Mobility Data Management

Yannis Theodoridis

InfoLab | University of Piraeus | Greece
infolab.cs.unipi.gr

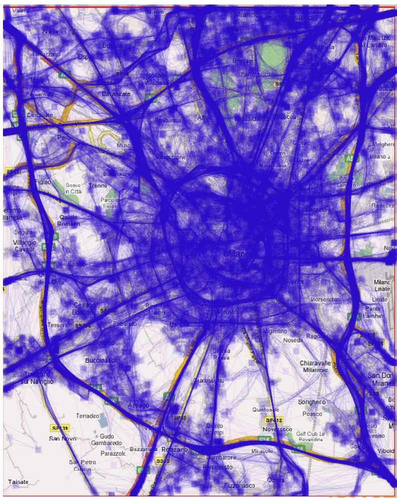
Univ. Cyprus, Nov. 2012

"πάντα ρει - Everything changes and nothing remains still"
Heraclitus (500 BC)

Our starting point: GPS data feeds

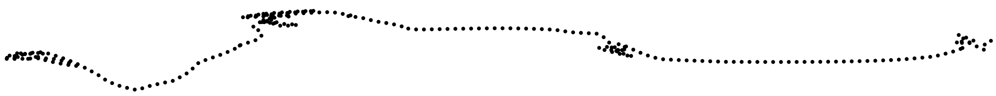
- Raw data: GPS recordings

```
N; Time; Lat; Lon; 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;
11;22/03/07 08:52:03;50.777317;7.205877; 68.8;
12;22/03/07 08:52:06;50.777185;7.206202; 68.1;
13;22/03/07 08:52:09;50.777057;7.206522; 67.9;
14;22/03/07 08:52:12;50.776925;7.206858; 66.9;
15;22/03/07 08:52:15;50.776813;7.207263; 67.0;
16;22/03/07 08:52:18;50.776780;7.207745; 68.8;
17;22/03/07 08:52:21;50.776803;7.208262; 71.1;
18;22/03/07 08:52:24;50.776832;7.208682; 68.6;
…
```

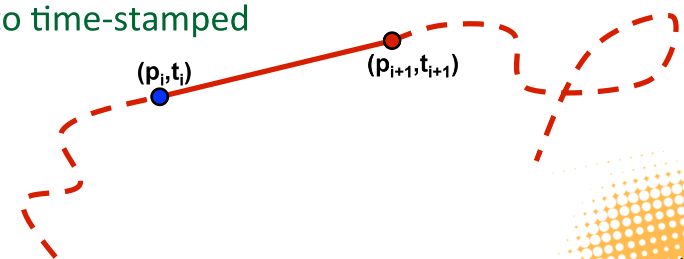


What is a (GPS-based) trajectory?

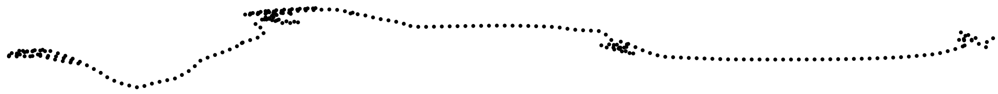
- A **trajectory** is a model for a motion path of a moving object (animal, car, human, …)
  - (due to discretization) a sequence of sampled time-stamped locations (p_i, t_i) where p_i is a 2D point (x_i, y_i) and t_i is the recording timestamp of p_i



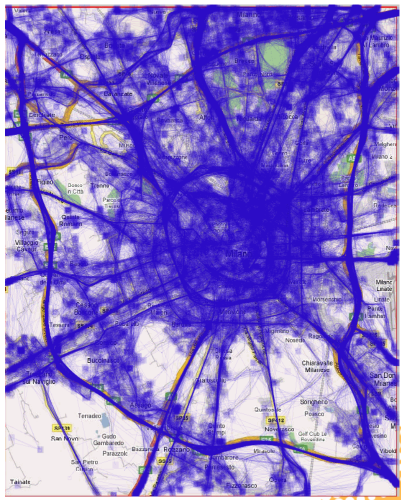
- A common representation in MOD is a **3D polyline** in the plane where vertices correspond to time-stamped locations (p_i, t_i)
  - and **linear interpolation** is assumed between (p_i, t_i) and (p_{i+1}, t_{i+1})



Key questions that arise



- How to **reconstruct a trajectory** from raw logs?
- How to **store and query trajectories** in a DBMS?
  - What is a "trajectory" data type?
    - simply a sequence of (x, y, t) tuples?



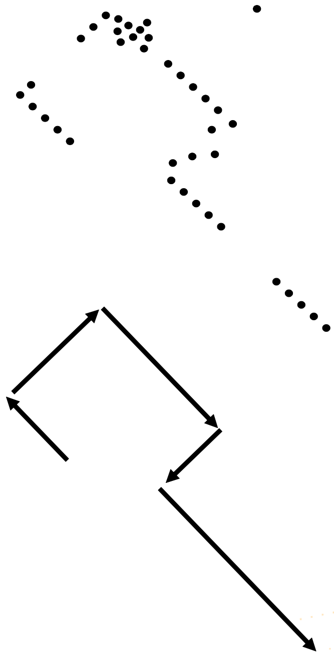
Acquiring trajectories from raw data

About mobility data

The trajectory reconstruction problem

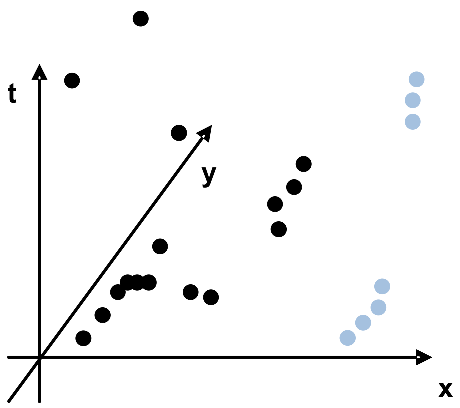
The trajectory reconstruction problem

- From raw data, i.e., time-stamped locations
  - Raw data (3D points) arrive either one-by-one or in bulks
- … to trajectory data, i.e., continuous evolutions
  - Linear interpolation is assumed between samples,
  - Redundancy is reduced, noise is removed,
  - etc.



