View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

ELSEVIER
Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia - Social and Behavioral Sciences 48 (2012) 1033 – 1046
Procedia
Social and Behavioral Sciences

Transport Research Arena– Europe 2012

A strategy on how to utilize smartphones for automatically reconstructing trips in travel surveys

Philippe Nitsche[a*], Peter Widhalm[a], Simon Breuss[a], Peter Maurer[a]

[a]*Austrian Institute of Technology, Giefinggasse 2, 1210 Vienna, Austria*

Abstract

The acquisition of travel data is currently based on cost- and time-intensive questionnaires and yields mostly an incomplete picture due to limited coverage and inadequate updates. There is an urgent need for technologically supported data acquisition tools. This paper introduces a novel approach to developing a large-scale travel survey by intelligently employing data from smartphones. Based on signals of the embedded accelerometers and GPS reveivers, an ensemble of probabilistic classifiers is trained for automatically reconstructing the individual trips composing a tour, including the mode choice. In the region of Vienna, Austria, 266 hours of travel data were collected to train and evaluate the models. Using a set of 72 features, the best classification results are achieved for detecting walks (92%) and bike rides (98%). Railway modes were correctly identified in 80% of all cases, which is subject to further research. In case of GPS losses only accelerometer data are used, which still shows promising results. This allows the method to incorporate places where there is normally only a weak or no GPS signal. Future smartphone applications are intended to spread the tool among traffic users, while the effort for them should be kept to a minimum i.e. no manual entries or questionnaires are necessary. Due to the increasing popularity of smartphones, the tool has the potential to be used on a wide-spread basis.

© 2012 Published by Elsevier Ltd. Selection and/or peer review under responsibility of the Programme Committee of the Transport Research Arena 2012 Open access under CC BY-NC-ND license.

Keywords: Smartphones; mobility data; travel survey; accelerometer; GPS, transport modes, mode detection

[*] Philippe Nitsche. Tel.: +43 50550 6517; fax: +43 50550 6599. *E-mail address*: philippe.nitsche@ait.ac.at.

1877-0428 © 2012 Published by Elsevier Ltd. Selection and/or peer review under responsibility of the Programme Committee of the Transport Research Arena 2012 Open access under CC BY-NC-ND license.
doi:10.1016/j.sbspro.2012.06.1080

1. Introduction

Data acquired from travel surveys provide essential information for traffic planners, public transport providers, infrastructure authorities or researchers. They are the basics for traffic modeling and optimization of transportation services and routing. Conventional methods for collecting data for travel surveys comprise computer-assisted telephone interviewing (CATI), personal interviews (CAPI), self-interviews (CASI), mail-back or web-based questionnaires, traffic counting on cross sections or intersections as well as analyses of transport schedule inquires. Regarding large-scale travel surveys, most of these methods are cost- and time-intensive and are often only applied once a decade. Moreover, non-response issues and underreported trips are well-known problems in surveys, as discussed by Brög (1982), Richardson (1996) or Zmud (1997). Since the late 1990s, novel technologies such as Global Positioning System (GPS) devices were utilized as a supplement to measure person travel. One of the first household surveys with a GPS subcomponent was conducted in 1997 in Austin, Texas, followed by many studies that were conducted to examine the application of GPS to determine travel behavior (cf. Murakami 1999; Pearson 2001; Wolf 2001; Stopher 2007; Bohte 2009; Gong 2011). In general, the results of these studies indicate the potential of GPS devices – often combined with Geographic Information Systems (GIS) – to replace or supplement traditional methods. Common devices are either wearable or mounted in household vehicles. Nevertheless, in the year 2011, the most common problems of GPS devices for travel surveys are still signal losses in shadowed areas such as urban canyons (cf. Gong 2011) or underground transportation systems, high energy consumption and the acceptance of users carrying the device in a daily travel.

The increasing popularity of smartphones brings novel opportunities for collecting data for travel surveys. Since the early stages of mobile phone spread, tracking survey methods were subject to research. Asakura (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 (2010) proposed a method for estimating route choice models from smartphone GPS data and achieved satisfactory results, although the sample size had to be enlarged. According to an article of Lane et al. (2010), the research area of mobile phone sensing will revolutionize many business sectors, including transportation. Besides a GSM/UMTS module, modern smartphones are equipped with Assisted-GPS, wireless LAN and embedded sensors such as accelerometers, magnetometers or gyroscopes. This brings a high number of data sources, which can be intelligently employed for detecting mobility behavior.

