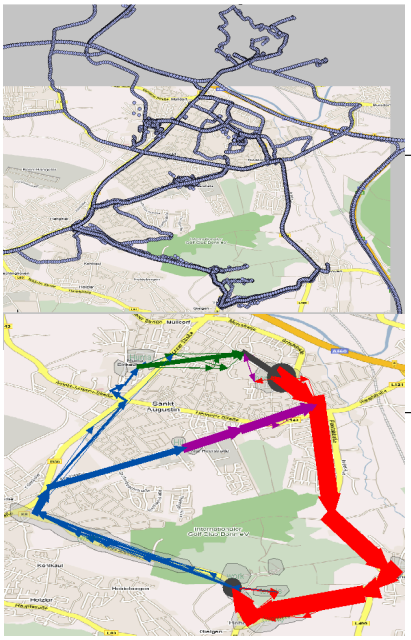
Visual Analytics of Movement Data



Fraunhofer Institut Intelligente Analyse- und Informationssysteme

Natalia Andrienko & Gennady Andrienko

www.ais.fraunhofer.de/and

www.geopkdd.eu

September 2007

Context

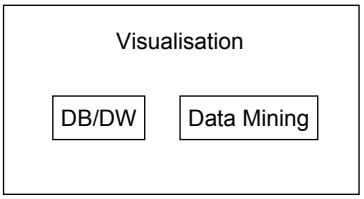
- EU-funded GeoPKDD project: Geographic Privacy Aware Knowledge Discovery and Delivery
- Goal: develop methods for analysis of movements of discrete objects

  { (object_id, time, position_in_space) }
- Consortium composition (disciplines): data mining + DB/DW + (geo)visualisation

Data miners' vision:

DB/DW → Data Mining → Visualisation

Our approach (Visual Analytics):

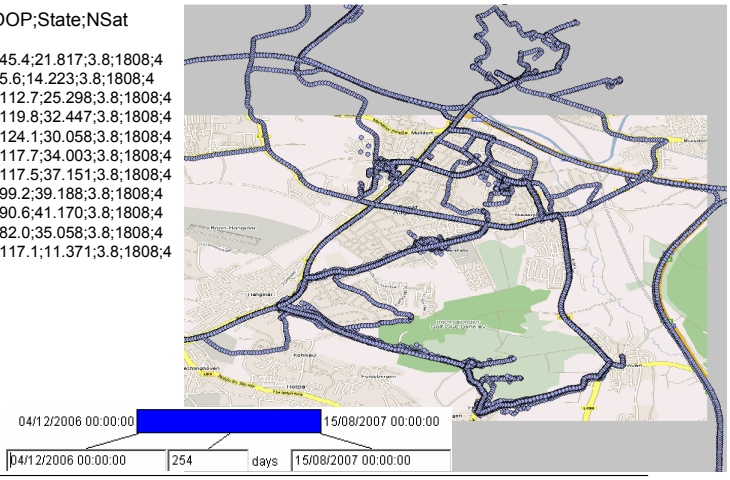


Data (example)

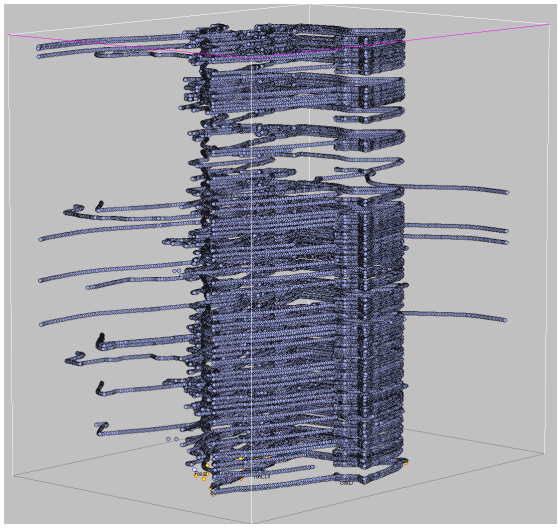
Positions of a private car from 03/12/06 till now

N;Time;Lat;Long;Height;Course;Speed;PDOP;State;NSat
…
8;22/03/07 08:51:52;50.777132;7.205580; 67.6;345.4;21.817;3.8;1808;4
9;22/03/07 08:51:56;50.777352;7.205435; 68.4;35.6;14.223;3.8;1808;4
10;22/03/07 08:51:59;50.777415;7.205543; 68.3;112.7;25.298;3.8;1808;4
11;22/03/07 08:52:03;50.777317;7.205877; 68.8;119.8;32.447;3.8;1808;4
12;22/03/07 08:52:06;50.777185;7.206202; 68.1;124.1;30.058;3.8;1808;4
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
16;22/03/07 08:52:18;50.776780;7.207745; 68.8;90.6;41.170;3.8;1808;4
17;22/03/07 08:52:21;50.776803;7.208262; 71.1;82.0;35.058;3.8;1808;4
18;22/03/07 08:52:24;50.776832;7.208682; 68.6;117.1;11.371;3.8;1808;4
…

About 90,000 positions



Data volume: about 90,000 time-referenced positions



Task (example)

- Investigate the movement behaviour of the car owner:
  - Detect and interpret typical trips: sources, destinations, routes, intermediate stops, purposes, …
  - Find out the typical times of the trips
  - If possible, interpret also atypical trips
- Problems:
  - no explicit trips in the data but just positions
  - no meaningful places but latitudes and longitudes

⇒ Places and trips have to be defined through data analysis!

✓ *We know that the positions were recorded only while the car moved. We can use this!*

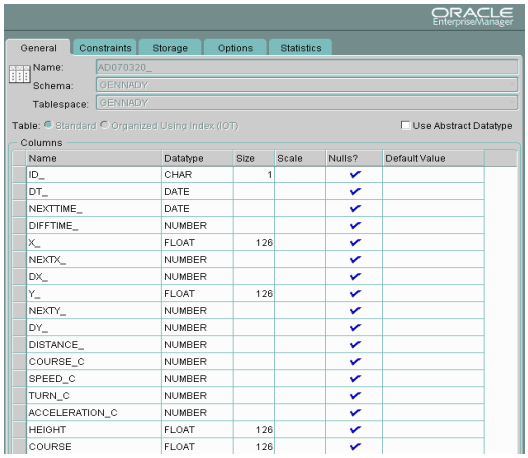
Data preprocessing (one-time operation in DB)

Original data:

- <t, x, y>

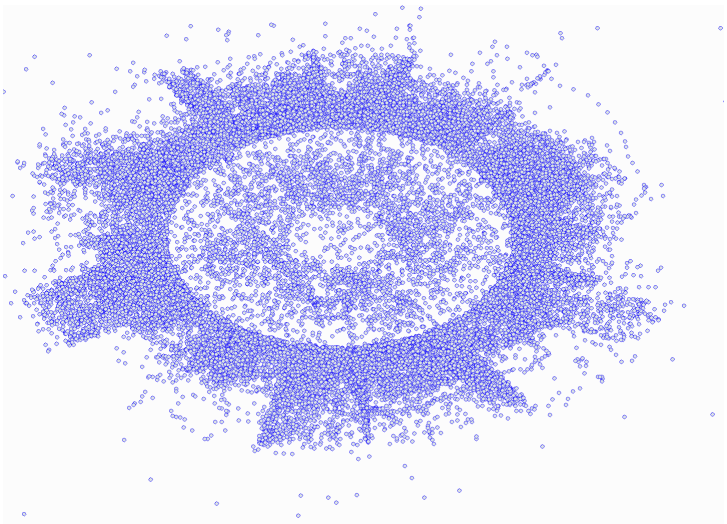
Results of preprocessing:

- Points are arranged into sequences (NEXTTIME field)
- DIFFTIME and DISTANCE to the next position, DX, DY
- 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…



We separated time-consuming data pre-processing (needs to be performed only once!) from rapid analysis and aggregation procedures (that take just few seconds)

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

Defining places

- We know that the positions were recorded only while the car moved.

