Sensing Urban Density using Mobile Phone GPS Locations: A case study of Odaiba area, Japan

Teerayut Horanont[1,2,*], Santi Phithakkitnukoon[3,*], Ryosuke Shibasaki[1]

[1]Department of Civil Engineering, The University of Tokyo, Tokyo, Japan
[2]Sirindhorn International Institute of Technology, Pathum Thani, Thailand
[3]Department of Computer Engineering, Chiang Mai University, Chiang Mai, Thailand
teerayut@siit.tu.ac.th, santi@eng.cmu.ac.th, shiba@csis.u-tokyo.ac.jp

Abstract. Today, the urban computing scenario is emerging as a concept where humans can be used as a component to probe city dynamics. The urban activities can be described by the close integration of ICT devices and humans. In the quest for creating sustainable livable cities, the deep understanding of urban mobility and space syntax is of crucial importance. This research aims to explore and demonstrate the vast potential of using large-scale mobile-phone GPS data for analysis of human activity and urban connectivity. A new type of mobile sensing data called "Auto-GPS" has been anonymously collected from 1.5 million people for a period of over one year in Japan. The analysis delivers some insights on interim evolution of population density, urban connectivity and commuting choice. The results enable urban planners to better understand the urban organism with more complete inclusion of urban activities and their evolution through space and time.

Keywords: GPS, mobile sensing, urban density, mobile phone locations, pervasive computing, urban computing.

1 Introduction

New technology can help cities manage guarantee and deliver a sustainable future. In the past few years, it has become possible to explicitly represent and account for time-space evolution of the entire city organism. Information and communication technology (ICT) has the unique capability of being able to capture the ever-increasing amounts of information generated in the world around us, especially the longitudinal information that enables us to investigate patterns of human mobility over time. Thus, the use of real-time information to manage and operate the city is no longer just an interesting experience but a viable alternative for future urban development.

In this research, the analysis of mobile phone location, namely "Auto-GPS", has been used to serve as frameworks for the variety of measures of effective city planning. More specifically, we explore the use of location information from Auto-GPS to characterize human mobility in two major aspects. First is the commuting statistics and second is the city activity, how the change of activities in part of urban space can be detected over times.

In general, a classic travel survey is frequently used to acquire urban connectivity and trip statistics. However, they truly lack of long-term observation and sample size is always the main limitation due to the highly cost and extra processing time. In this paper, we propose a novel approach that takes advantage of anonymous long-term and preciously collected spatial-temporal location generated by Auto-GPS function from ordinary mobile phone users. As of the best of our knowledge, this is the first time that large-scale GPS traces from the mobile phone have been observed and analyzed countrywide for travel behavior research.

To have evidence showing clearly how this would help planning and decision making, we selected one of the major active area in central Tokyo called Odaiba as our study area. Odaiba is a large artificial island in Tokyo Bay, Japan. It was initially built for defensive purposes in the 1850s, dramatically expanded during the late 20th century as a seaport district, and has developed since the 1990s as a major commercial, residential and leisure area. Odaiba is suitable for this analysis since it is isolated from other parts of Tokyo. It provides all urban amenities like a small city including hotels, department stores, parks, museums, office buildings and residential areas.

The rest of the paper is organized as follows: Section 2 outlines related work; Section 3 describes the datasets and the basics of Auto-GPS; Section 4 covers methodology; Section 5 explains the results from our analysis; and Section 6 provides conclusion.

2 Related Work

Location traces from mobile devices have been increasingly used to study human mobility, which is important for urban planning and traffic engineering. Several aspects of human mobility have been exploited. Human trajectories show a high degree of temporal and spatial regularity with a significant likelihood of returning to a few highly visited locations [1]. Despite the differences in travel patterns, there is a strong regularity in our mobility on a regular basis, which makes 93% of our whereabouts predictable [2]. Understanding mobility patterns would yield insights into a variety of important social issues, such as the environmental impact of daily commutes [3].

These recent studies have emphasized on modeling, prediction, and inter-urban analysis of human mobility, but not on the richer context of it such as the engaged activity in the location visited. There are many studies that use GPS records to identify trip trajectories. Most of these works begin with the segmentation of GPS logs into individual trips, usually when there is a significant drop in speed [4][5], or when GPS logs remain in one area for a certain amount of time [6][7].

