

## Is "Smart Census" Possible? An Exploratory Study

Eunyoung Shim<sup>1</sup>, Jiyun Tark<sup>1</sup> and Youngtae Cho<sup>1\*</sup>

<sup>1</sup> Center for Smart Technology and Public Health, Seoul National University

\* Corresponding author (youngtae@snu.ac.kr)

## Introduction

The increase of smartphone penetration rate in many countries is amazing. To take South Korea as an example, the penetration rate has reached close to 73% in 2013(Statista, 2013). Although overall smartphone penetration rate in the world is still about 16%(**need update**), it is not difficult to see quick increase of the figure in near future, since smartphones are expected to rapidly spread in Africa, Southeast Asia, India, and China (Strategy Analytics 2013). Another striking feature regarding a smartphone is the extremely high use time. For instance, a survey, based on 38,000 Korean adult respondents, on smartphone use conducted in March 2013 showed that Koreans use smartphones for 1 hour and 43 minutes per day on average (Korean Media Panel Research, 2012). Ericsson Consumer Lab Research Platform (Ericsson 2011) reported that over 34% of Americans bring their smartphone to bed. Furthermore, most users keep their smartphone very close to their body (i.e., in a pocket or a purse or on the table nearby) while they do not use it, even though we do not have statistic for it. It is not even an exaggeration to say that a smartphone is the most physically intimate item to its user.

Smartphones are equipped with various types of sensors. Although variations exist according to the maker and model, basically equipped sensors include GPS (global positioning system), WPS (wifi-based positioning system), light, gyro, accelerometer, sound, and/or magnetic sensors. Since a number of basic smartphone functions and phone applications utilize geographic information, the GPS and WPS sensors are equipped in all smartphones. Accuracy of smartphone's GPS and WPS sensor information has well been documented already (Bierlaire, Chen, and Newman 2012; Herrera et al. 2010).

Those three facts that (1) most people have smartphones, (2) most users carry and use smartphones every second, and (3) all smartphones are equipped with geographic sensors may imply a very important demographic meaning. Demography is basically interested in the

usual location of people (e.g, residence and work/school), and the most basic but important source of demographic information is census which counts the number of people who actually reside, not simply registered, in an administrative area. If geographic sensor data monitored by a smartphone are utilized to indicate user's residence and/or work locations, it may be comparable to the location information surveyed by census. Hinted at by such an idea, the current study aims (1) to examine if one's usual locations (residence and work/school) can be detected by the GPS or WPS data monitored and compiled through one's smartphone, and (2) to evaluate the results to see how accurately and effectively smartphone can function as a census taker. Here, we would like to define the method introduced in this study a *Smart Census*, since geographic data monitored by a smartphone are used and the usual locations are automatically estimated by the smartphone without requiring user's special attention or data input.

There are four very important conditions that we have seriously kept in mind when developing the *Smart Census*. First, *Smart Census* needs to detect usual places where users spend quite a large amount of time in their daily lives, since basically census is to enumerate the *de jure* population. Second, space, rather than a geographic spot, should be detected by *Smart Census*. Address itself is a geographic spot on the grid of longitude and latitude. However, the place represented by the address in fact is a space where people reside or work. Once space is detected, we can easily designate a geographic spot to represent the space. Third, to take full advantage of the smart features of a smartphone, monitoring of geographic information should be conducted in a non-intrusive way that does not interrupt user's daily life. That is, a smartphone user needs to neither alter his/her daily life nor technically operate anything with the smartphone, except for installing and opening the *Smart Census* application, due to the participation in the *Smart Census*. Lastly, technical procedures to monitor, store, archive, calculate, and analyze the geographic sensor data should be as simple as possible so

that all these procedures to be operated by the user's smartphone, and the only final results (i.e., locations of mostly residing places in daily life) are submitted. Since the geographic sensor data are a very important piece of personal information that should be protected, it is desirable to limit the information submitted as final as possible.