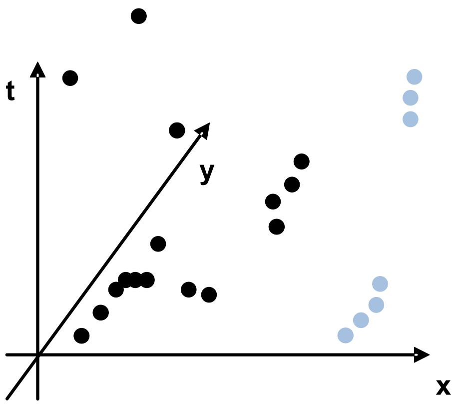
Reconstructing trajectories

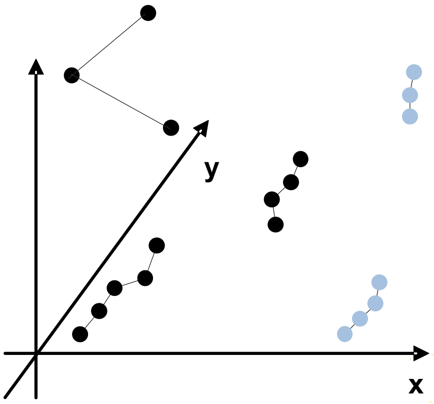
- Collected raw data represent time-stamped geo-locations
- Raw data (3D points) arrive either one-by-one or in bulks
- **Any idea?**



Reconstructing trajectories

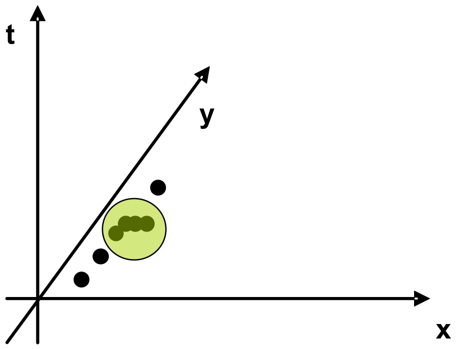
- [Marketos et al. 2008] proposes filters / thresholds that decide whether the new series of data is to
  - be appended to an existing trajectory, or
  - initiate a new trajectory, or
  - be considered as noise

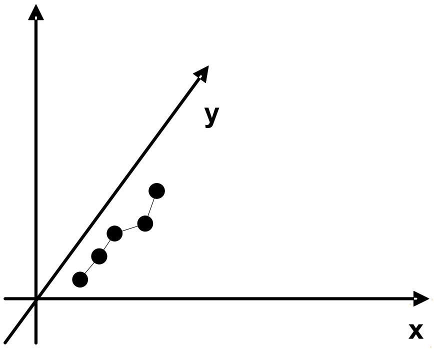




Reconstructing trajectories

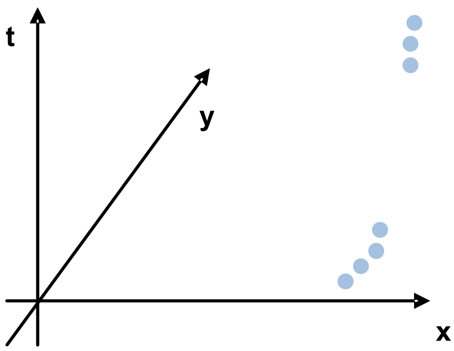
- 1st parameter: tolerance distance
  - The tolerance of the transmitted time-stamped positions
  - In other words: the **maximum distance between two consecutive time-stamped positions** of the same object in order for the object to be considered as **stationary**

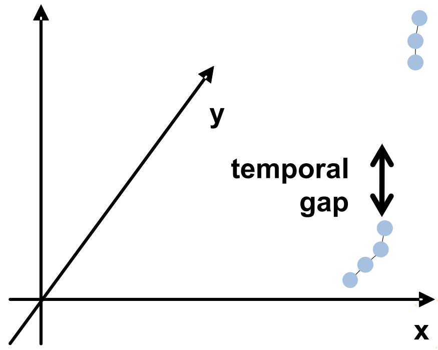




Reconstructing trajectories

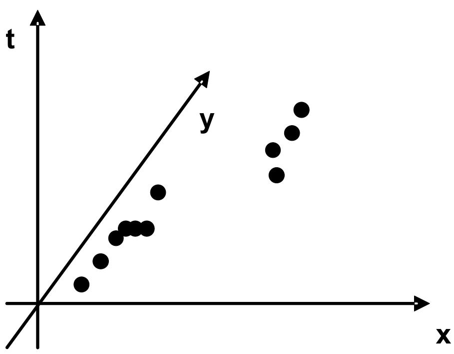
- tolerance distance
- 2nd parameter: temporal gap between trajectories
  - The **maximum allowed time interval** between two consecutive time-stamped positions of the same trajectory for a single moving object

