Enhancing the use of Global Positioning Systems data from wearable GPS devices as travel survey method: a case study in the Copenhagen Region

Thomas Kjær Rasmussen*
DTU Transport, Technical University of Denmark
Bygningstorvet 116B, 2800 Kgs. Lyngby, Denmark
Tel.: +45 45256562
E-mail: tkra@transport.dtu.dk

Jesper Bláfoss Ingvardson
DTU Transport, Technical University of Denmark
Bygningstorvet 116B, 2800 Kgs. Lyngby, Denmark
Tel.: +45 45256552
E-mail: jbin@transport.dtu.dk

Katrin Halldorsdottir
DTU Transport, Technical University of Denmark
Bygningstorvet 116B, 2800 Kgs. Lyngby, Denmark
Tel.: +45 45256563
E-mail: katha@transport.dtu.dk

Otto Anker Nielsen
DTU Transport, Technical University of Denmark
Bygningstorvet 116B, 2800 Kgs. Lyngby, Denmark
Tel.: +45 45251514
E-mail: oan@transport.dtu.dk

*Corresponding author

1. Introduction

In recent years much effort has been given to investigate the use of GPS data as the data source for travel surveys ([1-5], etc.), as this ideally provides the investigator with far more detailed travel information than traditional travel survey methods. Additionally, the data collection sets far less requirements to the respondents, as e.g. answering time-consuming questionnaires regarding route choices etc. are avoided. This enables larger sample sizes, but also sets higher requirements to work done in the post processing, where detailed travel information are derived from the raw data. GPS units (or smart-phones etc.) facilitates the collection of complete travel patterns of people, but there is the need for advanced methods to deduct these patterns through applications of e.g. algorithms identifying used modes of travel etc..

This paper presents a new post processing algorithm and its application to raw GPS data from a multi-day person-based GPS travel survey including five modes of transport. To validate the results, corresponding interview-based travel survey data were collected for one of the three days in the survey period.

2. Method

The method applied in this study is based on the trip and mode detection algorithms developed in Schüssler [6], however supplemented by additional more disaggregate components using GIS-software. Overall, the algorithm is a four-step sequential process allowing feed-back loops, consisting of: (i) GPS data cleaning; (ii) Trip and activity identification; (iii) Trip segmentation into single-mode trip legs; and (iv) Travel mode identification. The first three steps of the algorithm are, with minor modifications, based on Schüssler [6]. The major contribution of the present work is the new mode identification step described in the following.

The proposed three-fold mode identification process is shown in Figure 1.

