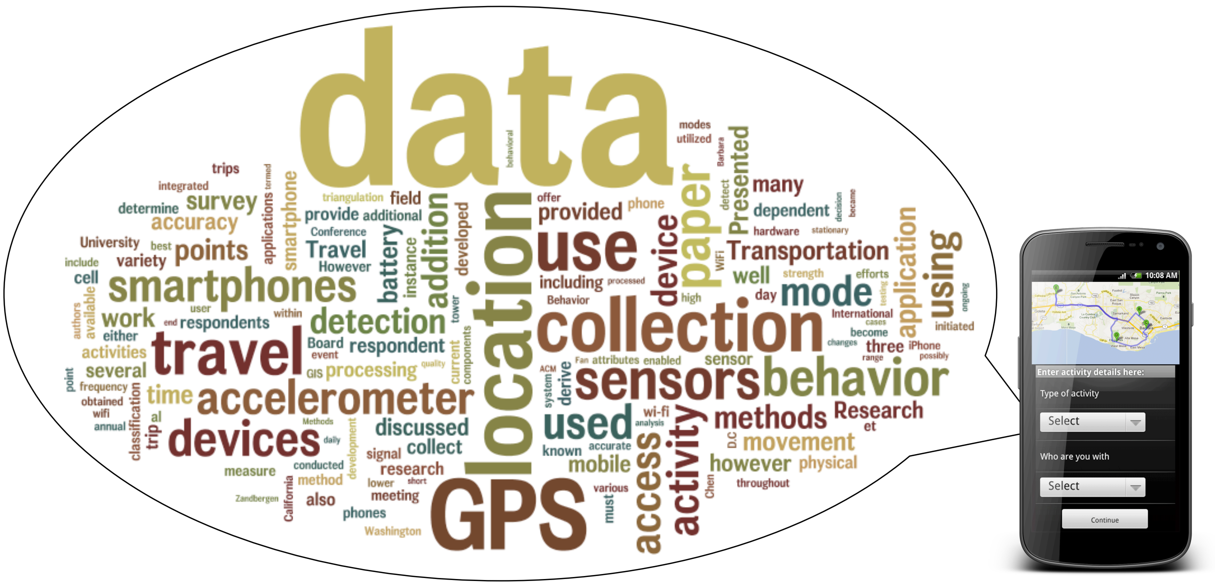

Examining the use of smartphones for travel behavior data collection

Kathleen Deutsch [1,2]
Grant McKenzie [1]
Krzysztof Janowicz [1]
Wenwen Li [1,2]
Yingjie Hu [1]
Konstadinos Goulias [1,2]

University of California, Santa Barbara [1]Geotrans Laboratory, [2]Department of Geography, [3]Center for Spatial Studies

Travel Behavior Research: Current Foundations, Future Prospects

13th International Conference on Travel Behavior Research
Toronto 15-20, July 2012

Examining the use of smartphones for travel behavior data collection

Kathleen Deutsch, PhD Candidate
Geotrans Lab
1832 Ellison Hall,
University of California, Santa Barbara
Santa Barbara, CA 93106
Phone: 1(805) 893-3867
E-mail: deutsch@geog.ucsb.edu

Abstract

The practices and standards of travel behavior data collection have changed significantly over the past decade with the introduction of mobile, location based technology. The use of GPS devices such as loggers has been a great enhancement to the field. However, with the increased ubiquity of smartphones, which come equipped with a variety of sensors useful to behavioral data collection, the possibilities and methods used in data collection are again shifting. These mobile devices offer several exciting opportunities to either collect more data from respondents with a similar amount of burden, or collect previously burdensome data (such as detailing time use by paper and pen survey) with little or no interaction from the respondent. In this paper, an overview of possible sensors is presented, as well as current research efforts in the field. A forthcoming analysis of sensor frequency, battery expenditure and accuracy of detection will also be discussed in the final version of this paper.

Keywords

Data collection, smartphone, GPS, accelerometer, sensor accuracy.

Preferred Citation

Deutsch, K.E,, G. McKenzie, K. Janowicz, W. Li , Y. Hu, and K. Goulias, (2012). Examining the use of smartphones for travel behavior data collection. Geotrans working paper. Presented at the Conference of the International Association for Travel Behavior Research, Toronto, Canada.

