WORKING PAPER

ITLS-WP-15-06

The challenge of obtaining ground truth for GPS processing

By
Peter Stopher, Li Shen, Wen Liu and Asif Ahmed

March 2015

ISSN 1832-570X

INSTITUTE of TRANSPORT and LOGISTICS STUDIES

The Australian Key Centre in

Transport and Logistics Management

The University of Sydney

Established under the Australian Research Council's Key Centre Program.

NUMBER:	Working Paper ITLS-WP-15-06
TITLE:	**The challenge of obtaining ground truth for GPS processing**

ABSTRACT:

The increasing use of GPS as a substitute or complement to conventional travel surveys has brought with it an increasing need for a reliable source of ground truth, i.e., information on the actual travel in which each respondent engaged, including the mode and purpose of such travel. Over the past decade or so, the main source that has been used for providing ground truth is the prompted recall survey. As the survey task has become less of a challenge, it has become evident that the prompted recall survey is subject to many of the same shortcomings as most self-administered surveys, with reporting error, misunderstanding of what constitutes a trip, etc. In this paper, we review some of the common problems encountered in prompted recall surveys, especially those undertaken at some period of time after the GPS survey was undertaken, and using self-report methods. Following this, the paper describes a recent experiment in using life-logging cameras to record a person's travel, together with GPS, thereby providing a new source of ground truth data that is not subject to processing issues of the GPS data, nor to self-report or other issues with prompted recall data.

KEY WORDS: ***challenge, ground truth, GPS, SenseCam, prompted recall.***

AUTHORS: **Stopher; Shen; Liu and Ahmed**

CONTACT:

INSTITUTE OF TRANSPORT AND LOGISTICS STUDIES (C13)

The Australian Key Centre in Transport and Logistics Management

The University of Sydney NSW 2006 Australia

Telephone:	+612 9114 1824
E-mail:	business.itlsinfo@sydney.edu.au
Internet:	http://sydney.edu.au/business/itls

DATE: March 2015

1. Introduction

Over the past two decades, GPS surveys have moved further and further into the mainstream of household travel surveys, having initially been used primarily as a validation tool for conventional diary surveys, until recently when there have been several instances of GPS-only household travel surveys being undertaken (e.g., Oliveira et al., 2011; Stopher et al., 2013a). There seems to be little debate now about the fact that GPS devices can collect much more accurate data about travel than any previous method of surveying the public. GPS devices are capable of recording the position of a person every second, and can therefore show accurately the time at which the travel took place, the route that was travelled, and (assuming the device is equipped with the appropriate capability) the speed of travel. All of these elements are collected with much greater detail and accuracy than has ever been possible from any form of interrogation of survey respondents. However, just as the GPS device can collect extremely accurate information on position, time, and speed, the device is not able to record information about when a trip starts and ends, what means of travel is used on each segment of a trip, what the purpose is of the trip, and who might be accompanying the survey respondent on the trip. Fortunately, over the time that GPS has emerged as an increasingly interesting device for collecting travel data, quite sophisticated methods have been developed to infer the missing information from the data collected by the GPS device, together with the reporting of some contextual and other data collected from respondents as part of their household, person, and vehicle attributes.

While software has become increasingly sophisticated over the past decade or so, there remains a persistent issue – how to assess the accuracy of the inferences. What is needed for this assessment is what has been termed “ground truth”. Ground truth is the knowledge of what the respondent really did while carrying the GPS device. It is essential if assessment of the accuracy and realism of inferential software is to be achievable. Not surprisingly, some of the reluctance to move away from conventional survey methods (self-administered and interviewer administered surveys) can be laid at the lack of objective assessment of the accuracy of the inferred characteristics from a GPS survey. In controlled experiments, it is possible to gain relatively good ground truth, by instructing those participating in the experiment on what travel activities to carry out while carrying the GPS devices. Even here, there can be a problem of a person deciding not to follow the exact instructions or encountering a situation in which the exact instructions cannot be followed. There then exists the unreliability of recounting what actually happened, especially with sufficient detail as to verify the correct ground truth of the travel. Outside of controlled experiments, however, the acquisition of ground truth becomes quite challenging. In the next section of this paper, we explore some of the potential sources of ground truth and discuss the extent to which these sources measure truth with sufficient accuracy and reliability to be used as a yardstick against which to assess the outcomes of inference from GPS data.

