Results of the break-out group: Similarity measures

Group discussion with Joachim Gudmundsson, Harvey Miller, Rodrigo Silveira, Mathias Versichele, and Stefan van der Spek

Many of the fundamental problems in trajectory analysis have one issue in common, namely calculating the similarity between two (sub)trajectories. Simplified, the problem is as follows. Given two (polygonal) curves P and Q, one would like to match them up in an optimal way (i.e., find a continuous mapping from the points of P to the points of Q, so that the mapping maps the endpoints of one curve to the endpoints of the other curve). Such a mapping is especially useful if it is associated with an appropriate metric between curves.

However, the choice of distance measure is problematic and crucially dependant on the application area. The most fundamental successful distance measure to this date is probably the Fréchet metric which is one of the most natural measures of this type: It requires finding a mapping f between the curves so that $W(f) = \max_{\{x \in P\}} |x - f(x)|$ is minimized. This is also known as the person-dog metric: imagine a person walking on P and a dog walking on Q. The Fréchet distance between P and Q is the shortest leash that enables both the person and the dog to travel along P and Q with a leash connecting them. However, there are many more fundamental measures in the literature such as the classical Hausdorff distance between the curves, the Longest Common Subsequence model, a combination of parallel distance, perpendicular distance and angle distance, the average Euclidean distances between paths and so on.

In the group discussions we discussed distance measures focussing on real world applications specifically on domain areas where trajectories have been generated by animals (birds, primates...) and humans in urban areas. The members immediately agreed that there is no one generic distance measure that can be used for all applications, instead it will highly dependent on the data quality, the context and the application area.

However, as a first step one should agree upon what properties a good similarity measure should have. After some discussions the group agreed that a good distance measure should fulfil the following properties (at a minimum):

- Comprehensible – easy to understand
- Complete – take all the relevant parameters into account
- Decomposable
- Minimal/non-redundant
- Measurable
- Meaningful/operational

The group then focussed its discussions on which parameters might be of interest for many applications.

- Location
- Velocity

- Sinuosity
- Shape
- Time (e.g. time of the day)
- Direction
- Context
    - activity
    - mode of transport
    - ...

As can be seen from this summary the problem of defining a good distance measure is very complex and requires both expert domain knowledge and knowledge about computability (what can be calculated and how can it be computed?).