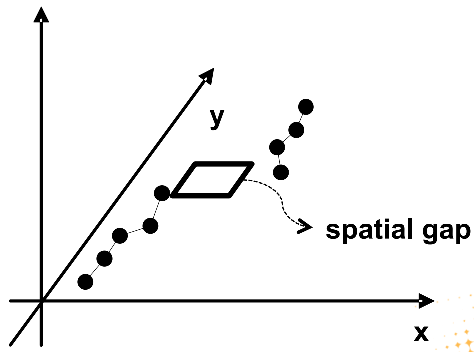




Reconstructing trajectories

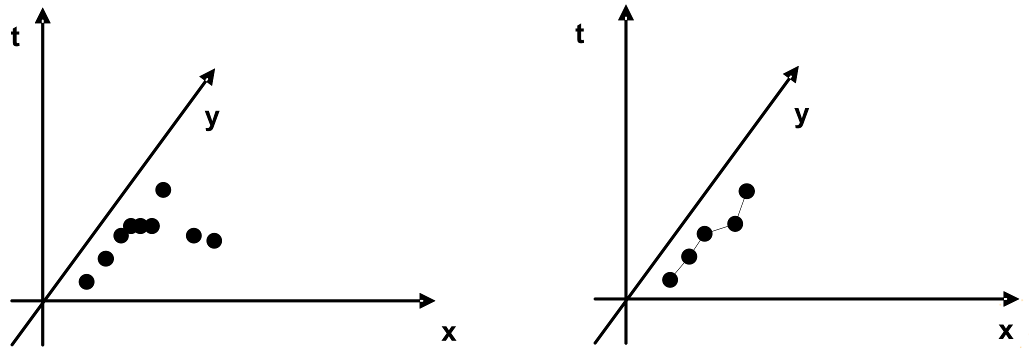
- tolerance distance, temporal gap
- 3rd parameter: spatial gap between trajectories
  - The **maximum allowed distance** in 2D plane between two consecutive time-stamped positions of the same trajectory





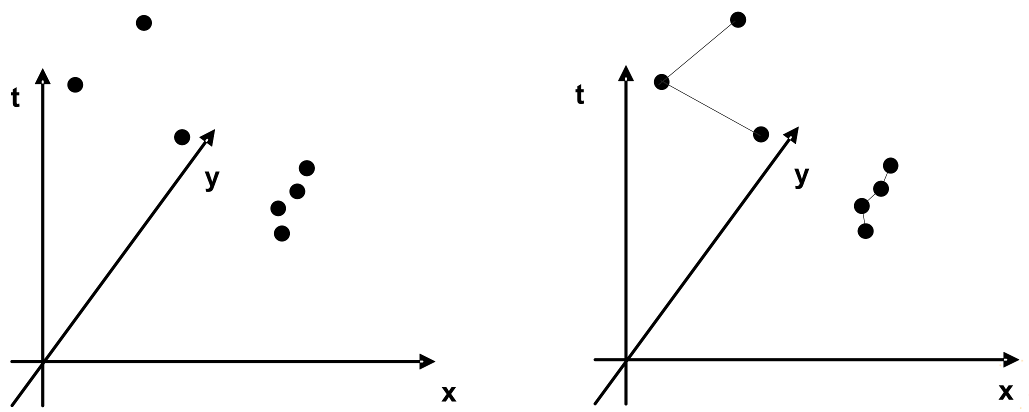
Reconstructing trajectories

- tolerance distance, temporal gap, spatial gap
- 4^{th} parameter: maximum speed
  - Decides whether a reported time-stamped location is **noise**, hence to be discarded from the output trajectory



Reconstructing trajectories

- tolerance distance, temporal gap, spatial gap, maximum speed
- 5^{th} parameter: maximum noise duration
  - The **maximum duration of a noisy part** of a trajectory. If 'noise' continues longer than $noise_{max}$, most probably it is not noise but, instead, the starting positions of a new trajectory!

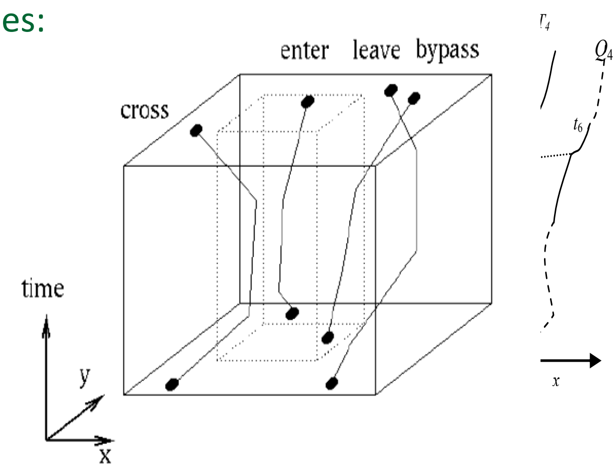


Location-aware querying

From primitive to advanced MOD queries

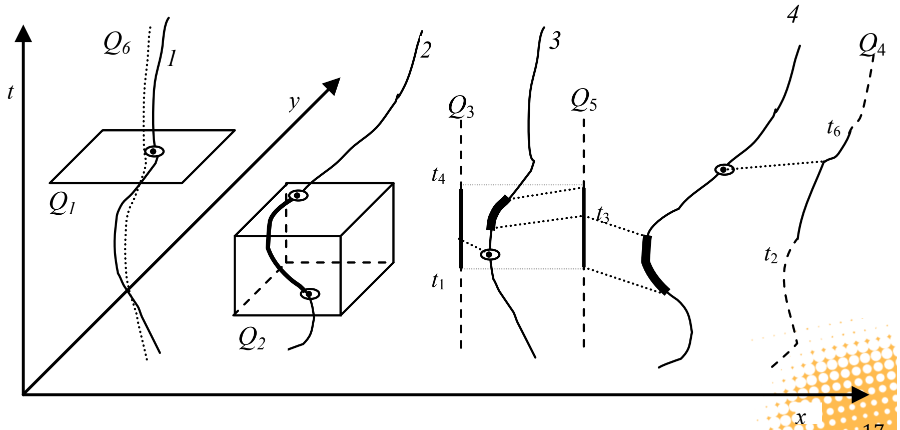
What kind of queries?