1. Introduction

The increased sophistication of current technology and the decrease in cost per unit of many devices has introduced many improvements to methodologies of collecting travel behavior data. Among these, the use of sensors detecting geographic position and attributes of travel such as mode and route choice have become heavily integrated in data collection efforts, primarily activity diaries. The reduction in size of GPS chips has allowed for the use of small portable devices. Discussion of the integration of GPS using various forms of hardware devices in household travel surveys has been ongoing for the last decade (Wolf, 2001; Stopher, 2006). GPS data loggers have been used in many data collection instances either as comparison to recorded activity diary data provided by respondents, or as a precursor and informant for online prompted recall surveys. The benefit of this GPS data has been realized in large-scale data collection efforts. Many planning organizations are including GPS in data collection either as a portion of the sample, or completely as a GPS only sample (for a review of several studies see Sen and Bricka, 2009). The use of GPS data to provide data through automation regarding the locations, and durations of activities and travel conducted by respondents allows for more detailed questioning of decision making or attributes of the activity such as social involvement, or lower respondent burden translating into possibly a higher response rate.

With the ubiquity of personal, GPS enabled devices such as smartphones and tablets, and the upward trend of ownership of these devices, data collection for travel behavior models has entered into an entirely new realm of possibilities. Providing respondents the option to use their own device for data collection can reduce survey costs, as well as increase the likelihood that the respondent will remember to carry the device. Applications developed for these smartphones or tablets can offer flexibility to either collect data passively with very little respondent burden, or interactively throughout the day utilizing the data network, real time data transmission, and local memory and storage options. This periodic in-situ interactive data collection can spread the time requirement of questionnaires throughout the day in smaller increments of time rather than at once at the end of the day.

With the increase of data collection using portable devices, we must be keenly aware of the systematic aspects of the devices and the sensors available for data collection. The use of smartphones introduces several mechanisms by which location and activity data can be obtained. Applications can use a variety of sensors for location retrieval such as cell tower or GPS triangulation, as well as WiFi access points. The selection of which method is used is dependent on parameters such as battery preservation and efficiency of the device to provide the location. The accuracy and reliability of these methods in deriving activities, locations and travel however varies. The inability of a device to locate three GPS satellites indoors can be overcome by using a WiFi access point or cell towers. However, WiFi access points are not always available given a lack of open networks, and in many instances cell tower triangulation is met with the possibility of lower accuracy. In addition, sensors can provide richer information regarding movement, by accessing data from the 3-dimensional accelerometer, and wifi signal strength from nearby access points. The use of accelerometer can provide further detail of movement signatures to derive mode, if sampled at a high enough frequency. In addition to these sensors, many smartphones also have built in sensors to measure environmental components such as sound, light, temperature, humidity and pressure, which could be utilized to collect further data dependent on specific research goals.

The quality and depth of behavioral data is of utmost importance for understanding the intricacies of daily and longer-term decision making for travel demand forecasting. Automating a portion of the data collection necessary allows for deeper investigation into the intricacies of daily behavior. As GPS enabled mobile devices such as smartphones and tablets become integrated into the lives of a larger percentage of the population, more flexible data collection opportunities become apparent for the field of travel behavior. These opportunities are however met with a set of difficulties and concerns that must be addressed to ensure a high quality in data collected. Quality of sensor data, system requirements such as battery, data transmission, and phone storage, as well as challenges with respect to processing sparse data to extract an accurate snapshot of a respondent's activity participation must be addressed.

This paper is largely meant as a resource paper, detailing the types of data that can be collected from various sensors, processing of this data, and research within the travel behavior field utilizing smartphones for data collection. Following an extensive review of sensors and methods developed to deal with these data, examples of implementation will be discussed. Though this paper has a large focus on reviewing methodologies already implemented, the paper is set within the framework of current development conducted by the authors, and will also be presented. In addition, it must be noted that the discussion of this paper is broadly focused and can be applied to a variety of smartphones, but the development by the authors is specifically tailored to the Android environment, and thus certain discussion pertains more so to devices using this operating system.