2. Potential Sources of Ground Truth

For the purposes of this section of the paper, we ignore controlled experiments, because the most important aspect of determining ground truth must continue to be determining it in situations where there is no control. In other words, the most important context within which ground truth needs to be measured is that of the practical sample survey.

2.1 Diary Surveys

Not a few researchers have reported an assessment of proposed software procedures against data collected by a standard conventional travel diary, usually collected for a single day (Tsui, 2005; Feng and Timmermans, 2013) when the person is carrying the GPS device. Unfortunately, the early applications of GPS to validating diary surveys showed the unreliability of the conventional diary as a mechanism to collect data about travel. Assessed against the attributes that GPS collects unquestionably, it was found that people omitted trips (about 20 percent seems to be the average rate of omission), reported trips as taking more time than they do in reality, and also reported trips as being a shorter distance than is the reality. In addition, a more detailed study of the discrepancies between GPS records and diary records (Stopher and Shen, 2011), shows that some trips are reported to take place at a different time of the day than was actually the case, sometimes are even on a different day, may also have gone to a different place than reported in the diary, or may not have taken place at all. Most of these studies have been restricted to looking only at the trip and not at mode of travel and purpose of the trip. However, it can be expected that there will be similar errors in mode and purpose, resulting largely from respondents completing diaries retrospectively and relying on faulty memory to do so.

Furthermore, if GPS is being put forward as a more accurate method of collecting travel data than the conventional self-report travel diary, then it seems to be somewhat illogical to turn to diaries as a way to check the accuracy of inferred results from the GPS survey. It is, perhaps, indicative of the challenge and accompanying frustration to achieve ground truth that has led a number of researchers to rely on such a source for ground truth.

2.2 The Prompted Recall Survey

Close to the beginning of the interest in using GPS to collect data, the idea of a prompted recall survey was put forward and tested (Bachu et al., 2001). The idea of the prompted recall survey is fundamentally to provide respondents with that data from the GPS device that each has carried and then to ask questions about the travel recorded. It is theorised that the playback of the actual travel recorded by the GPS device acts as a prompt to memory, allowing respondents to recall more accurately what actually happened. The prompted recall survey has become probably the most frequent mechanism for producing ground truth about the GPS-recorded data. It is usually carried out on a subsample of respondents, providing thereby an assessment of the accuracy of software inferences, and also providing information for potential improvement of the software.

In its original form, the prompted recall survey involved printing out a day's worth of the GPS recorded travel and inserting the maps of the travel into a paper-based survey to be sent to respondents within a few days of the GPS data collection. However, it was recognised fairly early on that the design of the questionnaire was complex, because there was a necessity to allow respondents to correct such things as whether or not the trip ends were correct or that certain trip ends might have been missed (e.g., a stop to pick someone up or drop someone off) while other trip ends might have been added, when they were, in reality, just traffic stops. In addition, information was desired on the travel modes, trip purposes, and other household members, friends, or business associates who might have travelled with the respondent. Not only did this involve a challenging design for a paper-based survey, but it also could rapidly become overly burdensome.

As a result of the problems with the paper-based survey, web-based surveys were developed to replace them (Giaimo et al., 2010). Through the use of drop-down menus and lists, the web-based prompted recall could be completed with less apparent complexity to the respondent and with greater richness in what the respondent could tell about their travel.

In another recent example of using web-based prompted recall (Montini et al., 2013), the authors noted that the level of corrections obtained form the prompted recall were quite diverse, with some being very good, while others were not useful or required extensive correction by survey staff. They also noted that there were problems created for the process when data from the GPS were not of a high standard, presumably because of using minimal and rapid processing before displaying results to respondents.

Probably the latest version of a Prompted Recall survey was conducted in Jerusalem (Oliveira et al., 2011), in which GPS data were downloaded in the field to laptop computers, processed and immediately shown to respondents to obtain their inputs on the data items not recorded by GPS. Unfortunately, the only published paper on this (Oliveira et al., 2011) provides very little information on the actual outcomes of this process, although they do indicate that the prompted recall data were able to assist in improving the GPS data significantly. Specifically, Oliveira et al. (2011) presented a comparison of the 2010 Jerusalem (Israel) Travel Habits Survey (JTHS) (3,000 households) with the 1996 traditional HTS survey. They used CAPI technology facilitated by GPS-based prompted recall for automatic edit checks for completeness of individual daily patterns and also, for identification of joint activities and trips and, carrying information obtained from one household member to other household members' diaries during the survey. The additional attributes and confirmation of missed activities captured by the GPS gathered in the PR samples was then used to generate trip correction factors. It was then used to build and calibrate models that imputed additional details for the GPS-only households.

