Overview

1. BOGORNY, V., and S. SHEKHAR. "Tutorial on Spatial and Spatio-Temporal Data Mining." *Part ii-Trajectory*.

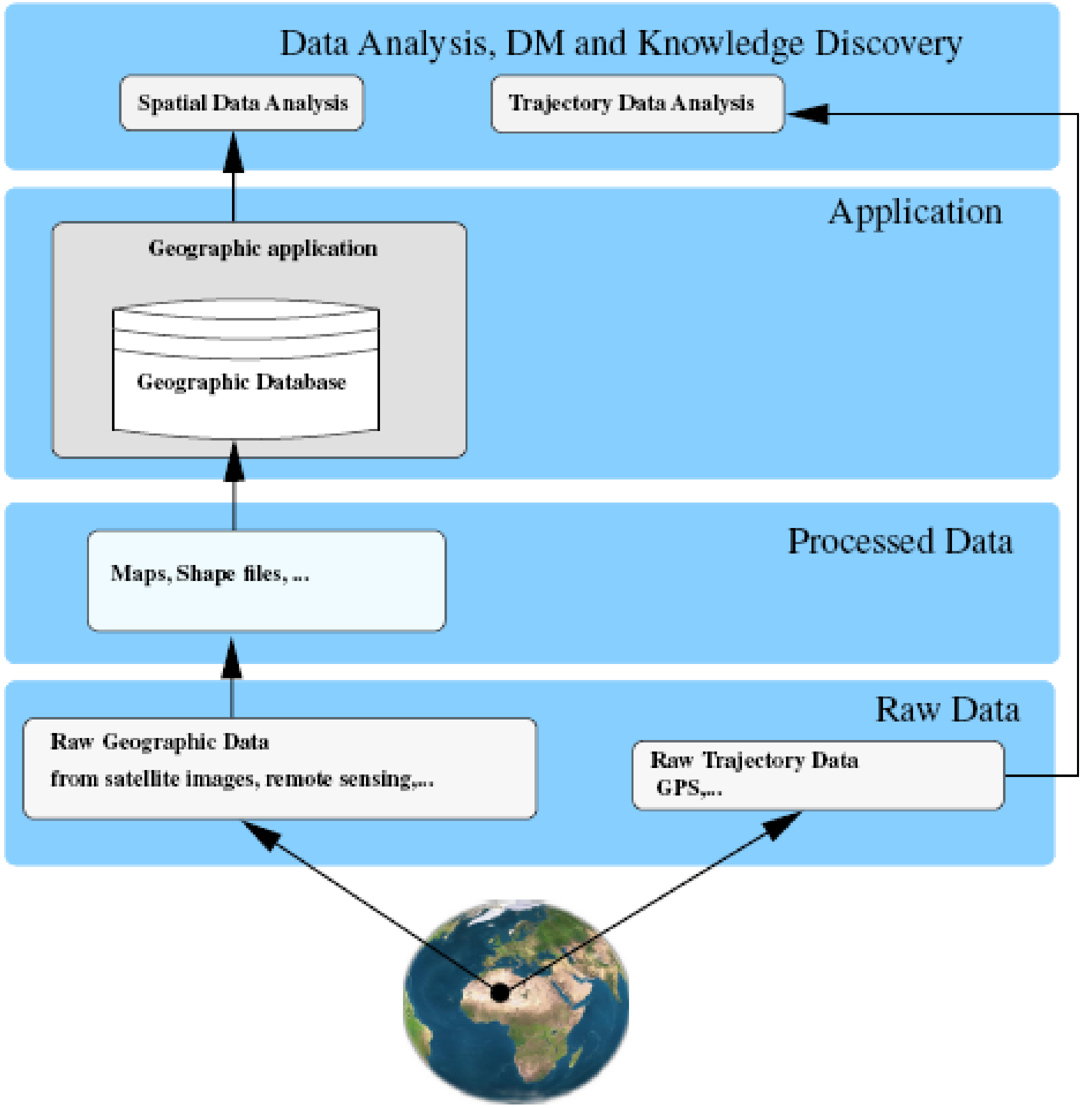
Spatial-Temporal Data

- Spatial-temporal data is a special case of time series where (one of) the information recorded at each time point is the location of an object
- A time series over spatial locations is also called “trajectory”
- Often, there is additional information on time slots (e.g. semantic information on the location such as “museum” or “airport” …)
- We review the some of the recent trends in mining spatial-temporal (aka: spatio-temporal) data