2. Review of existing work

2.1. Current data collection efforts

The potential use of mobile communication devices in travel behavior has been discussed within the field for many years. Early instances of the use of GPS equipped phones include the work by Shinji and Hato (2006) and Clark and Doherty (2008). The increased market share of smartphones in recent years and rapidly increasing adoption rate has accelerated the integration of this technology in data collection within the field. Since these early studies, numerous researchers have contributed to the amassing knowledge of smartphone use in travel behavior data collection. The work can be divided into two main groups of use cases, which for this paper will be termed *user initiated event detection* and *background event detection* methods.

In applications termed *user initiated event detection*, user interaction with the application is required to detect and log significant events that occur throughout the day. For instance, CycleTracks developed and used by the San Francisco County Transportation Authority was developed and used to log trips made by bicyclists once the user initiated a tracking session (Charlton et. al, 2011). This required a respondent to begin and end the tracking and recording session at the beginning and end of the trip. Following this work, NuStats has used the CycleTracks application and modified it to produce an extension (“Pacelogger”) as a proof of concept application. In addition to this, a second application “RouteScout” was developed as a next version data collection app (for more details of these two applications, see the April 19, 2012 NuStats presentation in the FHWA webinar on smartphones and travel behavior). Both of these applications again require the respondent to initiate the logging of trips, and upon the completion of the trip add additional survey questions to collect supplementary data including mode, purpose and destinations.

The applications termed *background event detection* use sensors and services in the background of the application to collect a constant stream of data. The incoming sensor data is then handled using a variety of signal detection methods and processing techniques depending on the specific interests in data collection. This streaming data and processing provides an automated detection of significant events such as activities and trips. Sensors can also be utilized to detect attributes such as mode, and additional data sources can be used to provide a finer definition of the activity location. Recent work taking advantage of smartphone sensors for automatic detection and travel detail derivation include the app of the UbiActive (Chen and Fan, 2012), and Quantified Traveler (Jariyasunant, et al. 2012) research projects. The UbiActive application uses three smartphone sensors (accelerometer, GPS and magnetic sensor) to record data in order to derive travel distance, duration and mode of the trips conducted by respondents, as well as physical activity conducted throughout the day. In addition, this work uses the Experience Sampling framework which includes a short survey triggered upon trip completion. The monitoring and data collection provides feedback to users about the physical activity, calories expended and additional possibilities to partake in active modes of transportation. In order to initiate a survey following a trip, the data is processed in real time. The Quantified Traveler likewise utilizes GPS and accelerometer to derive location and mode used during trips. The incoming stream of data is processed and translated into trip footprints (a quantification of environmental, health and monetary measures). This information is provided on a website for users to review, including comparisons of their behavior to others, and feedback, in addition to an attitudinal survey.

2.2. Sensor types and processing methods

The year 2007 was the beginning of a great shift in the mobile telecommunication industry. The mobile phone became more than a communication devise as the device became integrated with various entertainment components and services such as navigation and Internet capability. Notably, Apple’s iPhone made its debut, the blackberry became GPS enabled, and Google announced their intentions of distributing the Android platform at the Open Handset Alliance. As ownership of smartphones has become more widespread, uses and applications for the device and the accompanying hardware have also diversified. Because of this wide range of use cases, these devices have a range of built in sensors, which can be utilized for unobtrusive sensing of behavior. Each of these sensors however have a range of accuracy and application for deriving components of behavior, and therefore require a variety of processing methods.

2.2.1. Location sensors

The manners in which geographic latitude and longitude coordinates can be obtained by smartphone hardware are dependent on the operating system of the device. Methods of location acquisition using GPS can either be by GPS only or by AGPS (assisted GPS). Within an Android environment, the developer has freedom over the specification of methods used to obtain the location fix. Location can be procured by specifying for the location manager to access the fine-grained location (GPS triangulation only), course-grained (cell tower and wifi access point translateration), or best location provider (a combination of these three sensors dependent on criteria such as accuracy, battery requirement and fix retrieval time). Each of these options has tradeoffs, which will be discussed further below. The choice of which location provider is use is not however available with the current iOS application specifications which will automatically choose the best location provider.

