Challenges of Context-Aware Movement Analysis – Lessons learned about Crucial Data Requirements and Pre-processing

Christian Gschwend, Patrick Laube

Department of Geography, University of Zurich, Winterthurerstrasse 190, CH – 8057 Zurich
christian.gschwend@geo.uzh.ch, www.geo.uzh.ch/~cgschwen
patrick.laube@geo.uzh.ch, www.geo.uzh.ch/~plaube

Summary: This paper reports on initial insights gained from a project aimed at the development of methods for context-aware movement analysis. We report on two case studies (animals and pedestrians) where we aimed to relate basic derived movement properties (such as speed, turning angle, sinuosity) to the geographic context embedding this movement. We present our lessons learned with respect to data requirements (granularity, accuracy) and pre-processing (segmenting, map matching).

KEYWORDS: Movement analysis, moving objects, movement parameters, geographic context

1. Introduction

GIScience has seen significant progress in analysing *second order effects* (O'Sullivan and Unwin, 2010) in movement analysis, such as arrangement patterns (e.g. flocks or leadership patterns, Laube *et al.*, 2005, Andersson *et al., 2007*) or trajectory similarity and clustering (Buchin *et al.*, 2009, Pelekis *et al.*, 2007). Much less work has been done investigating *first order effects*, assuming that movement properties and patterns also emerge due to the variability of the embedding geographical context – for example, a timid deer may speed up when crossing a forest clearing, but leave a sinuous slow trace when foraging. This paper reports on initial insights gained from a project developing methods for context-aware movement analysis. We report on two case studies (trajectories of animals and shoppers) where we related basic derived movement properties (such as speed, turning angle, sinuosity) to the geographic context embedding this movement. Here we present our lessons learned with respect to data requirements and pre-processing.

2. Problem Statement

On the *movement* side, we use GPS localization that allows for quasi-continuous tracking of moving individuals in space-time (Van der Spek *et al*., 2009). GPS trajectories allow derivation of fine-grained descriptive movement parameters, such as speed, sinuosity, or turning angle (Figure 1). The *geographic context* enabling and constraining movement is clearly application dependent. For wild animals, relevant context might be habitat type or terrain, for shoppers it might include spatio-temporal properties of the urban transit network and personal points of interest (home, work, gym, Figure 1). Note, we do not want to identify *what* context factors are important for a given movement process but rather quantify the movement-context interrelation when we assume we have access to expertise capable of identifying relevant context (i.e. habitat type for a foraging animal).

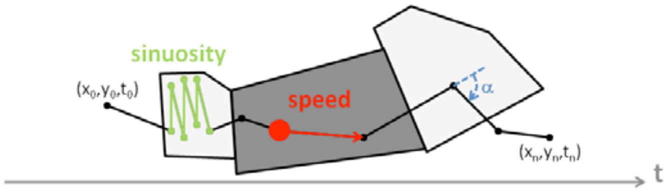

Figure 1. Movement trajectory with derived movement parameters embedded in geographic context.

In this paper we investigate minimal data requirements and crucial pre-processing steps for content-aware movement analysis. In detail, we address the following questions:

- What are *crucial data pre-processing* steps, for movement and context data, enabling context-aware movement analysis?
- Given movement trajectories and distributions of the habitat types (land use) with respect to their constituting fixes: Are basic *exploratory statistics* relating computed movement properties (speed, turning angle, sinuosity) to habitat types an adequate means for context-aware movement analysis?
- What are *minimal requirements* for movement data and geographic context data for the above analysis (with respect to granularity, accuracy, metadata)?

3. Data and Experiments

Case studies were selected from urbanism and behavioural ecology, featuring data with differing properties in terms of temporal resolution and movement space (Table 1). First, we analysed the movement properties of finely sampled trajectories of pedestrians moving in the urban network space of the city of Delft, NL. Here, people leaving a parking deck in the centre of Delft were given a GPS device and their trips through the city were recorded. We used both raw GPS data as well as pre-processed trip data where stationary phases were manually removed. Second, movement data of chamois foraging in the Swiss National Park were used to perform an experiment relating speed to the underlying habitat type. This data set reflects typical data from monitoring studies in behavioural ecology, where technical constraints may dictate rather coarse sampling rates.