From recent data collection by the authors, an opportunity arose to assess how accurately this type of prompted recall survey can produce ground truth. The results of this study are provided in the following section of this paper.

2.2.1 A Case Study for Prompted Recall

In the Cincinnati, Ohio GPS-only household travel survey, GPS data were collected from a sample of about 2,000 households for an average of about three days (Stopher et al., 2013a). A non-random subsample of households was asked to complete a web-based prompted recall survey some days after the GPS devices were returned and after preliminary processing was completed on the data. (Preliminary processing is important because GPS devices may collect data that can be defined as "spurious trips", resulting from a stationary device solving positioning information with insufficient satellites in view, and may also collect "clouds" of points when the device is stationary, such as at the end of a trip. Removal of these poor records, as well as other occasional unreliable records, is important to give respondents confidence in the GPS results.)

In the pilot survey for this case study, the web design allowed respondents to change the times at which trips started and ended, even though we know that the GPS is very accurate in this. The purpose of doing this was to allow respondents to fix the occasional "cold start" problem (where the GPS fails to get a fix on position immediately when the trip starts), and also to allow correction if the signal was lost somewhere during the day. However, it was found that respondents often rounded the times to the nearest 5, 10 or 15 minutes (e.g., the GPS showed that a trip started at 9:07:18 and the respondent changed this to 9:00 or 9:05). In some instances, however, respondents would change the time of a trip from the morning to the afternoon, or vice versa. Because of a large number of such changes to times, almost all of which were, in fact, clearly wrong, the main survey did not allow respondents to make changes to the times at which trips occurred. These changes that respondents made in the pilot are a clear indicator that people do not recall their travel correctly or accurately, so that any method that relies on memory is likely to be flawed and therefore to not collect ground truth.

For the web-based prompted recall survey, respondents were asked to review just one of the days of travel that was recorded on the GPS device that they had carried. Of the 2,059 households that provided complete GPS data, 601 households with 989 GPS-carrying respondents completed the prompted recall survey. (Because the survey was web-based, households without access to the Internet, and households that lacked confidence to undertake a web-based survey did not take part in the prompted recall survey.) From the 601 households, a random sample of 100 households was drawn for the purposes of conducting an in-depth analysis of the prompted recall responses.

The 100 households that were sampled had a total of 834 trips that could be analysed for the purposes of this paper, i.e., these trips matched between the prompted recall and the GPS survey. For these trips, we analysed what the GPS software produced by way of mode and purpose identification. Of the 834 trips, the prompted recall and GPS did not match on mode for 201 trips or 24 percent. Similarly, there was not a match on purpose for 496 trips or 59.5 percent. These are both rather high percentages. It is certainly the case that a number of the failures to match are a result of imperfect inference. However, what is more of concern to this paper is where the prompted recall results seem to be at variance with logic.

For the case of mode, Table 1 summarises the causes of a mismatch between the GPS and the Prompted Recall result.

Table 1: Documentation of Reasons for a Mismatch between PR and GPS on Mode

Reason for Mismatch	**Number**	**Percent of Unmatched**	**Percent of Total**
1. PR wrong based on the speed information	58	28.9%	7.0%
2. PR mistakenly joined or separated trips	10	5.0%	1.2%
3. PR put "other"	8	4.0%	1.0%
4. Loss of GPS information	26	12.9%	3.1%
5. Bus/car mismatch (on bus routes)	53	26.4%	6.4%
6. Bus/car mismatch (not on bus routes)	25	12.4%	3.0%
7. PR corrected the wrong GPS processing results	21	10.4%	2.5%
Total	201	100%	24.1%

The first three reasons for mismatch are problems that arise from people incorrectly responding to the prompted recall survey. The first one of these, and the most serious, accounting for almost one-third of the mismatches, are entirely trips that the respondent identified as walk, where the speed of travel both on average and on a point-by-point basis was much too high for walking or even jogging or running (in excess of 20 km/h). The second reason resulted from respondents incorrectly joining or separating trips. Most often, this arose when respondents considered that a round trip was a trip, rather than two separate trips, such as a trip to drop a child at school, where the trip to school and returning from school was joined into a single trip. The third reason is a bit more perplexing. Apart from the fact that the modes did not include skateboarding, it is curious that some of the trips could have been conducted by a mode that was not in the prompted recall list. It is also possible that people either do not understand the mode concept or that they do not recall what mode they used, so the insert “other” because of one or other of these situations. These three reasons together account for almost 40 percent of the mismatches on mode and slightly less than 10 percent of all of the trips that matched between GPS and PR.