The best method for triangulating position is dependent on the circumstances under which the location is being questioned. It is well known that GPS performs poorly in situations where satellite communication is obstructed (for example indoors or in urban settings). Cell tower derived location is known to be the least accurate, and as such has contributed to the naming convention of "coarse" grained location. Wi-fi on the other hand, can be much more accurate; however it is dependent on density of known wi-fi access point locations. In addition to existing known access points initially provided by the device services, databases such as that provided by skyhook (http://www.skyhookwireless.com/) can be used to enhance accuracy of this location data, which as advertised can provide location fixes accurate to 20 meters. Zandbergen (2009) however reported disparity in this report while utilizing this service for comparison testing on a 3G iPhone. In addition to testing wifi, Zandbergen examined location error of each of these three sensors discussed. Reported error from this work is provided in Table 1. In addition, Table 1 details the advantages and disadvantages of using these three positioning methods with respect to the drain on the battery, ability to triangulate and accuracy issues.

Table 1: Positioning Methods Comparison

	Location error* (min, max, median)	**Battery drain**	**Indoor fixes**	**First fix response**	**Requirements**	**Impediment to accuracy**
Cell tower	30m, 2,731m, 599m	lower	yes	quicker	Three or more cell towers	Cell towers not always available (rural areas), accuracy depends on density and bandwidth of cell towers and service
Wi-Fi AP	16m, 562m, 74m**	lower	yes	quicker	Wi-Fi signal nearby (not necessarily connected)	Best coverage in populated areas
GPS	.4m, 16.6m, 6.9m	higher	unreliable	Slower (clear line of sight to 3 satellites needed)	View of three or more (4 required for higher "z" dimension accuracy) GPS satellites	Better outdoors, obstruction (indoors, urban canyons, etc.) reduce accuracy, start up time to first fix is slower

*As reported in Zandbergen (2009) Accuracy of iPhone Locations… Transactions in GIS, 13(s1), 5-26
**The findings of this study directly contradict the reports by companies such as skyhook, stating a 20m accuracy for Wi-Fi AP based locations

2.2.2. Accelerometer

Largely due to the attention of gaming and entertainment offered by smartphone platforms, more recent hardware devices have improved accelerometers. First generation phones such as the Apple iPhone and early android phones included an accelerometer, but this accelerometer only measured in one dimension. Current smartphones include 3 axis accelerometers however, allowing measurement in the x, y and z dimensions. The use of this 3-dimensional movement data can be used to detect changes in the phone status, initiated for instance by a trip. Primary purposes of accelerometer measurements include the derivation of mode as well as more general identification of movement, indicating the participation in at activity at a new location. Previous work using sensors to collect travel behavior data relied primarily on GPS data to derive mode. Work included that of Zhang (2008) and Stenneth (2011) use GPS or a combination of GPS and GIS data in the later of the two to derive mode. However, GPS fixes must be obtained at a high frequency in order to differentiate between modes, which is impractical due to the high demand of power and relatively short battery life of smartphones when compared to dedicated GPS devices. Much research has been conduced on the methods by which accelerometer data should be processed. For instance, Chen and Fan (2012) followed Bouten et al. (1997) and integrated each of the three dimensions of the accelerometer and used the sum of these to determine an activity count measure, which provided an indicator of physical activity. Both lower level and higher level statistics have been implemented in classifying accelerometer data as discussed in Baek, et al., 2004. Reddy et al. (2010) discuss the process of feature extraction, including many of these lower level statistics as well as the Fourier transform coefficient for classification. They then detail the testing of several classification schemes, including the best performing method of using decision trees followed by a discrete hidden markov model. In addition to this, the authors compare different location methods (cell, GPS and Wifi) with the addition of accelerometer to illustrate the necessity of data fusion to detect mode. It is important to note as this paper illustrates, that the data obtained from the accelerometer should be fused with additional sensor data to most accurately derive activities or modes. In the instance of mode detection, the authors illustrate that omitting GPS leads to a 10.4% decrease in accuracy of mode classification. In addition, if accelerometer is utilized to determine a Boolean measure of stationary vs. non stationary, relying on the accelerometer alone could possibly lead to many misidentified activities. For instance, a person might be moving around in one location (possibly doing chores at home with their phone in their pocket), which would be identified as movement even though the location is still fixed. For this reason, location data should be used to determine whether these changes are actual changes in location.