- The nature of trajectory data provides us with the ability to query them with a variety of operators.
- Primitive queries on trajectories:
  - Coordinate-based
    - Range, NN
  - Trajectory-based
    - Topological, Directional



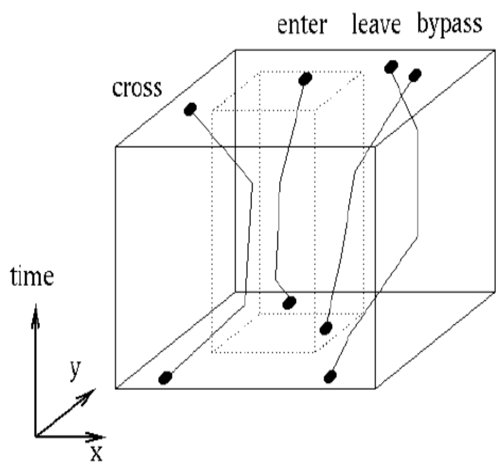
Coordinate-based queries

- Spatial (range or NN) search
  - "Find all trajectories that were inside area A at time instant t (or time interval I)" or
  - "Find the trajectory that was closest to point B at time instant t (or time interval I)"



Trajectory-based queries

- Topological / directional search
  - "Find all trajectories that entered (crossed, left, bypassed, etc.) or were located west (south, etc.) of an area" or
  - "Find all trajectories that crossed (met, etc.) or were located left of (right of, in front of, etc.) a query trajectory TQ



Taxonomy of location-aware queries

- Dimensions [Mokbel & Aref, 2007]:
  - **Type**: range, NN, reverse NN, closest-point, …
  - **Time**: past, present, future
  - **Duration**: snapshot, continuous
  - **Query (reference) object**: stationary, moving
  - **Data objects**: stationary, moving

- Queries: every (?) possible combination of the above dimensions

Examples

slides from [Mokbel & Aref, 2007]

Where are my nearest McDonalds for the next hour?

- *Type:* Nearest-Neighbor query
- *Time:* Future
- *Duration:* Continuous
- *Query:* Moving
- *Object:* Stationary

Send E-coupons to all cars that I am their nearest gas station

- *Type:* Reverse NN query
- *Time:* Present
- *Duration:* Snapshot
- *Query:* Stationary
- *Object:* Moving

Examples

slides from [Mokbel & Aref, 2007]

Continuously report the number of cars in the freeway

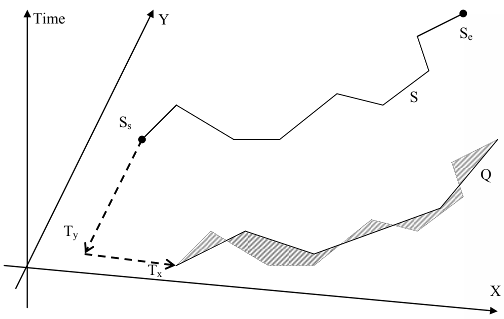
- *Type:* Range query
- *Time:* Present
- *Duration:* Continuous

What was the closest dist. between Taxi A & me yesterday?

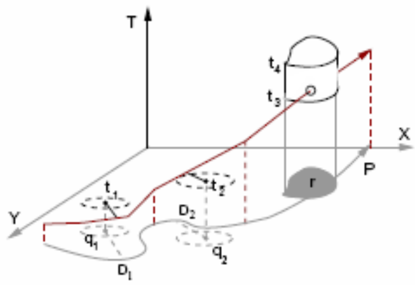
- *Type:* Closest-point query
- *Time:* Past
- *Duration:* Snapshot
- *Query:* Moving
- *Object:* Moving

Advanced trajectory-based queries

- **Trajectory similarity** queries
  - [Frentzos et al. 2007] "Given a query trajectory *Q*, find the k- most similar trajectories to *Q* (perhaps, constrained is space and/or time)"

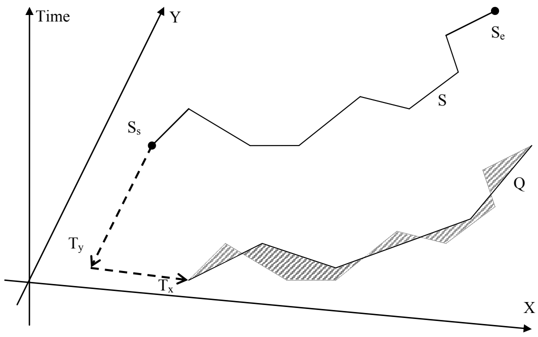


- **Spatio-temporal pattern** queries
  - [Hadjieleftheriou et al. 2005] e.g. "Find objects that crossed through region *A* at time t_1, came as close as possible to point *B* at a later time t_2 and then stopped inside circle *C*



Trajectory Similarity Queries

- Key question:
  - How do we measure **distance** or **(dis-)similarity** between two trajectories?
- Similarity variations:
  - [Pelekis et al. 2007, 2011]
    Similarity in space and/or time,
    or wrt. derived information (e.g. speed or direction)



- Similarity queries have been studied extensively in time-series literature
  - But, things are different here! Both <u>where</u> and <u>when</u> are important

Trajectory Similarity Queries (cont.)

- Different points of view: **Moving clusters**, **Flock queries**
  - What is a flock?
    - a large enough subset of objects moving along paths close to each other for a certain time
  - In the flock, identify **leaders** and **followers**
- Solutions:
  - [Benkert et al. 2008],
    [Gudmundsson & van Kreveld, 2006]

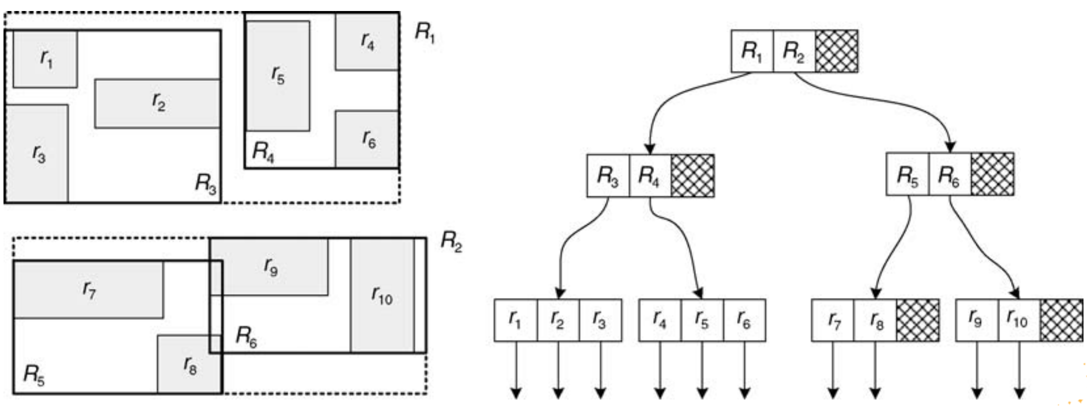
Efficiently trajectory indexing and storage in MODs