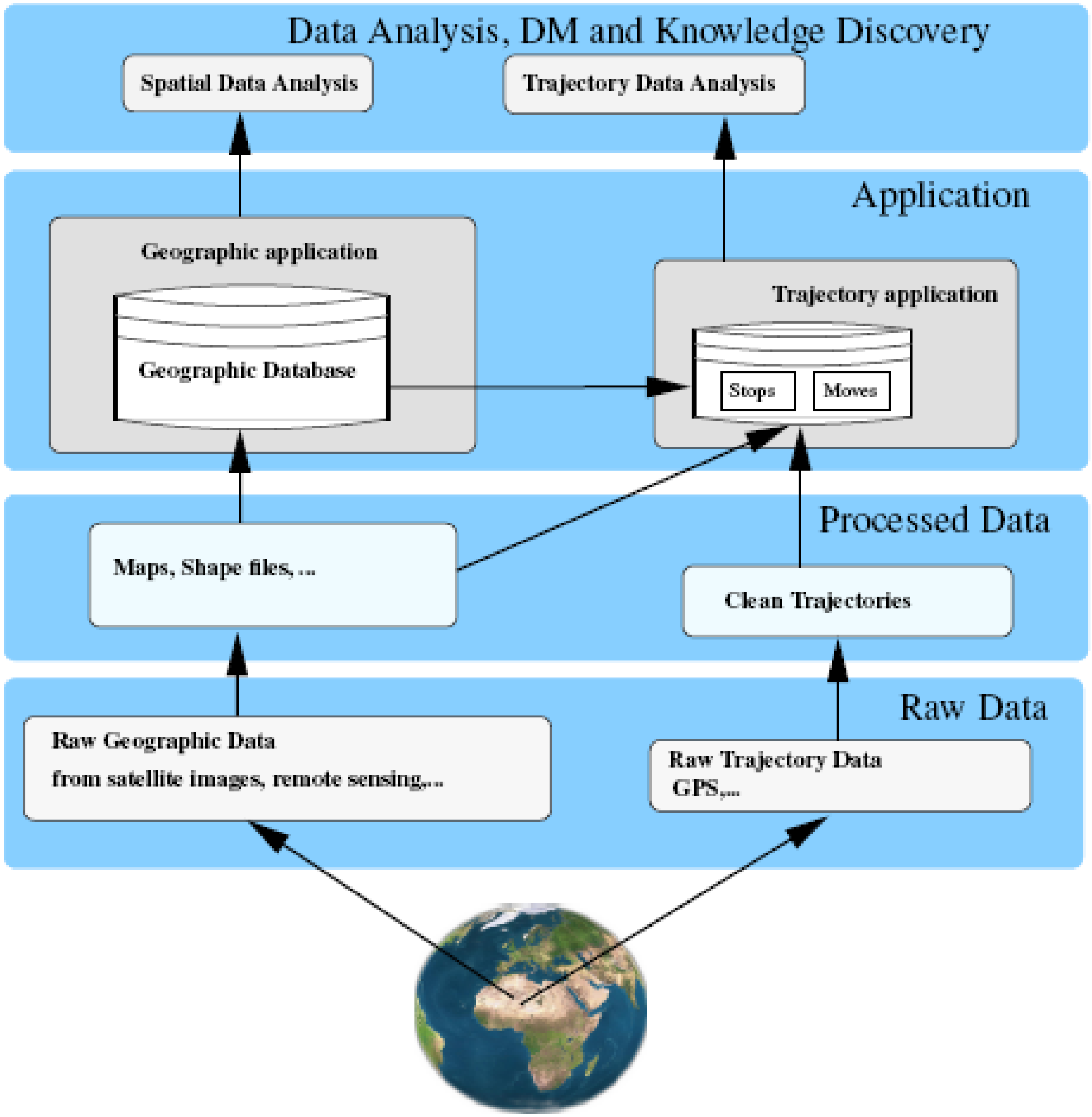
Spatial-Temporal Data

- In general, there are two major approaches to trajectory mining:
  - **Geometry-based methods** consider only geometrical properties of trajectories; they focus on "location-based" similarity
  - **Semantic-based methods** compute patterns based on the semantics of the data and are somewhat independent of the specific spatial locations

Geometry-based approach



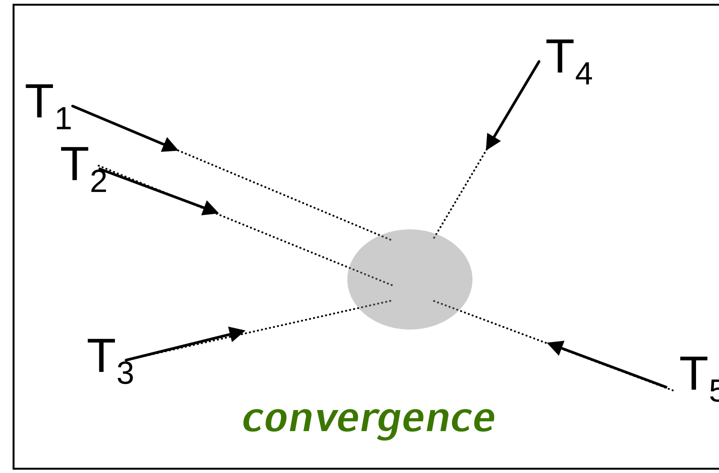
Semantic-based approach



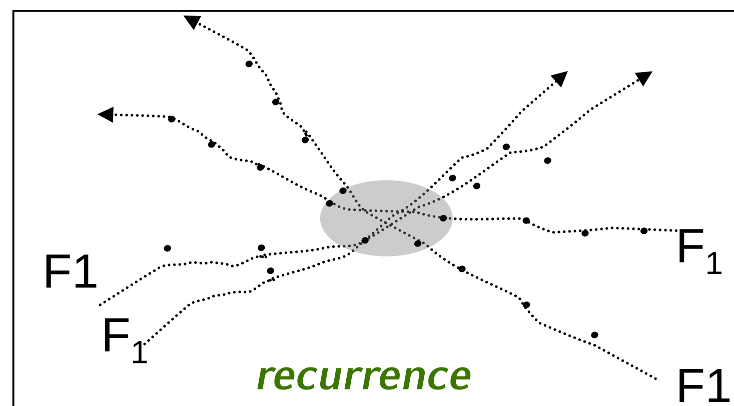
Geometry-based Trajectory Mining

- Laube et al. 2004 proposed five patterns based on location, direction, and/or movement:

1. **Convergence**: At least *m* entities pass through the same circular region of radius r (regardless of time)