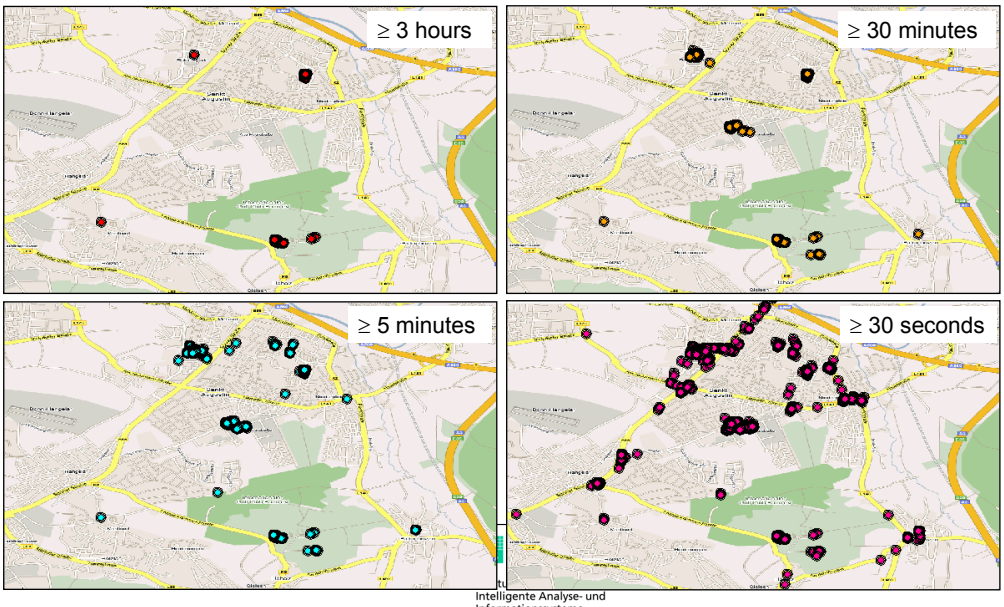
⇒ Large time gaps between successive positions indicate stops.

⇒ We can query the database for the positions where DIFFTIME (distance in time to the next position) exceeds a specified threshold

- The possible meanings of the extracted places will differ depending on the threshold
  - Large gaps: places where the person spends much time (home, work)
  - Medium gaps: shops, doctors, sport facilities, …
  - Small gaps: traffic lights, street crossings, …
- Places separated by medium to large gaps may be destinations and sources of trips

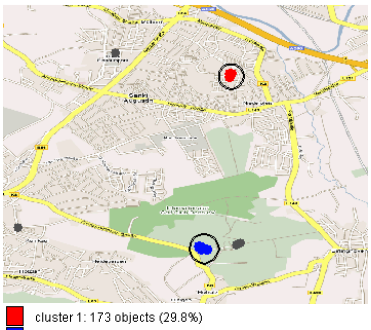
  ⇒ Sequences of positions between the gaps represent the trips

Extraction of the places of stops (DB query + GUI + Visualisation)



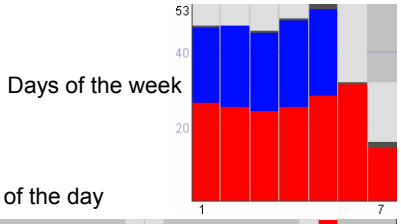
Interpretation of the places of stops
A) Long stops (≥ 3 hours)

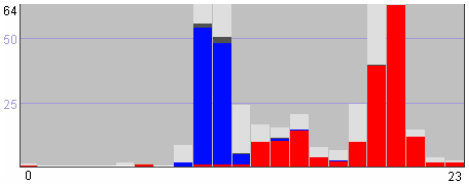
1) Spatial clustering: find repeated stops



cluster 1: 173 objects (29.8%)
cluster 2: 109 objects (18.8%)
noise: 8 objects (1.4%)

2) Look at the days and times of the occurrence

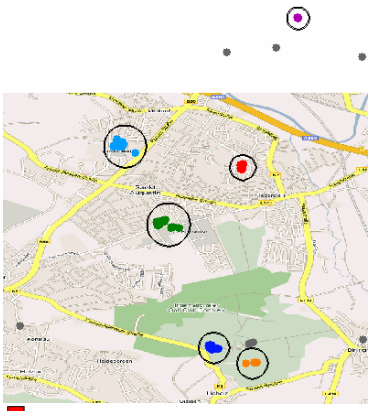




⇒ **Red: home, blue: work**

Interpretation of the places of stops
B) Medium stops (≥ 30 minutes)

1) Spatial clustering: find repeated stops

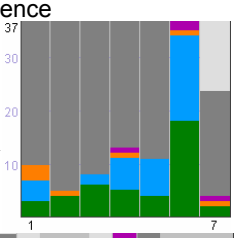


cluster 1: 184 objects (22.7%)
cluster 2: 117 objects (14.4%)
cluster 3: 42 objects (5.2%)
cluster 4: 35 objects (4.3%)
cluster 5: 7 objects (0.9%)
cluster 6: 4 objects (0.5%)
noise: 17 objects (2.1%)

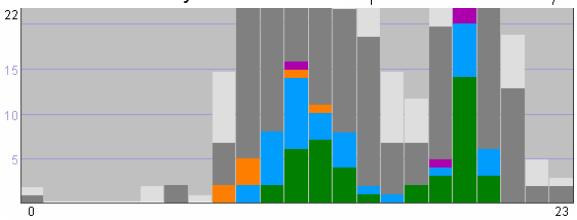
2) Look at the days and times of the occurrence

(clusters 1&2 excluded)

Days of the week



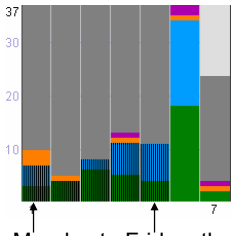
Hours of the day

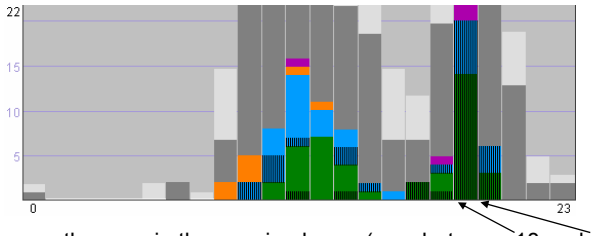


⇒ **Green & light blue: probably shopping**

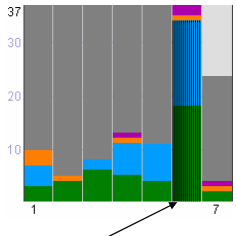
Fraunhofer IAIS Institut Intelligente Analyse- und Informationssysteme

