ARTICLE IN PRESS

Transportation Research Part C xxx (2013) xxx–xxx

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Supporting large-scale travel surveys with smartphones – A practical approach

Philippe Nitsche *, Peter Widhalm, Simon Breuss, Norbert Brändle, Peter Maurer

Austrian Institute of Technology, Giefinggasse 2, 1210 Vienna, Austria

ARTICLE INFO

Article history:
Received 15 December 2012
Received in revised form 2 September 2013
Accepted 5 November 2013
Available online xxxx

Keywords:
Smartphones
Mobility data
Travel survey
Accelerometer
GPS
Transport modes
Mode detection

ABSTRACT

Collection of travel data is a key task of transportation modeling. Data collection is currently based on costly and time-intensive questionnaires, and can thus only provide limited cross-sectional coverage and inadequate updates. There is an urgent need for technologically supported travel data acquisition tools. We present a novel approach for supporting travel surveys using data collected with smartphones. Individual trips of the person carrying the phone are automatically reconstructed and trip legs are classified into one of eight different modes of transport. This task is performed by an ensemble of probabilistic classifiers combined with a Discrete Hidden Markov Model (DHMM). Classification is based on features extracted from the motion trajectory recorded by the smartphone's positioning system and signals of the embedded accelerometer. Our approach can cope with GPS signal losses by including positioning data obtained from the mobile phone cell network, and relies solely on accelerometer features when the trajectory cannot be reconstructed with sufficient accuracy. To train and evaluate the models, 355 h of probe travel data were collected in the metropolitan area of Vienna, Austria by 15 volunteers over a period of 2 months. Distinguishing eight different transportation modes, the classification results range from 65% (train, subway) to 95% (bicycle). The increasing popularity of smartphones gives the proposed method the potential to be used on a wide-spread basis and can complement existing travel survey methods.

1. Introduction

Data collected with travel surveys provide essential information for traffic planners, public transport providers, infrastructure authorities and transportation scientists. Travel data are the basis for transportation modeling and optimization of transportation services and routing. Conventional methods for collecting data for travel surveys comprise computer-assisted telephone interviews and personal interviews, computer-assisted self-interviews, mail-back questionnaires, web-based questionnaires, traffic counting at cross sections or intersections and analyses of transport schedule inquires. Most of the above conventional methods are costly and time-intensive. Hence large-scale travel surveys have been often conducted only once in a decade. Continuous data collection has already been discussed by Edmonston and Schultze (1995) for the US census. According to Stopher and Greaves (2007), a continuous survey collects data of a certain sample of households on a continuous basis. Acquired data is then averaged over a pre-defined period and provides up-to-date information about travel behavior in the region of interest.

* Corresponding author. Tel.: +43 50550 6517; fax: +43 50550 6599.
E-mail addresses: philippe.nitsche@ait.ac.at (P. Nitsche), peter.widhalm@ait.ac.at (P. Widhalm), simon.breuss@ait.ac.at (S. Breuss), norbert.braendle@ait.ac.at (N. Brändle), peter.maurer@ait.ac.at (P. Maurer).

0968-090X/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.trc.2013.11.005

In addition to the cost factor, conventional surveys are also affected by non-response issues and underreported trips, see Abraham et al. (2006), Brög et al. (1982), Groves and Couper (1998), Groves (2006), Richardson et al. (1996) or Zmud and Arce (2000). Since the late 1990s, technologies such as Global Positioning System (GPS) devices have been utilized as a supplement to measure people travel. One of the first household surveys with a GPS subcomponent was conducted in 1997 in Austin, Texas, followed by many studies to examine the application of GPS for inferring travel behavior (Bohte and Maat, 2009; Gong et al., 2012; Marchal et al., 2011; Murakami and Wagner, 1999; Pearson, 2001; Stenneth et al., 2011; Stopher et al., 2008; Wolf et al., 2001a, 2001b). The results of these studies indicate the potential of GPS devices (portable or mounted in household vehicles) to replace or supplement traditional methods – often combined with Geographic Information Systems (GIS). Nevertheless, the most common problems of GPS devices for travel surveys are still signal losses in shadowed areas such as urban canyons (Gong et al., 2012) or underground transportation systems, high energy consumption and the acceptance of users carrying the device during daily travel.

The increasing popularity of smartphones opens novel opportunities for collecting data for travel surveys. Asakura and Hato (2004) studied the use of the location positioning function of cellular phone systems for tracking individual travel behavior and demonstrated the feasibility of mobile phone sensing for travel surveys. Bierlaire et al. (2010) proposes a method for estimating route choice models from smartphone GPS data and achieved satisfactory results, although the sample size was limited. A typical smartphone contains several internal MEMS (micro-electro-mechanical systems) sensors (e.g. accelerometer, magnetometer, gyroscope) and two different positioning services. They offer precise self-positioning via Assisted-GPS and approximate network positioning using GSM Cell IDs and WIFI SSIDs. This means a rich set of data sources, which can be exploited for classifying mobility behavior.

Existing approaches for utilizing smartphones can be summarized as follows: (1) they require an uninterrupted GPS lock to guarantee high quality positioning and speed information available throughout the data collection (Reddy et al., 2010; Stenneth et al., 2011), and (2) they can distinguish only a very small number of transportation modes such as "still", "walk" and "motorized" (Anderson and Muller, 2006; Ayu et al., 2011; Gonzalez et al., 2008; Liao et al., 2007; Sohn et al., 2006; Zheng et al., 2008). However, a practical approach with true potential to replace conventional surveys must cope with imperfect GPS data: during normal daily activities, the integrated GPS receiver of cell phones is often severely shielded, e.g. when the phone is carried in a bag or covered by clothing. Furthermore, a practical approach should infer the modes of all trip legs of a trip chain. Fig. 1 illustrates the concept of trip legs, which are defined as segments of a trip separated by transport mode changes or intervening activities with a short dwell time (cf. McGuckin and Nakamoto, 2004).