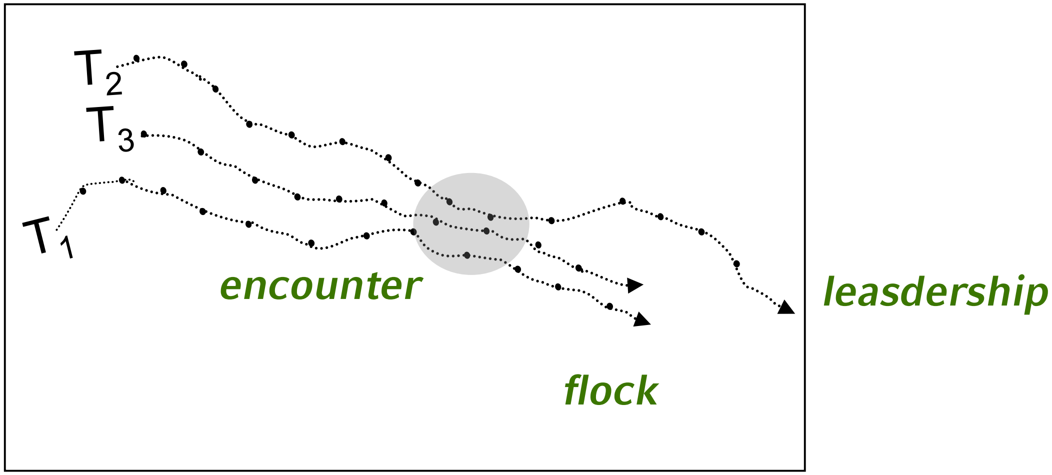


2. **Recurrence**: at least *m* entities visit a circular region at least *k* times



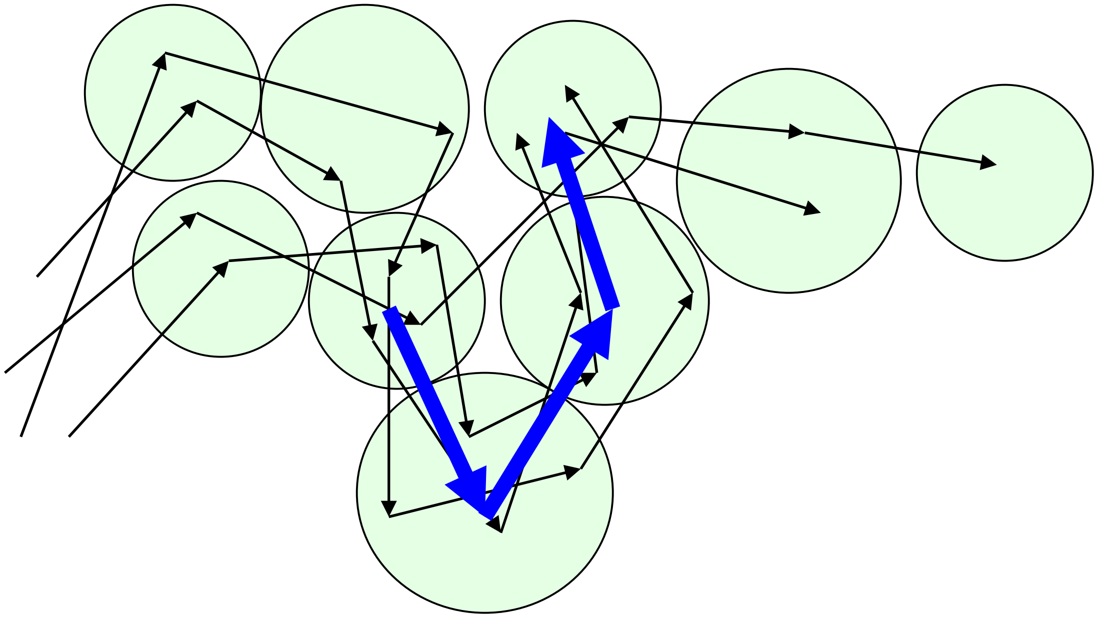
Geometry-based Trajectory Mining

3. **Flock pattern**: At least *m* entities are within a region of radius *r* and move in the same direction during a time interval >= *s* (e.g. traffic jam)
4. **Leadership**: At least *m* entities are within a circular region of radius *r*, they move in the same direction, and at least one of the entities is heading in that direction for at least *t* time steps. (e.g. bird migration)
5. **Encounter**: At least *m* entities will be concurrently inside the same circular region of radius *r*, assuming they move with the same speed and direction. (e.g. traffic jam at some moment if cars keep moving in the same direction)



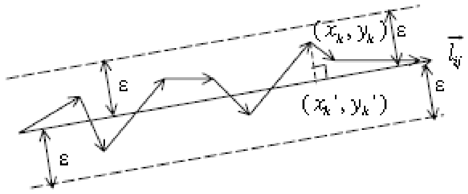
Geometry-based Trajectory Mining

- Frequent patterns: frequent followed paths/frequent sequential patterns



Geometry-based Trajectory Mining

- Computing frequent sequential patterns (e.g. Cao 2005):
  1. Transforms each trajectory in a line with several segments
     - A distance tolerance measure is defined
     - All trajectory points inside this distance are summarized in one segment

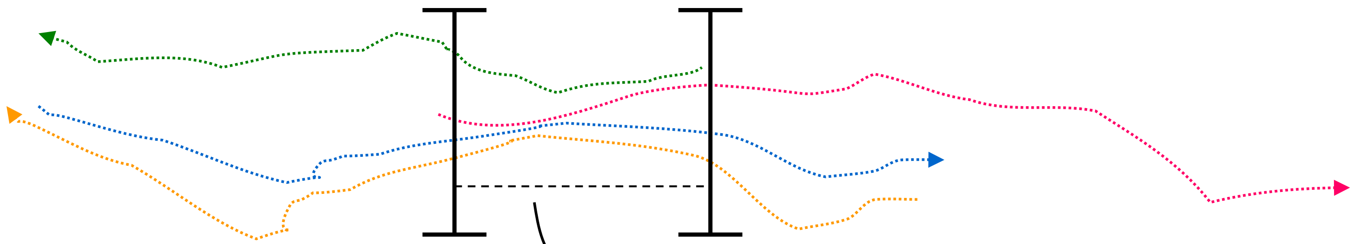


(a) Segment complies with $\vec{l_{ij}}$

  2. Similar segments are grouped
     - Similarity is based on the angle and the spatial lenght of the segment
       - Segments with same angle and length have their distance checked based on a given distance threshold
     - From the resulting groups, a medium segment is created
       - From this segment a region is created
  3. Frequent sequences of regions are computed considering a minSup threshold