With the advance of today's ICT technologies, it is possible to realize a sort of socio-technical super-organism to support high levels of collective "urban" intelligence and various forms of collective actions [8]. It therefore becomes our interest in this work, by building on our previous research [9,10], to investigate on how to use large-scale, long-term GPS data from mobile phones to extract valuable urban statistics and to project the real world information.

3 Datasets

There were two datasets used in this study. The main dataset was collected from approximately 1.5 million mobile phone users through a certain mobile service provided by a leading mobile phone operator in Japan. Under this service, handsets provide a regular stream of highly accurate location data, and thereby enable support services that are closely linked with the user's behavior. Technically, an Auto-GPS-enabled handset position is measured within five minutes and sent through a network of registered services (Fig. 1). The data was recorded from August 2010 to October 2011. In order to preserve user privacy, Auto-GPS data was provided in a completely anonymous form to ensure privacy of personal information. It is important to acknowledge that there was some selection bias in this dataset, as participants were limited to users of a specific mobile phone service. The distribution of user type was estimated from 50,000 online surveys and about 2.6 percent (1,356 respondents) replied to use this service.

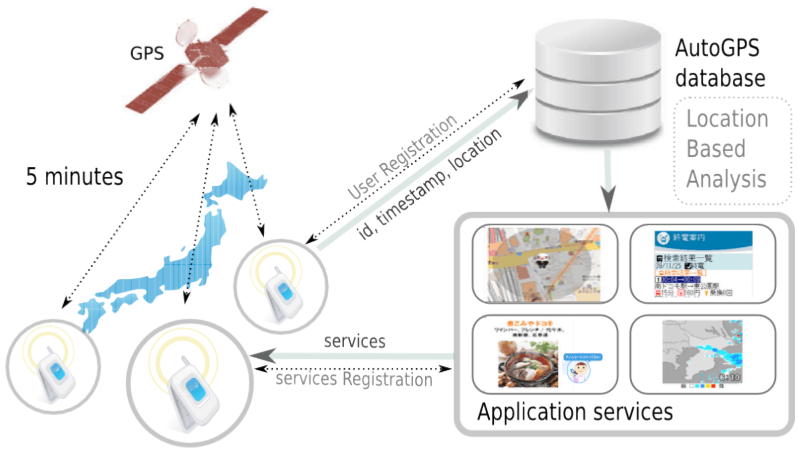

Fig. 1. Flow diagram illustrating the Auto-GPS services supported by Japanese handsets since 2009

Figure 2 shows a graph of the average number of GPS points per day in this dataset. A small sample of the raw data is shown in Fig 3. Each record in the dataset has six-tuple information that includes: User ID, Timestamp, Latitude, Longitude, Error rate, and Altitude. The approximated error rate was identified into three levels: 100 meters, 200 meters, and 300 meters, based on the strength of GPS signal available to the handset.

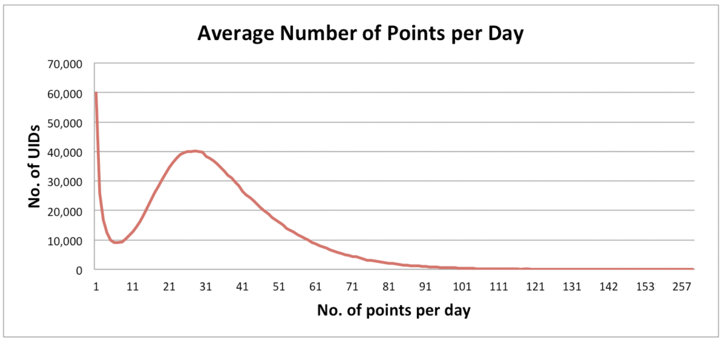

Fig. 2. The average number of GPS points per day is 37, indicating that the users spent approximately three hours traveling each day.

Dummy-ID	Time	Latitude	Longitude	Error	Altitude
00862690	2010-08-01 12:01:09	34.69888	135.534146	1	64.00
00862754	2010-08-01 21:10:13	39.703028	141.146445	2	176.94
00886354	2010-08-01 12:48:23	34.33872	135.600167	3	165.73
00862690	2010-08-01 14:46:09	34.709877	135.591781	1	64.00
00169966	2010-08-01 18:19:52	35.534478	140.304336	3	39.64
00169966	2010-08-01 18:24:52	35.527892	140.312319	3	17.83

Fig. 3. A sample of Auto-GPS data that includes an anonymous dummy-id, timestamp, geo-location, error level, and altitude. The error level indicates the strength of the GPS signal available to the handset.

