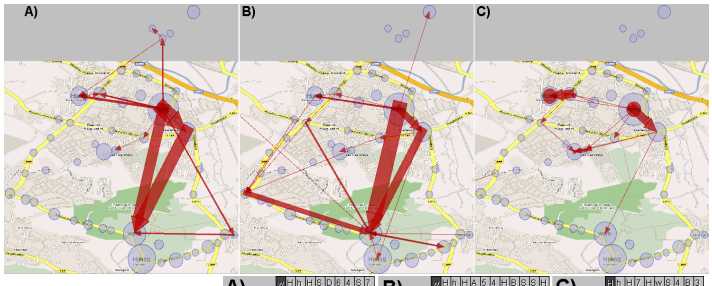
An artefact of the straight line filtering (threshold = 20 meters) by the data collection software

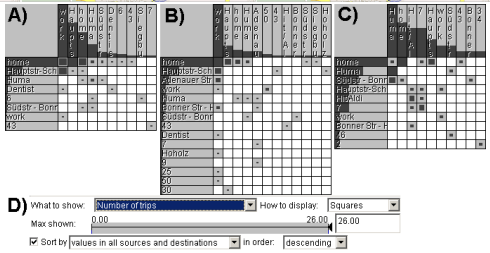
3. Visual Analytics Methods

Analysis (1):
- Interactive division of the point stream into trajectories or segments according to (a) temporal gap; (b) spatial gap; (c) characteristic intervals within time cycles
Example: starts and ends of trajectories defined according to temporal thresholds 2 hours (A), 5 minutes (B) and 30 seconds (C)

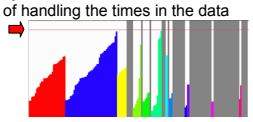


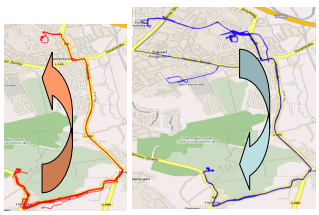
Analysis (2):
- Defining regions of interest interactively or by clustering of the stop locations;
- Aggregating trajectory segments by common starts and stops
Example: total numbers of trajectories between ROIs; the trajectories in A, B, C have been defined using 3 different values for the temporal gap





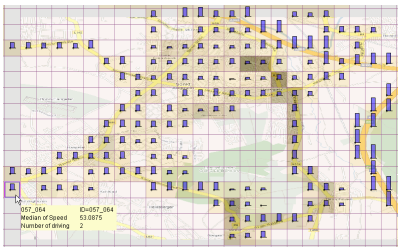
Analysis (3): **interactive clustering** (joint work with S.Rinzivillo, Univ. Pisa)
- generic clustering algorithm OPTICS (Ankerst, Breunig, Kriegel, and Sander 1999)
- the algorithm has been implemented so that cluster building is separated from distance and neighbourhood computation
- several variants of distance measures designed specially for trajectories
- various ways of handling the times in the data





Analysis (4):
-Spatial and temporal aggregation

For selected time interval and given spatial resolution the system computed median speed (shown by bars) and density of movements (shown by shading)



4. Further work

All proposed methods are designed to be scalable and applicable to much larger data sets!

We plan to:
1. Continue the development of out-of-memory methods for visual data analysis
2. Build a library of trajectory transformations and distance functions
3. Take into account speeds, directions, accelerations, turns, properties of space, time, and moving objects
4. Develop interactive client-server spatio-temporal aggregation
5. Experiment with different data sets

GeoPKDD - Geographic Privacy-Aware Knowledge Discovery
ICT 014915, 2005-2008
www.geopkdd.eu

Natalia Andrienko and Gennady Andrienko
Fraunhofer Institute IAIS
53754 Sankt Augustin, Germany
http://www.ais.fraunhofer.de/and

Fraunhofer IAIS Institut Intelligente Analyse- und Informationssysteme