This paper introduces a novel approach to utilizing these smartphone sensors for large-scale mobility surveys. To cope with the problems of GPS satellite losses, accelerometers have been investigated in detail. Since they measure rotational and translatory movements, places with a weak or even no satellite coverage can be incorporated. To achieve this, multivariate parametric models were fitted to the distribution of feature vectors extracted from a training set of both GPS and accelerometer data. Recognizing specific patterns in the frequency and time domain of the accelerometer signals allows the identification of transport mode changes. The goal of this research is to automatically reconstruct individual trips including the mode choice. In this context, automatically means that travelers do not have to answer a questionnaire about their mobility behavior. The method identifies the sequence of trips without manual entries. Due to the high spread of smartphones, the proposed approach permits a region-wide acquisition of mobility data and leads to an improved data basis for traffic planning. Moreover, it enables a novel way for continuous measurements, which have already been discussed by Edmonston (1995) for the US census. According to Stopher (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. An application for smartphone operating systems that automatically captures trips and transport modes would provide an innovative tool for continuous surveys as long as users are willing to download it. Incentives that are useful to raise interest are discussed in Section 6.

This paper describes the models developed and is divided into six main sections. Section 2 defines important terms that are used throughout the paper. In Section 3, the requirements for a large-scale mobility survey are explained and constraints regarding smartphones are discussed. Section 4 deals with the collection of data in order to test and validate the models for detecting trips and transport modes. This includes the development of a data collection tool used by test persons at their daily travel as well as the correction and reduction of the high amount of data acquired. Section 5 explains the models developed to identify trips and mode changes, before the classification results are presented. Discussions on a future application of the proposed method conclude this paper.

2. Important definitions

Among various studies and publications, there is no formal agreement on the definition of travel elements and trip chains. Therefore, definitions of the most important terms are given, which are mainly but not completely derived from a paper by McGuckin (2004). An illustration of the definitions and their interrelation is depicted in Fig. 1.

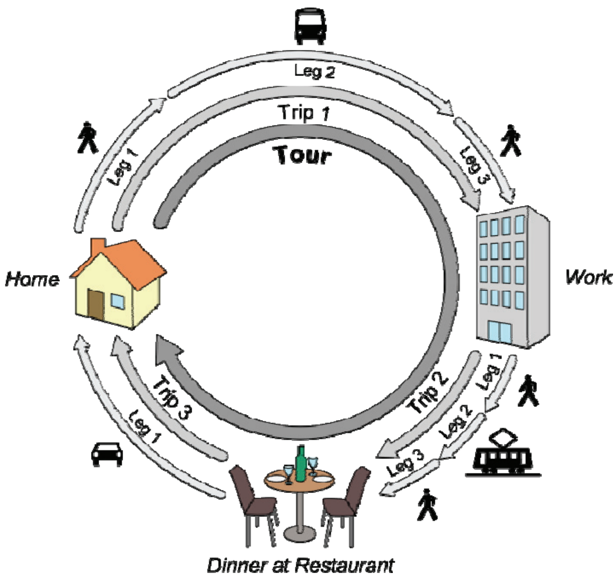

Fig. 1: Illustration of a tour, trips and trip legs

Tour	Total travel between source and destination, including both directions of travel, such as home-work-home
Trip	Direct travel between two destinations with one or more transport modes. An example for a trip is home to work.

Trip leg	Segment of a trip, which is separated by a change of transport mode or an intervening stop with a short dwell time (e.g. stop for a coffee, public transport transfers).
Activity	Each trip is travelled for the purpose of a certain activity. Living, working, shopping or studying are examples for common activities.

3. Requirements for a large-scale mobility survey

The following section deals with general requirements for travel surveys, which must be fulfilled to achieve useful results. It must be stated that a community-based method using smartphones does not fulfill all general requirements for a large-scale mobility survey. Therefore, constraints and limits of the utilization of smartphones are discussed and considered in the following sections of this paper.