2.2.2. Wi-fi signal strength

Wi-fi has been previously discussed as a location method provided that access points with known physical location are used in triangulation. However, a less explored use of wi-fi sensors is that of providing a mechanism to monitor stationary behavior that is less power consuming. The use of signal strength from various wi-fi access points can offer an additional measure of stability, as access points rapidly dropping in and out over short periods can indicate higher speeds of movement. The access points used in this instance do not need to be open access points, and the device does not need to be connected to the wi-fi network broadcasted by these beacons. Additionally, the physical location (although useful to have) is not necessary for detection of movement. All access points broadcast a mac address and often times have a name associated with them, and each smartphone device is equipped to measure several components including whether the device is connected, the SSID, MAC address and signal strength of each nearby access point. The use of wi-fi data is limited however to the simple classification of movement, rather than the more complex classification of mode enabled by accessing the accelerometer. Work using this method on smartphones has been limited. However, researchers have developed a methodology of detecting movement and location using computers (Krumm and Horvitz, 2004), which can be translated to smartphone devices. Although this is limiting in the devices ability to unobtrusively sense behavioral attributes, the use of wi-fi does offer a possible mechanism to reduce the battery drain experienced by the respondent in running the application on his or her phone.

3. Analysis of sensor frequency, activity inference and battery drain

At the time of this paper submission, the analysis for this paper is ongoing. The analysis of this paper involves collecting data at a variety of frequencies to determine optimal value for measurements while also monitoring the battery life of the mobile device. The tradeoff between data quality and completeness and the system requirements will be discussed in the final paper, thus providing recommendations of collection rates under several use cases (for instance activity detection versus mode detection). The full paper will be prepared by the time of the meeting, and print copies will be available for distribution. Alternatively, the full paper will be available upon completion at www.escholarship.org.

4. Ongoing development

The motivation behind this work is the ongoing development of an application to be used for data collection in which respondents' activities will be detected in real time, initiating a survey focused on several aspects of the current activity. While the focus of this paper is to determine the most optimal set of parameters for activity detection using an Android smartphone, the entire system focuses on the detection of activities, geolocation of these activities to likely places using spatio-temporal signatures, and prompting of questions related to likely activities taking place at these locations. Figure 1 illustrates the overall system for collecting this activity information.