Indexing techniques

MOD engines

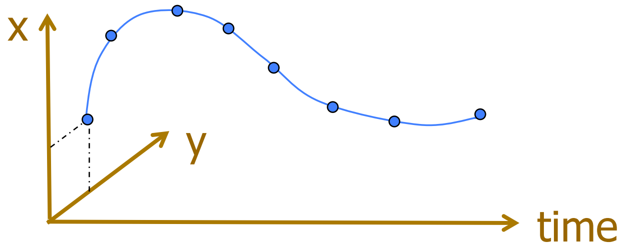
R-trees for spatial data

- For d-dimensional point or region data
- Is it portable to mobility data?
  - In other words, is space + time simply a 3D space?



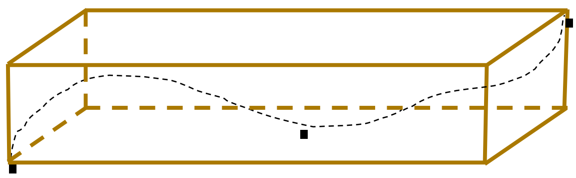
What ?? for mobility data

- Challenges:
  - Both space and time are equally important!
  - But! time is not simply a 3rd dimension
    - e.g. it is monotonously increasing
  - How does a trajectory approximation look?
    - Approximation is necessary for indexing

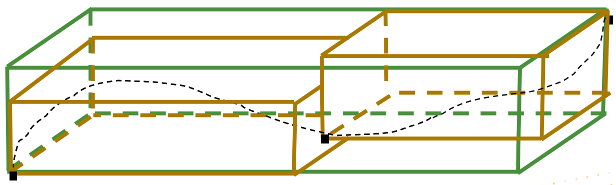


Two approaches: native vs. parametric space

- Typically approximate using MBRs; then index these MBRs
  - we can use R-trees etc. ☺
  - trajectories are lines, thus MBRs add extensive empty space ☹



- How many MBRs per trajectory?
  - One MBR per trajectory (too much empty space…) or one MBR per segment (too many MBRs…)
- Can we do anything better?
  - Smart "partitioning" for MBRs [Hadjieleftheriou et al. 2002]



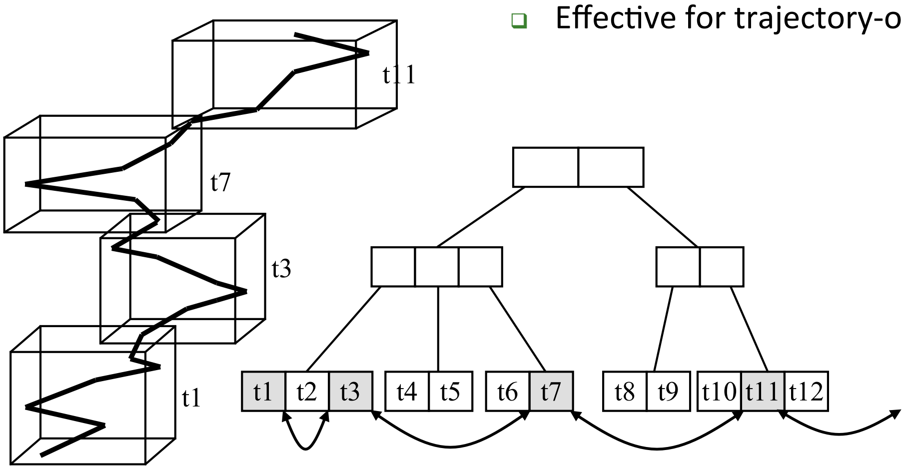
MOD indexing techniques

- Indexing the past (= trajectories)
  - unconstrained movement : the trajectory-bundle tree (TB-tree) [Pfoser et al. 2000]
  - network-constrained movement: the fixed-network-restricted tree (FNR-tree) [Frentzos, 2003]
- Indexing the present (and anticipated future)
  - Data partitioning: TPR-tree [Saltenis et al. 2000], TPR*-tree [Tao et al. 2003]
  - Space partitioning: B^x-tree [Jensen et al. 2004], ST^2B-tree [Chen et al. 2008]
- (Hybrid solution for) Indexing the past & present
  - R^{PPF}-tree [Pelanis et al. 2006]

- We focus on historical tracks of moving objects (trajectories) ➔ concentrate on the 1st group

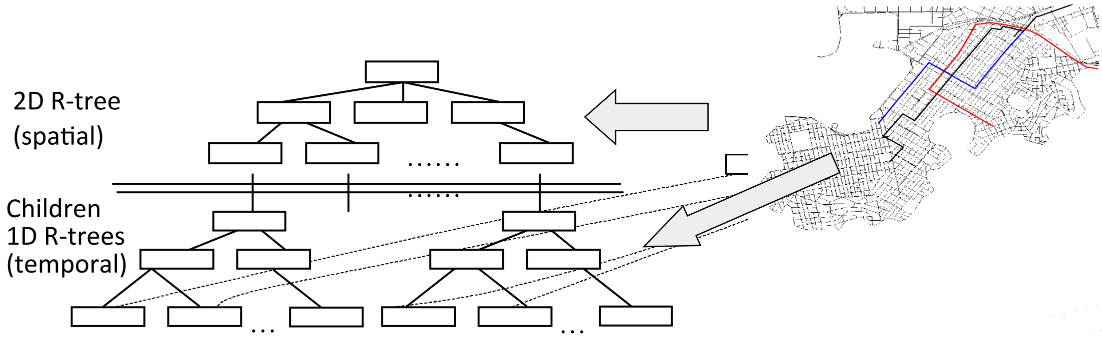
TB-tree

- [Pfoser et al. 2000] Maintains the ‘trajectory’ concept
  - Each node consists of segments of a single trajectory
  - nodes corresponding to the same trajectory are linked together in a chain
  - Effective for trajectory-oriented queries