We present a novel practical approach for supporting travel survey that (1) alleviates the problems of GPS satellite losses by including positioning data obtained from the cellular network and accelerometer readings and (2) infers a richer set of transport mode categories. Accelerometers provide rotational and translational movement information. Based on this data, our approach can reconstruct trips even with weak or no satellite coverage. GPS and accelerometer features are extracted to train a classifier, similar to the work by Kwapisz et al. (2011), Sun et al. (2010), Wang et al. (2010) or Yang (2009). In contrast to previous research, we distinguish between eight different, fine-grained transport mode categories and deal with real-world smartphone data without requiring the respondent to wait for a GPS lock or removing trips with GPS signal losses in a pre-filtering step. Recognizing specific patterns in the frequency and time domain of the accelerometer signals allows

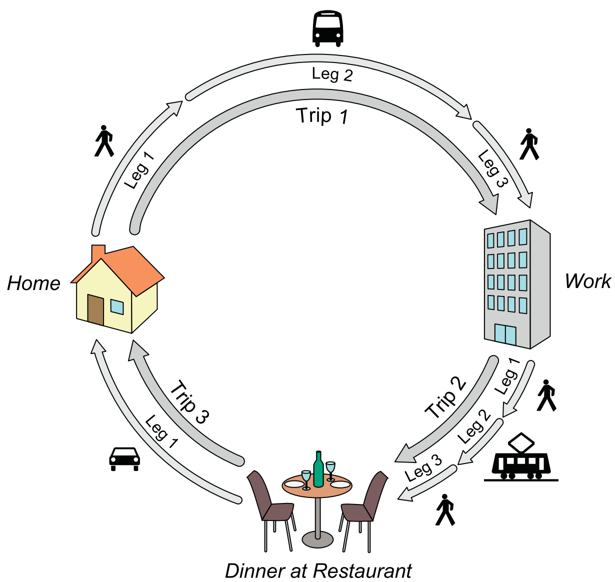

Fig. 1. Illustration of an exemplary trip chain with trips and trip legs.

the detection of transport mode changes. To collect all data necessary for a complete survey, additional user entries are needed, e.g. the trip purpose.

Due to the high spread of smartphones, our approach permits a region-wide collection of mobility data, leads to an improved data basis for transportation planning and enables a novel way for continuous measurements.

This paper is organized as follows: Section 2 describes the requirements for large-scale travel surveys and the constraints with regard to smartphones. Section 3 describes the technique for automatic mode detection, including test data collection, feature extraction, the models for identifying trip legs and experimental results. Section 4 proposes a strategy for integrating the method into existing travel survey procedures. Recommendations for how to overcome major barriers to future applications in large-scale surveys are discussed in Section 5.

2. Requirements for a travel survey

This section recaptures the general requirements for travel surveys, which must be fulfilled to achieve useful results. A community-based method using smartphones does not fulfill all general requirements for a large-scale travel survey. We therefore discuss constraints and limitations of utilizing smartphones.

2.1. Representativeness

Travel surveys provide information about a chosen sample of a population and must be representative with regard to five aspects: (1) spatial representativeness ensures that the results are valid for the entire area of interest, meaning that the collected data must be spatially distributed instead of concentrated to a certain area; (2) temporal representativeness ensures that trips are temporally distributed; (3) travel surveys must cover appropriate shares of socio-demographic aspects such as gender, age, education or size of household; (4) representativeness regarding transport modes, i.e. all kinds of non-motorized and motorized modes must be considered; (5) representativeness of the trip purpose that helps to split up the given population in behaviorally homogenous groups.

When rolling out the data collection application to a set of volunteers, it is not ensured that the user sample is representative regarding all of the five requirements mentioned above. There might be a bias towards younger people or phone users with a personal technical interest. Roux et al. (2009) discussed the advantages of GPS-based surveys and argued that participants may have a particular profile, i.e. the participation correlates with higher education or greater mobility. It is clear that not all socio-demographic groups can be reached without specific incentives for them. However, this limitation does not concern surveys that target specific populations, for example users on a university campus, for which a sampling base may be available. Given this fact, we argue that a smartphone-based survey method should not be seen as a replacement but as a supplement to conventional methods. A further discussion on incentives is provided in Section 4.3.

As the representativeness of smartphone-based surveys remains an open issue, so does the representativeness of the trip purposes. This requirement can only be fulfilled in future developments, because it was not goal of our research to detect trip purpose. Encouraging research in automatic trip purpose recognition has been carried out by Wolf et al. (2001a), Bohte and Maat (2009) or Chen et al. (2010).

2.2. Comparability

Travel survey results are typically compared to previous surveys or other surveys in order to determine trends and differences. Consequently, travel surveys must be conducted in a way that allows standardized data analyses and post-processing. The contents covered by a survey and the data formats must be harmonized with other surveys. The detected trip times and modes can be processed in a manner harmonized to previous or current conventional surveys. However, since automatic detection of trip purposes is not part of the presented approach and left for future research, the respondents are required to provide this information manually.

2.3. Accurate reconstruction of trips

Underreported trips are a common problem in travel diaries or surveys. The trip information ideally contains sources and destinations, which can also be clustered into larger regions. Survey data also requires temporal information for each trip, i.e. the start, end time, and its duration. This allows determining travel times, average trip durations or daily traffic load curves. A smartphone-based survey also covers very short trips that are often omitted in a conventional survey method.