Reason 4 arose mainly where some GPS information was missing, so that the prompted recall may be right, but also could be wrong. These are instances where it simply cannot be determined whether the Prompted Recall result was correct. Reasons 5 and 6are generally cases where one of the two sources indicated bus and the other indicated car, but where it is not possible to determine which one is correct. Where the trip is on a bus route, it is usually a case that the Prompted Recall respondent indicated car, while the GPS software indicated bus, and the trip was made entirely along bus routes and included a number of stops at bus stop locations. Of course, because most bus stops in the US are at intersections, these could also be traffic stops for a car. However, the results are again indeterminate as to which is actually correct. Reason 6 arose because there is a part of the study area for which bus route and bus stop GIS information was not provided, so the trips were identified by the GPS software as car, and by the prompted recall survey as bus. Again, there is not a clear outcome in these instances, but it is reasonable to suppose that the prompted recall may be correct.

Finally, the last reason is the clear benefit of the prompted recall. These are cases where the prompted recall was clearly right and the GPS processing result was wrong. For example, the GPS may have identified a particular trip as being by bicycle, when it was, in fact, a car trip, and the respondent identified it as a car trip. These constitute just over 10 percent of all the mismatches and about 2.5 percent of the total number of matched trips.

From the point of view of mode, it cannot be stated with any confidence that the prompted recall survey has provided ground truth. Rather, on trips that could be matched between the GPS and Prompted Recall, the Prompted Recall was certainly wrong in around 9 percent of cases and possibly or probably wrong in another 10 percent of cases. A level of error of between 10 and 20 percent is not acceptable for ground truth on mode.

For purpose, it is first necessary to state that information for purpose coding was quite incomplete. The primary sources of information were the reported addresses for work, school and the most frequently used grocery stores. There was no useful GIS of land use, coded to a level that would permit identification of purpose. One of the problems for purpose coding is illustrated by the failure of people to provide their workplace addresses. For the total sample (not just those analysed in this section) there were 2,868 workers, of whom 2,160 provided sufficient workplace address information to allow a geocode to be determined for their workplace. 50 provided partial information, but insufficient to allow geocoding, while 658 provided not information at all.

Misreporting of purpose was very common. Anecdotally, we see an adult from a household who takes a child to school or picks a child up from school and reports this as a school trip for the adult. This also happens for some other purposes, where an obvious pick-up or drop-off activity is reported as having the purpose of the person picked up or dropped off, rather than the purpose of pick up/drop off. We also find numerous instances of other respondents who make a pick up or drop off trip and combine the two trips into a single trip from home back to home. In other cases, trips with much longer activity durations are also combined into a single trip and the purpose information is not provided correctly. Table 2 provides the statistics for the 496 mismatches on purpose.

Table 2: Documentation of Reasons for a Mismatch between PR and GPS on Purpose

Reason for Mismatch	**Number**	**Percent of Unmatched**	**Percent of Total**
1. PR wrong based on known addresses	80	16.1%	9.6%
2. PR misunderstood the definition of trip purpose	5	1.0%	0.6%
3. PR refused to provide the information	51	10.3%	6.1%
4. PR mistakenly joined or separated trips	15	3.0%	1.8%
5. Lack of land-use information	141	28.4%	16.9%
6. Mismatch due to the trip purpose of "other"	89	17.9%	10.7%
7. Multi-function places	51	10.3%	6.1%
8. Unknown reason for the mismatch	39	7.9%	4.7%
9. PR corrected the wrong GPS processing results	25	5.0%	3.0%
Total	496	100.0%	59.5%

The first four reasons for a mismatch are all clearly prompted recall errors. The first reason arises in instances where the information on the maps is quite clear that this is wrong. For example, a person indicates that a location is "home" when it is clearly non-residential and does not match the information provided in the survey as to the home address. The second reason is where respondents clearly indicated a purpose that does not match what can be seen on examination of the GPS record. Situations like a person who drives a child to school claiming the trip purpose to be school for the driver typify the situation for this outcome. The third reason is self-explanatory and is often repeated in conventional surveys where a percentage of respondents simply leave purpose blank. The fourth reason arises where respondents incorrectly joined two trips or separated a trip into more than one and thereby misidentified the purpose. These four reasons account for almost one-third of the mismatches on purpose, and also for over 18 percent of the total trips analysed.

