Understanding the Coverage and Scalability of Place-centric CrowdSensing

Yohan Chon[†], **Nicholas D. Lane**[‡], **Yunjong Kim**[†], **Feng Zhao**[‡], **Hojung Cha**[†]
[†]Yonsei University, [‡]Microsoft Research Asia

ABSTRACT
Crowd-enabled *place-centric* systems gather and reason over large mobile sensor datasets and target everyday user locations (such as stores, workplaces, and restaurants). Such systems are transforming various consumer services (for example, local search) and data-driven organizations (city planning). As the demand for these systems increases, our understanding of how to design and deploy successful crowdsensing systems must improve. In this paper, we present a systematic study of the coverage and scaling properties of place-centric crowdsensing. During a two-month deployment, we collected smartphone sensor data from 85 participants using a representative crowdsensing system that captures $\approx$ 48,000 different place visits. Our analysis of this dataset examines issues of core interest to place-centric crowdsensing, including place-temporal coverage, the relationship between the user population and coverage, privacy concerns, and the characterization of the collected data. Collectively, our findings provide valuable insights to guide the building of future place-centric crowdsensing systems and applications.

Author Keywords
Mobile Crowdsourcing, Smartphone Sensing

ACM Classification Keywords
H.4.m Information Systems Application: Miscellaneous.

INTRODUCTION
The proliferation of sensor-enabled smartphones makes it increasingly feasible to build crowdsensing systems that gather large-scale mobile sensor data. An important and growing class of these crowd-enabled systems are *place-centric* – that is, they are designed to provide place-related information and focus on locations that participants routinely visit (for example, cafes, supermarkets, offices, homes, or schools) as targets for data collection and analysis. Numerous research prototypes have been demonstrated, enabling, for example, improved local search and recommendation services [24, 11, 7], or even the estimation of the number of customers in a coffee shop [20, 25]. In addition, early stage commercial examples of place-centric crowdsensing are already emerging through companies such as Gigwalk [4] and FieldAgent [1]. These systems recruit users for tasks such as gathering photos of store interiors (to assess product displays or stock levels) or real estate. To maintain this rapid progress, we must continue to deepen our understanding of place-centric crowdsensing systems in all areas. On the surface, these systems can appear deceptively simple: contributors collect data which is then analyzed for a variety of purposes. However, to design and deploy successful systems, a diverse set of technical and social challenges must be met simultaneously.

In this paper, we report on the systematic study of a large-scale deployment using a representative place-centric crowdsensing system built with commodity smartphones. Our deployment is one of the largest examples of mobile crowdsensing yet studied. We recruited 85 people to collect data for two months while making $\approx$ 48,000 place visits in Seoul, Korea. Each day, contributor phones periodically collected a range of standard smartphone sensors (viz. WiFi, GPS, camera, or microphone); the resulting dataset includes $\approx$ 22,000 audio clips and $\approx$ 6,200 photos among other data types. Our aim is to investigate fundamental issues related to place-centric crowdsensing's coverage and scalability. The issues we survey include the coverage achieved with respect to places, place categories, and temporal patterns; the relationship between the number of participants and system coverage; user privacy concerns and how they affect coverage; and, finally, a detailed characterization of the data collected as a means to understand future potential place-centric applications[1]

Based on our deployment experiences and analysis of the collected dataset, we report findings that are relevant to builders of crowdsensing systems. These findings include: (1) Even with a small number of contributors (85), we find that crowdsensing can provide relatively high coverage levels for select place categories or groups (for example, 15% of places in the top 1% of popularity), especially given the city's large size (nearly 100,000 places in total). Our results identify the specific place categories with high coverage and quantify how often these places are visited; such information is important in understanding the viability of monitoring desired place combinations. (2) We find empirical distributions that model important factors underpinning the relationship between the size of a crowdsensing user population and place coverage. For instance, we find that in our crowdsensing study the number of place visits by participants follows a power-law distribution. Based on these results, we develop a simple but effective generative model of place coverage based on participant population and city characteristics that we can use to estimate how

[1]In the interest of experimental reproducibility the data needed to produce the results reported in this study will be available here: [30].