Geometry-based Trajectory Mining

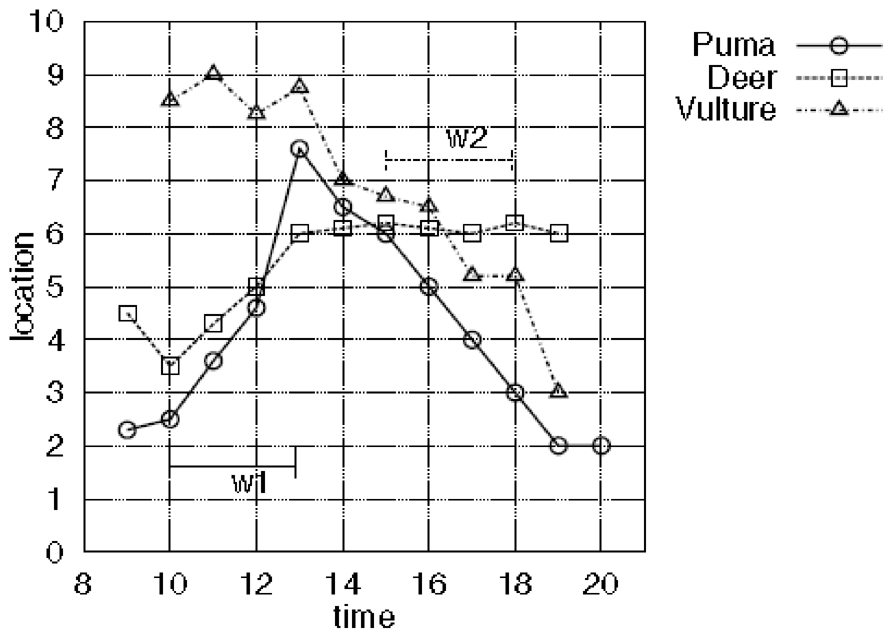
- Frequent mobile group patterns (Hwang 2005):
  - A group pattern is a set of trajectories close to each other (with distance less than a given *minDist*) for a minimal amount of time (*minTime*)
  - Direction is not considered
  - Use Apriori algorithm to compute frequent groups



Group pattern: *time, distance, and minsup*

Geometry-based Trajectory Mining

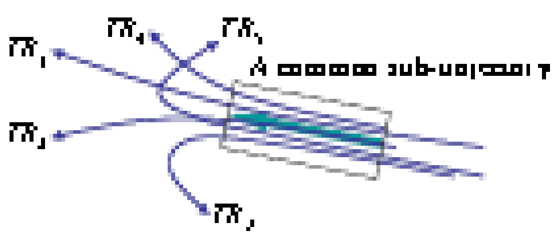
- Co-location Patterns (Cao 2006):
  - Co-location episodes in spatio-temporal data
  - Trajectories are spatially close in a time window and move together



Geometry-based Trajectory Mining

- Trajectory Clustering (Han 2007):
  - Algorithm TraClus: Group sub-trajectories using a density based clustering algorithm
  - 2 step approach
    1. Partition each trajectory in line segments with a user defined length L
    2. Cluster similar line segments based on spatial proximity of the time points
  - Similarity of line segments: Euclidean distance between segments (sub-trajectories); in theory: could be anything else

    => however, time is not considered in this approach

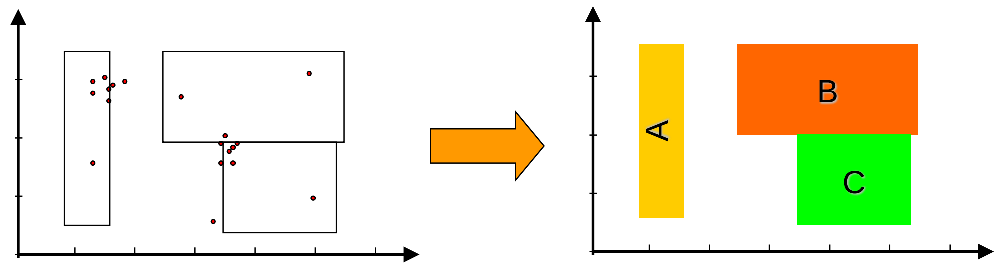


Geometry-based Trajectory Mining

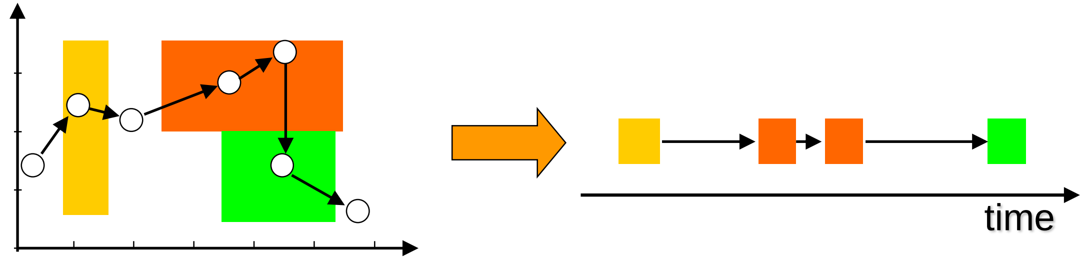
- Sequential Trajectory Pattern Mining (T-Patterns; Giannotti 2007):
  - Considers both space and time
  - Describes frequent behavior in terms of visited regions (ROIs)
  - Three-step approach
    1. Compute regions of interest (ROIs), i.e., regions with many trajectories (regardless of time)
    2. Transform trajectory into sequence of ROIs: select trajectories intersecting at least two regions in a sequence and annotate the time traveled between regions
    3. Compute T-Patterns, i.e., sequences of regions visited during the same time intervals

Geometry-based Trajectory Mining

– Visualization of the idea of T-Patterns:
  - Regions of interest (ROIs)