3.1. Representativeness

Travel surveys deliver information about a preliminary chosen sample of a population. Results need to be representative regarding five aspects. Spatial representativeness ensures that the results are valid for the entire area of interest. In other words, the data collected must be spatially distributed instead of concentrated to a certain area. The second aspect, temporal representativeness ensures that trips are covered at different months, days and daytimes. Travel surveys also need to consider appropriate shares of socio-demographic aspects such as sex, age, education or size of household. One should also achieve a representativeness regarding transport modes, i.e. all kinds of non-motorized and motorized modes must be considered. The fifth aspect of representativeness is the trip purpose that helps to split up the given population in behaviorally homogenous groups.

When considering those requirements for a smartphone-based survey, it must be admitted that not all of these five aspects can be covered with the proposed approach. When spreading the data collection application to the smartphone community, it is not ensured that the user sample is representative. There might be a bias towards younger people or phone users with a personal technical interest. Roux (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 can be assumed that not all socio-demographic groups can be reached without specific incentives for them. Considering this constraint, the research team of this paper agreed that the proposed survey method does not serve as a replacement but as a supplement to conventional methods. Since the mobile sensing community is still in its infancy, socio-demographic representativeness can only be achieved by developing incentives for underrepresented user groups. A further discussion on incentives is given in Section 6. 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.

Another constraint for a smartphone-based survey method is the non-representativeness of trip purposes. Since it is not a goal of this research to detect trip purposes, this requirement can only be fulfilled in future developments. Encouraging research in this field has been carried out by Wolf, Guensler & Bachman (2001), Bohte & Maat (2009) or Chen et al. (2010).

3.2. Comparability

Results of travel surveys are typically compared to previous or other ones in order to determine trends and differences. Consequently, travel surveys must be conducted in a way that allows standardized data analyses and post-processing. Ideally, the contents covered by a survey as well as the data formats must be harmonized with other surveys. Comparability is also considered requirement for the smartphone-based survey method proposed in this work. The detected trip times and modes can be processed in a manner harmonized to previous or current conventional surveys.

3.3. Precise detection of trips

As mentioned before, underreported trips are a common problem in travel diaries or surveys. Trip chains and individual trips must be precisely acquired in a survey. The trip information collected should ideally contain source and destination, which could also be clustered into bigger regions. Survey data also requires temporal information for each trip, i.e. the start, end time, and further 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 forgotten in a conventional survey method. The acquisition of transport modes and stops leads to a segmentation of the trip into trip legs. The proposed method for travel surveys detects these legs by identifying transport mode changes.

3.4. User acceptance

When considering novel technologies such as phones or GPS devices as survey tools, the aspect of user acceptance plays an important role. Roux (2009) developed a model that shows 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. It can be assumed that these biases also occur in a smartphone-based survey. Consequently, the following requirements should be considered: The user interface of a smartphone application to collect travel data must be as simple 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 are necessary. 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 calculations and tasks necessary for collecting data need to be optimized and reduced. The methods developed in this project automatically reconstruct daily trips and trip legs without any questionnaire. The idea is to collect sensor and GPS data in the background, while major computation tasks are carried out on a server.

3.5. Data protection

In mobility surveys, personal data about individual travel behavior is collected. Therefore, data protection 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. A common approach is to obtain a written declaration of consent from the persons involved. However, various basic data protection issues regarding the national data protection laws must be considered in the planning and conduction of a travel survey. Initially, these include the preliminary avoidance or minimization of 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.

3.6. Technical requirements

In the initial project phase, minimal technical requirements have been defined, which comprise the operating system, computation performance as well as the availability of sensors. The research team agreed on using Android-based smartphones with Assisted-GPS and a triaxial accelerometer. Moreover, WLAN and GSM is necessary for combined positioning. It must be stated that continuous capturing of GPS and accelerometer signals may result in higher battery consumption. Therefore, smartphones known for their high battery duration should be chosen as test devices.

4. Collection of test data

Probe data was necessary to calibrate and validate the models for detecting trips and transport modes. The following section deals with the devices chosen for collecting data and testing the models and describes the procedure of data acquisition.

4.1. Phones and Sensors

Technical requirements were defined to choose appropriate devices for data collection. Based on these requirements, the following smartphone models were utilized: the Samsung I-9000 Galaxy S, the HTC Desire HD, and the HTC Desire Z (see Fig. 2). In general, they mainly differ in size, handling, GPS accuracy and power consumption, which brings a variety of usage and therefore more representative data. All selected phones contain an acceleration sensor manufactured by Bosch. These are the SMB380 or the almost identical BMA150 sensor, which measures the acceleration in all three spatial axes at 100 Hz. The assisted GPS of the devices calculates the position, speed, bearing and accuracy of the position once per second. In addition, the phone estimates its position based on the available GSM Cell Ids and WLAN cells at specific intervals.