Reason 5, which is the single largest problem for mismatching arises where respondents did not provide requested address information for schools, workplaces, and shopping locations, with the result that the GPS software was unable to identify the probable purpose of the trip. This failure to provide address information affected almost 17 percent of all matched trips and accounts for almost one third of the total mismatches on purpose.

In the Cincinnati survey, work, school, home, and grocery shopping were the only purposes that could be identified with some degree of expected correctness by the inference software. However, an attempt was made to infer purposes such as social-recreation, eat meal, and other. Incorrect identification of these non-work, non-school, non-shopping purposes is the primary cause for the mismatches on reason 6. Similarly, reason 7 arises where the destination is something such as a shopping centre or home, and the purpose may not be shopping at the shopping centre (but could be eat meal, personal business, social, etc.) or home where the purpose is work, because the person is working from home.

Reason 8 are mismatches that we cannot explain, while reason 9, accounting for 5 percent of the mismatches and about 3 percent of the total data in this test, are cases where the prompted recall clearly offered correction to the inferences produced by the GPS software. Again, as in the case of mode, only around 3 percent of the total matched trips can be clearly corrected by the prompted recall survey, while the prompted recall survey would lead to incorrect changes in 18 percent of cases. Again, it has to be concluded that a prompted recall survey of this type is not capable of providing ground truth on purpose.

Further to the analysis of mode and purpose mismatches between PR and GPS, respondents' misunderstanding of the concept of a trip and a trip segment, as used in the profession, might be an issue that would interfere with the ability to gain relatively good ground truth. For example, if a person goes to school, but spends a couple of minutes to buy breakfast on the way, there will be two trip segments shown on the prompted recall website, but the person might think the two trip segments should be joined into one trip from home to school. Given this assumption, it was felt to be interesting to do a detailed analysis regarding people's edits on trips. In the Cincinnati survey, respondents were shown trip segments, so as to enable the collection of details about each trip segment, where a segment is defined as that part of a trip that is carried out on a single mode of travel (Axhausen, 1995). There are four possible actions for respondents to edit the trip segments (add/join/separate/delete). Table 3 provides the statistics for these four possible actions.

Table 3: Trip Identification Analysis

Types of edits	**PR correct**	**GPS correct**	**Not sure**	**Total**	**Percent of total**
1. PR added or extended a trip	34	8	2	44	30.1%
2. PR joined a trip	10	25	2	37	25.3%
3. PR separated a trip	7	0	0	7	4.8%
4. PR deleted a trip	6	51	1	58	39.7%
Total	57	84	5	146	100.0%
Percent of Total	39.0%	57.5%	3.4%	100.0%	

It can be seen clearly that the original GPS data are more correct than the prompted recall data, 57.5 percent compared to 39.0 percent. In those 84 GPS correct trips, 33 of them (39.3%) were due to the respondents' misunderstanding of the concept of a trip, either wrongly adding or joining a trip. A typical case is that many respondents joined two trip segments neglecting that they are different modes, such as a walking trip followed by a bus trip, and they most likely joined the two trips

together. It demonstrates that people have a problem to provide ground truth. On the other hand, for the 57 correctly edited trips, in nearly half of them people extended the trips to a correct place, which fixed the GPS signal loss issue. Besides, it can be seem from the table that respondents are most likely to delete trips and least likely to separate a trip into two or more, (possibly because to separate a trip is a bit difficult for them, although all seven separated trips are correct edits).