Table 1. Characteristics of case study data.

	Pedestrians Delft	**Chamois Swiss National Park**
Temporal resolution	2sec	10min
Space	Network, OpenStreetMap	Euclidean unconstrained
Moving Objects	Pedestrians (*Homo sapiens s.*)	Chamois (*Rupicapra rupicapra*)
Context	Shopping and leisure points of interest (points)	Habitat types (polygons)
Data source	TU Delft, Stefan van der Spek	Swiss National Park
Date	18.11.2009	04.12.2002 – 31.05.2010
Number of points	2'300	29'100

3.1 Case study #1: Filtering and Map Matching

The first case study investigated effects of pre-processing movement data in an urban context. Speed values provided by the GPS device were compared with different ways of computing speed from location fixes, both for raw GPS data and manually filtered trip data (Figure 2). First, speed was calculated from the distance moved within consecutive fixes (sampling rate of 2 seconds, few longer intervals). Second, speed was computed after a naïve map matching (c.f. Bernstein and Kornhauser, 1996, White *et al.*, 2000) technique was applied. For the naïve map matching, fixes were snapped to the closest network edge, with a maximal snapping threshold of 15 meters (Figure 2).

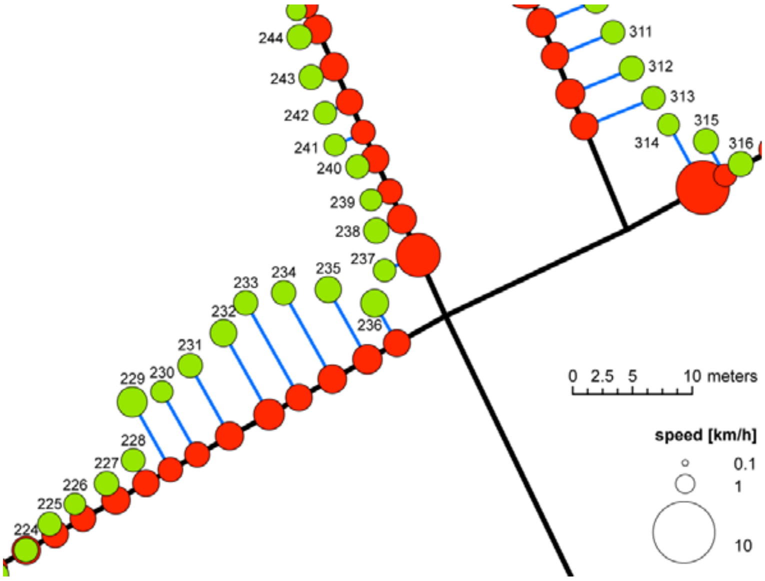

Figure 2. Example trajectory section for a pedestrian in Delft, without (green) and with naïve map matching (red), fix indices at sampling rate of 2 seconds.

3.2 Case study #2: Relating Speed and Habitat Type

The second case study aimed to relate speed to the underlying habitat type embedding the movement of eleven GPS-tracked chamois in the Swiss National Park (Figure 3). A dataset with a temporal resolution of 10 minutes was chosen to investigate whether movement data with such a coarse temporal granularity could be used to relate movement and context. Again, speed was calculated assuming constant speed between two consecutive fixes. Here, raw GPS data was segmented into stops (removed) and moves, using a simple algorithmic approach (Laube and Purves, 2011). Raw and filtered movement data was then related (point-in-polygon) to three habitat types aggregated from a detailed habitat data set (www.habitalp.de).