2.4. User acceptance

When considering novel technologies as survey tools, user acceptance plays an important role. Roux et al. (2009) developed a model for the willingness to participate in a GPS-based travel survey. Results indicate that households with a higher income, a high number of cars and with high-tech equipment correlate with participation. Moreover, younger, male and healthy people are more willing to be involved. As explained in Section 2.1, such biases are likely to occur in a

Please cite this article in press as: Nitsche, P., et al. Supporting large-scale travel surveys with smartphones – A practical approach. Transport. Res. Part C (2013), http://dx.doi.org/10.1016/j.trc.2013.11.005

smartphone-based survey. Hence, the following requirements should be taken into account: The user interface of a smartphone application to collect travel data must be as simple and intuitive as possible. It should neither distract users in their daily phone activities nor cause any limitations of the phone performance. Ideally, data collection should run as a background task, while the effort for the participants should be kept to a minimum i.e. no manual entries or questionnaires. An obvious problem of modern smartphones is the battery performance that can dramatically decrease when many tasks and sensors are active. To cope with this problem, the computations and tasks necessary for collecting data need to be optimized and reduced.

2.5. Privacy requirements

Travel surveys collect personal data about individual travel behavior. Therefore, data privacy should be a major issue in all phases of a survey including collecting, analyzing and publishing travel data. Information about trip sources, destinations or dwell times at certain places may reveal sensitive personal data concerning ethnical background, political and religious views, health or sexual life. Consequently, it must be ensured that all data is made anonymous to avoid direct references to a person. Besides, it is a common approach to obtain a written declaration of consent from the persons involved. Various privacy issues regarding the relevant national data protection laws must be considered in the planning and conduction of a travel survey. Initially, data protection laws include the preliminary avoidance and minimization of collecting and storing individual-related data. Furthermore, it must be ensured that the data is protected against access, utilization or modification by unauthorized persons. Data sets that are not used any more should be deleted. In order to foster transparency for the involved persons, they need to be informed about the purpose of data utilization as well as about collecting and processing methods. In the research project described in this paper, all data protection requirements mentioned above are fulfilled. Data is made anonymous and test persons have to sign a declaration of consent to allow the collection and utilization of their mobility data. In future smartphone-based survey software, users should also declare their consent before they start collecting travel data.

3. Automated trip reconstruction and travel mode classification

Automated travel mode classification consists of segmenting a tour into trips and trip legs and inferring the transport mode used by the person carrying the phone. This section explains details about the collection of sample data, feature extraction from the smartphone data and mode classification. A detailed description can also be found in Widhalm et al. (2012).

3.1. Probe data collection

The travel mode classifier is based on a data-driven statistical model describing the conditional distribution of sensor data recorded by the smartphone given the mode of transport. In order to estimate such a model and to evaluate classification accuracy, probe data was collected and manually annotated with the transport mode used.

We used the following Android based smartphone devices: the Samsung I-9000 Galaxy S, the HTC Desire HD, and the HTC Desire Z (see Fig. 2). In general, the devices mainly differ in size, handling, GPS accuracy and power consumption. All selected phones contain a MEMS acceleration and magnetic field sensor measuring acceleration and magnetic field in all three spatial axes at 100 Hz. The assisted GPS of the devices computes the position, speed, bearing and accuracy of the position with 1 Hz and the approximate network position with a non-uniform sampling interval.

We developed a logging application for the selected Android smartphones. The application collects data from every embedded sensor with the respective frequency and synchronizes this data with location and speed information. Data was collected by 15 volunteers in the metropolitan area of Vienna, Austria over a period of 2 months. Each volunteer was

Fig. 2. Test devices used for probe data collection and validation.

equipped with a smartphone with the installed data logging application. During travel, the volunteers annotated the current transport mode (see Fig. 3(a)). The annotations distinguish between the following modes: walk, bicycle, motorcycle, car, bus, electric tramway, metro, train, and wait. Simultaneously traveling and annotating every mode change can be hectic in real life, producing errors. We have therefore developed a post processing tool enabling the volunteers to correct annotations and to enhance the quality of the ground truth data (see Fig. 3(b)).

The data were collected under realistic circumstances and during normal daily activities. No instructions were given on how to carry the phone or which routes and transport modes to choose. In contrast to other studies, no measures were taken to guarantee a constant GPS connection, such as initially obtaining a GPS lock before starting a trip or filtering out trips with poor positioning data quality. A volunteer had to start the tracking application on the phone when beginning a new trip and to annotate the current mode of transport for each trip leg. The total amount of ground truth data comprises 355 h of travel time with a high variety of transport mode shares (cf. Table 1). It was of major importance that the shares be equally distributed and close to the representative regarding the transport system in the Vienna region.

3.2. Preprocessing and feature extraction

The movement tracks recorded by the data logging application are noisy, and their accuracy depends on the availability and quality of the GPS signal. The tracks are pre-processed with a Kalman filter (Kalman, 1960), i.e. computing accurate and smooth trajectories by combining the raw GPS and cell location data with predictions of a linear motion model assuming zero mean, Gaussian-distributed accelerations. Fig. 4 shows an example of a recorded movement track and the filtering result. The readings of the tri-axial accelerometer are transformed to a rotation-invariant signal by discarding the direction of the three-dimensional acceleration vector and retaining only its magnitude.