2.2.2 Conclusions on Prompted Recall

While the case study on prompted recall shows the problems that arise with self-report methods even of validating the times at which travel took place, it is acknowledged that there are other methods of administering a prompted recall survey. As discussed earlier in the paper, some researchers have used a laptop or tablet method to undertake a prompted recall, where GPS data are downloaded and partially processed immediately on a laptop or tablet computer, and shown to the respondent during the GPS retrieval activity (Oliveira et al., 2011). Because this process allows interviewer intervention in the completion of the prompted recall survey, which could eliminate some of the self-report errors discussed in this paper, it could be expected that this process may lead to a somewhat better acquisition of ground truth. However, while the authors of this paper have no experience with this method, it still raises several questions again on the adequacy of the resulting ground truth data. Our experience is that automated processing of the GPS data stream into trips is still subject to some error, due to GPS device properties, as reported also by Montini et al. (2013). For this reason, we insist on undertaking a manual map-editing step before data are provided to respondents for a prompted recall (Stopher et al., 2013b), which delays when the data can be presented to the respondent. Because this process clearly cannot be implemented in the field, the data presented in the field to the respondent may be of questionable quality. This seems likely to lead to lack of confidence in the GPS results for some respondents, as well as potential confusion about what the GPS has recorded. There may also be a situation in which entire trips are missing from the record, and for which information cannot be obtained in the interviewer-based prompted recall.

Second, the costs associated with a face-to-face interview and GPS processing during the retrieval activity seem likely to add substantially to survey costs and could make the prompted recall survey unaffordable in a number of instances. It also requires a significant workforce in the field to conduct such a prompted recall and we would question whether the value of the resulting data is sufficient to justify the expense.

3. A New Alternative for Collecting Ground Truth

One of the developments of the early 21st century is that of life logging. One class of life logging is called visual life-logging (Wang and Smeaton, 2013). This process can be supported by a wearable life-logging camera that takes still photos (e.g., SenseCam and Narratives), or videos (e.g., GoPro). Because taking videos requires more battery consumption for a continuing daily record and may cause more ethical issues, photos are typically recorded as a product of life logging. Most life-logging cameras take photos at a pre-specified frequency or can be triggered to take a photo by changes of sensors or wearer intervention, such as tapping the device a prescribed number of times. A fish-eye lens with a wide angle is used to capture the view from the wearer. One day’s activities can therefore be “logged” into thousands of pictures. Like most digital devices, battery life is always a concern for users. Since the purpose of the camera is to capture one’s life, most life-logging cameras currently on the market can last approximately two to three days, but can be recharged at the end of each day, so that battery life does not become an issue.

Such cameras have already been used in research in several fields, mostly in aspects of health research. For example, Silva et al. (2013) report use of such life-logging cameras for memory rehabilitation in patients who have lost memory due to accidents or illnesses. Life-logging cameras have also been reported in other health-related studies, such as studying the sedentary behaviour of subjects (Kerr et al., 2013) and also undertaking dietary analysis (O’Loughlin et al., 2013). In addition to health studies, life-logging cameras are also used in social reflection (Fleck and Fitzpatrick, 2009) where professionals can review their previous practice and reflect on what they did and how they might improve their practice in future. Kelly et al. (2011) investigated active and sedentary travel behaviour by using life-logging cameras. Although their work was initially focused on public health, it shows a potential that these state-of-the-art cameras can help to collect travel information that travel diaries usually report.

Shen and Stopher (2013) used the SenseCam to collect experimental ground truth data for GPS processing. Each picture taken by the SenseCam is time-stamped so that the pictures, taken every 30 seconds, can be matched to a GPS record. Assuming that respondents wear the camera on the upper front of the body, the pictures provide reliable information about what the person is doing and what the surroundings are. When travelling, it is usually easy to recognise whether the person is driving a car, is a passenger in a car, a passenger in a bus, a passenger on a train, or is bicycling, or walking, etc. If the person is engaged in an activity, the surroundings give clues as to whether the person is shopping, at work, in a social gathering, eating a meal, etc.

Some examples of pictures taken with life logging cameras are shown in Figures 1 through 6. Each of these is readily identifiable to a mode of travel or an activity, as noted in the caption of each.