The second dataset is the census data, which was used for validation purpose. The census data was released in 2008, which represents census information for a grid size of one square kilometers. This data was provided by the National-Land Information Office [11].

4 Methodology

Our Auto-GPS data was considered as big data, having a total of 9.2 billion GPS records. Our data was handled and pre-processed using Hadoop./Hive on a computer cluster of six slave nodes. We first processed our Auto-GPS data to estimate the density of people who reside within specific areas (i.e., area population density), more specifically, finding their home locations. First, we extracted the locations of daily rest, or "stay points," for each individual trajectory. Let P represent a set of sequential traces of the user such that $P = \{p(1), p(2), p(3),\ldots, p(i),\ldots\}$ where $p(i)$ is the i^{th} location of the user and $p(i) = \{\text{id, time, lat, lon}\}$. A stay point is defined as a series of locations in which the user remains in a certain area for a significant period of time, where distance in space and difference in time between observed points are applied as constrained multi-criteria in the detecting method i.e., Distance(p_{start}, p_{end}) $< D_{threh}$ and TimeDiff (p_{start}, p_{end}) $> T_{threh}$,, where D_{threh} and T_{threh} are adjustable parameters; D_{threh} is the maximum coverage threshold of movement in which an area is considered as a stay point, and T_{threh} is the required minimum amount of time that the user spends in a stay point.

We recruited 15 subjects to carry a smartphone for one month with an application that allowed the subjects to identify stops that they made each day. With this ground truth information, we found that the spatial and temporal criteria [12] to identify stay points most accurately were 196 meters and 14 minutes, as shown in our experimental results in Figs. 4 and 5.

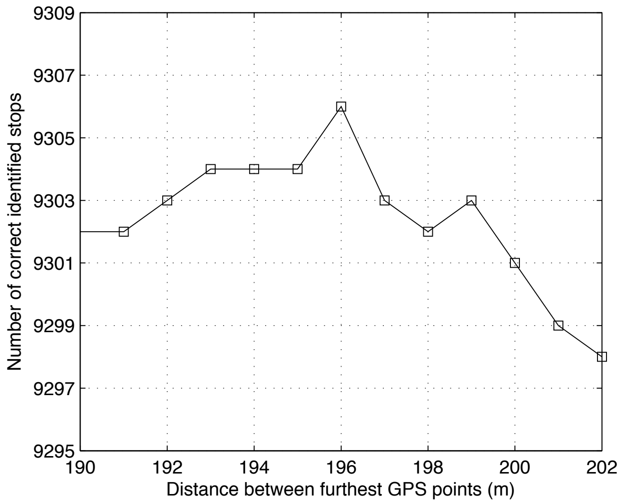

Fig. 4. Stop detection accuracies for different distance threshold values.

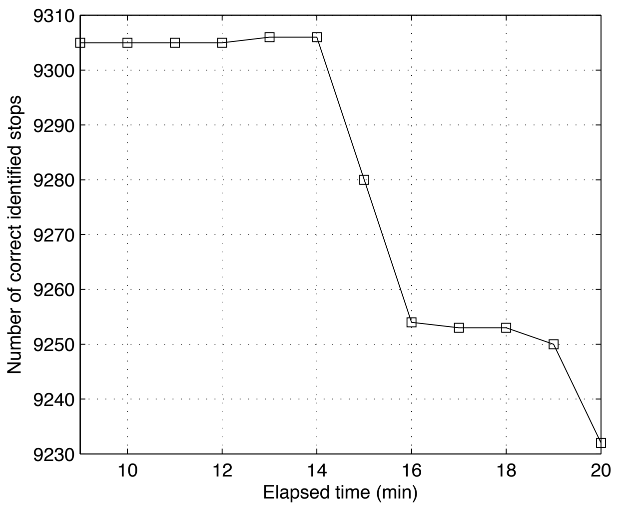

Fig. 5. Stop detection accuracies for different time threshold values.