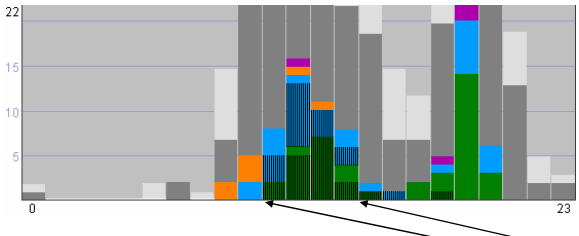
Interpretation of the clusters 3 and 4 (continued)





Monday to Friday: the stops mostly occur in the evening hours (max between 18 and 19)





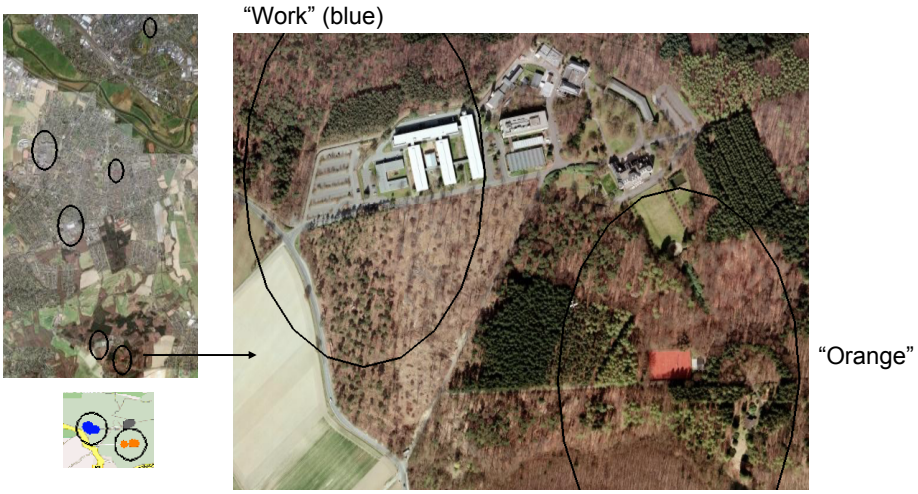
Saturday (max stops among all days): the stops mostly occur between 10 and 14

Fraunhofer IAIS Institut Intelligente Analyse- und Informationssysteme

Interpret the places with the help of Google Earth

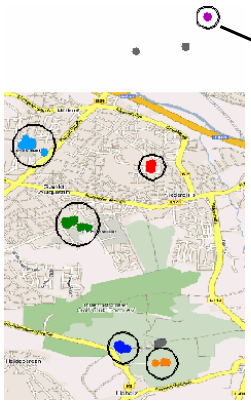
Look like big shopping centres

Interpret the places with the help of Google Earth (cont.)



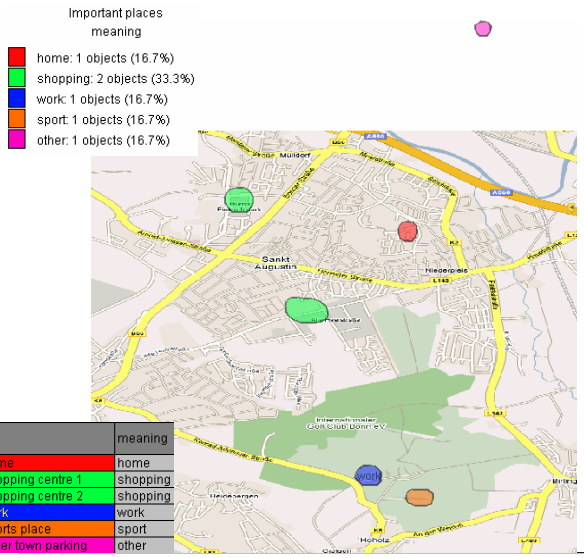
"Orange" seems to be a place for sports or leisure

Interpret the places with the help of Google Earth (cont.)

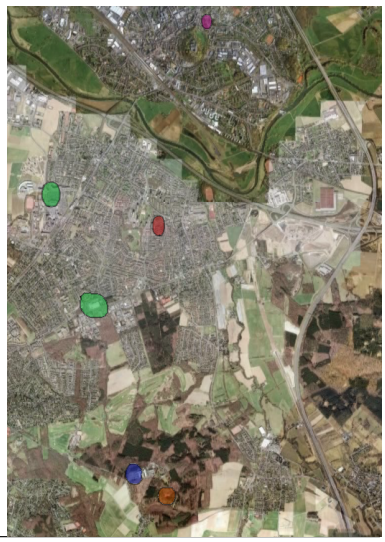


"Magenta" is a parking place in the neighbouring town

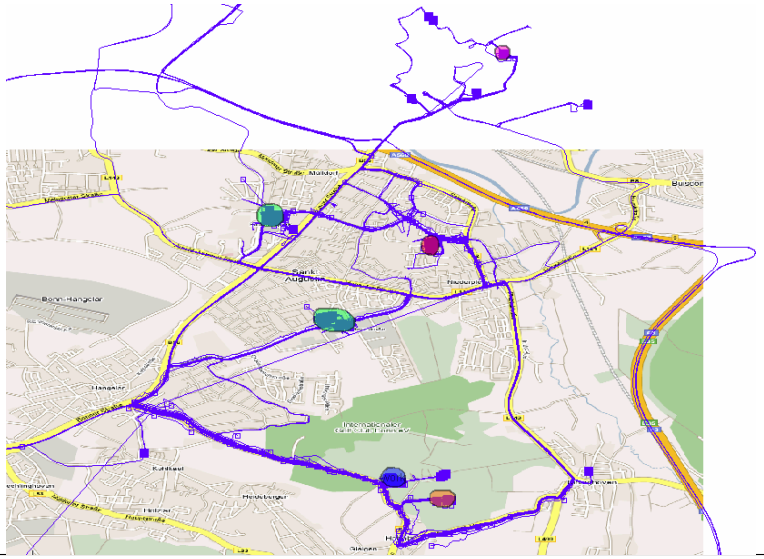
The interpreted places



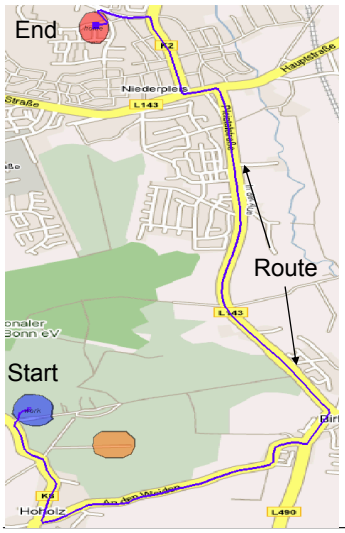
	meaning
home	home
shopping centre 1	shopping
shopping centre 2	shopping
work	work
sports place	sport
Other town parking	other



Extraction of trips (DB operation + GUI + Visualisation)