Our classification approach involves a Discrete Hidden Markov Model (DHMM), which assumes that consecutive observations are statistically independent, given the hidden state. If the hidden states correspond to the individual modes of transport and features are extracted from small time windows of only a few seconds, the assumption of statistical independence will be violated. Large time windows of more than a minute could reduce such dependencies, but could also include multiple modes of transport. Therefore, the time window must be adaptively chosen in a way such that it is homogeneous with respect to the mode of transport and large enough to minimize the dependency between the current and the previous time window. To avoid coverage of multiple modes of transport, we assume that transport mode changes occur during a spatially stationary period or a period when a volunteer starts or stops walking. Hence a new time window starts whenever either speed drops below a threshold, high amplitudes of the accelerometer signal indicate that the user might be walking or the maximum window size of 2 min is exceeded.

For every time window the following features are extracted, summing up to a total of 77 features:

- 5th, 50th and 95th percentile of speed, accelerations, decelerations and direction change per unit time as computed from the GPS readings,
- standard deviation of the high-frequency accelerometer magnitudes and
- power Spectrum of the accelerometer signal for frequencies $i\omega/128$ Hz with $i = 1,\ldots,64$ and the sampling frequency $\omega = 50$ Hz.

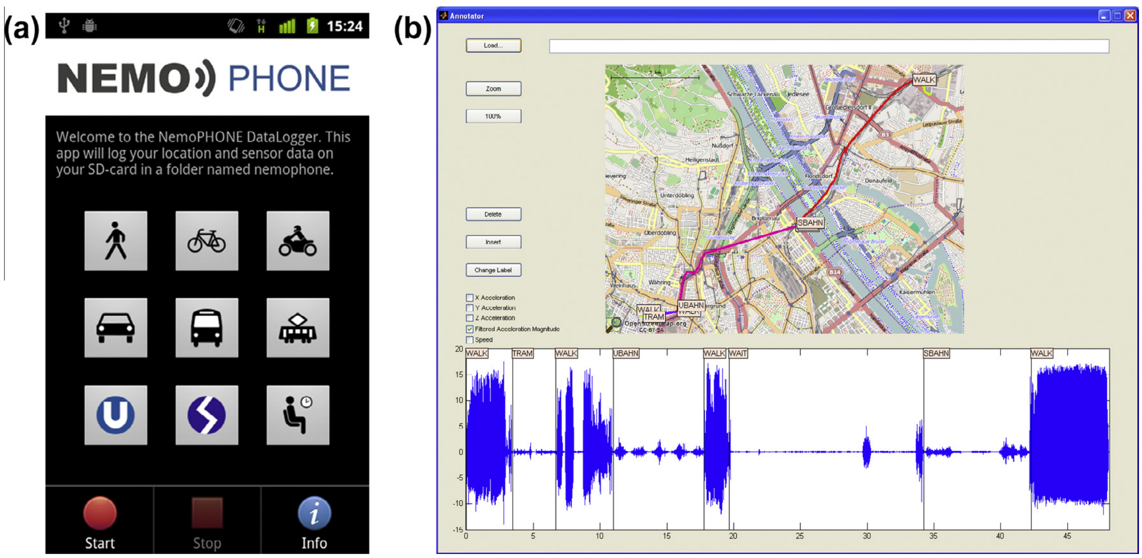

Fig. 3. (a) Screenshot of the probe data logging application and (b) Screenshot of the correction tool.

Please cite this article in press as: Nitsche, P., et al. Supporting large-scale travel surveys with smartphones – A practical approach. Transport. Res. Part C (2013), http://dx.doi.org/10.1016/j.trc.2013.11.005

Table 1
Share of trips, travel times and GPS-coverage per transport modes.

Mode	Trips [#]	Time [h]	GPS-coverage (%)
Walk	1058	74	60
Bike	180	58	90
Motorcycle	50	13	53
Car	215	72	80
Bus	99	15	77
Electric Tramway	52	7	41
Metro	212	36	34
Train	223	82	20
Overall	2089	355	57

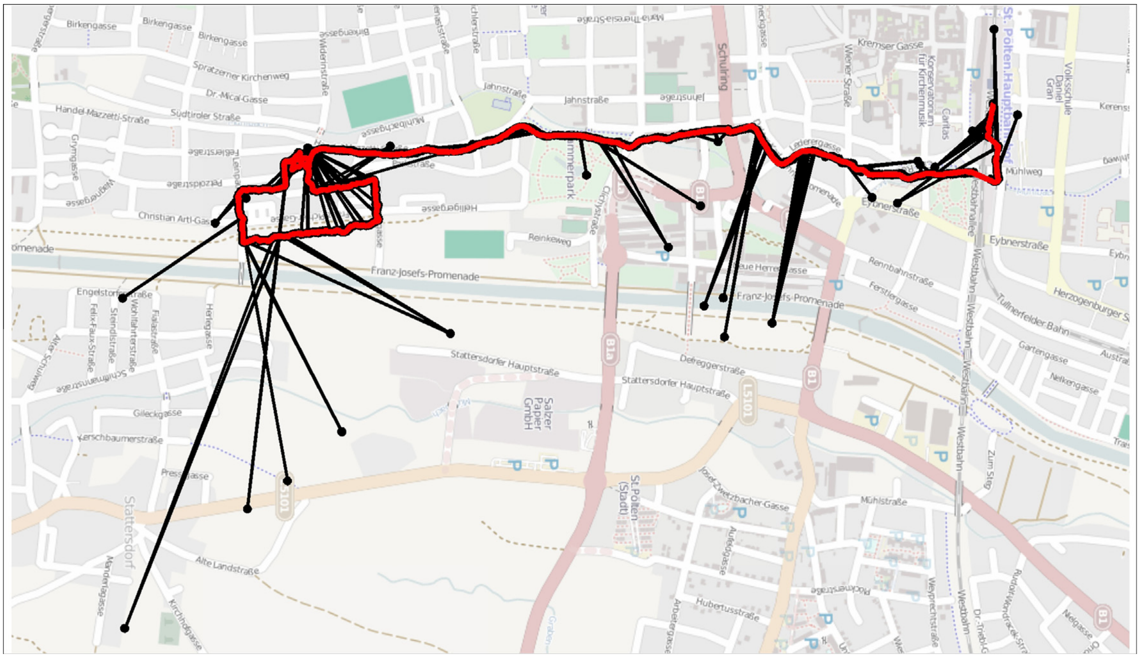

Fig. 4. The raw movement track recorded by the data logging application (black dots and lines) and the filtered trajectory computed with the Kalman Filter (red line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

An example of the velocity and acceleration signal along with its time–frequency spectral decomposition is shown in Fig. 5.