Fig. 2: Test devices utilized for data collection and validation

4.2. Test data collection and correction

A data logging application has been developed for the Android operating system. It acquires data from all embedded sensors at the highest possible sampling frequency and synchronizes this data with the available location and speed information. Data was collected in the city of Vienna, Austria, as well as in surrounding districts. During travel, the test persons added annotations containing the current travelling mode (see Fig. 3(a)). It was differentiated between the modes walk, bicycle, motorcycle, car, bus, electric tramway, metro, train, and wait. Since travelling and annotating at every mode change can be hectic in real life, errors occured. Therefore, a correction tool was implemented in MATLAB to correct these errors manually and to ensure the quality of the recorded data (see Fig. 3(b)). Using this tool, each person checked the mode changes of the trips and corrected missing or wrong annotations.

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

4.3. Data amount

Data was collected by 14 test persons over a period of two months. Each subject was equipped with a smartphone and the preinstalled data logging application, which had to be turned on during their daily trips. When starting a new trip, the test person had to start the tracking application on the phone and annotate the current mode of transport for each trip leg. The total amount of ground truth data comprises 266 hours of travel time with a high variety of transport mode shares (cf. Tbl. 1). It was of major importance that the shares are equally distributed and representative regarding the transport system in the Vienna region. At the time of this paper completion, data collection is still ongoing and data of underrepresented transport modes (e.g. motorcycle, bus and electric tramway) will be enhanced. The models developed will then be fitted with the enhanced data set in order to improve their accuracy. However, preliminary results in Section 5 are based on the 266 hours data.

Tbl. 1: Share of travel times per transport modes

Mode	Time [h:mm]
Walk	30:41
Bike	37:16
Motorcycle	11:07
Car	59:44
Bus	9:05
Electric Tramway	4:04
Metro	28:32
Train	70:53
Wait	14:49
Overall	266:11

5. Methodology for detecting trips and classifying modal choice

The proposed method involves filtering the data and extracting suitable features, employing statistical models to periodically calculate probabilistic classifications and recombining these individual classifications into the most likely sequence of transportation modes used. In this section, each of these steps is explained in more detail.

5.1. Feature extraction

The smartphones are equipped with a GPS receiver and acceleration sensors for all three spatial axes. If the GPS signal is lost, a rough location estimate can be derived from the cell network. The positioning system of the smartphone also provides a measure of accuracy for every location record. This allows retaining only the most accurate location estimates within a sliding time window. The resulting trajectory is smoothed and resampled evenly in time so that GPS speed and accelerometer data can be calculated at a desired frequency, which is chosen to be 1 Hz.

A feature vector is extracted every five seconds from data of a certain time window. In order to obtain homogeneous data it is important to detect and filter out stops and retain only data from periods where the user is moving. Stops are defined as moving slower than 0.5 m/s. The size of the time window is continuously adapted to separate stops and periods of motion. The data are presented to the classifier as soon as the time window is large enough to provide meaningful information about the transportation mode. If there is not enough location data for the stop detection, the time window is set to a minimal size of 20 seconds.

In total, 72 different features are used for classification, whereas seven features are extracted from GPS data, 64 features are derived from the accelerometer signals and one feature represents the detected phase of motion that is described in Section 5.2. The features comprise the minimum, median and maximum of speed and acceleration. Further, the maximum angle defined by three successive points of the trajectory is calculated. The accelerometer provides three-dimensional acceleration vectors, corresponding to the three spatial axes. The magnitude of the acceleration vectors is calculated and the signal is resampled to 50 Hz. By applying a Fourier transform to the signal, the amplitude spectrum is computed from 128 samples for frequencies ranging from 0 Hz to 25 Hz. The autocorrelation functions as well as the characteristic spectra of the signals produced by a walking person and a user taking the bus is shown in Fig. 4.