Based on the detected stay points, we estimated home location of each subject as the location with the highest number of stay points between midnight and 6 a.m. This yielded a fairly accurate estimation of home locations, which is comparable (R^2 = 0.79) to the population density information of the Census Data provided by the Statistics Bureau, Ministry of Internal Affairs and Communications, as shown in Fig. 6.

According to the result in Fig. 6, we considered our "stay points" to be reliable for our further analysis. The stay points and home locations were used as inputs for calculation of various urban indicators and statistics across different spatial and temporal levels in the next section.

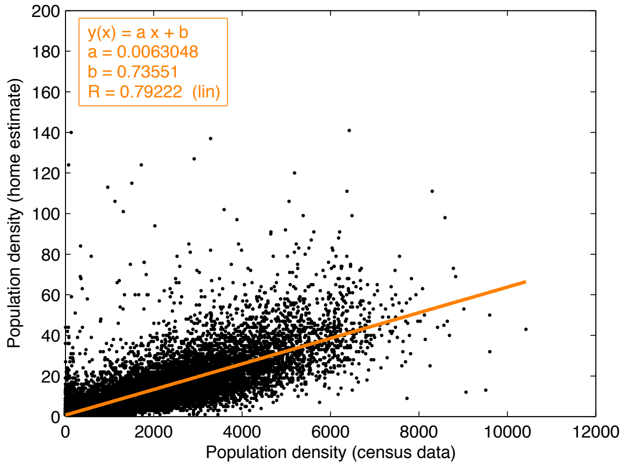

Fig. 6. A comparison of the estimated home locations from the Auto-GPS data against the Census Data of one-square kilometer grids.

5 Results

Finding urban descriptive knowledge of the people who use urban spaces is one of the most important information for urban planners. Our first result attempts to explain the origin of people flow. We constructed multiple criteria to define visitors in an area. We used the minimum stay of 30 minutes and excluded people who have home and work location in the area. (Note that work location was derived in a similar way we did for the home location.) The maximum annual visit is set to eight times as it is the third quartile of the entire dataset (Fig. 8). The annual total of visitors to Odaiba area was estimated at 80,463 people from 1.5 million total samples or 5.36% of the population. Figure 7 shows the choropleth map of estimated yearly visitors. As expected, the nearer the prefecture is to the Odaiba area, the more visitors are coming from. There are some exception for the big city such as Nagoya, Osaka, Fukuoka, and Hokkaido where air transport services are operated frequently.

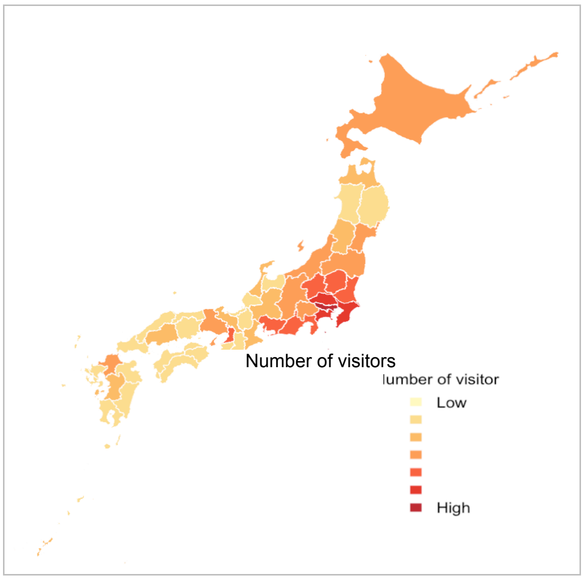

Fig. 7. Estimated annual visitors to Odaiba area by prefecture.

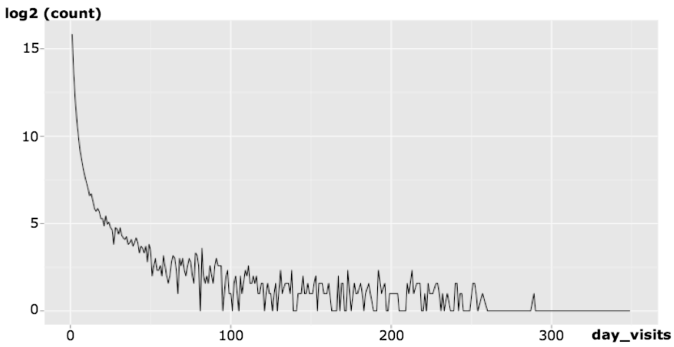

Fig. 8. Count of the number of daily visits to Odaiba area.