3.3. Classification

In order to tackle classification in the 77-dimensional feature space with noisy sample data, we chose a classification scheme described in Ho (1998). This approach randomly selects subspaces of the original feature space and trains an ensemble of classifiers on the subspaces. Given a feature vector of a single time window, the estimated posterior class probabilities are first averaged over the individual classifiers in the ensemble. Subsequently, a DHMM combines the instantaneous classifier output with transition probabilities between hidden states. The hidden states correspond to the used modes of transport and govern the distribution of the observable features. Travel mode classification selects the most probable sequence of hidden states, given the observed sequence of feature vectors and modeled transition probabilities. This sequence is computed with the Viterbi algorithm (Viterbi, 1967). A schematic overview of this classification technique is given in Fig. 6.

3.4. Experimental results

We evaluated classification performance with class size normalized *precision* and *recall* and a cross validation scheme, where one trip at a time was left out from training and used for validation. The recall ratio, also called sensitivity, is defined as the fraction of correctly recognized instances of a certain mode. Recall can be interpreted as the probability that a

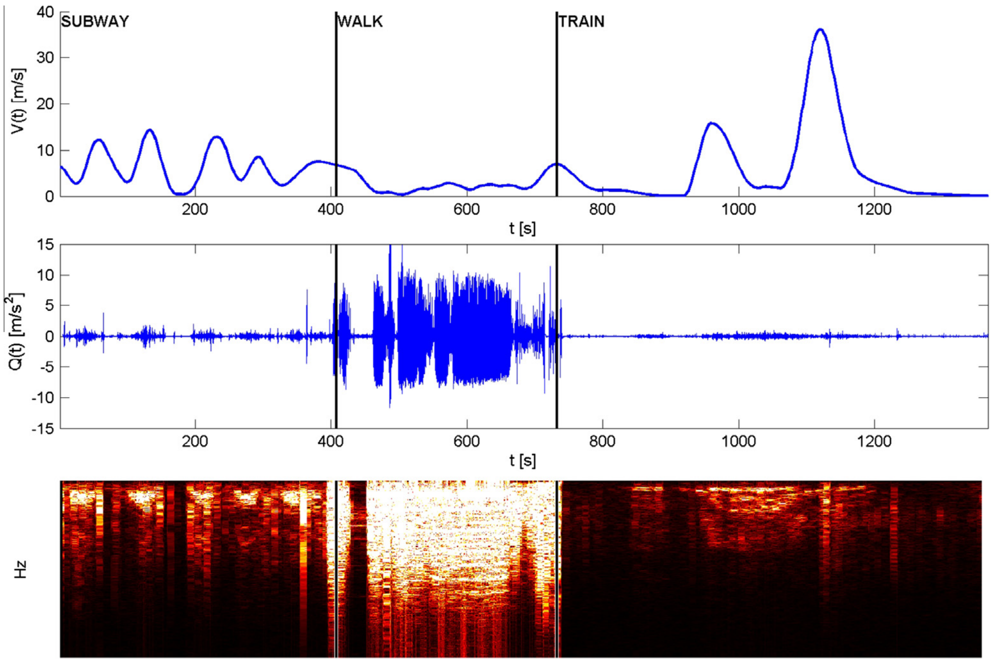

Fig. 5. Example of velocity computed from a Kalman-filtered trajectory (top row), magnitude of the high-frequency accelerometer readings (middle row) and frequency spectrum of the accelerometer signal (bottom row).

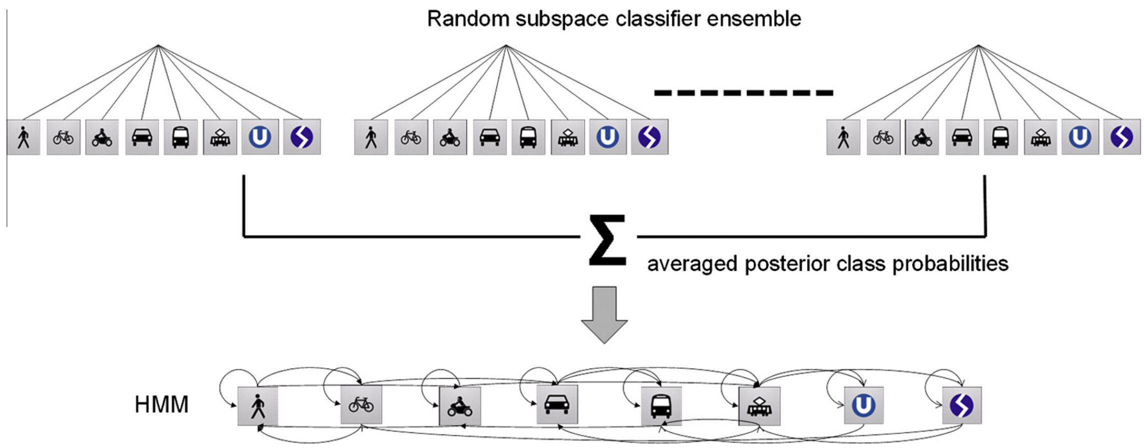

Fig. 6. Schematic overview of the classification technique.

randomly selected instance of a mode will be correctly classified. The class size normalized precision ratio, also called as positive predictive value, serves as an estimate of the probability that a randomly selected instance classified as a certain mode is classified correctly (i.e. it actually *is* an instance of the mode), given that all modes have equal prior probabilities.