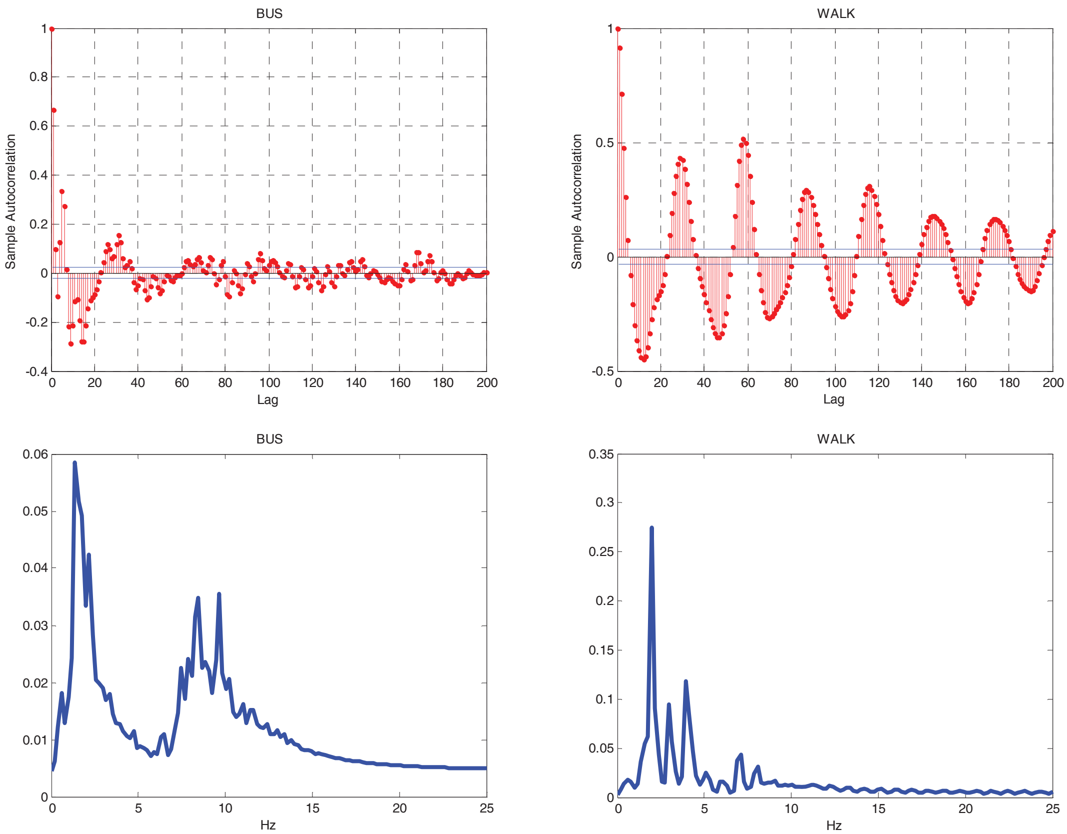

Fig. 4: Autocorrelation (top row) and amplitude spectrum (bottom row) of a signal produced by a person going by bus (left) and a walking person (right).

5.2. Modelling and classification

The collected sample data allows estimating the class conditional distribution of the extracted features for each mode of transport. The distributions were modeled as a mixture of multivariate Gaussian kernels, where the mixture components correspond to the following phases of the user's trajectory: a) stop, b) acceleration, c) deceleration and d) moving at roughly constant velocity. Probabilistic classifications can be calculated from the class posteriors according to the modeled class conditional distribution. Since the feature space is high dimensional and the training sample size is limited, the Random Subspace Method (Ho 1998) was adopted, where the original feature space is split into a number of subspaces and classifiers are trained over each of these subspaces. By combining the posteriors of the classifiers, a single classification is obtained. This concept is visualized in Fig. 5.

Every five seconds, a new feature vector is extracted and a classification is computed. However, since the classification will have a nonzero error rate, the resulting sequence of classifications may switch too frequently between different modes of transport and thus exhibit unreasonably short periods of using one mode before changing to another. Hence, transition probabilities are estimated from the distribution of the time between mode changes in the training data and the Viterbi algorithm (Viterbi 1967) is employed to identify the most likely sequence of classifications.