Figure 9 shows the number of daily visitors to Odaiba area from which it appears that the area is more popular during summer. This is because of the big event arranged by the Fuji Television Network whose station is based in Odaiba. The highest number of visits to Odaiba was on the 14^{th} of August when the Tokyo Bay Grand Fireworks Festival was held. The second most visits were on the Christmas Eve, as the area is known as a dating place. We notice a significant distinct drop in the number of visitors dramatically on 11th March. It was the day when an earthquake of magnitude 9.0 hit Japan in 2011, followed with the "radiation leakage" of two nuclear power plants in Fukushima prefecture. We observed 3 weeks of an abnormal reduction in the number of visits of the area before it returned to normal.

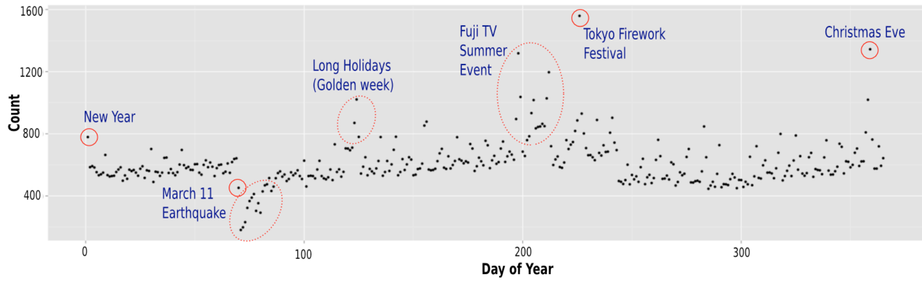

Fig. 9. Detected daily visits to Odaiba area. The magnitude of anomalies can vary greatly between events, and this could lead to composite dominated by a few major events.

Next, we visualized how different activities between weekdays and weekend are by overlaying weekdays stay points over the weekends' (Fig. 10). Surprisingly, there are several clusters that highly dominate over each others in particular locations. By incorporating prior knowledge of the area and collection of news, it reveals a clear evidence of how the patterns are created. The locations marked with "a" are complex buildings where shopping malls, hotels, and restaurants are situated. This yields a similar distribution of both weekdays and weekends. The "b" marker is for an open space where we can observe that half of the areas are more active during the weekends than weekdays. This is because of the special events are usually held only on weekends. The areas in upper part are served as outdoor parking spaces and are the main area of Fuji TV summer events. This event is usually held for three months in the summer both weekends and weekdays. The "c" areas are event spaces that are mainly occupied during weekends. The "d" areas are office buildings that are more active on the weekdays. Please note that the "d_1" area is the construction area during our data collection period.

In addition, Fig. 11 provides the visitor count information in each building for the entire year. The height of the building corresponds to the number of visitors. It is clear that shopping malls and restaurant complex type buildings are the most popular destinations in the Odaiba area. Both of them have approximately 10 times more visitors than the office areas, which is relatively intuitive.

6 Conclusion

This preliminary research explores the potential of using mobile phone GPS locations in a new context and broader advances towards the understanding of today's excessive mobility. The finding of this remarkable dataset is to capture the urban evolution from the real movement of people. The results display the findings of a comprehensive and creative process of the use of Auto-GPS data. Finally, the importance of this research ultimately lies on how it can be practically applied and utilized for future sustainable urban developments.

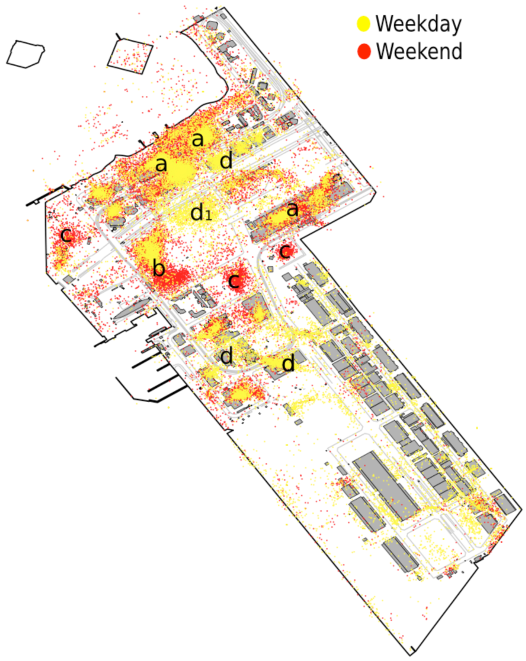

Fig. 10. Comparison of estimated visitor density between weekdays (yellow) and weekends (red).