FNR-tree

- (Frentzos, 2003) a forest of 1D (temporal) R-trees on top of a 2D (spatial) R-tree
  - There is an additional “Parent” 1D R-tree which indexes the temporal intervals of the 1D R-trees leaf nodes



Moving Objects Database Systems

- From traditional DBMS to **Moving Object Database (MOD) engines**
  - Data types, indices, query processing & optimization strategies for trajectories
- Spatial and temporal dimensions are considered as first-class citizens.
- State-of-the-art prototype MOD engines
  - **SECONDO** (Güting et. al.) IDEAS’00, ICDE’05, MDM’06
  - **HERMES** (Pelekis et. al.) EDBT’06, SIGMOD’08

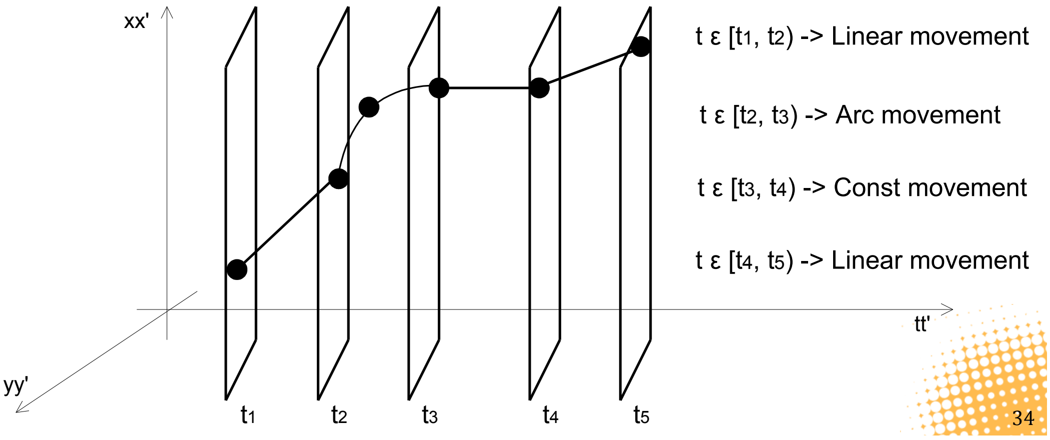
SECONDO

http://dna.fernuni-hagen.de/Secondo.html/index.html

- A generic DBMS framework that can be filled with implementation of various data models (R, OR, XML) and data types (spatial data, moving objects)
  - Built on top of Berkeley DB.
- A MOD is a set of SECONDO objects of the form *(name, type, value)*, where *type* is one of the implemented algebras
- About 20 implemented algebras
  - standard algebra, relational algebra, R-Tree algebra, spatial algebra, etc.
- Query optimizer includes optimization of conjunctive queries, selectivity estimation, and implementation of an SQL-like query language

http://infolab.cs.unipi.gr/hermes/

- 2 implementations: Oracle Spatial vs. PostgreSQL
- Supports several data types: moving point, moving line, moving polygon, etc.
- 3DR-tree and TB-tree indexing



Summary

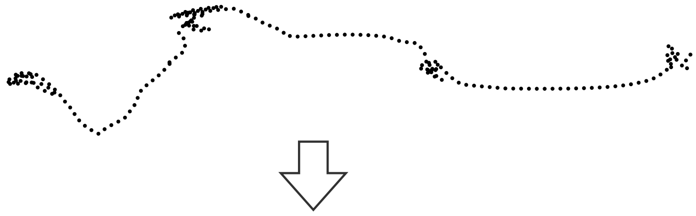
Summary on Mobility Data Management

- From (stationary) spatial to moving object databases

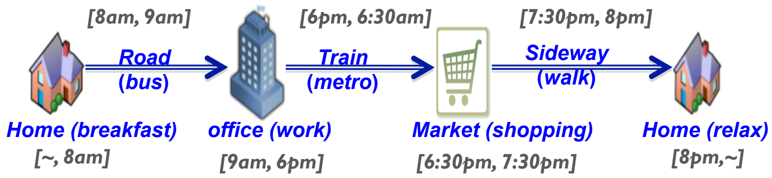
- Current research agenda:
  - **Modeling** issues
    - From 'raw GPS' to semantically-annottated trajectories
  - **Benchmarking** issues
    - Generating realistic synthetic trajectory data
    - Querying based on emerging **location-based social networking** (LBSN) apps
  - **Advanced MOD** implementations
    - centralized vs. distributed vs. noSQL architectures

From “raw” to “semantic” trajectories

raw mobility data
sequence (x,y,t) points
e.g., GPS feeds



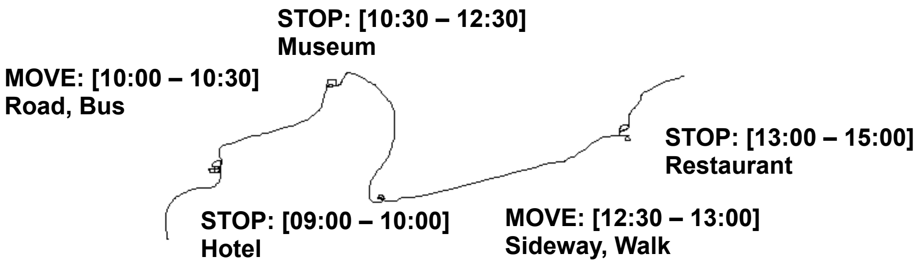
meaningful mobility tuples
<place, $time_{in}$, $time_{out}$, tags>