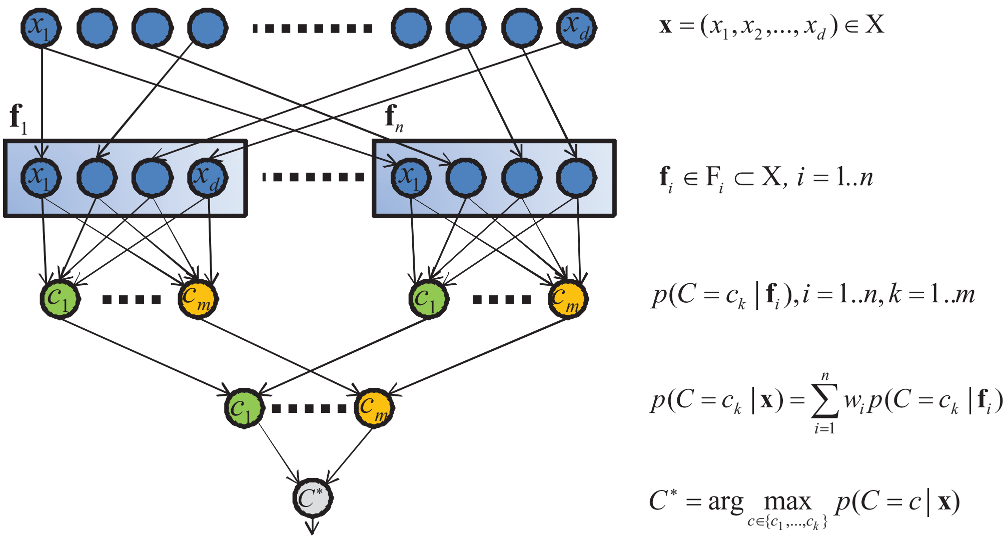

Fig. 5: The original feature space X is split into axis-parallel subspaces F_i and probabilistic classifiers are trained over each of these subspaces. The classifiers calculate posterior probabilities that the observed features were generated by a particular transport mode. Finally, the posteriors are combined using combining coefficients **w**, and the transportation mode with the highest posterior probability is chosen as classification result C*.

5.3. Preliminary results

For these experiments, 75 percent of the ground truth data is used to train the models and the remaining 25 percent as a test set for evaluation. There was no GPS signal available in 17 percent of the test data set. The confusion matrix is calculated for the complete test set as well as for the subset where no GPS data was available. In the latter case, only accelerometer features are used for classification. The results are given in Tbl. 2 and Tbl. 3, where the rows represent classification results C* and the columns represent the true labels. The number in each cell gives the percentage of predicted class labels with respect to the actual class. Due to integer rounding, the sum of each column does not always equal 100 percent.

As seen in the confusion matrix of the complete test set (Tbl. 2), the transport modes "bike" and "walk" can be identified with an accuracy of respectively 98 and 92 percent. Combining the three classes "Tramway", "Train" and "Metro" to a single class "Rails" brings an improved accuracy of 80 percent. The problems in differentiating between those three rail modes could be caused by a too small sample size. Moreover, one metro line in Vienna uses trains of the tramway network, which hinders an exact classification due to similar dynamic characteristics. 36 percent of all metro trips have been identified as walking. The source of this error can be found in unreported cases where the subjects were walking in the train.

The confusion matrix in Tbl. 3 indicates a higher accuracy for classifying bus and bike trips, while the identification of car, motorcycle and walk trips are more erroneous. Since for this subset only accelerometer signals are used, this indicates that bike and bus modes are unambiguous regarding their acceleration behavior. It must be mentioned that the subset size of motorcycle trips is too low to achieve useful results. Only six percent of the motorcycle data have GPS lacks. Furthermore, the accelerometer feature set for motorcycles seems to be similar to those for bus and walking, which causes a false classification. Further collection of underrepresented trip modes is planned to increase accuracy. However, the results are promising and demonstrate the feasibility of the proposed method.

Tbl. 2: Confusion matrix of the complete test set

Predicted class		Actual class: Bus	Car	Bike	Rails: Tramway	Train	Metro	Rails	Walk	Motor-cycle
	Bus	**77%**	3%	0%	0%	3%	2%	2%	0%	40%
	Car	0%	**76%**	0%	0%	0%	0%	0%	0%	0%
	Bike	0%	0%	**98%**	0%	0%	0%	0%	0%	0%
Rails	Tramway	20%	12%	0%	43%	52%	23%	45%	2%	2%
Rails	Train	0%	0%	0%	37%	27%	8%	23%	4%	0%
Rails	Metro	0%	2%	0%	0%	8%	31%	13%	1%	0%
Rails		20%	13%	0%	80%	87%	61%	**80%**	7%	2%
	Walk	3%	0%	2%	20%	7%	36%	14%	**92%**	8%
	Motorcycle	0%	8%	0%	0%	4%	1%	3%	0%	**50%**