Figure 1: Pictures Indicating Train as Mode of Travel

Figure 2: Travelling by Bus

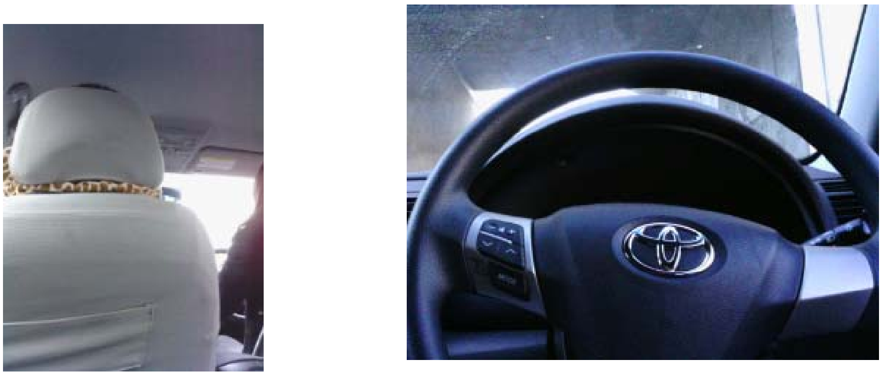

Figure 3: Car Passenger (Left) and Car Driver (Right)

Figure 4: Walking

Figure 5: Activity of Shopping

Figure 6: Activity of Eating

The paper also suggests that missing GPS data is the main issue for GPS surveys, and the SenseCam can record these missing trips, resulting in a 20 percent increase of accuracy for trip identification. According to the images from SenseCam, the majority of the missing GPS trips are walk trips, which indicates that GPS may record trips by other modes more accurately than walk trips. Given the main reasons for missing GPS data (i.e., cold start, short duration travel, and travelling in urban canyons), it is a reasonable result. For those trips not missing in GPS records, the accuracy of mode and purpose detection results can be increased by 10 percent and 15 percent, respectively by the life-logging camera.

As life-logging cameras become more widely accepted by the public for their own use, we anticipate that they will also become possible to use to collect 'ground truth' for GPS surveys. They are clearly superior to any self-report or interview method. Currently, work is in progress to automate recognition of the features in a picture, to allow computer processing of the images to identify mode and possibly purpose. However, we also anticipate that some type of artificial intelligence procedure is likely to be needed in the end to make the use of the photos effective. It will also be necessary to remove some of the restrictions on the photos and their accompanying data that are currently imposed by some manufacturers. Should these restrictions be removed, then the use of life-logging cameras on a small subsample of a GPS survey should prove very effective in providing ground truth.

4. Conclusions

While many researchers have depended on various prompted recall methods to procure ground truth for the purposes of validating software processing of GPS data and to assist in improvements to such software, the authors of this paper are strongly of the opinion that the results of all such surveys are deficient in defining ground truth. Until recently, there has been no reliable source of ground truth, especially for general surveys of the public. Even carefully designed and controlled experiments have lacked an ability to produce reliable ground truth.

This paper has illustrated some of the issues that arise with attempting to use prompted recall surveys to obtain ground truth. These include misunderstanding of the concept of a trip, as used in the profession, failure to identify mode or purpose for a trip, claiming mode or purpose that is not consistent with the factual data, etc. With all of these problems, some of which can be ameliorated to a degree by using face-to-face interrogation, it is concluded that prompted recall cannot provide sufficiently accurate ground truth for the purposes of validating software inferences for GPS records, nor can it be used to assist in significant improvements to inferential software.

In place of the prompted recall survey, it is suggested that the use of life-logging cameras, which are becoming increasingly prevalent, could offer a useful and much more accurate method of obtaining ground truth. While the cameras have their own limitations, such as difficulty of use in low light conditions, and the relative infrequency of picture taking (compared to the frequency of GPS positioning usually employed), these limitations seem to be relatively unimportant for the purposes of validating GPS inferences. Further, life-logging cameras have the potential to provide better data for training through artificial intelligence, thus allowing improvements in the inferential software and providing a more acceptable assessment of the accuracy of inference.

References

Axhausen, K.W. (1995). Travel Diaries: An Annotated Catalogue 2nd Edition, Draft Working Paper, Institute für Straßenbau und Verkehrsplannung, Leopole-Franzens Universität, Innsbruck, Austria.

Bachu, P., R. Dudala, and S. Kothuri (2001), ‘Prompted Recall in Global Positioning Survey: Proof of Concept Study’, Transportation Research Record Number 1768, National Academy of Science, Washington, DC, 106-113.

Feng, T. and H.J.P. Timmermans (2013). Transportation mode recognition using GPS and accelerometer data. Transportation Research Part C Emerging Technologies 01/2013; 37:118–130.