- **Semantic Trajectory: T={e_{first},…,e_{last}}**
- **Episode: e_i = (STOP | MOVE, t_{from}, t_{to}, *place, tag*)**

From “raw” to “semantic” trajectories

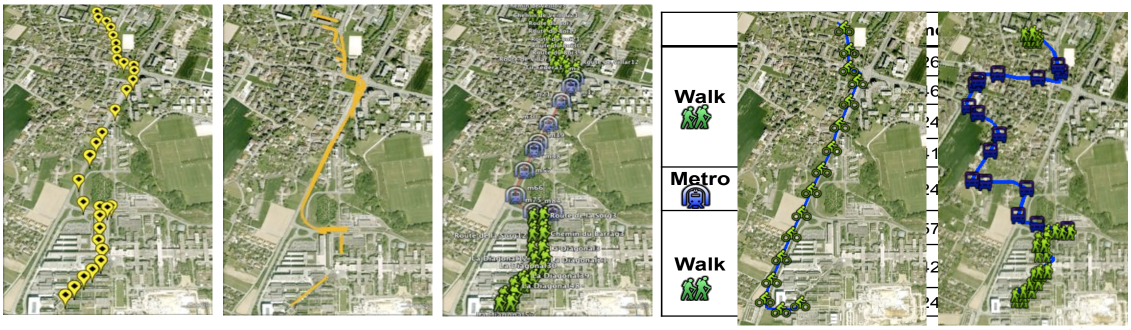
- **Stops** are places (points, regions) where the object stays “static”
- **Moves** are the parts of the object’s trajectory in between two Stops, i.e. where the object is “moving”
- **Tags** are meta-data associated with Stops and Moves
  - information about (at least…) **when? where?** (also…) **how? what? why?**



Why semantic trajectories?

- Detection of homogenous fractions of movement,
  - Trajectory is reconstructed **as a sequence of episodes (stops/moves)**
    - E.g., home, shopping, move by bus, biking, …
- Semantic data abstraction & compression (efficiency/effectiveness)

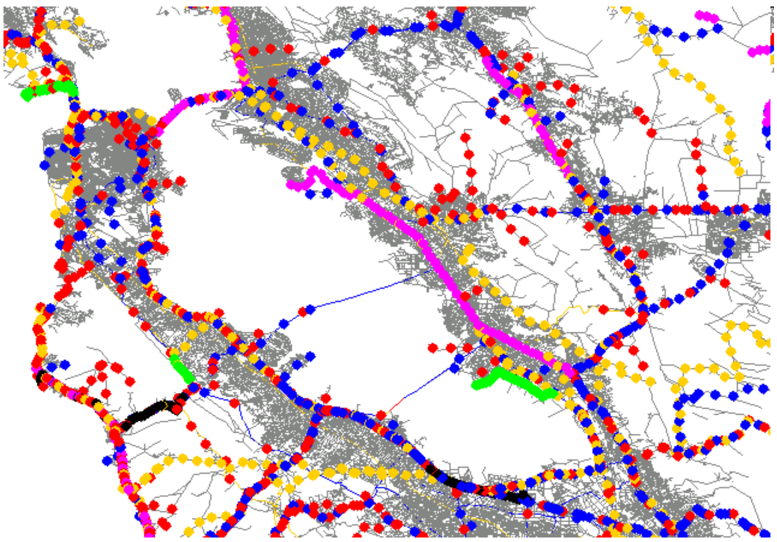
Home-office trajectory examples



Raw GPS Points | Trajectory Notion of Segments | Semantic-Aware Trajectory | (a) HomeOffice via Bike | (b) HomeOffice via Bus

Synthetic trajectory data generators

- Objectives:
  - flexibility,
  - realism,
  - scalability,
  - …
- State-of-the-art
  - Brinkhoff
  - BerlinMOD (on top of SECONDO)
- Challenge:
  - Generate-by-example



source: www.fh-oow.de/institute/iapg/personen/brinkhoff/generator

Questions

Reading list

Mobility data modeling & MOD engines (1)

- de Almeida, V.T. et al. (2006) Querying Moving Objects in SECONDO. Proceedings of MDM.
- Behr, T. and Güting, R.H. (2005) Fuzzy Spatial Objects: An Algebra Implementation in SECONDO. Proceedings of ICDE.
- Brinkhoff, T. (2002) A Framework for Generating Network-based Moving Objects. GeoInformatica, 6(2): 153-180.
- Cao, H. and Wolfson, O. (2005) Nonmaterialized Motion Information in Transport Networks. Proceedings of ICDT.
- Dieker, S. and Güting, R.H. (2000) Plug and Play with Query Algebras: SECONDO – A Generic DBMS Development Environment. Proceedings of IDEAS.
- Düntgen, C. et al. (2009) BerlinMOD: a Benchmark for Moving Object Databases. The VLDB Journal, 18: 1335-1368.
- Güting, R.H. et al. (2000) A Foundation for Representing and Querying Moving Objects. ACM Transactions on Database Systems, 25(1):1-42.
- Güting, R.H. et al. (2006) Modeling and querying moving objects in networks. VLDB Journal, 15(2): 165-190.

Mobility data modeling & MOD engines (2)

- Marketos, G. et al. (2008) Building real-world trajectory warehouses. Proceedings of MobiDE.
- Pelekis, N. et al. (2006) Hermes - A Framework for Location-Based Data Management. Proceedings of EDBT.
- Pelekis, N. et al. (2008) HERMES: aggregative LBS via a trajectory DB engine. Proceedings of ACM SIGMOD.
- Theodoridis, Y. (2003) Ten Benchmark Database Queries for Location-based Services. The Computer Journal, 46(6): 713-725
- Theodoridis, Y. and M. Nascimento (2000) Generating Spatiotemporal Datasets on the WWW. SIGMOD Record, 29(3): 39-43.