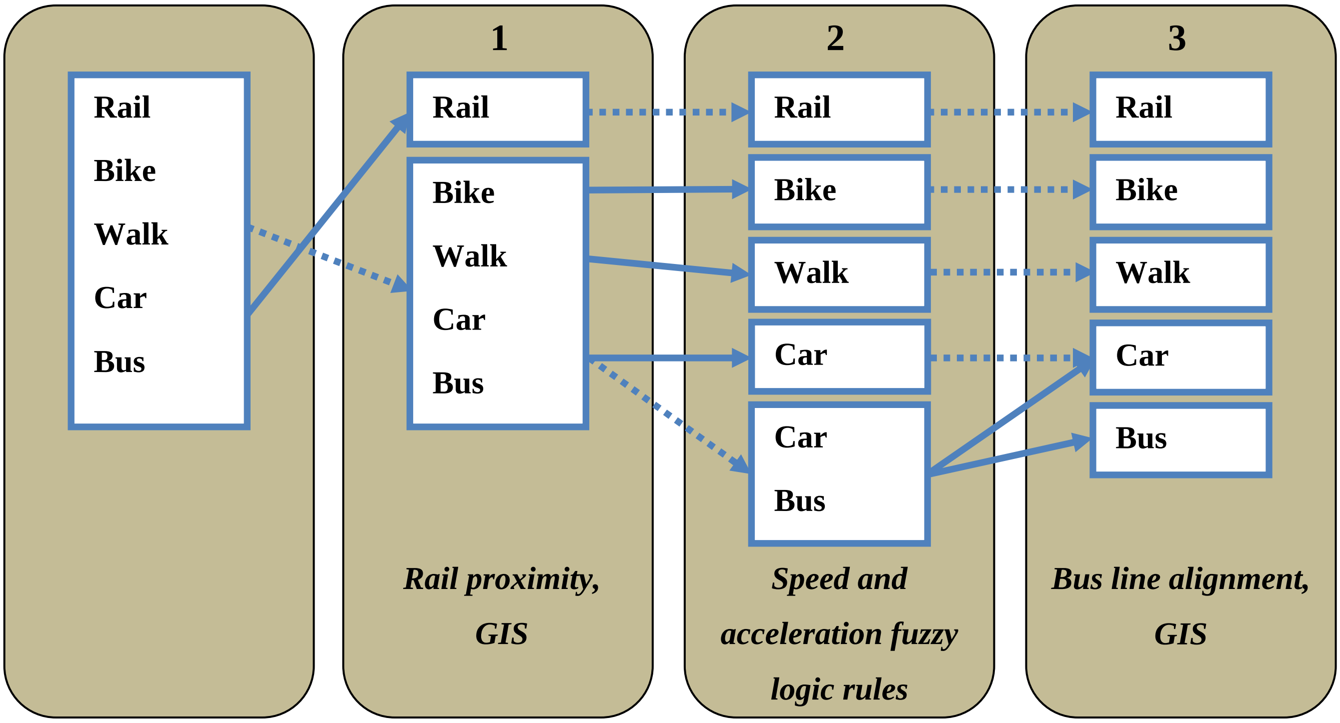


Figure 1: The step-wise mode classification algorithm.

The first step identifies rail trips based on the proximity of observations to the rail network. In the second step fuzzy logic rules ([6]) are applied to speed and acceleration profiles of the remaining trip legs in order to classify the most probable mode among the remaining modes. This successfully classifies walking and biking trip legs, while many car trip legs are wrongly classified as bus trip legs, and vice versa. To improve results we propose a new approach based on GIS-analysis to distinguish between car and bus (Step 3). This approach is based on an analysis of coherence between GPS carrier stopping locations and bus line bus stops; all trip legs are analysed to identify whether they follow the stopping pattern of a bus route. If so, and if the origin and destination of the trip leg are in the proximity of a bus stop of the bus lines identified, then the trip leg is classified as a bus trip leg.

Following the mode identification, a feed-back algorithm detecting irregular mode sequences is run. This is based on various logic rules, and identifies e.g. illogical car-bus-car-trips as well as merges trip legs incorrectly split by the trip identification algorithm. As many trip legs identified by the algorithm were found to be random scatter instead of actual trips, a map matching algorithm was run to eliminate such trip legs. Hence, if the GPS observations of a trip leg cannot be matched to constitute a trip on the network, the trip leg is removed from the dataset.

3. Results

The collected GPS raw data contains 644 person days of travel (equalling 4553 trip legs identified by the algorithm), of which corresponding travel diaries are available for 101 person days. Connecting the data yields 745 (or 464) trip legs identified by the algorithm when excluding (or including) the map matching algorithm. This compares to the 521 trip legs reported in the travel diary.

The feed-back algorithms improved the results of the trip identification notably, cf. Figure 2. As can be seen, removing scatter trip legs through map matching causes the distribution of the trip lengths to represent the reported trip lengths better.

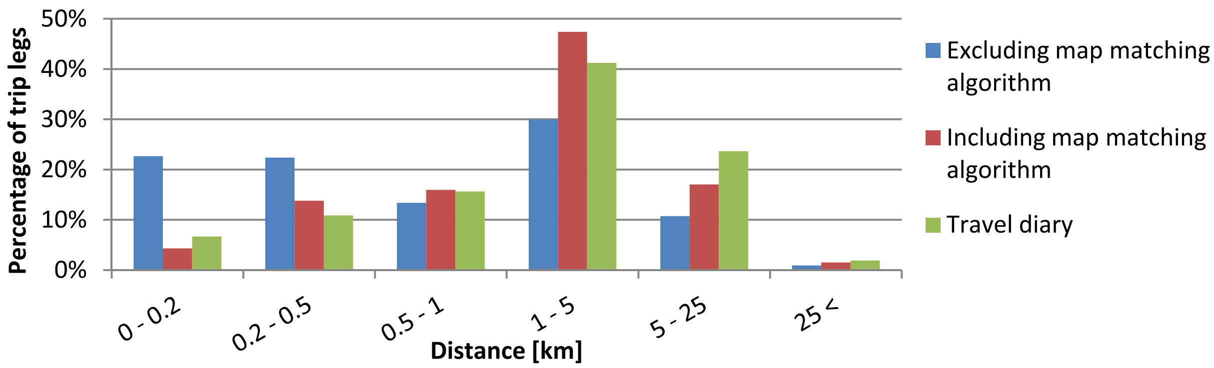


Figure 2: Distribution of trip length for identified trip legs compared to stated travel survey trip lengths.