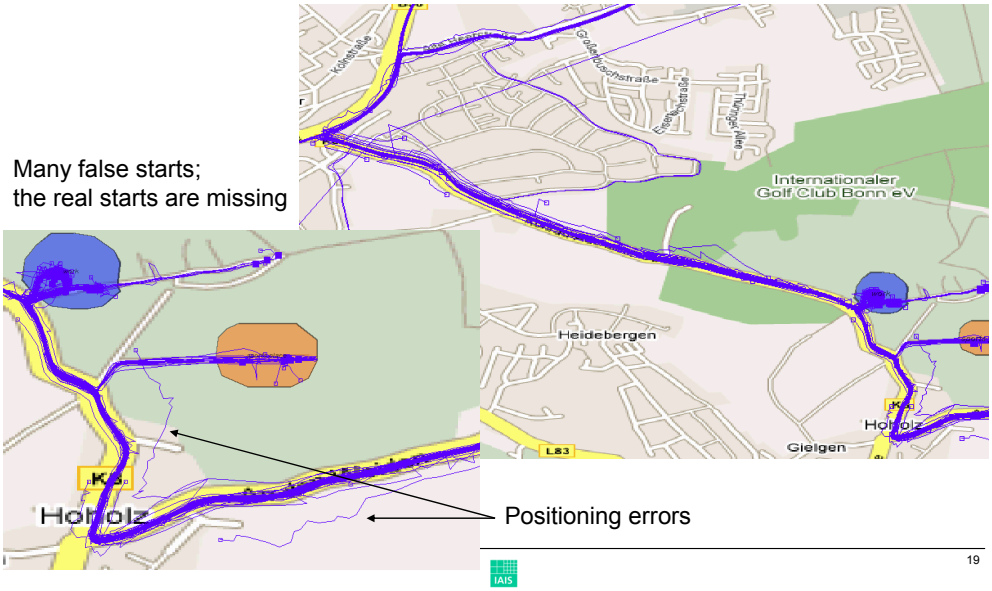


A trip in more detail



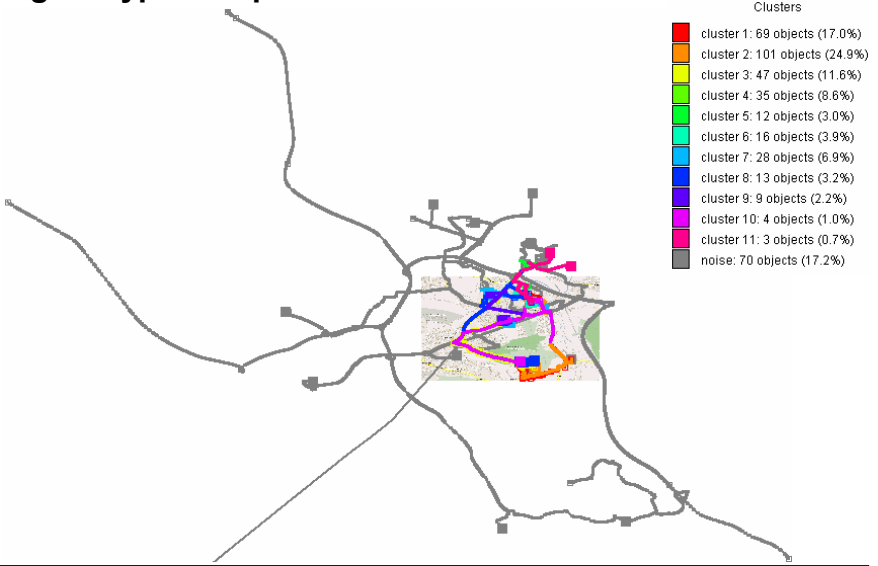
Problems with data quality…

Many false starts;
the real starts are missing

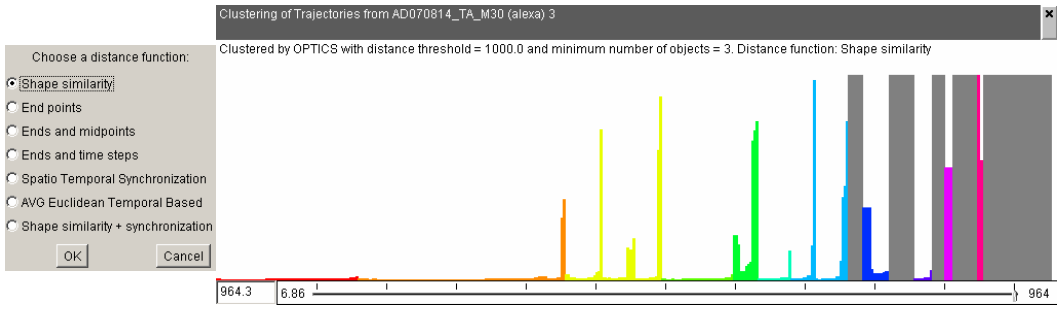


Positioning errors

Looking for typical trips

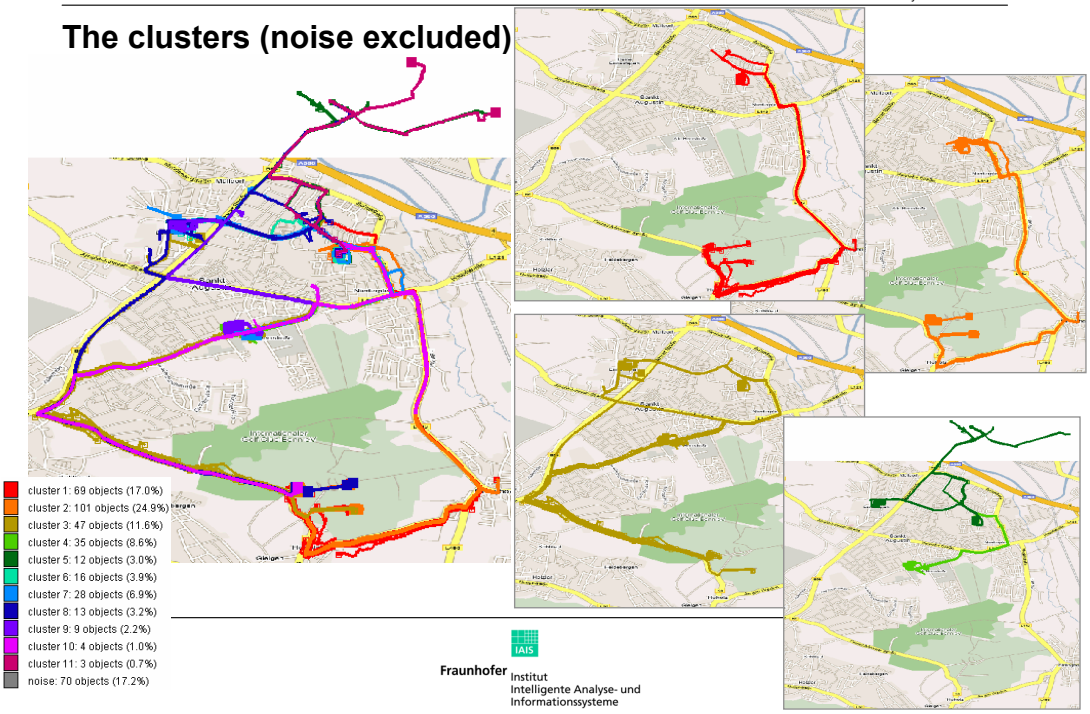


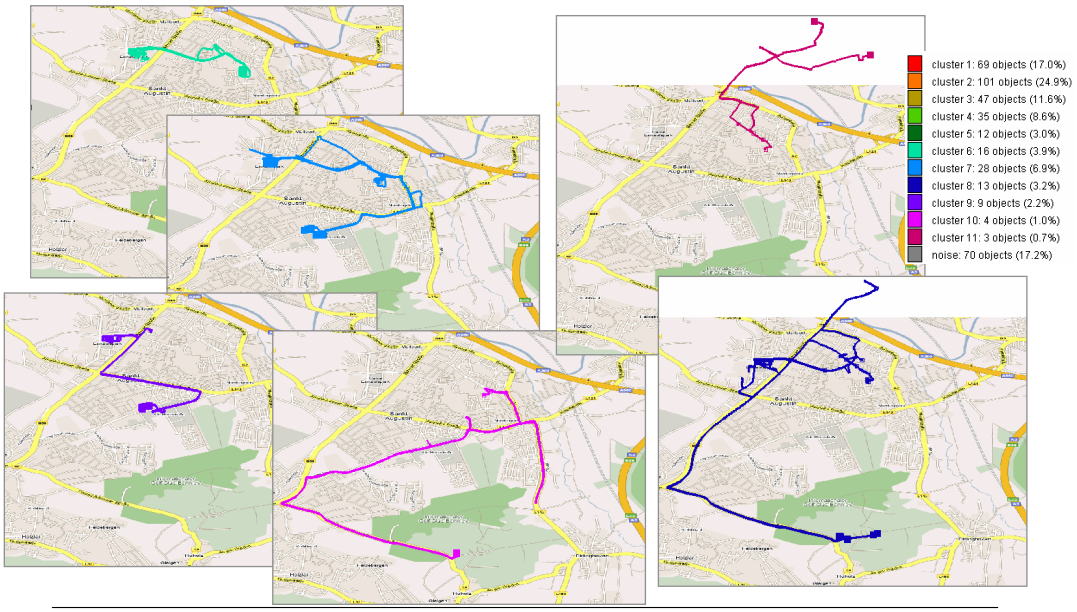
Interactive tool for spatial clustering



Joint work with S.Rinzivillo, Univ. Pisa:
- spatial clustering algorithm OPTICS (Ankerst, Breunig, Kriegel, and Sander 1999)
- our implementation: cluster building is separated from distance and neighbourhood computation
- benefit: an analyst may try different distance measures
- we have developed several variants of distance measures designed specially for trajectories
- various ways of handling the times in the data