Table 2 shows the resulting confusion matrix and the precision and recall ratios. The rows of the matrix represent the true travel modes according to user annotations, and the columns represent the predicted travel mode classes. The numbers in the cells indicate the conditional frequency (in percent) each mode was predicted, given the respective user annotation. This means that, up to integer rounding effects, each row sums up to 100%.

For every transport mode the classification result is significantly better than random guessing, which would yield 0.125% recall accuracy (given that there are 8 distinct classes). The classifier correctly annotated more than 80% of the time traveled by bus, bike, motorcycle or per pedes (walk).

The confusions between above-ground train and subway can be explained by the similarity between these classes with respect to velocity and accelerations. Incorporating contextual information such as the positions of stations, public transport routes and schedules could reduce these errors.

Please cite this article in press as: Nitsche, P., et al. Supporting large-scale travel surveys with smartphones – A practical approach. Transport. Res. Part C (2013), http://dx.doi.org/10.1016/j.trc.2013.11.005

Table 2
Confusion matrix computed from the collected sample data following a leave-one-trip-out cross-validation scheme.

		Predicted mode							
		Walk (%)	Bike (%)	Motor-cycle (%)	Car (%)	Bus (%)	Tram (%)	Subway (%)	Train (%)
Ground truth	Walk	82	4	1	3	1	1	5	4
	Bike	3	95	0	0	1	0	0	0
	Motorcycle	2	2	80	2	10	1	2	1
	Car	3	0	3	67	6	3	6	11
	Bus	6	2	1	2	81	2	4	1
	Tram	12	4	0	1	3	72	5	4
	Subway	16	1	0	3	5	2	65	9
	Train	7	1	1	6	3	6	11	65
	Recall	82	95	80	67	81	72	65	65
	Precision	62	88	91	79	74	83	66	68

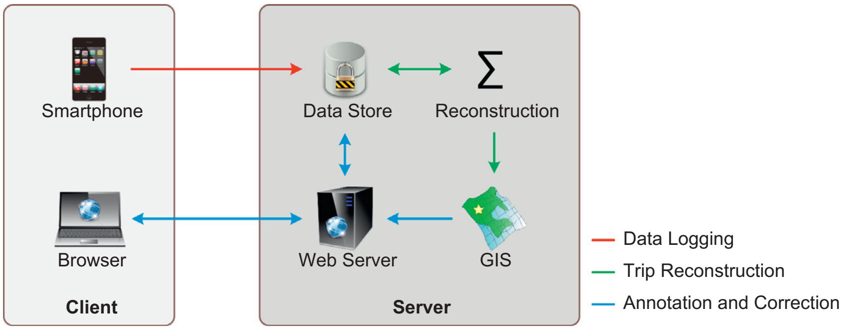

Fig. 7. Proposed client server architecture for a semi-automatic trip survey.

Misclassifications between car and train as well as between motorcycle and bus can be attributed to variability resulting from different driving styles, vehicles, routes and traffic states. To address this problem, future research will include modeling the heterogeneity of the defined transportation modes based on unsupervised learning techniques. Further potential for improving the classification results lies in adapting the classifier to user characteristics by semi-supervised incremental updates of the models.

Remarkably, there are several misclassifications of tram, train and subway as walking. We believe that these classification errors are caused by ambiguities (walking inside a train) or inaccurate user annotations. To address the latter, we plan to develop robust learning algorithms, which better cope with outliers and label noise.

4. Proposed implementation strategy

The proposed approach for recognizing travel behavior with smartphones shows promising results, although there is room for improvement. This section proposes an implementation strategy, which deals with the technical and organizational integration of a smartphone-based survey tool.

4.1. Compatibility to and integration into existing survey systems

The proposed survey method can be used as a supplement to other methods. Current research projects deal with the development and design of user-friendly and adaptive questionnaires to collect travel data with smartphones. In future large-scale surveys, the method presented in this paper can be combined with adaptive and/or questionnaires to avoid misclassifications or to cover additional information such as the trip purpose or household and participant characteristics. Currently, due to a bias of technology-based survey methods regarding representativeness, only a mix of modern and conventional methods with predefined subsamples leads to improved large-scale travel surveys.

4.2. Data handling

We suggest implementing a semi-automatic trip survey using the client–server model shown in Fig. 7. By doing so, the extensive computation of the trip legs and travel modes can be performed on server hardware and will not drain the user's battery. A second advantage is the simple and small app on the smartphone itself. This makes it easier to integrate new

Please cite this article in press as: Nitsche, P., et al. Supporting large-scale travel surveys with smartphones – A practical approach. Transport. Res. Part C (2013), http://dx.doi.org/10.1016/j.trc.2013.11.005

versions and other operating systems (e.g. iOS, Windows Phone). A major drawback of the client–server model is the data communication. When the computation is performed on a server, all sensor data has to be transferred. A protocol has been designed and implemented to reduce traffic usage of the smartphone app, which compresses the sensor data size from more than 30 MB to approx. 6 MB per hour. Using this encoding scheme, all data can be buffered on the device and transferred at a later point using a WIFI connection if available.