Tbl. 3: Confusion matrix of the test instances where the GPS signal was lost

Predicted class		Actual class: Bus	Car	Bike	Rails: Tramway	Train	Metro	Rails	Walk	Motor-cycle
	Bus	**99%**	0%	0%	0%	0%	0%	0%	0%	60%
	Car	0%	**27%**	0%	0%	0%	0%	0%	0%	0%
	Bike	0%	0%	**100%**	0%	0%	0%	0%	1%	0%
Rails	Tramway	0%	73%	0%	4%	44%	20%	38%	4%	0%
Rails	Train	0%	0%	0%	84%	47%	20%	44%	15%	0%
Rails	Metro	0%	0%	0%	0%	6%	0%	5%	0%	0%
Rails		0%	73%	0%	88%	98%	40%	**87%**	19%	0%
	Walk	1%	0%	0%	12%	2%	58%	13%	**80%**	40%
	Motorcycle	0%	0%	0%	0%	0%	2%	0%	0%	**0%**

5.4. Limitations and future work

There are several potential improvements, which can enhance classification accuracy. Firstly, the class conditional distributions of the feature vectors are not Gaussian. Empirical analysis showed that many features follow a heavy-tailed and asymmetric distribution. Furthermore, since accuracy of measurement and annotation are limited, real world data contains outliers that affect the sample estimates of the distribution parameters. To improve the algorithms, different classes of distribution and robust estimators will be estimated that allow a better model fit without overfitting in the presence of limited amount of sample data. A promising candidate approach is to identify clusters in the data, decorrelate the features for every cluster using robust Principal Component Analysis and model the distribution along the first few Principal Axes individually based on sample quantiles.

Currently, only GPS data is used for stop detection and separation of the different phases of the users trajectory. If the GPS signal is lost, the size of the time window is reduced in order to prevent mixing different phases in one feature vector. However, decreasing the window size increases noise and the variability of the extracted features, which affects classification accuracy. Future investigations will therefore include using the acceleration sensors for stop detection and separation of the distinct phases of the trajectory. A model for phase detection based on accelerometer data can be built from samples where both GPS and acceleration data are available. When the GPS signal is lost, phase separation will be carried out using accelerometer data only.

6. Discussion on a future application

The proposed method for collecting mobility data from smartphones shows promising preliminary results, although it is still subject to research. Future smartphone applications are intended to spread the tool among traffic users, who can thus collect data for a continuous travel survey. Since not all requirements of large-scale travel surveys can be fulfilled, the method 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. First, 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 must be encouraged to participate by providing incentives. On the one hand, explicit incentives such as financial payments, vouchers or coupons are a good choice, 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, it must be mentioned that these explicit incentives may cause another bias towards specific user groups. This leads to the idea of implicit incentives which can either replace or supplement explicit ones (Sloof 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.

As mentioned before, the proposed survey method and the automatic reconstruction of trips, trip legs and transport modes can be used as a supplement to other methods. So-called mobile surveys are increasingly used as a tool for collecting data by questionnaires on mobile devices. Current research projects deal with the development and design of user-friendly and intelligent questionnaires to collect travel data with smartphones. In future large-scale surveys, the automatic trip reconstruction method presented in this paper can be combined with intelligent mobile questionnaires to cover additional information such as the trip purpose or household and participant characteristics. Generally, the authors of this paper agree that currently only a mix of different survey methods with predefined subsamples leads to improved continuous large-scale surveys.

7. Conclusions

This paper introduces a novel approach to utilizing smartphones for travel surveys. Instead of developing a questionnaire-based mobile tool, the proposed method automatically reconstructs trips, trip legs and transport modes by simply carrying the smartphone during a daily travel. The research focuses on the analysis of data captured by the embedded accelerometer combined with speed information from the GPS receiver. Multivariate parametric models are fitted to the distribution of feature vectors extracted from a training set of both GPS and accelerometer data. The collected sample data allows estimating the class conditional distribution of the extracted features for each mode of transport. Transition probabilities are estimated from the distribution of the time between mode changes in the training data. If GPS was available, 72 different features are used to train and evaluate the classifier. This results in an accuracy of 92 percent for identifying walks and 98 percent for detecting bike rides. The classification of bus, car and rail trips also shows satisfactory results, although higher accuracy will be achieved with a higher sample size. In case of GPS lacks, only accelerometer data is used, which leads to different classification rates. Modes that cause unambiguous accelerometer signals, such as busses and bikes, can be classified with superior accuracy. In a travel survey, this allows incorporating places where there is normally only a weak or no GPS signal.