-some of the measures can tolerate the specific errors in the car driving data

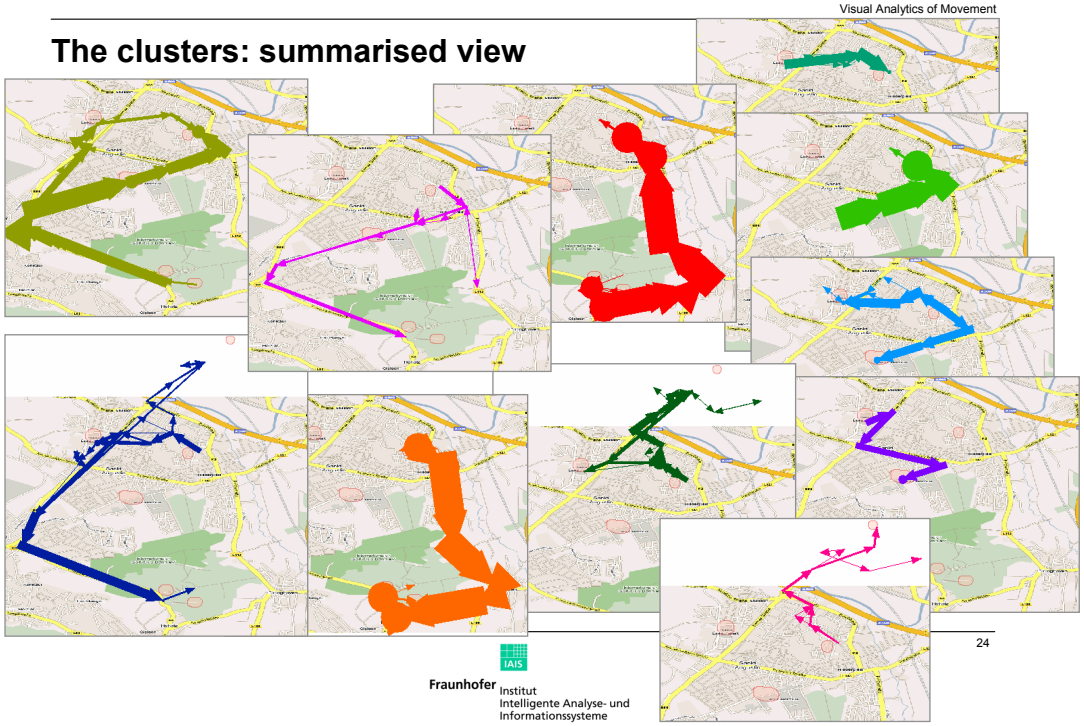
The clusters (noise excluded)





Fraunhofer Institut
Intelligente Analyse- und
Informationssysteme

The clusters: summarised view



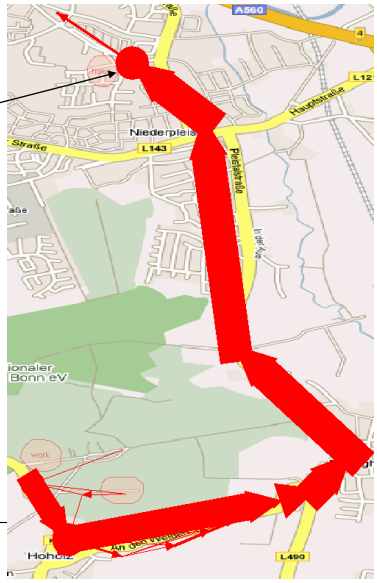
Fraunhofer Institut
Intelligente Analyse- und
Informationssysteme

The summarised view in more detail

Major flow

Variations

Many small moves

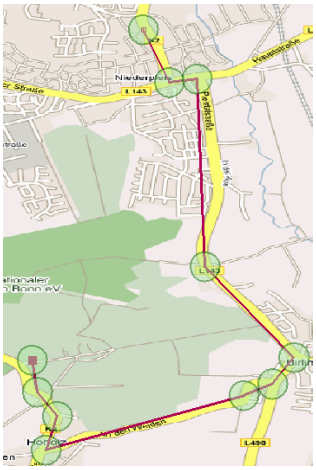


Summarisation: how it is done (1)

Original trajectory

Simplified trajectory

Generalised positions

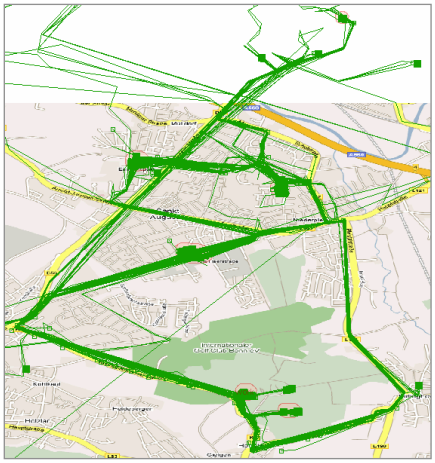


Minimum angle of direction change (degrees):	30
Minimum duration of a stop:	60
Minimum radius around a position:	100.0

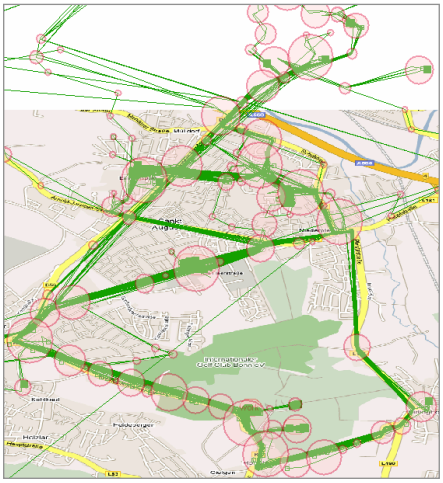
← Specifies which positions to keep

26

Summarisation: how it is done (2)