If a user wants to participate in the travel survey, he or she has to agree to the data privacy statement. Subsequently, the participant can download and install the app onto his smartphone. The automatic data collection can be started in the background without any user interaction. The sensor data is buffered on the device and uploaded via WIFI or UMTS connection. The reconstruction algorithm splits the data into trips and legs and computes the corresponding travel mode. If the data contains GPS information, additional GIS data can be used to improve the accuracy of mode detection. The resulting trips and legs are saved in the encrypted server data store.

The user can use a browser to login on the web interface. The system shows a list and map with his personal trips and legs based on the reconstruction algorithm, where he can add additional information such as personal data or trip purpose and correct wrong or missing legs. Any corrections are saved into the data store of the server and used to improve the reconstruction algorithm by personalizing the general model toward the special user characteristics.

4.3. Recruiting volunteers

Since not all requirements of large-scale travel surveys can be fulfilled, our approach is mainly intended to supplement conventional methods by drawing a subsample of the population. There are two possibilities to cope with the problem of biases due to an inhomogeneous distribution of participants. Firstly, the subsample only contains younger persons with a personal technical interest, e.g. students at a certain campus. This allows user-specific travel surveys with the limitation of a socio-economic non-representativeness. Secondly, the underrepresented user groups could be encouraged to participate by providing incentives. Explicit incentives such as financial payments, vouchers or coupons are an option, because the participants can easily understand the incentive and use the application regardless of the size of the community. Examples for possible explicit incentives are free tickets for public transport or parking areas, coupons for gas stations and smartphones at a lower price. On the other hand, we want to stress that such explicit incentives may influence travel behavior and/or cause another bias towards specific user groups. This leads to the idea of implicit incentives, which can either replace or supplement explicit ones (Sloof and Sonnemans, 2011). Examples are social incentives that may encourage participants to be an active member of a community such as social networks. For the particular case of a community-based travel survey with smartphones, sharing their daily routes with the possibility to optimize travel times could be an idea. Moreover, secondary information can be generated from the travel data, e.g. daily number of steps, distances covered or estimated burnt calorie. Such relational incentives require a more comprehensive smartphone application that provides this additional individual information.

5. Conclusions

This paper introduced a novel approach to utilizing smartphones for travel surveys. Instead of developing a questionnaire-based mobile tool, the proposed method automatically reconstructs trips and trip legs including transport modes by simply carrying the smartphone during a daily travel. The research focused on the analysis of data captured by the embedded accelerometer combined with speed and location information from the GPS receiver. An ensemble of probabilistic classifiers combined with a Discrete Hidden Markov Model (DHMM) compute the most likely transport mode and mode changes. The best classification results were achieved for detecting walk, bus and bike trips. Differentiating between different railway modes (over-ground train, tramway and subway) was found to be more erroneous due to the similarity of vibration response and frequency behavior. In general, modes that cause unambiguous accelerometer signals could be classified with superior accuracy. In a travel survey, this allows incorporating places where there is normally only a weak or no GPS signal.

A strategy for implementing the method into existing survey procedures was recommended. The smartphones serve as thin clients, whereas main calculations are done on a central server. In the sense of a semi-automatic survey, participants can manually check and/or correct their reconstructed travel diary or enter additional information such as the trip purpose.

In future large-scale travel surveys, the proposed method can supplement conventional data collection methods. A combination of modern technology-based data collection tools with conventional solutions seems to be a good choice to meet all requirements of a large-scale survey such as representativeness, comparability and data quality.

Acknowledgements

This work was performed in the scope of the Austrian research project NEMO-PHONE (Novel Methods for Collecting Mobility Data with Smartphones), which was funded by the Austrian Federal Ministry of Transport, Innovation and Technology. The authors would like to thank all members of the project consortium and give gratitude to all persons responsible for the programme line ways2go.

References

Abraham, K.G., Maitland, A., Bianchi, S.M., 2006. Nonresponse in the American time use survey who is missing from the data and how much does it matter? Public Opinion Quarterly 70, 676–703.

Anderson, I., Muller, H., 2006. Practical Activity Recognition Using GSM Data (Technical Report). Department of Computer Science, University of Bristol.

Asakura, Y., Hato, E., 2004. Tracking survey for individual travel behaviour using mobile communication instruments. Transportation Research Part C: Emerging Technologies 12, 273–291.

Ayu, M.A., Mantoro, T., Matin, A.F.A., Basamh, S.S.O., 2011. Recognizing user activity based on accelerometer data from a mobile phone. In: 2011 IEEE Symposium on Computers Informatics (ISCI). Presented at the 2011 IEEE Symposium on Computers Informatics (ISCI), pp. 617–621.

Bierlaire, M., Chen, J., Newman, J., 2010. Modeling Route Choice Behavior From Smartphone GPS data (Report TRANSP-OR 101016).

Bohte, W., Maat, K., 2009. Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: a large-scale application in the Netherlands. Transportation Research Part C: Emerging Technologies 17, 285–297.

Brög, W., Erl, E., Meyburg, A.H., Wermuth, M.J., 1982. Problems of nonreported trips in survey of nonhome activity patterns. Transportation Research Record 891, 1–5.

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: Policy and Practice 44, 830–840.

Edmonston, B., Schulze, C., 1995. Modernizing the U.S. Census, National Academy Press.

Gong, H., Chen, C., Bialostozky, E., Lawson, C.T., 2012. A GPS/GIS method for travel mode detection in New York City. Computers, Environment and Urban Systems 36, 131–139.