  - Transform trajectory into a sequence of ROIs

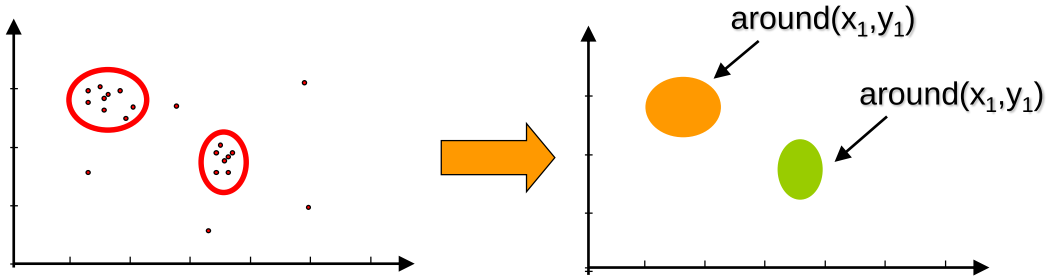


  - Sample pattern:

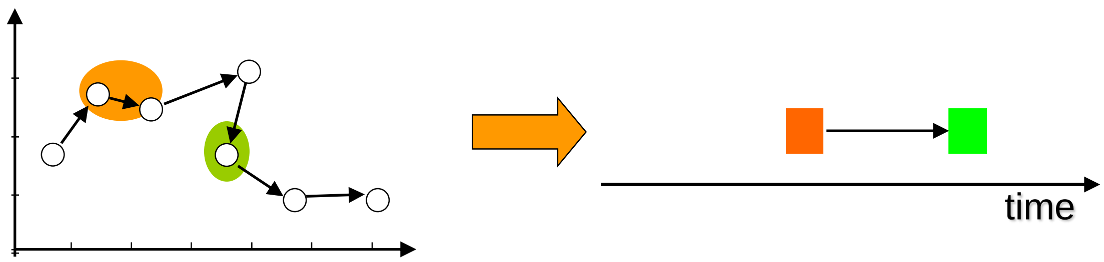
$$A \xrightarrow{20\,\text{min.}} B$$

Geometry-based Trajectory Mining

- Visualization of the approach
  - Step 1: detection of ROIs



  - Step 2: transformation

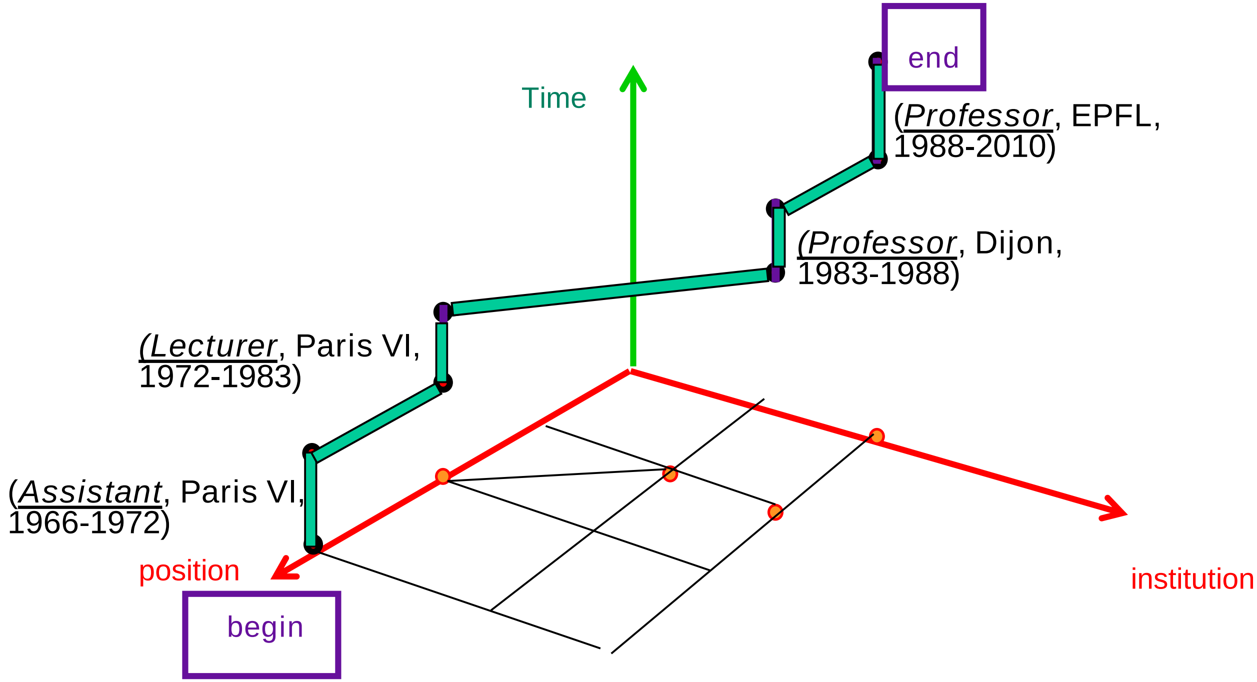


  - Compute pattern:

$$around(x_1, y_1) \xrightarrow{20\text{min.}} around(x_2, y_2)$$

Semantic based method

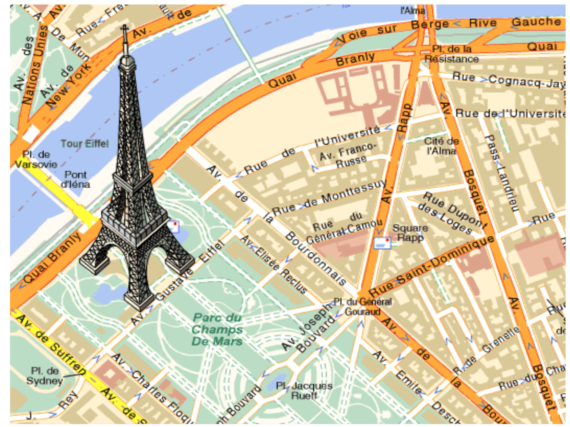
- A Conceptual View on Trajectories (Spaccapietra 2008)
  - Trajectory is a spatio-temporal object that has generic features (independent of the application) and ***semantic*** features (depend on the application
  - Trajectory = travel in abstract space, e.g. 2D career space:



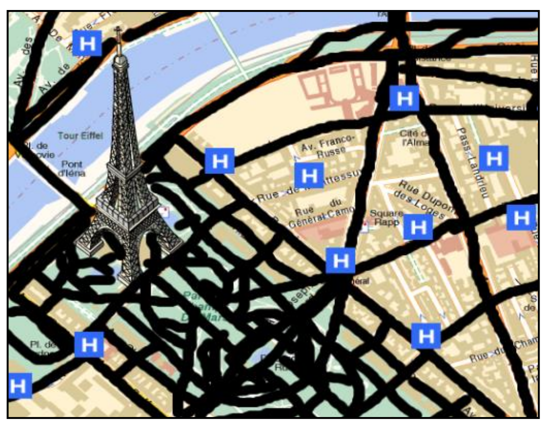
Semantic based method

- Semantic trajectories = geo data + trajectory data

Trajectory Samples (x,y,t)



Geographic Data



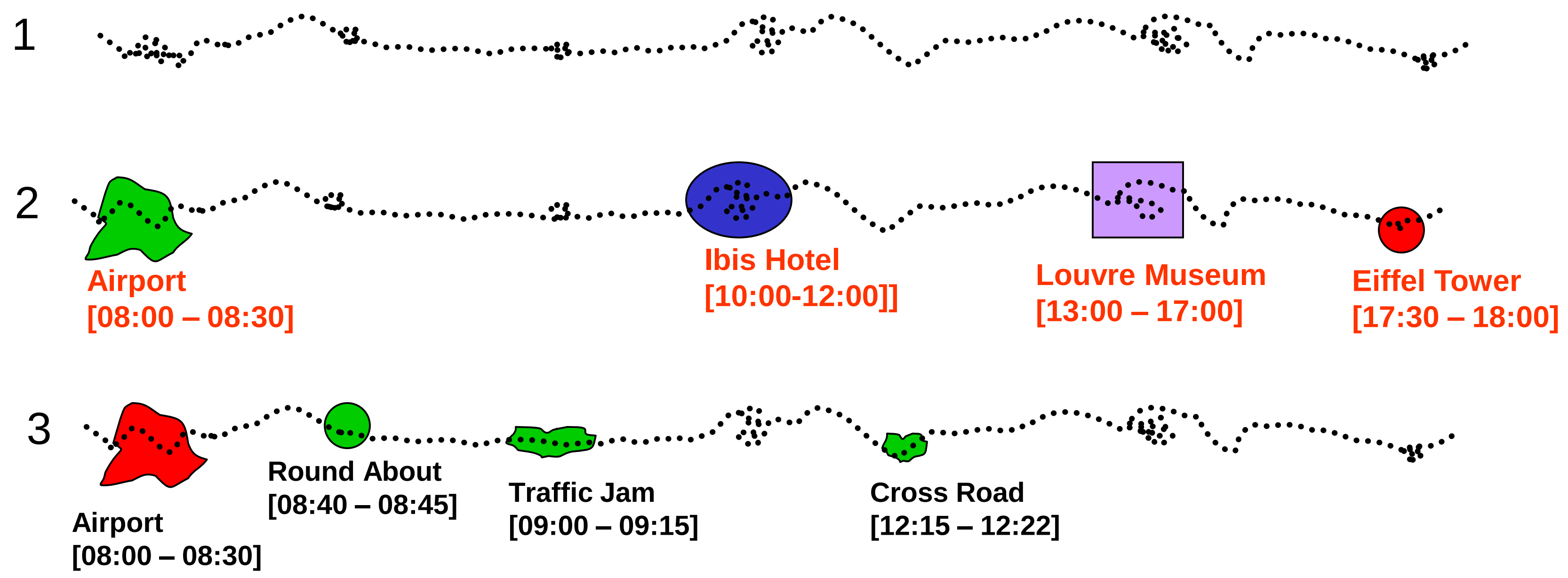
Geographic Data + Trajectory Data = **Semantic Trajectories**

Semantic based method

- Difference between stops and moves
  - STOPS
    - Important parts of trajectories
    - Where the moving object has stayed for a minimal amount of time
    - Stops are application dependent
    - Tourism application: Hotels, touristic places, airport, …
    - Traffic Management Application: Traffic lights, roundabouts, big events…
  - MOVES
    - Are the parts that are not stops

Semantic based method

- Stops and moves are independent of the application



Semantic based method

- Geometric Patterns enriched by semantics (Bogorny 2008):
  - Very little semantics in most trajectory mining approaches (geometry-based approaches)