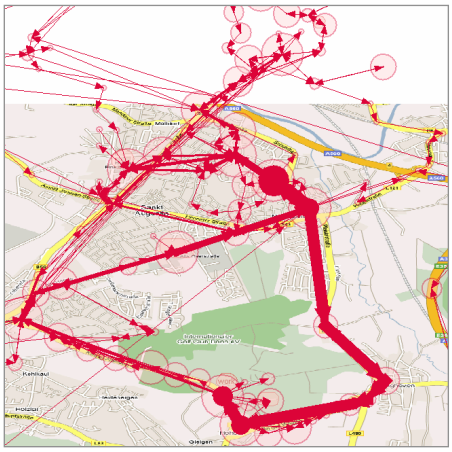


All trajectories simplified

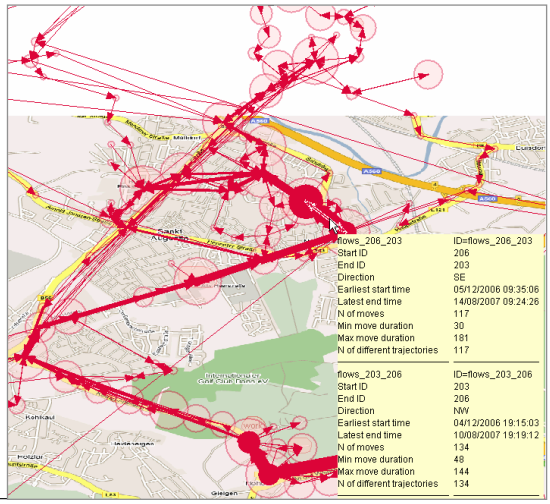


Generalised positions (circles, for simplicity) are built around spatially close positions of all trajectories

Summarisation: how it is done (3)

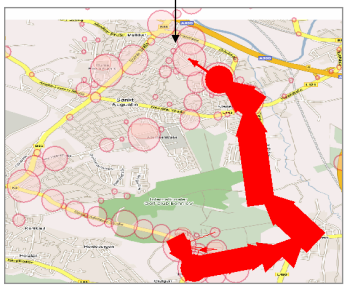


Numbers of moves between the positions are counted and represented by arrows of varying width

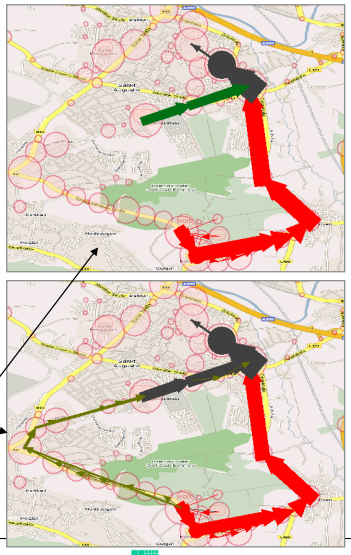


Representation of clusters

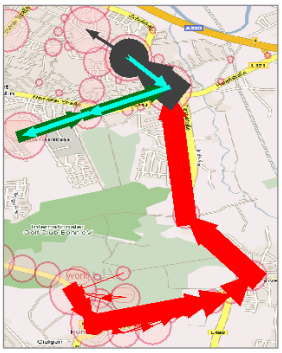
A single cluster selected: the moves from this cluster only are counted and shown



2 or more clusters selected: the moves from these clusters are counted and shown; dark grey corresponds to common moves



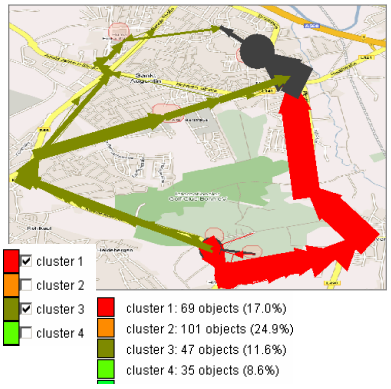
Disadvantage: some arrows may cover other arrows…



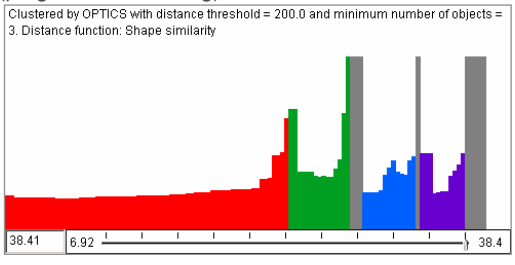
In the future, we'll try to put them in a 3D view (e.g. Google Earth)

Let's return to our analysis…

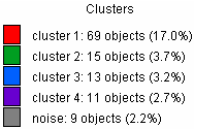
The trips starting at work (clusters 1 and 3):



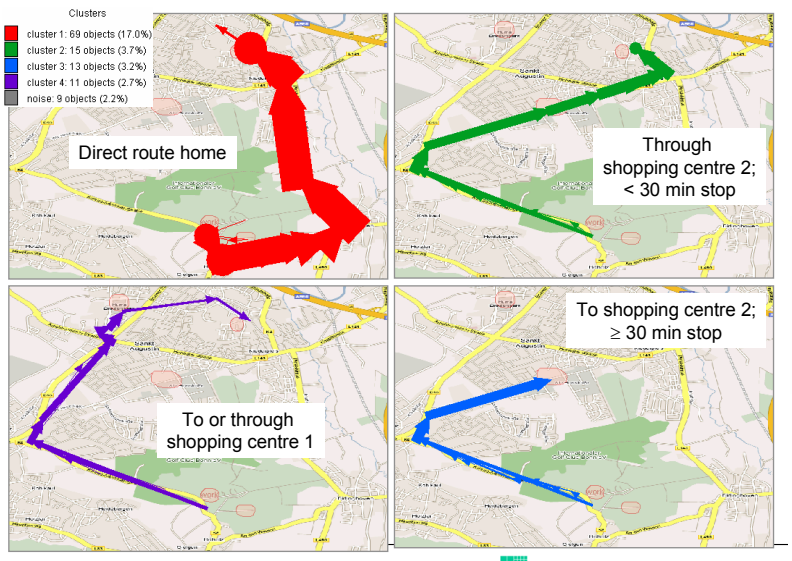
These clusters can be further refined (progressive clustering):



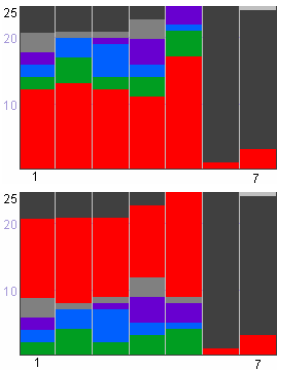
Applied only to selected trajectories!