Gonzalez, P.A., Weinstein, J.S., Barbeau, S.J., Labrador, M.A., Winters, P.L., Georggi, N.L., Perez, R.A., 2008. Automating Mode Detection Using Neural Networks and Assisted GPS Data Collected Using GPS-Enabled Mobile Phones. Presented at the 15th World Congress on Intelligent Transport Systems and ITS America's 2008 Annual Meeting, New York, NY.

Groves, R.M., 2006. Nonresponse rates and nonresponse bias in household surveys. Public Opinion Quarterly 70, 646–675.

Groves, R.M., Couper, M.P., 1998. Nonresponse in Household Interview Surveys, first ed. Wiley-Interscience.

Ho, T.K., 1998. The random subspace method for constructing decision forests. IEEE Transactions on Pattern Analysis and Machine Intelligence 20, 832–844.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. Journal of Basic Engineering 82, 35–45.

Kwapisz, J.R., Weiss, G.M., Moore, S.A., 2011. Activity recognition using cell phone accelerometers. SIGKDD Exploration Newsletter 12, 74–82.

Liao, L., Patterson, D.J., Fox, D., Kautz, H., 2007. Learning and inferring transportation routines. Artificial Intelligence 171, 311–331.

Marchal, P., Madre, J.-L., Yuan, S., 2011. Postprocessing procedures for person-based global positioning system data collected in the French national travel survey 2007–2008. Transportation Research Record: Journal of the Transportation Research Board 2246, 47–54.

McGuckin, N., Nakamoto, Y., 2004. Trips, Chains, and Tours—Using an Operational Definition. Presented at The NHTS Conference 2004.

Murakami, E., Wagner, D.P., 1999. Can using global positioning system (GPS) improve trip reporting? Transportation Research Part C: Emerging Technologies 7, 149–165.

Pearson, D., 2001. Global Positioning System (GPS) and travel surveys: results from the 1997 Austin Household Survey. Presented at the 8th Conference on the Application of Transportation Planning Methods, Corpus Christi, Texas.

Reddy, S., Mun, M., Burke, J., Estrin, D., Hansen, M., Srivastava, M., 2010. Using mobile phones to determine transportation modes. ACM Transactions on Sensor Networks 6, 1–27.

Richardson, A.J., Ampt, E.S., Meyburg, A.H., 1996. Nonresponse issues in household travel surveys. In: Transportation Research Board Conference Proceedings. Presented at the Conference on Household Travel Surveys: New Concepts and Research Needs.

Roux, S., Marchal, P., Armoogum, J., 2009. Acceptability of the use of new technologies by interviewees in surveys. Presented at the NTTS 2009 – New Techniques and Technologies for Statistics, Brussels, Belgium.

Sloof, R., Sonnemans, J., 2011. The interaction between explicit and relational incentives: an experiment. Games and Economic Behavior 73, 573–594.

Sohn, T., Varshavsky, A., Lamarca, A., Chen, M.Y., Choudhury, T., Smith, I., Consolvo, S., Hightower, J., Griswold, W.G., Lara, E.D., 2006. Mobility detection using everyday GSM traces. In: Proceedings of the Eighth International Conference on Ubiquitous Computing (Ubicomp 2006). Springer, pp. 212–224.

Stenneth, L., Wolfson, O., Yu, P.S., Xu, B., 2011. Transportation mode detection using mobile phones and GIS information. In: Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS '11. ACM, New York, NY, USA, pp. 54–63.

Stopher, P., FitzGerald, C., Zhang, J., 2008. Search for a global positioning system device to measure person travel. Transportation Research Part C: Emerging Technologies 16, 350–369.

Stopher, P.R., Greaves, S.P., 2007. Household travel surveys: where are we going? Transportation Research Part A: Policy and Practice 41, 367–381.

Sun, L., Zhang, D., Li, B., Guo, B., Li, S., 2010. Activity recognition on an accelerometer embedded mobile phone with varying positions and orientations. In: Yu, Z., Liscano, R., Chen, G., Zhang, D., Zhou, X. (Eds.), Ubiquitous Intelligence and Computing. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp. 548–562.

Viterbi, A., 1967. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE Transactions on Information Theory 13, 260–269.

Wang, S., Chen, C., Ma, J., 2010. Accelerometer based transportation mode recognition on mobile phones. In: 2010 Asia-Pacific Conference on Wearable Computing Systems (APWCS). Presented at the 2010 Asia-Pacific Conference on Wearable Computing Systems (APWCS), pp. 44–46.

Widhalm, P., Nitsche, P., Brändle, N., 2012. Transport mode detection with realistic Smartphone sensor data. In: Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012). Tsukuba, Japan, pp. 573–576.

Wolf, J., Guensler, R., Bachman, W., 2001a. Elimination of the travel diary: experiment to derive trip purpose from global positioning system travel data. Transportation Research Record 1768, 125–134.

Wolf, J., Loechl, M., Myers, J., 2001b. Trip Rate Analysis in GPS-enhanced Personal Travel Surveys. Presented at the International Conference on Transport Survey Quality and Innovation, Kruger Park, South Africa.

Yang, J., 2009. Toward physical activity diary: motion recognition using simple acceleration features with mobile phones. In: Proceedings of the 1st International Workshop on Interactive Multimedia for Consumer Electronics, IMCE '09. ACM, New York, NY, USA, pp. 1–10.

Zheng, Y., Li, Q., Chen, Y., Xie, X., Ma, W.-Y., 2008. Understanding mobility based on GPS data. In: Proceedings of the 10th International Conference on Ubiquitous Computing, UbiComp '08. ACM, New York, NY, USA, pp. 312–321.

Zmud, J.P., Arce, C.H., 2000. Item nonresponse in travel surveys: causes and solutions. In: Transportation Research Circular. Presented at the Transport Surveys: Raising the Standard.