

Transportation Activity Analysis Using Smartphones

Yu Xiao*, David Low*, Thusitha Bandara*, Parth Pathak*, Hock Beng Lim*,
Devendra Goyal[†], Jorge Santos[†], Caitlin Cottrill[†], Francisco Pereira[†], Chris Zegras[‡], Moshe Ben-Akiva[‡]

*Intelligent Systems Center, Nanyang Technological University, Singapore

[†]Singapore-MIT Alliance for Research and Technology, Singapore

[‡]Massachusetts Institute of Technology, Cambridge, MA, USA

I. INTRODUCTION

Transportation activity surveys investigate when, where and how people travel in urban areas to provide information necessary for urban transportation planning. In Singapore, the Land Transport Authority (LTA) carries out such a survey amongst households every four years. The survey is conducted through conventional questionnaires and travel diaries. However, the conventional surveys are problematic and error-prone. We are developing a smartphone-based transportation activity survey system to replace the traditional household surveys, which can potentially be used by LTA in future.

Smartphones equipped with various sensors can be used to collect GPS, GSM, accelerometer and other contextual information for place detection [1], trajectory tracking [2] and transportation mode detection [3], [4], [5]. Our smartphone-based survey system is designed to seamlessly capture the daily transportation activity profiles of users, including their modes of transportation and the places they visit. We develop algorithms to process the data collected from smartphones for transportation mode detection and stop detection.

Our transportation mode detection algorithm uses the speed statistics derived from GPS and cellular network information, and the standard deviation of the magnitude of force obtained from accelerometer samples. Our algorithm can distinguish more transportation modes than previous works [4], [5]. For example, our algorithm refines the motorized transportation modes into bus, MRT (Mass Rapid Transit)¹ and taxi, which is essential for reporting public transportation activity.

Our stop detection algorithm can identify when and where the user stops and can also infer whether the stops are public transportation transfers. The algorithm is based on geographical moving distance in time windows, the pattern of the change in cellular identification, and the context of transportation mode.

Our algorithms address the following issues in a novel manner:

- *Missing GPS signals in indoor places and outdoor places surrounded by high-rise buildings.* Unlike previous works such as SenseLoc [2] and GeoLife [3] that derive speed and location information only from GPS fixes, we augment GPS-based positioning with network-based localization to reduce the impact of missing GPS signals.

¹The MRT is a rapid transit system that forms the backbone of the public transportation system in Singapore.

- *Coarse-granularity of network-based localization which results in low localization accuracy when the smartphone is roaming within the coverage of a cellular cell.* We first identify a stop as a stay within a certain area for a certain duration. After that, we use the pattern of the change in cell identification to filter out the false positives caused by the errors of the network-based localization.
- *The occurrence of missing GPS signals can be used for detecting the entrance into indoor places, but it may also occur at outdoor areas.* We use the transportation mode context to filter out such false negative detections.

The results of trip analysis using our algorithms can be accessed via our project web portal. Using the web portal, survey participants can validate the travel activities detected automatically by the system. Compared with traditional manual surveys, our system can provide more accurate results while minimizing the effort required from the users. For example, the survey participants do not need to manually log their status during trips.

II. TRANSPORTATION ACTIVITY ANALYSIS

Our survey system consists of smartphones that collect GPS, GSM and accelerometer data, and servers that run data analysis algorithms and host the survey web application. In this paper, we focus on the data analysis algorithms.

A. Data pre-processing

We pre-process the raw data to obtain information for stop and transportation mode detection, e.g. calculate the force based on acceleration in x, y and z dimensions, and obtain the position estimate corresponding to GSM cell identification and signal information from remote location servers.

B. Trace generation

We create traces of positions using GPS fixes and network-based localization. The idea is to fill the intervals of GPS fixes which are caused by missing GPS signals with the GSM-based position estimates if available. A position included in a trace can be represented by a four-element tuple: $\{timestamp, latitude, longitude, source\}$, where *source* indicates whether the location is obtained from GPS fixes. All the positions included in a trace are sorted by timestamp.