## **Methods**

### *Data Collection*

We collected geographic sensor monitored data (GPS and WPS) from Seoul residents aged 20 to 49 who use smartphones daily. We limited age groups from the 20s to the 40s for the sample selection, since the smartphone penetration is the highest in these age groups. Sample selection was carried out by the Gallup Korea, the largest social survey company in Korea. The Gallup Korea maintains their sample pools who are representative of Seoul residents. Among them, those who used Samsung Galaxy S3, Galaxy Note 1, Galaxy Note 2 models were selected as the study participants to minimize the possible errors caused by technical differences across different smartphone brands. Initially 360 samples were collected, but finally 336 participated in the study. Even though those participants were representative samples, we did not give pay much attention to it since the purpose of our exploratory study did not include population estimation from the sample.

Samples were requested to download and install an application called "Smart Census" to their smartphones. The "Smart Census" application was developed by authors only for the use of the current study. When the application was installed and first opened, users were informed about this study and the details of data collection procedures. Participants could choose to proceed or leave. When they chose to proceed, they were asked to provide basic demographic information including the addresses of residence and work/school. They were also asked if the residing address was the same as their registered address. Then the

application ran automatically to gather GPS and WPS information once in every five minutes for seven days. Therefore a total of 2,016 times (7days\*24hours\*12times) of GPS and WPS sensor data could be collected, although there were variations across samples due to uncontrollable mechanical or technical problems. We designed the "Smart Census" application to store all collected GPS and WPS data in the participants' own smartphones. When 7 days of data collection was completed, the application automatically showed a pop-up message to the participants informing the end of participating in Smart Census and asked if they agreed to send the GPS and WPS sensor data collected by and stored in their smartphones to the study server which was secured and protected by the Statistics Korea. Thus, informed consent to take part in this study were taken twice from the study participants, just after the installing the application and right before the submission of collected data, and the entire study design and procedure was reviewed for its ethical issues by the IRB committee of the School of Public Health, Seoul National University.

Participants could join the *Smart Census* any time during March 11-17, 2013. Data submission from the last participant was done on March 22, 2013. Since the "Smart Census" application was developed only for those participants, we sent a text message with a link that they can download the application rather, than to have them use general application stores for Android phones. When GPS and WPS data were submitted from all participants to the secured server, Statistics Korea reviewed and cleaned them first to remove any types or kinds of information that could violate the Personal Information Protection Act of Korean Law.

### *Analysis*

As noted above, what we would like to know utilizing smartphone geographic sensor data is space, rather than a spot on the grid of latitude and longitude, that a user spends quite a large amount of time in his/her daily life. To figure out the most densely residing space, we

employed the kernel density estimation (KDE) technique. The KDE is a conventionally used spatial analysis method (Anderson 2009). The definition of KDE function is quite simple, and the density estimation can be easily conducted by the general statistical packages (e.g., SAS or STATA) or spatial analysis packages (e.g., Arc GIS, R-spatial analysis, Q GIS). The result of spatial analysis using KDE can be variable by the selection of bandwidth, and there have been a number of studies on the optimum selection of bandwidth (e.g., Jones, Marron and Sheather 1996; Kile 2010). We designated one kilometer as the bandwidth value in this study which is the default value of Arc GIS (ver. 9.3), since we were interested only in one or two most densely residing space (i.e., home and work/school) and the selection of bandwidth value did not affect the finding of this space. Further, due to the exploratory nature of this study, we tried to design a method as simple as possible.

As aforementioned, there were two types of geographic information monitored by smartphones: GPS and WPS. GPS information is monitored when the smartphone catches the satellite signal, which means GPS data can be missing when the smartphone is located where the satellite signal is weak (e.g., inside building). WPS information is monitored when the smartphone catches the wifi signals. Since Seoul, the study site, is second to none regarding its internet and mobile connectivity, wifi signals are almost ubiquitous, although it is possible that wifi signal is weak or zero in some places like inside elevators or basement of new building where wifi transmitter is not equipped yet. For the spatial analysis, we used GPS information primarily and WPS information as a backup.

To find the most densely residing places and to increase accuracy, we designed a two-stage KDE method here. In the first stage, KDE was used to detect the locations of the high density place based on the entire record of GPS or WPS information. In a graphic presentation of density contour map, the most densely places looked like a mountain or a peak. The top point of the mountain is the hotspot of the density contour. In the second stage,

we chose GPS or WPS data only within one standard deviation, in terms of density, from the hotspot of each mountain, and adapted KD technique again to increase the estimation accuracy by deleting outliers.