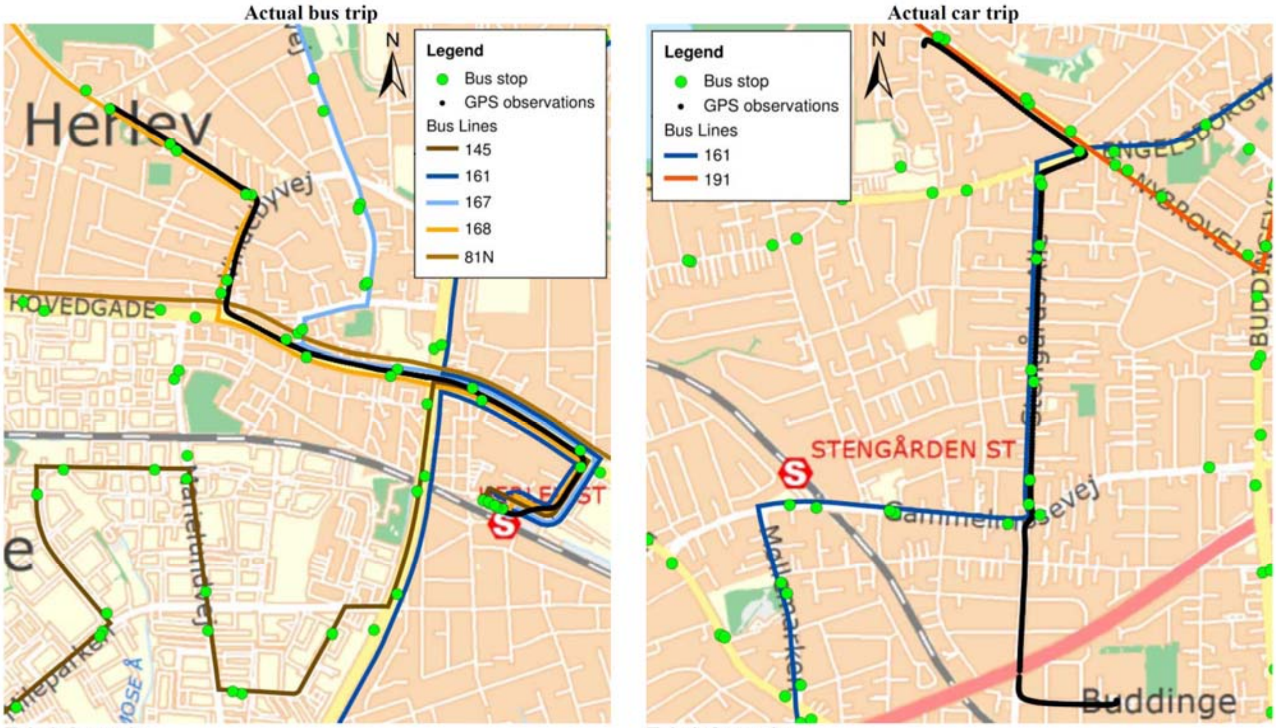
The final results of the mode identification including feed-back and map matching algorithms are shown in Table 1. It can be seen that, on average, more than 92% of travel diary trip legs are correctly identified by the mode identification algorithm. When including wrongly identified trip legs, e.g. scatter trips, 84% of the trip legs identified by the algorithm are assigned the correct mode.

Observed / Algorithm	Walk	Bike	Bus	Car	Rail	Non-trips	Confidence rate
Walk	75	6	1	1	-	13	78.1%
Bike	1	104	-	5	-	3	92.0%
Bus	-	-	27	-	-	-	100.0%
Car	1	7	7	152	1	19	81.3%
Rail	-	-	-	-	33	2	94.3%
Other	1	1	-	1	-	1	-
Total	78	118	35	159	34	38	**84.6%**
Success rate	96.2%	88.1%	77.1%	95.6%	97.1%	-	**92.2%**

Table 1: The results when including map matching algorithm

The success rates achieved seems to be high when comparing to general success rates of 82.6% in Gong et al. [4], 79.1% in Chen et al. [7], 87.4% (own calculation from numbers available) in Bolbol et al. [5] and 91.7% in Chung & Shalaby [8]. Especially, the identification of bus trips is notably better than the rates of 62.5% achieved in Gong et al. [4], 53.3% achieved in Chen et al. [7], and 58.3% achieved in Bolbol et al. [5].