Fleck, R. and G. Fitzpatrick, Teachers’ and tutors’ social reflection around SenseCam images, *International Journal of Human-Computer Studies*, 67 (2009) 1024–1036

Giaimo, G., R. Andersen, L. Wargelin, and P. Stopher (2010). Will It Work? Pilot Results from the First Large-Scale GPS-Based Household Travel Survey in the United States, *Transportation Research Record No. 2176*, pp. 26-34

Greaves, S., S. Fifer, R. Ellison, and G. Germanos (2010), ‘Development of a Global Positioning System Web-Based Prompted Recall Solution for Longitudinal Travel Surveys’, Transportation Research Record Number 2183, National Academy of Science, Washington, DC, 69–77.

Kelly, P., A. Doherty, E. Berry, S. Hodges, A. Batterham, and C. Foster. (2011) Can we use digital life-log images to investigate active and sedentary travel behaviour? Results from a pilot study. *International Journal of Behavioral Nutrition and Physical Activity* 2011,8:44.

Kerr, J., S. J. Marshall, S. Godbole, J. Chen, A. Legge, A. R. Doherty, P. Kelly, MS, M. Oliver, H. M. Badland,and C. Foster. (2013). Using the SenseCam to Improve Classifications of Sedentary Behavior in Free-Living Settings. *American Journal of Preventive Medicine*. 2013;44 (3):290–296

Moiseeva, A., J. Jessurun and H. Timmermans (2010). Semiautomatic Imputation of Activity Travel Diaries: Use of Global Positioning System Traces, Prompted Recall, and Context-Sensitive Learning Algorithms, *Transportation Research Record Number 2183*, National Academy of Science, Washington, DC, 60–68.

Montini, L., N. Reisser-Schüssler, and K.W. Axhausen (2013). Field Report: One-Week GPS-based Travel Survey in the Greater Zurich Area, 13th Swiss Transport Research Conference, Monte Verità / Ascona, April 24 – 26, 2013.

Oliveira, M.G.S., P. Vovsha, J. Wolf, Y. Birotker, D. Givon, and J. Paasche (2011), ‘Global Positioning System-Assisted Prompted Recall Household Travel Survey to Support Development of Advanced Travel Model in Jerusalem, Israel’, *Transportation Research Record Number 2246*, National Academy of Science, Washington, DC, 16-23.

O'Loughlin, G., S. Cullen, A. McGoldrick, S. O'Connor, R. Blain, S. O'Malley and G. D. Warrington. (2013). Using a Wearable Camera to Increase the Accuracy of Dietary Analysis. *American Journal of Preventive Medicine.* 2013;44(3):297–301

Shen, L. and P.R. Stopher (2014), Using SenseCam to Pursue 'Ground Truth' for GPS Travel Surveys, *Transportation Research Part C.* vol.42, pp. 76-81

Silva, A., Pinho, S., Macedo, L. and Moulin, C. (2013).Benefits of SenseCam Review on Neuropsychological Test Performance.*American Journal of Preventive Medicine* 2013; 44(3). Pp. 302–307

Stopher, P.R., P. Bullock, and Q. Jiang (2003). Visualising Trips and Travel Characteristics from GPS Data, Road and Transport Research, 12 (2), pp. 3-14.

Stopher, P. and L. Shen (2011). An In-Depth Comparison of GPS and Diary Records, *Transportation Research Record No. 2246*, pp. 32-37, Washington, DC.

Stopher, P.R., C. Prasad, L. Wargelin, and J. Minser (2013a). Conducting a GPS-Only Household Travel Survey, in J. Zmud, M. Lee-Gosselin, M. Munizaga and J.A. Carrasco (eds), *Transport Survey Methods: Best Practice for Decision Making*, Bingley, UK: Emerald Group Publishing Limited, 91-113.

Stopher, P.R., S.P. Greaves and L. Shen (2013b), Comparing two processing routines for GPS traces – Lessons learnt, paper to be presented to *36th Annual Australasian Transport Research Forum ATRF 2013*, Brisbane, Australia, 2-4 October.

Tsui, S. Y. A. (2005) An enhanced system for link and mode identifications for GPS-based personal travel surveys, Dissertation, Graduate Department of Civil Engineering, University of Toronto

Wang, P. and A. F. Smeaton. (2013). Using visual lifelogs to automatically characterize everyday activities. *Information Sciences*, 230 (2013) 147–161