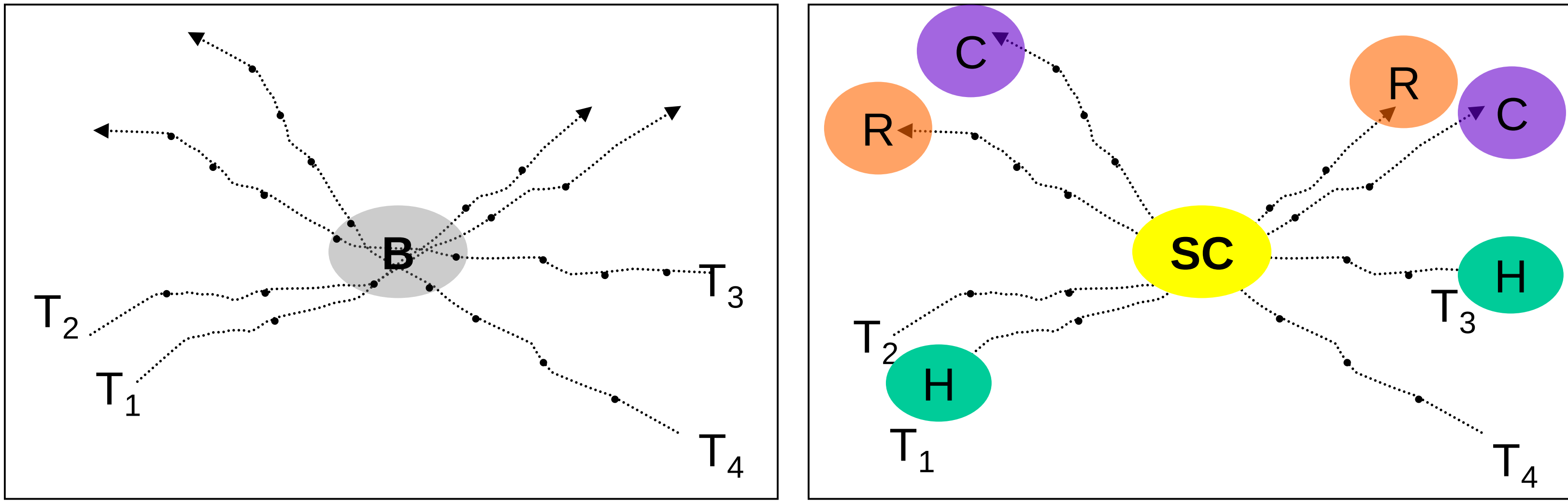
  Thus:
  - Patterns are purely geometrical
  - Hard to interpret

  Thus:
  - Enrich geometric patterns with semantic information

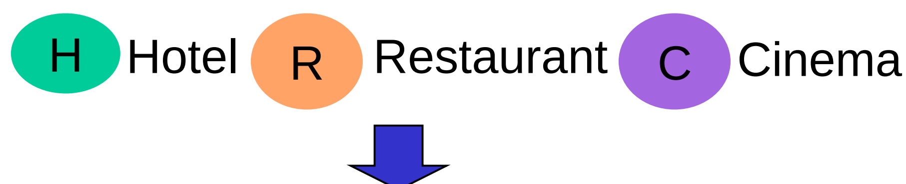
    (stimulated many approaches on how to add semantics to trajectories)

Semantic based method

- Semantic Enrichment (Example):



Geometric Pattern



Semantic trajectory Pattern

(a) Hotel to Restaurant, passing by SC
(b) go to Cinema, passing by SC

Semantic based method

- Stop and Move computation: SMoT (Alvares 2007a)
  - A *candidate stop C* is a tuple (R_C, Δ_C), where
    - R_C is the geometry of the candidate stop (spatial feature type)
    - Δ_C is the *minimal time duration*

    E.g. [Hotel - 3 hours]

  - An application *A* is a finite set

    A = {C_1 = (R_{C1} , Δ_{C1}), …, C_N = (R_{CN} , Δ_{CN})} of *candidate stops* with non-overlapping geometries R_{C1}, … ,R_{CN}

    E.g. [Hotel - 3 hours, Museum – 1 hour]

Semantic based method

- Stop and Move computation: SMoT (Alvares 2007a)

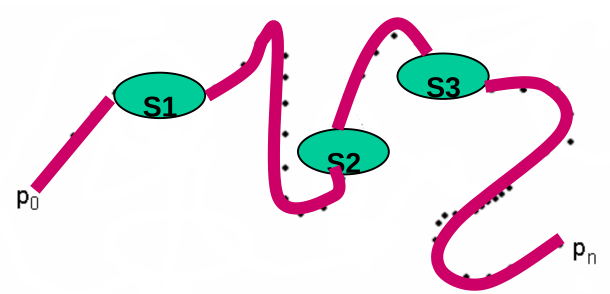
A *stop* of a trajectory T with respect to an *application A* is a tuple (R_{Ck}, t_j, t_{j+n}), such that a maximal subtrajectory of

$$T\{(x_i, y_i, t_i) \mid (x_i, y_i) \textit{ intersects } R_{Ck}\} = \{(x_j, y_j, t_j), (x_{j+1}, y_{j+1}, t_{j+1}), \ldots, (x_{j+n}, y_{j+n}, t_{j+n})\}$$

where R_{Ck} is the geometry of C_k and $|\ t_{j+n} - t_j\ | \geq \Delta_{Ck}$

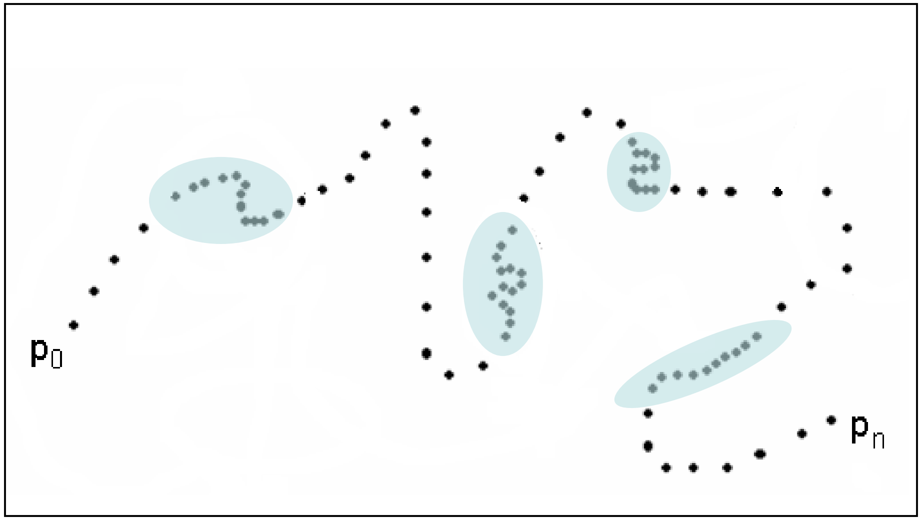
A *move* of T with respect to A is:

- a maximal contiguous subtrajectory of T:
  - between the starting point of T and the first stop of T; OR
  - between two consecutive stops of T; OR
  - between the last stop of T and the ending point of T;
- or the trajectory T itself, if T has no stops.