In Figure 3 two examples of the application of the Bus Line alignment analysis is visualised. To the left is a GPS trace which coincides with numerous bus lines on parts of the trip leg and ends where these bus lines terminate. Additionally, the trace originates at a bus stop on Line 168, and seems to follow the alignment and stopping pattern of this bus line on most of the trip leg. Consequently, the trip leg is (correctly) classified as a bus trip on Line 168. In the example to the right, the overlap is lower and the bus line followed on parts of the trip does not stop at the origin or destination. Consequently, the trip leg is (correctly) classified as a car trip.



Trip information

Date:	27-10-2011
Starting time:	08:40:38
Ending time:	08:49:37
Trip length:	8 min 59 sec

Bus line stop overlap	# stops	Hit%
145	4/5	80%
161	3/4	75%
167	8/11	72%
168	8/11	72%
81N	3/4	75%

Origin stop:	168
Destination stop:	67/145/161/167/168
Classified as:	Bus trip
Actual bus line:	168

Trip information

Date:	09-11-2011
Starting time:	06:42:26
Ending time:	06:49:07
Trip length:	6 min 41 sec

Bus line stop overlap	# stops	Hit%
161	3/5	60%

Origin stop:	None
Destination stop:	None
Classified as:	Car trip
Actual bus line:	Car trip

Figure 3: Example of results of the bus detection algorithm.

4. Conclusions

This study proposes a combined fuzzy logic- and GIS-based algorithm to process raw GPS data. The algorithm is applied to GPS data collected in the highly complex Copenhagen Region transport network, and detects trip legs and distinguishes between five modes of transport. The application shows, compared to other studies, promising results through a 92% success rate in identifying the correct mode. There are however room for improvements; it was found that the trip leg identification process should be improved further, as e.g. many trip legs are wrongly split into several parts. An improvement of this process will improve the performance of the overall GPS processing algorithm.

5. Acknowledgement

This work has been undertaken as a part of the project “Analyses of activity-based travel chains and sustainable mobility” (ACTUM), and has thereby been funded by the Danish Council for Strategic Research. The authors would furthermore like to thank Nadine Schüssler from ETH Zurich for providing support of the Java-based software used in parts of the analysis.

6. References

[1] Wolf, J., 2000. Using GPS Data Loggers To Replace Travel Diaries In the Collection of Travel Data. Ph.D. Thesis,Georgia Institute of Technology, United States of America

[2] Stopher, P., FitzGerald, C., Zhang, J., 2008. Search for a global positioning system device to measure person travel. Transportation Research Part C 16 (2008), 350-369

[3] Mavoa, S., Oliver, M., Witten, K., Badland, H., 2011. Linking GPS and travel diary data using sequence alignment in a study of children’s independent mobility. International Journal of Health Geographics. 10:64

[4] Gong, H., Chen, C., Bialostozky, E., Lawson, C., 2012. A GPS/GIS method for travel mode detection in New York City. Computers, Environment and Urban Systems 35 (2012), 131-139

[5] Bolbol, A., Cheng, T., Tsapakis, I., Haworth, J., 2012. Inferring hybrid transportation modes from sparse GPS data using a moving window SVM classification. Computers, Environment and Urban Systems 36 (2012), 526-537

[6] Schüssler, N., 2010. Accounting for similarities between alternatives in discrete choice models based on high-resolution observations of transport behavior. Ph.D. Thesis, ETH Zurich, Switzerland

[7] Chen, C., Gong, H., Lawson, C., Bialostozky, E., 2010. Evaluating the feasibility of a passive travel survey collection in a complex urban environment: Lessons learned from the New York City case study. Transportation Research Part A 44 (2010), 830-840

[8] Chung, E.-H., Shalaby, A., 2005. A Trip Reconstruction Tool for GPS-based Personal Travel Surveys. Transportation Planning and Technology, 28:5, 381-401