Typical trips of the person from the work



Distribution by days of week

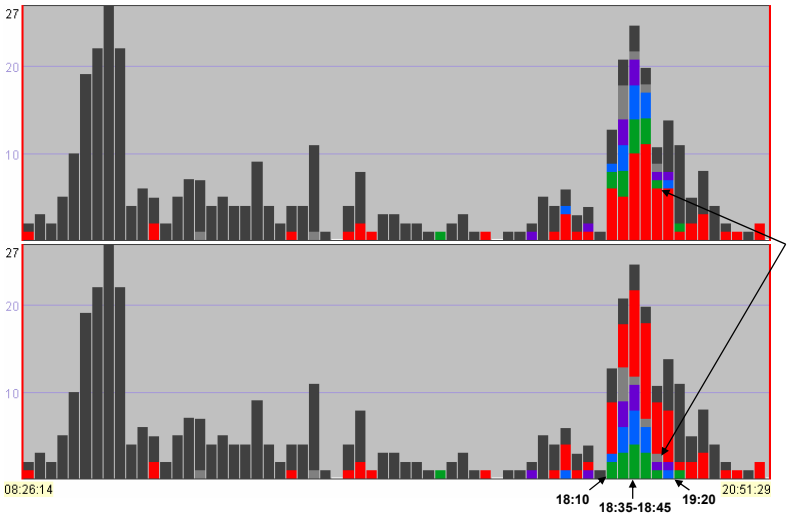


The way through shopping centre 1 is chosen more often on Thursdays and Fridays than on other days.

Maximum visits of shopping centre 2 for 30 min or more are on Wednesday.

Fraunhofer Institut Intelligente Analyse- und Informationssysteme

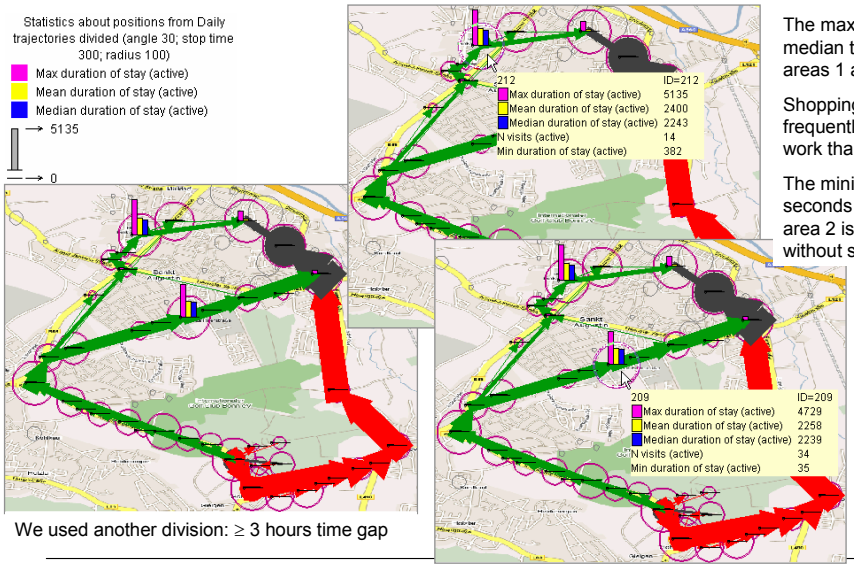
The trips by times of the day



The trips to/through shops rarely begin after 18:55.

Fraunhofer Institut Intelligente Analyse- und Informationssysteme

How much time is spent in the shopping areas?



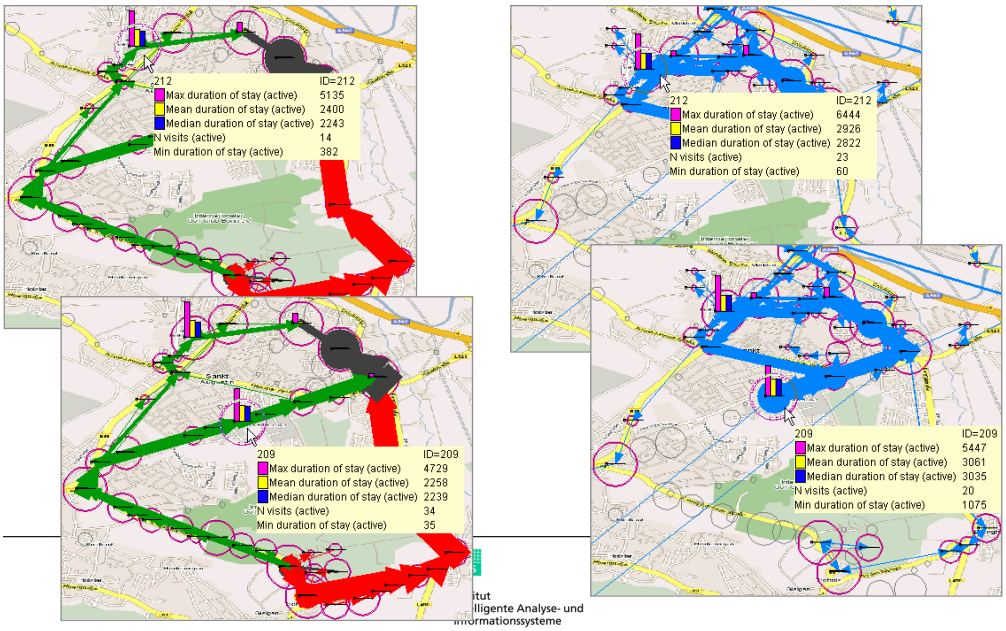
The maximum, mean and median times spent in shopping areas 1 and 2 are similar.

Shopping area 2 is visited more frequently on the way from the work than shopping area 1.

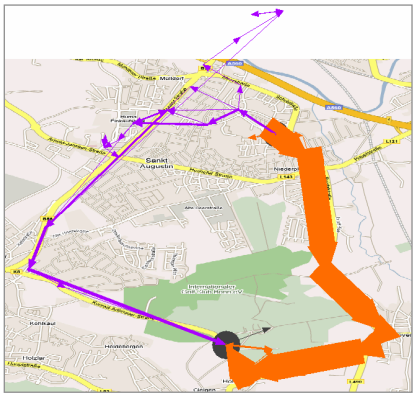
The minimum duration of stay 35 seconds indicates that shopping area 2 is sometimes passed without stopping.

We used another division: ≥ 3 hours time gap

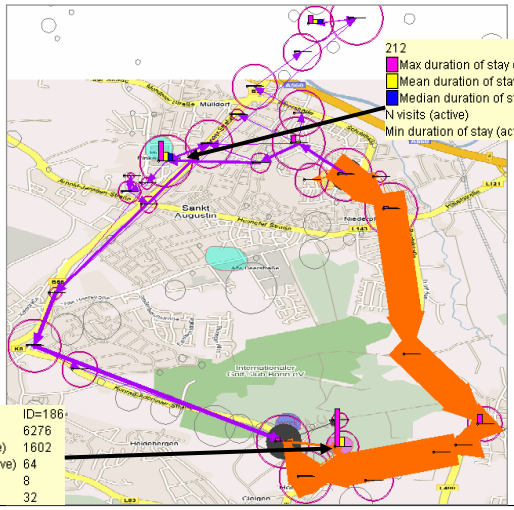
Comparison with the visits of the shops on Saturday