Future effort will focus on using acceleration signals for stop detection and separation of the distinct phases of the trajectory. A model for phase detection based on accelerometer data can be built from samples where both GPS and acceleration data are available. 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 carried out in the scope of the Austrian research project NEMO-PHONE (Novel Methods for Collecting Mobility Data with Smartphones), which is 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

Asakura, Y., & Hato, E. (2004). Tracking survey for individual travel behavior using mobile communication instruments. *Transportation Research Part C*, 12, 273-291.

Bierlaire, M., Chen, J., & Newman, J. (2010). *Modeling Route Choice Behavior From Smartphone GPS data*. Report TRANSP-OR 101016. Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne, October 2010.

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*, 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*, 44, 830-840.

Edmonston, B., & Schultze, C. (1995). Modernizing the US Census. National Academy Press, Washington, DC.

Gong, H., Chen, C., Bialostozky, E., & Lawson, C.T. (2011). A GPS/GIS method for travel mode detection in New York City. *Computers, Environment and Urban Systems*, In Press, Corrected Proof, Available online 16 June 2011.

Ho, T.K. (1998). The random subspace method for constructing decision forests, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20, Issue 8 (Aug 1998), 832-844.

Lane, D. et al. (2010). A Survey of Mobile Phone Sensing. *IEEE Communications Magazine*, September 2010, 140-150.

McGuckin, N., & Nakamoto, Y. (2004). *Trips, Chains and Tours – Using an Operational Definition*. Paper submitted for the NHTS Conference, November 2004.

Murakami, E., & Wagner, D.P. (1999). Can using global positioning system (GPS) improve trip reporting?. *Transportation Research Part C*, 7, 149-165.

Pearson, D. (2001). *Global Positioning System (GPS) and Travel Surveys: Results from the 1997 Austin Household Survey*. Paper presented at the 8th Conference on the Application of Transportation Planning Methods, Corpus Christi, Texas.

Richardson, A., & Ampt, E. (1996). Nonresponse Issues in Household Travel Surveys. *Conference Proceedings 10 – Household Travel Surveys: New Concepts and Research Needs*. National Academy Press, Washington D.C., USA, 79-114.

Roux, S., Marchal, P., & Armoogum, J. (2009). *Acceptability of the use of new technologies by interviewees in surveys*. Paper presented at the NTTS 2009, New Techniques and Technologies for Statistics, Brussels, Belgium, February 2009.

Sloof, R., & Sonnemans, J. (2011). The interaction between explicit and relational incentives: An experiment. *Games and Economic Behavior*, In Press, Corrected Proof, Available online 26 March 2011.

Stopher, P., & Greaves, S. (2007). Household travel surveys: Where are we going? *Transportation Research Part A*, 41, 367-381.

Stopher, P., Fitzgerald, C., & Zhang, J. (2007). Search for a Global Positioning System Device to Measure Person Travel. *Transportation Research Part C*, 16, 350-369.

Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13, Issue 2 (Apr 1967), 260-269.

Wolf, J., Guensler, R., & Bachman, W. (2001). *Elimination of the Travel Diary: An Experiment to Derive Trip Purpose From GPS Travel Data*. Paper presented at the 80th Annual Meeting of the Transportation Research Board, January 7-11, 2001, Washington, D.C.

Wolf, J., Loechl, M., & Myers, J. (2001). *Trip Rate Analysis in GPS-enhanced Personal Travel Surveys*. Paper presented at the International Conference on Transport Survey Quality and Innovation, Kruger Park, South Africa, August 2001.

Zmud, J.P., Arce, C.H. (1997). *Item Nonresponse in Travel Surveys: Causes and Solutions*. Paper presented at the International Conference on Transport Survey Quality and Innovation, Grainau, Germany, May 1997.