Once the most densely residing places are estimated, we checked the probability density function (PDF) of the contour line that included addresses of the residence and work/school that study participants provided. Using the distances between the responded address and the actually recorded GPS or WPS positions, we calculated the mean distance and standard deviation for each study participant. By doing this, we were able to decide which levels of PDF were acceptable for residence and/or work/school, respectively on what accuracy level.

## **Results**

The following figure shows the estimation results for residence places. Among 336 participants, effective cases for the current study were 287. Reasons for omission are described in the figure. The green shaded part presents cumulative statistics by corresponding probability density functions. When PDF was set for 95%, the residence addresses of almost 87% of the effective cases were included within the corresponding contour line, and the mean of the distances between the responded address and each of GPS or WPS spots was 34.5 meters. When 90% of PDF was considered, the corresponding figures chances to 94% and 40.8 meters, respectively.

## ► Residence

	N	%
Total	336	100
Effective	287	85.4
Non-Effective	49	14.6
Real residence different from addressed residence	31	63.3
No address recorded	6	12.2
Residence impossible to distinguish from work/school	10	20.4
Not enough GPS/WPS information	2	4.1

PDF	99%	95%	90%	85%	Other	All
N	158	92	19	5	13	287
%	55.1	32.1	6.6	1.7	4.5	
Mean (m)	14.9	68.1	124.2	157.2	153.6	48.0
std	9.5	42.8	81.1	90.1	114.5	59.5
Cum N		250.0	269.0	274.0	287.0	
Cum %		87.1	93.7	95.5	100.0	
Cum Mean		34.5	40.8	43.0	48.0	
Cum std		37.2	47.5	50.8	59.5	

## Conclusion

One of the major differences between conventional feature phones and smartphones is that smartphone is able to collect, calculate and operate data, just as a personal computer does. As its penetration rate dramatically increase worldwide, and as it becomes indispensable to most users throughout their daily lives, the capability of smartphones to detect the life context of users should be paid more attention by demographers. It is because what smartphone can produce may be comparable to census or social surveys that attempt to learn the various aspects of individual's life context. The current study showed a one possibility that geographic raw data monitored by smartphones can be converted to a value-added piece of information that can detect the location of user's home and work/school.

Of course, we do not simply argue that 'Smart Census' will replace current types of census (i.e., mail-in census, census by census takers, or computer-based census). However, as



census-related environment is getting worse in most countries, ‘Smart Census’ can be considered a possible supplementary solution in near future. For instance, Korea will change its census method in 2015 from the conventional household-visit census to registration-based one. Since Korea’s registration statistics have been managed well and maintain high quality, we are sure that the new method will save a large amount of budget and time, while producing a descent quality census result. But as the number and share of single-dwelling household increase dramatically in recent years, there are so many Koreans whose registered address and actual residence are not congruent. In fact, our sample showed that almost 30% of the 20s, 20% of the 30s, and about 10% of the 40s did not have the consistent addresses of actual residential place and registered place. As we all know, census should be based on *de jure* population. We believe that the current exploratory study on ‘Smart Census’ can suggest at least a direction to meet the new changes and challenges on the census-related environment.

## References

- Bierlaire, M. J. Chen, J. Newman. (2013). "A Probabilistic Map Matching Method for Smartphone GPS data." *Transportation Research Part C: Emerging Technologies* 26: 78-98.
- Ericsson (2011) "From APPS to Everyday Situations." Ericsson Consumer Insight Summary.
- Herrera, J. C., D. B. Work, R. Herring, X. Ban, Q. Jacobson, and A. M. Bayen. (2010) "Evaluation of Traffic Data Obtained via GPS-enabled Mobile Phones: The Mobile Century Field Experiment." *Transportation Research Part C: Emerging Technologies* 18: 568-583.
- Korea Information Society Development Institute (2012) “Korean Media Panel Research”
- Strategy Analytics (2013) "Smartphone Sales Surge in North America and Asia but Western Europe Fights Growing Trend." Press release.  
(<http://www.strategyanalytics.com/default.aspx?mod=pressreleaseviewer&a0=5376>)