Semantic based method

- Improvement: CB-SMoT (Palma 2008)
  - Cluster based: cluster trajectories based on speed
  - Low speed => important place
  - Algorithm similar to SMoT but clusters trajectory points first and adds semantics to clusters



Semantic based method

- Comparison: SMoT vs. CB-SMoT (Application: transportation)

Readings on Spatial-Temporal Data

- Geometric based methods

Huiping Cao, Nikos Mamoulis, David W. Cheung: Discovery of Periodic Patterns in Spatiotemporal Sequences. IEEE Trans. Knowl. Data Eng. 19(4): 453-467 (2007)

Panos Kalnis, Nikos Mamoulis, Spiridon Bakiras: On Discovering Moving Clusters in Spatio-temporal Data. SSTD, 364-381 (2005)

Florian Verhein, Sanjay Chawla: Mining spatio-temporal patterns in object mobility databases. Data Min. Knowl. Discov. 16(1): 5-38 (2008)

Florian Verhein, Sanjay Chawla: Mining Spatio-temporal Association Rules, Sources, Sinks, Stationary Regions and Thoroughfares in Object Mobility Databases. DASFAA, 187-201 (2006)

Changqing Zhou, Dan Frankowski, Pamela J. Ludford, Shashi Shekhar, and Loren G. Terveen. Discovering personally meaningful places: An interactive clustering approach. ACM Trans. Inf. Syst., 25(3), 2007.

Cao, H., Mamoulis, N., and Cheung, D. W. (2005). Mining frequent spatio-temporal sequential patterns. In ICDM ’05: Proceedings of the Fifth IEEE International Conference on Data Mining, pages 82–89, Washington, DC, USA. IEEE Computer Society.

Readings on Spatial-Temporal Data

- Geometric based methods (cont.)

Laube, P. and Imfeld, S. (2002). Analyzing relative motion within groups of trackable moving point objects. In Egenhofer, M. J. and Mark, D. M., editors, GIScience, volume 2478 of Lecture Notes in Computer Science, pages 132–144. Springer.

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

Laube, P., van Kreveld, M., and Imfeld, S. (2005b). Finding REMO: Detecting Relative Motion Patterns in Geospatial Lifelines. Springer.

Lee, J.-G., Han, J., and Whang, K.-Y. (2007). Trajectory clustering: a partition-and-group framework. In Chan, C. Y., Ooi, B. C., and Zhou, A., editors, SIGMOD Conference, pages 593–604. ACM.

Li, Y., Han, J., and Yang, J. (2004). Clustering moving objects. In KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 617–622, New York, NY, USA. ACM Press.

Nanni, M. and Pedreschi, D. (2006). Time-focused clustering of trajectories of moving objects. Journal of Intelligent Information Systems, 27(3):267–289.

Readings on Spatial-Temporal Data

- Geometric based methods (cont.)

Verhein, F. and Chawla, S. (2006). Mining spatio-temporal association rules, sources, sinks, stationary regions and thoroughfares in object mobility databases. In Lee, M.- L., Tan, K.-L., and Wuwongse, V., editors, DASFAA, volume 3882 of Lecture Notes in Computer Science, pages 187–201. Springer.

Gudmundsson, J. and van Kreveld, M. J. (2006). Computing longest duration flocks in trajectory data. In [de By and Nittel 2006], pages 35–42.

Gudmundsson, J., van Kreveld, M. J., and Speckmann, B. (2007). Efficient detection of patterns in 2d trajectories of moving points. GeoInformatica, 11(2):195–215.

Hwang, S.-Y., Liu, Y.-H., Chiu, J.-K., and Lim, E.-P. (2005). Mining mobile group patterns: A trajectory-based approach. In Ho, T. B., Cheung, D. W.-L., and Liu, H., editors, PAKDD, volume 3518 of Lecture Notes in Computer Science, pages 713–718. Springer.

Cao, H., Mamoulis, N., and Cheung, D. W. (2006). Discovery of collocation episodes in spatiotemporal data. In ICDM, pages 823–827. IEEE Computer Society.

Readings on Spatial-Temporal Data

- Semantic based method

Bogorny, V. ; Wachowicz, M. A Framework for Context-Aware Trajectory Data Mining. In: Longbing Cao, Philip S. Yu, Chengqi Ahang, Huaifeng Zhang. (Org.). Domain Driven Data Mining: Domain Problems and Applications. 1 ed. : Springer, 2008a.

Bogorny, V., Kuijpers, B., and Alvares, L. O. (2008b). St-dmql: a semantic trajectory data mining query language. International Journal of Geographical Information Science. Taylor and Francis, 2008.

Palma, A. T; Bogorny, V.; Kuijpers, B.; Alvares, L.O. *A Clustering-based Approach for Discovering Interesting Places in Trajectories.* In: 23rd Annual Symposium on Applied Computing, (ACM-SAC'08), Fortaleza, Ceara, 16-20 March (2008) Brazil. pp. 863-868.

Spaccapietra, S., Parent, C., Damiani, M. L., de Macedo, J. A., Porto, F., and Vangenot, C. (2008). A conceptual view on trajectories. Data and Knowledge Engineering, 65(1):126–146.

Alvares, L. O., Bogorny, V., de Macedo, J. F., and Moelans, B. (2007a). Dynamic modeling of trajectory patterns using data mining and reverse engineering. In Twenty-Sixth International Conference on Conceptual Modeling - ER2007 - Tutorials, Posters, Panels and Industrial Contributions, volume 83, pages 149–154. CRPIT.

Alvares, L. O., Bogorny, V., Kuijpers, B., de Macedo, J. A. F., Moelans, B., and Vaisman, A. (2007b). A model for enriching trajectories with semantic geographical information. In ACM-GIS, pages 162–169, New York, NY, USA. ACM Press.