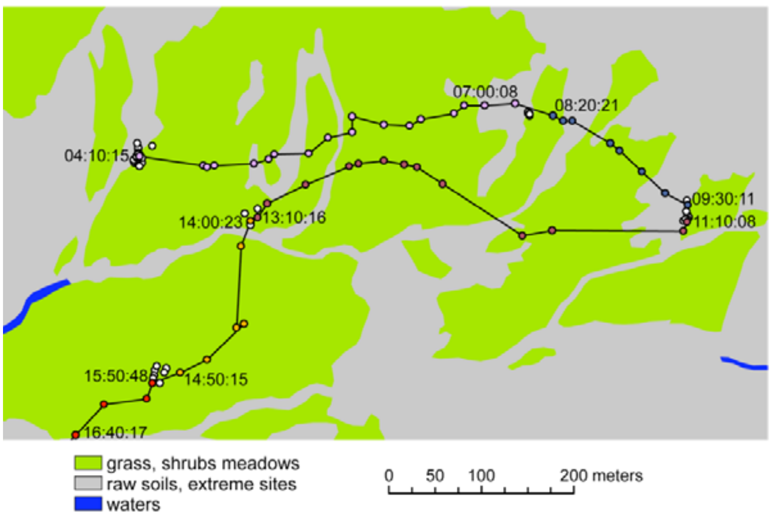

Figure 3. Example trajectory of chamois with habitat context.
Stationary fixes (white), moves in various colours, time of day (hh:mm:ss).

4. Results

For both case studies, speed values were binned and each bin resulted in an item on the ordinate of the box whisker plots (Figure 4). The box whisker plots show medians (horizontal bar), 25th and 75th

percentiles enclosing the middle 50% of the data (boxes, also interquartile range, IQR), minimum and maximum values (whiskers), and outliers (data points more than 1.5 times the IQR from either end of the box). Figure 4a shows results for the Delft pedestrians. The first two items describe speed measurements calculated by the GPS device itself, first for raw (r) and second for filtered data (f). Then follow computed speeds for raw (r), filtered (f), and both filtered and map matched data (f,mm). Figure 4b illustrates variable speed values over three different habitat types (grass, raw soils, forest). Here, for every habitat type raw GPS trajectories are compared to segmented and filtered data (stops removed).

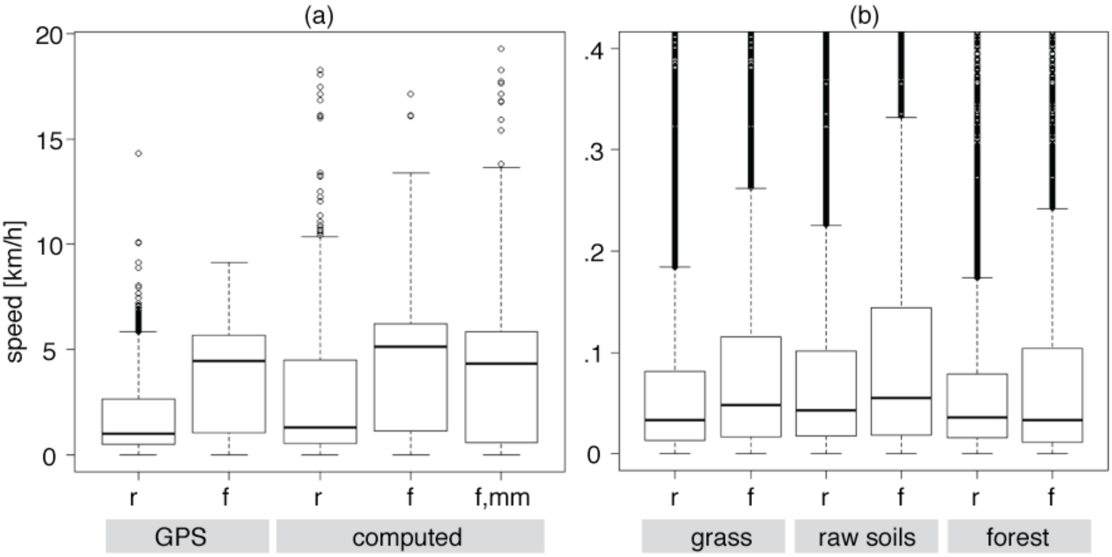

Figure 4. (a) Case study #1, speed for pedestrians; GPS vs. computed; raw data (r), filtered (f) and map matched (mm). (b) Case study #2, speed vs. habitat for chamois; raw data (r) vs. filtered (f).