Trips from home to work



Stops and times spent



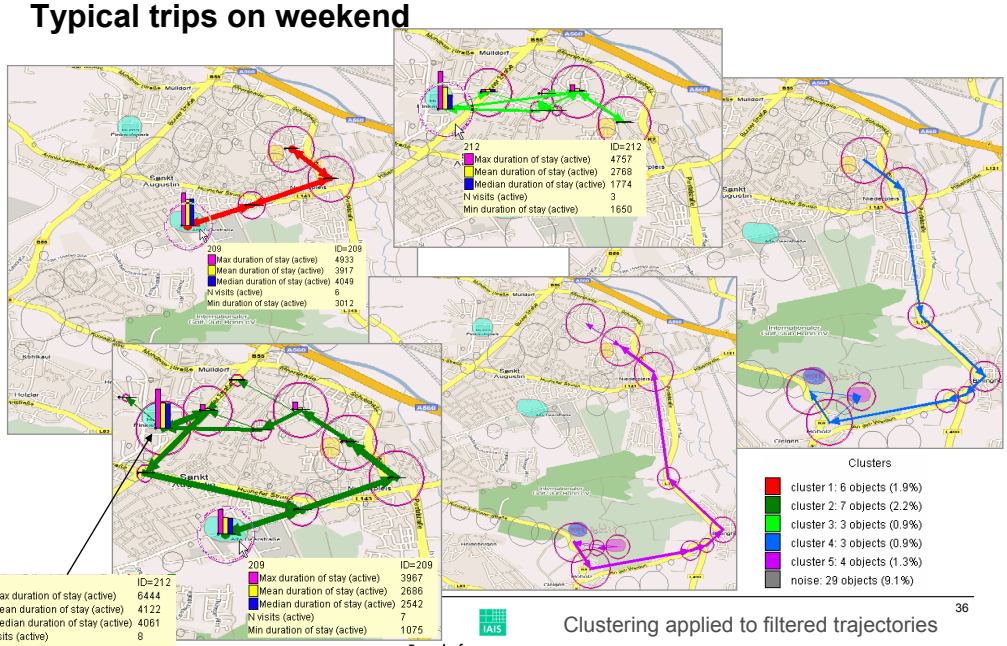
cluster 2: 105 objects (32.9%)

cluster 8: 14 objects (4.4%)

186 ID=186
Max duration of stay (active) 6276
Mean duration of stay (active) 1602
Median duration of stay (active) 64
N visits (active) 8
Min duration of stay (active) 32

Fraunhofer Institut Intelligente Analyse- und Informationssysteme

Typical trips on weekend

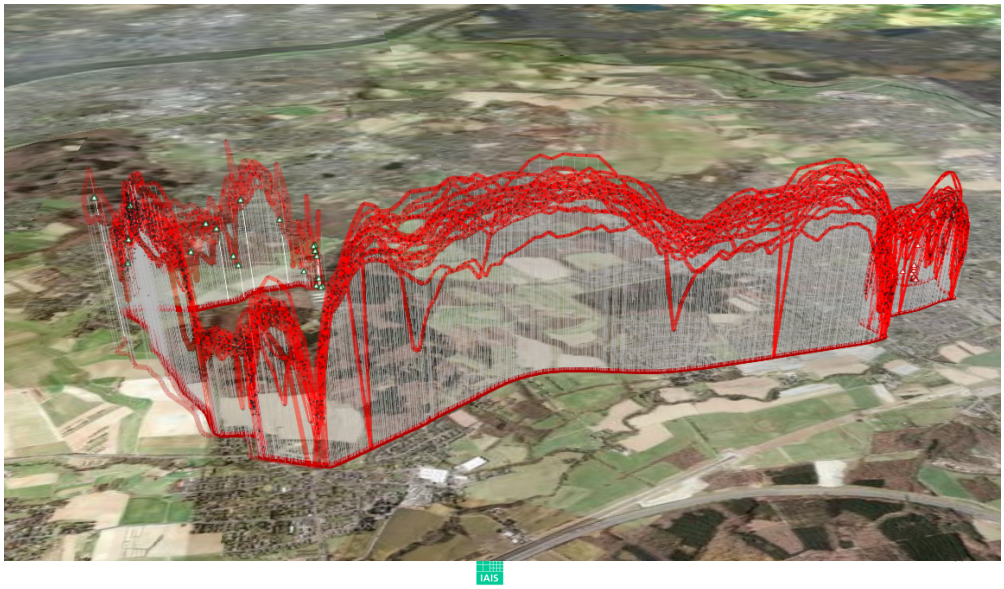


Clustering applied to filtered trajectories

Fraunhofer Institut Intelligente Analyse- und Informationssysteme

Cluster 1: comparison in 3D view (Google Earth)

Comparison of speed dynamics in 3D view (Google Earth)



What we have learned about the car owner:

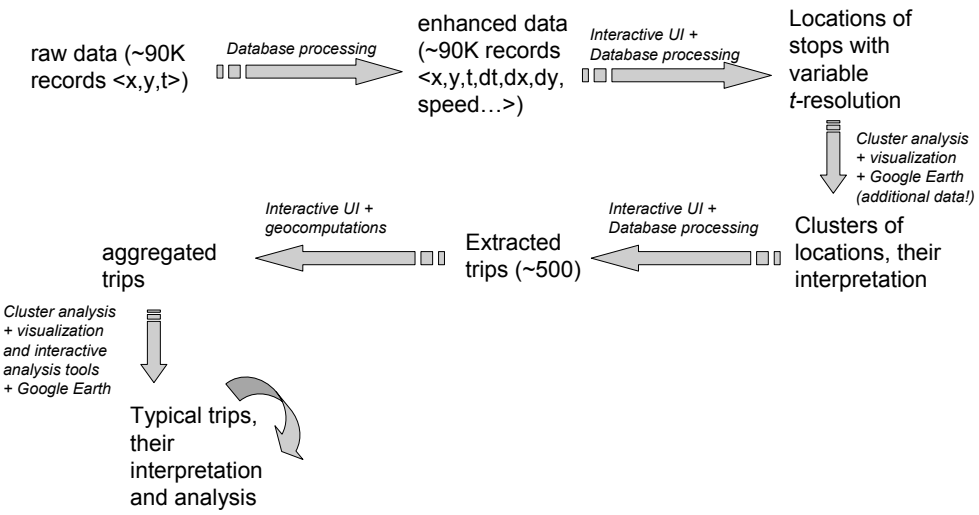
- The places where the person lives, works, and shops
- The typical routes from home to work, from work to home, to the shopping areas
- The places where the person frequently stops on the way from work to home
- The places where the person may stop on the way from home to work
- The durations of the stops, times spent for visiting the shopping areas
- The times of the trips and of the stops
- How the chosen routes are distributed over the days of the week

The tools we used:

- Database processing
- Interactive clustering
- Computational generalisation and summarisation
- Interactive displays
- Dynamic filtering

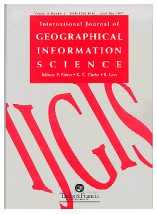
} **Visual analytics toolkit**

Reconstruction of the analytical process

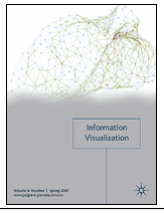


International Cartographic Association
Commission on Visualization

- http://kartoweb.itc.nl/icavis/index.html - to be updated soon
  meanwhile: http://www.ais.fraunhofer.de/and
  welcome to join our activities!



- GIScience 2006 workshop outcomes:
  Special issue on
  "GeoVisual Analytics for Spatial Decision Support",
  including "Setting the Research Agenda" paper
  Int.J.GIScience, 2007, v.21(8)



- AGILE 2008 workshop (May 2008, Girona, Spain)
  GeoVisualization of Dynamics, Movement and Change
  forthcoming special issue of
  Information Visualization

Fraunhofer Institut Intelligente Analyse- und Informationssysteme