MOD query processing (1)

- Benetis, R. et al. (2002) Nearest Neighbor and Reverse Nearest Neighbor Queries for Moving Objects. Proceedings of IDEAS.
- Benkert, M. et al. (2008) Reporting Flock Patterns. Computational Geometry, 41: 111-125.
- Frentzos, E. et al. (2005) Nearest Neighbor Search on Moving Object Trajectories. Proceedings of SSTD.
- Frentzos, E. et al. (2007) Index-based Most Similar Trajectory Search. Proceedings of ICDE.
- Gedik, B., and Liu, L. (2004) MobiEyes: Distributed Processing of Continuously Moving Queries on Moving Objects in a Mobile System. Proceedings of EDBT.
- Gudmundsson, J. and M. van Kreveld (2006) Computing longest duration flocks in trajectory data. Proceedings of ACM-GIS.
- Jensen, C.S. et al. (2003) Nearest Neighbor Queries in Road Networks. Proceedings of ACM-GIS.
- Li, F. et al. (2005) On Trip Planning Queries in Spatial Databases. Proceedings of SSTD.

MOD query processing (2)

- Mokbel, M.F. and W.G. Aref (2007) Location-aware Query Processing and Optimization: A Tutorial. Proceedings of MDM.
- Papadias, D. et al. (2003) Query Processing in Spatial Network Databases. Proceedings of VLDB.
- Pelekis, N. et al. (2007) Similarity Search in Trajectory Databases. Proceedings of TIME.
- Porkaew, K. et al. (2001) Querying Mobile Objects in Spatio-Temporal Databases. Proceedings of SSTD.
- Shekhar, S. and Yoo, J. S. (2003) Processing In-Route Nearest Neighbor Queries: A Comparison of Alternative Approaches. Proceedings of GIS.
- Sankaranarayanan, J. et al. (2005) Efficient Query Processing on Spatial Networks. Proceedings of ACM-GIS.
- Tao, Y. et al. (2002) Continuous Nearest Neighbor Search. Proceedings of VLDB.
- Xia, T. and Zhang, D. (2006) Continuous Reverse Nearest Neighbor Monitoring. Proceedings of ICDE.

MOD Indexing (1)

- Cai, Y. and Ng, R.T. (2004) Indexing Spatio-Temporal Trajectories with Chebyshev Polynomials. Proceedings of ACM SIGMOD.
- Chen, S. et al. (2008) ST^{2}B-tree: A Self-Tunable Spatio-Temporal B+-tree Index for Moving Objects. Proceedings of ACM SIGMOD.
- Frentzos, E. (2003) Indexing Objects Moving on Fixed Networks. Proceedings of SSTD.
- Hadjieleftheriou, M. et al. (2006) Indexing Spatio-temporal Archives. VLDB Journal, 15(2): 143-164.
- Jensen, C. S. et al. (2004) Query and update efficient B$^+$-tree based indexing of moving objects. Proceedings of VLDB.
- Kollios, G. et al. (2001) Indexing Animated Objects Using Spatiotemporal Access Methods. IEEE Trans. Knowledge and Data Engineering, 13(5): 758-777.
- Myllymaki, J. and Kaufman, J. (2003) High-Performance Spatial Indexing for Location-Based Services. Proceedings of WWW.
- Ni, J. and Ravishankar, C.V. (2007) Indexing Spatio-Temporal Trajectories with Efficient Polynomial Approximations. IEEE Trans. Knowledge and Data Engineering, 19(5): 663-678.

MOD Indexing (2)

- Pelanis, M. et al. (2006). Indexing the past, present, and anticipated future positions of moving objects. ACM Trans. Database Systems, 31(1): 255-298.
- Pfoser, D. et al. (2000) Novel Approaches to the Indexing of Moving Object Trajectories. Proceedings of VLDB.
- Rasetic, S. et al. (2005) A Trajectory Splitting Model for Efficient Spatio-Temporal Indexing. Proceedings of VLDB.
- Saltenis, S. et al. (2000) Indexing the Positions of Continuously Moving Objects. Proceedings of ACM SIGMOD.
- Saltenis, S. and C.S. Jensen (2002) Indexing of Moving Objects for Location-Based Services. Proceedings of ICDE.
- Tao, Y. and Papadias, D. (2001) MV3R-Tree: A Spatio-Temporal Access Method for Timestamp and Interval Queries. Proceedings of VLDB.
- Tao, Y. et al. (2003) The TPR*-tree: an optimized spatio-temporal access method for predictive queries. Proceedings of VLDB.

Semantic trajectories

- Baglioni, M. et al. (2009). Towards Semantic Interpretation of Movement Behavior. Proceedings of AGILE .
- Parent, C. et al. (2013). Semantic Trajectories Modeling and Analysis. ACM Computing Surveys, to appear
- Spaccapietra, S. et al. (2008). A Conceptual View On Trajectories. Data and Knowledge Engineering, 65(1).
- Spinsanti, L. et al. (2010). Where you stop is who you are: understanding peoples' activities. Proceedings of BMI.
- Xie, K. et al. (2009). From trajectories to activities: a spatio-temporal join approach. Proceedings of GIS-LBSN.
- Yan, Z. et al. (2010). A Hybrid Model and Computing Platform for Spatio-Semantic Trajectories. Proceedings of ESWC
- Yan, Z. et al. (2011). SeMiTri: A Framework for Semantic Annotation of Heterogeneous Trajectories. Proceedings of EDBT.
- Yan, Z. et al. (2011). SeTraStream: Semantic-aware Trajectory Construction over Streaming Movement Data. Proceedings of SSTD.
- Zheng, Y.et al. (2010). Understanding transportation modes based on GPS data for Web applications. ACM Transaction on the Web, 4(1)