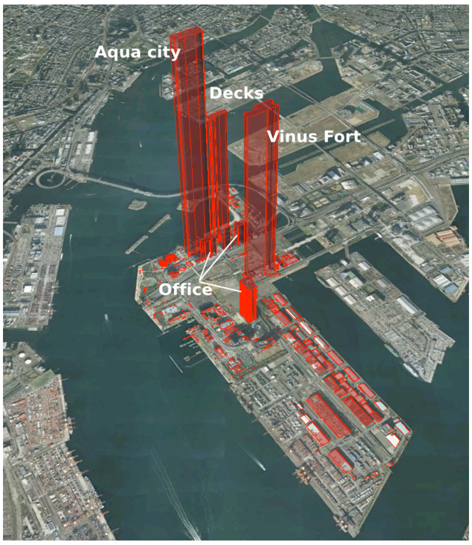

Fig. 11. Detected annual visits of each building in the Odaiba area. The height of building represents the number of visitors.

Acknowledgements. This research carried out under the support of DIAS (Data Integration and Analysis System) by MEXT (Ministry of Education, Culture, Sports, Science and Technology). We would like to thanks ZENRIN Data Com (ZDC) CO., LTD. for archiving the dataset used in this research.

References

1. Gonza ́lez, M. C., Hidalgo, C. A., Baraba ́si, A.-L.: Understanding individual human mobility patterns. Nature, vol. 458, pp. 238–238 (2009)
2. Song, C., Qu, Z., Blumm, N., Baraba ́si, A.-L.: Limits of predictability in human mobility. Science, vol. 327, no. 5968, pp. 1018–1021 (2010)
3. Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Rowland, J., Varshavsky, A.: A tale of two cities. In: Proceedings of the Eleventh Workshop on Mobile Computing Systems & Applications (2010)
4. de Jong R., Mensonides, W.: Wearable GPS device as a data collection method for travel research. In: ITS-WP-03-02, Institute of Transport Studies, University of Sydney (2003)
5. Li, Z., Wang, J., Han, J.: Mining event periodicity from incomplete observations. In: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (2012)
6. Ashbrook, D., Starner, T.: Using GPS to learn significant locations and predict movement across multiple users. Personal and Ubiquitous Computing, vol. 7, pp. 275–286 (2003)
7. Gong, H. Chen, C. Bialostozky, E., Lawson, C.: A GPS/GIS method for travel mode detection in New York City. Computers, Environment and Urban Systems (2011)
8. Fontana, D., Zambonelli, F.: Towards an Infrastructure for Urban Superorganisms: Challenges and Architecture. In: IEEE International Conference on Cyber, Physical and Social Computing (2012)
9. Horanont, T., Witayangkurn, A., Sekimoto, Y., Shibasaki, R.: Large-Scale Auto-GPS Analysis for Discerning Behavior Change during Crisis. IEEE, Intelligent Systems, vol.28, no.4, pp.26,34, July-Aug. (2013)
10. Horanont, T., Phithakkitnukoon, S., Leong, T.W., Sekimoto, Y., Shibasaki, R.: Weather Effects on the Patterns of People's Everyday Activities: A Study Using GPS Traces of Mobile Phone Users. PLoS ONE, Vol. 8, No. 12, e81153 (2013)
11. National-Land Information Office. Retrieved January 27, 2013, from http://www.mlit.go.jp/kokudoseisaku/gis/index.html
12. Zheng, Y., Li, Q., Chen, Y., Xie, X., Ma, W.-Y.: Understanding mobility based on GPS data. In: Proceedings of the 10th ACM International Conference on Ubiquitous Computing (UbiComp'08), New York, NY, USA, pp. 312–321 (2008)