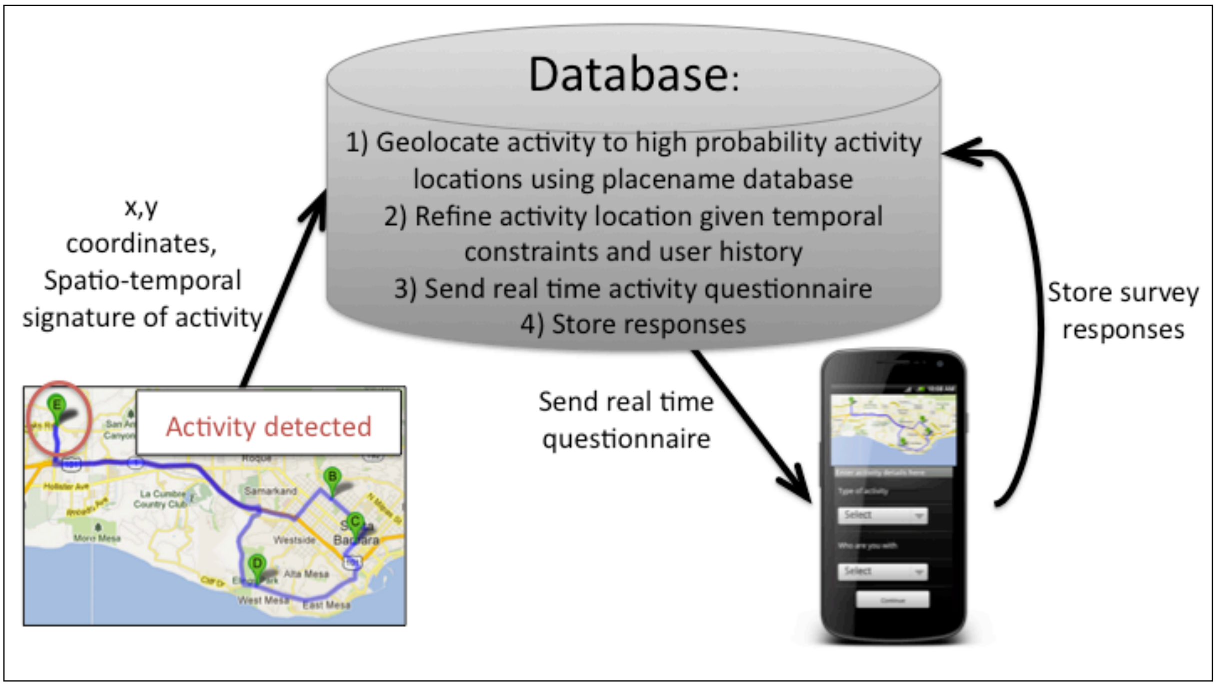

Figure 1: System Framework for Real Time Activity Survey

Acknowledgements

Funding for this research was provided by: The University of California Transportation Center, The University of California Office of the President and The Eisenhower Fellowship through the Department of Transportation. This paper does not constitute a policy or regulation of any Local, State, or Federal agency.

References

Baek, J., Lee, G., Park, W., Yun, B.J. (2004). Accelerometer Signal Processing for User Activity Detection. In: Negoita, M.G., Howlett, R.J., Jain, L.C. (eds.) KES 2004. LNCS (LNAI), vol. 3215, pp. 1611-3349. Springer, Heidelberg.

Bouten CV, Koekkoek KT, Verduin M, Kodde R & Janssen JD (1997). A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *IEEE Trans. Biomed. Eng.* **44**(3), 136–147.

Charlton, B., E. Sall, M. Schwartz and J. Hood, (2011). Bicycle Route Choice Data Collection using GPS-Enabled Smartphones. Presented at the 90th annual meeting of the Transportation Research Board, Washington D.C.

Chen, Q. and Fan, Yingling (2012*).* Smartphone-Based Travel Experience Sampling And Behavior Intervention. Presented at the 91st annual meeting of the Transportation Research Board, Washington D.C.

Clark, A. and S. Doherty (2008). Use of GPS to Automatically Track Activity Rescheduling Decsions. Presented at the 8th International Conference on Survey Methods in Transport, Annecy, France.

Itsubo S. and E. Hato (2006). A study of the effectiveness of a household travel survey using GPS-equipped cell phones and a WEB diary through a comparative study with a paper based travel survey. Presented at the 85th annual meeting of the Transportation Research Board, Washington D.C.

Jariyasunant, J., A Carrel, V. Ekambaram, D.J. Gaker, K. Thejovardhana, R. Sengupta and J. Walker. The Quantified Traveler: Using personal travel data to promote sustainable transport behavior. Presented at the 31st annual meeting of the Transportation Research Board, Washington D.C.

Krumm, J., and E Horovitz (2004). LOCADIO: Inferring motion and location from Wi-Fi Signal Strengths. Presented at Mobiquitus: The First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, Boston, MA.

Rodriguez, S. (2012). Smartphone Applications: "Pacelogger" and "RouteScout." Presented during the FHWA webinar on Smartphones and travel behavior, April 19, 2012 http://tmiponline.org/Clearinghouse/Items/20120419_-_Webinar_on_SmartPhones_and_Travel_Behavior.aspx

Reddy, S., M. Mun, J. Burke, D. Estrin, M. Hansen, and M. Srivastava (2010). Using mobile phones to determine transportation modes. ACM Trans. Sensor Network. **6**(2), 13.

Sen, S., and S. Bricka. (2009) *Data Collection Technologies—Past, Present and Future.* Presented at the International Association of Travel Behavior Research Conference, Jaipur, India.

Stenneth L., W Wolfson, P. Yu and B. Xu, (2011). Transportation Mode Detection using Mobile Phones and GIS Information. ACM SIGSPATIAL GIS '11, Chicago IL.

Stopher, P., C. FitzGerald and J. Zhang (2006) *Advances in GPS Technology for Measuring Travel.* Institute of Transport and Logistics Studies, Paper ITLS-WP- 06-15, University of Sydney.

Wolf, J., R. Guensler and W. Bachman. (2001). *Elimination of the Travel Diary: An Experiment to Derive Trip Purpose From GPS Travel Data.* Presented at the 80th Annual Transportation Research Board Meeting.

Zandbergen, P. (2009). Accuracy of iPhone Locations: A Comparison of Assisted GPS, WiFi and Cellular Positioning *Transactions in GIS,* **13**(s1): 5-26

Zhang, Y., Q. Li, Y. Chen, X. Xie and W. Ma, (2008). Understanding mobility based on GPS data. *Ubiquitous Computing*, ACM New York, pp 312-321.