5. Discussion

Figure 4a first illustrates that separating moves from stops has an important influence on computed speeds (median ~1km/h vs. ~5km/h). Second, the median of all three filtered speed categories (GPS, filtered, filtered & map matched) is in the same order of magnitude. Third, the median for map matched is slightly below the uncorrected signal. We argue that for this result two effects must be considered (Figure 2): (i) Map matching introduces error at intersections, where unrealistically large speed values result from the distorted geometry (fixes 236 to 237 or 313 to 314). (ii) Shadow effects in a 3D urban setting result in positive speed artefacts due to positional inaccuracy of the GPS signal at building transitions (e.g. fixes 228 to 229) – an error removed through map matching. We argue that in our case, the latter effect (building transitions) outnumbers the first (intersections), hence the lower median for (f,mm).

Figure 4b shows no significant difference in speed depending on the embedding habitat types. For grass and raw soils filtering out stops results again in higher speeds. The signal for forest is more complex, with a lower median but a larger range. A reason for this mixed signal could be that in forest, animals move more slowly in general. However, averaged values of speed over time intervals of 10 minutes are in general very low. We argue that such low, averaged speed values do not represent actual instantaneous speed of moving animals. For instance, in Figure 3 the first segment between 04:10:15 and 07:00:18 shows several transitions between habitat types where the granularity of the trajectory hardly allows for a conclusive link between speed and habitat type.

6. Conclusions and Outlook

From these initial experiments linking movement parameters to the embedding geographic context, we conclude with the following list of lessons learned:

- Removing (filtering) stops is a paramount pre-processing step, as pseudo-movement introduced by inaccurate GPS fixes of stationary objects swamps any signal.
- For network bound movement, we argue that there is an unavoidable catch-22 between computing derived movement parameters from unmatched fixes (which may not lie on network edges and are hence erroneous) or from map matched fixes (which must have an altered geometry and hence can't represent the 'true' movement).
- When the temporal granularity of movement data is so coarse that the interval between two consecutive fixes could include several stops and moves, computing instantaneous speed is not suitable, and hence establishing a link between such derived speed properties and movement context is not suitable either.

From these lessons learned we shape our next steps. We intend to continue with the Delft pedestrian data, but will apply more sophisticated map matching techniques that correct for the error sources identified above. Furthermore we started using animal tracking data with a finer temporal granularity. One promising data source is avian navigation research with ample data sampled at sub-second sampling rates.

7. Acknowledgements

The authors thank Stefan van der Spek, Delft University of Technology, Faculty of Architecture, Department of Urbanism, for providing the GPS-tracking data of the pedestrians in Delft (Urbanism on Track project) and his valuable comments. Furthermore, the authors thank Ruedi Haller and Flurin Filli, Swiss National Park, for providing the chamois data and the detailed habitat data. This work was partly supported by the Swiss National Science Foundation, project 200021_129963.

8. References

Andersson M, Gudmundsson J, Laube P and Wolle T (2008). Reporting Leaders and Followers among Trajectories of Moving Point Objects. *Geoinformatica*, 12(4), 497-528.

Bernstein D and Kornhauser A (1996). An introduction to map matching for personal navigation assistants. *Tech. rep.*, TIDE Center, New Jersey.

Buchin K, Buchin M, Van Kreveld M and Luo J (2009). Finding long and similar parts of trajectories. *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '09*, 296-305, ACM Press, New York.

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

Laube P and Purves R S (2011). How fast is a cow? Cross-scale Analysis of Movement Data, *Transaction in GIS*, 15(3), 401-418.

O'Sullivan D and Unwin D J (2010). *Geographic Information Analysis*. 2nd edition, John Wiley and Sons, Hoboken, NJ.

Pelekis N, Kopanakis I, Marketos G, Ntoutsi I, Andrienko G and Theodoridis Y (2007). Similarity Search in Trajectory Databases. *TIME'07: Proceedings of the 14th International Symposium on Temporal Representation and Reasoning*, 129-140, IEEE Computer Society, Washington DC.

Van der Spek S, Van Schaick J, De Bois P and De Haan R (2009). Sensing Human Activity: GPS Tracking. *Sensors*, 9(4), 3033-3055.

White C, Bernstein D and Kornhauser A L (2000). Some map matching algorithms for personal navigation assistants. *Transportation Research Part C: Emerging Technologies*, 8(1-6), 91-108.

Biography

Christian is a PhD student with the GIScience Centre at the University of Zurich. His PhD project aims at the development of GISc methods for context-aware movement analysis. Patrick is a lecturer with the Department of Geography, University of Zurich.