C. Stop detection

Given a trace of N positions $P = \{P_t, t = [t_0, t_{N-1}]\}$, where $t_i \leq t_j$ if $i \leq j$, we define a function $f(t_i, t_j)(i, j \in [0, N-1])$ to return the distance between P_{t_i} and P_{t_j} . Positions within a moving time window not smaller than T are clustered into a stop, if the maximum distance between any two of these positions is smaller than a threshold D . In addition, we detect the positions where the GPS signal gets lost or connected as potential stops in order to detect the possible entry/exit into/from indoor places. The stop detection algorithm is shown below.

```

while (i < N - 1) {
  j = i + 1;
  while (t_j - t_i ≤ T)
    j ++;
  if ((P_{t_i}.source! = P_{t_j}.source) && (j == i + 1))
    Mark P_{T_j} as a potential stop;
  while ((t_j - t_i ≥ T) && (P_{t_i}.source == P_{t_j}.source)
    && (max(f(t_k, t_l)(k, l ∈ [i, j])) ≤ D)) {
    P_{t_j} is included in the stop starting from t_i;
    j ++;
  }
  i = j;
}

```

D. Stop Merging

We merge successive stops if they are close to each other by time and distance, or if they are invalid ones caused by the errors of GSM-based position estimate. We merge the stops if the cell identification remains the same, or if at each stop the cell identification switches between two identifications and the two stops share the same set of identifications.

E. Detect the transportation modes between stops

Given a time window, functions $stdev()$, $maxV()$, and $avgV()$ represents the standard deviation of the magnitude of force, the maximum moving speed, and the average moving speed within the time window, respectively. We set the time window to cover exactly the trip between the two successive stops. Our mode detection algorithm uses decision-tree rules as below. We will work on the classification between bus, taxi and running using the stop rate per unit distance.

```

if (stdev() ≥ MOTO_ACCE)
  it is not in a motorized mode;
else if (maxV() ≥ TRAIN_MIN_SPEED)
  it is MRT;
else if ((maxV() ≥ MOTO_MIN_SPEED) &&
  (avgV() ≥ MOTO_AVG_SPEED))
  it is a moving bus or taxi.

```

F. Stop filtering based on the context of transportation mode

Given three successive stops S_0, S_1, S_2 , we compare the transportation modes between each two successive stops, and use the following algorithm to remove the stops that are not of interest. Let $m(S_i, S_j)(j = i + 1)$ indicate the transportation mode during the trip between stops S_i and S_j .



Fig. 1. Screenshot of a transportation activity report. The stops are marked on the map.

```

if ((m(S_0, S_1) == m(S_1, S_2) &&
  both modes are motorized ones &&
  stdev(S_1.starttime, S_1.endtime) ≤ MOTO_ACCE) {
  the vehicle may be waiting for traffic light or for
  passengers to alight at S_1. Filter out S_1 if it is not of interest.
}

```

After completing the above steps, we get the stop positions, the duration and the transportation mode between stops. This information is displayed on a map as shown in Fig 1.

III. DEMONSTRATION HIGHLIGHTS

The demo setup will include two laptops and several Android phones and iPhones. We will demonstrate the data collection applications on the phones, the process of automatic data analysis on a laptop, and the validation of analysis results via the web portal on the other laptop connected to the Internet. The analysis results can be displayed and edited on the web portal as shown in Fig 1. We will show some traces that we have collected via pilot trials during the previous months.

ACKNOWLEDGMENT

This research is supported by the Singapore National Research Foundation (NRF) through the Singapore-MIT Alliance for Research and Technology (SMART) Future Urban Mobility program.

REFERENCES

- [1] D. H. Kim, J. Hightower, R. Govindan, and D. Estrin, "Discovering semantically meaningful places from pervasive RF-beacons," *Proc. of the 11th International Conference on Ubiquitous Computing (Ubicomp 2009)*, pp. 21–30, 2009.
- [2] D. H. Kim, Y. Kim, D. Estrin, and M. B. Srivastava, "Sensloc: Sensing everyday places and paths using less energy," *Proc. of the 8th ACM Conference on Embedded Networked Sensor Systems (SenSys 2010)*, pp. 43–56, 2010.
- [3] Y. Zheng, Y. Chen, Q. Li, X. Xie, and W.-Y. Ma, "Understanding transportation modes based on GPS data for web applications," *ACM Transactions on the Web*, vol. 4, no. 1, pp. 1–36, Jan 2010.
- [4] E. Miluzzo, N. D. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S. B. Eisenman, X. Zheng, and A. T. Campbell, "Sensing meets mobile social networks: The design, implementation and evaluation of the Cenceme application," *Proc. of the 6th ACM Conference on Embedded Networked Sensor Systems (SenSys 2008)*, pp. 337–350, 2008.
- [5] S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, and M. Srivastava, "Using mobile phones to determine transportation modes," *ACM Transactions on Sensor Networks*, vol. 6, pp. 1–27, 2010.