UbiComp'13, September 8–12, 2013, Zurich, Switzerland.
Copyright © 2013 ACM 978-1-4503-1770-2/13/09...$15.00.
http://dx.doi.org/10.1145/2493432.2493498

many users are required for certain levels and types of coverage. (3) Although our crowdsensing study collects sensitive sensor modalities (including audio and images), and users are clearly cautious, this does not prevent the system from constructing a large-scale dataset during the study. For example, we find that participants allow audio collection in 93% of food-related places and 82% across all place types – a promising result for crowdsensing techniques that require a microphone. (4) Unsurprisingly, we find crowdsensed data to be extremely noisy. However, because of the volume of data collected, we also find that carefully tuned conventional audio- and image-based classifiers (such as sound or object classification) can mine a large number of diverse, place-related contexts – for example, thousands of spoken and written words; overheard instances of music or coughing; or observations of people, cars, and furniture. By detailing our experiences in mining context from this dataset – including how well each classifier works, which classifier-tuning approaches were effective, and the frequency and spread of contexts relative to place categories – our results can assist in identifying new potential place-centric crowdsensing scenarios.

STUDY DESIGN

In the following section, we describe the design of our study and the region in which it was performed.

Overview. Participants receive a smartphone app that collects sensor data (largely audio clips and images). Based on the subject's privacy settings, data is selectively transmitted to a server for storage and analysis. Collectively, the app and server comprise a representative crowdsensing system designed to support the study objectives.

Subjects are told that the system's aim is to gather large-scale data regarding the city that will be used for various urban *sensor-related* applications – we use the umbrella term "smart city" because many participants understand this phrase. Example applications described include building noise and pollution maps, modeling traffic flows, or improving local search with enhanced information about places.

Subjects use their own Android smartphones, onto which we install the smartphone app. This phone must remain their primary phone for the duration of the study. We request that participants carry this phone with them at all times, while also keeping the device charged and active most of the day.

Incentives. We compensate the study population in two ways depending on their performance: (1) *bonus payments* (BP) and (2) a *data competition* (DC). Everyone receives the same baseline payment for their participation, namely, the equivalent of 100 USD. Similarly, within each scheme, members are ranked based on a simple metric that is based on the number of images they capture (with a safety measure not reward images taken all in the same place). The key difference between the two compensation schemes is that under BP participants receive a bonus payment (equivalent of 20 USD) if they are in the top five participants at the end of the study. In comparison, under DC participants receive no additional money but have their performance feedback presented as a game in which they can compete against each other.

Privacy Controls. Importantly, subjects are absolutely free to remove any data they collect without this negatively affecting study compliance. All incentives are driven by data collection, but not necessarily by sharing that data with the server (or researchers). Participants can use any of the numerous privacy controls built-in to the smartphone app that either remove data or prevent similar data from being sampled in the future.

Protection of Human Subjects. We followed the policy of the National Research Foundation of Korea, which carefully examined this study's design prior to providing its approval (equivalent to the IRB in Korea). Collected data is accessible only to those researchers who are part of the project, with all data access audited to allow later review if necessary. Strong anonymization procedures are applied to any study data before it is released publicly.

Deployment Location. We conducted our study in Seoul, Korea. Seoul is the capital and largest metropolis in Korea (605.21km^2), with a population of 10,442,000. The city has good characteristics for our crowdsensing study: (1) 67% of people use smartphones, which is the one of the highest adoption rates in the world. Smartphone owners most actively use their phones for multimedia, location-based services, or commerce activities such as web browsing and online banking [35]. (2) The WiFi network is available almost everywhere in Seoul, even in metro buses and the subway. This characteristic enables high coverage of place recognition using WiFi fingerprints. (3) People are active around the city from early morning to late at night. Public transportation operates for the entire day (including a night bus), and public safety is very high.

Recruitment. To recruit subjects, we distributed advertisements via Facebook, in addition to using posters placed around the Yonsei University campus. Interested parties receive a comprehensive description of the study enabling them to make an informed decision regarding their participation.

Instructions. We requested that participants collect sensing data constantly throughout their daily lives. We informed them that the payments do not depend on the coverage of location sampling, and that incentives are tied into the number of images (which require user intervention to capture), not the number of visited locations. We expect that this instruction does not significantly alter participants' daily routines. However, they are informed that better-quality data will result from them visiting a variety of areas of the city and physically entering a diverse set of places while repeatedly returning to places they prefer. Since image capture is entirely subject-driven, they are instructed to capture any type of entity (building, objects, or people), at any time they wish.

CROWDSENSE@PLACE

In the following section, we describe the design and implementation of the crowdsensing system deployed in our study. We extended our CrowdSense@Place system in [7] by adopting various techniques from the literature, including [24, 6, 8, 22, 18] (when describing component details later, we highlight specific work when appropriate). As a result, the de-

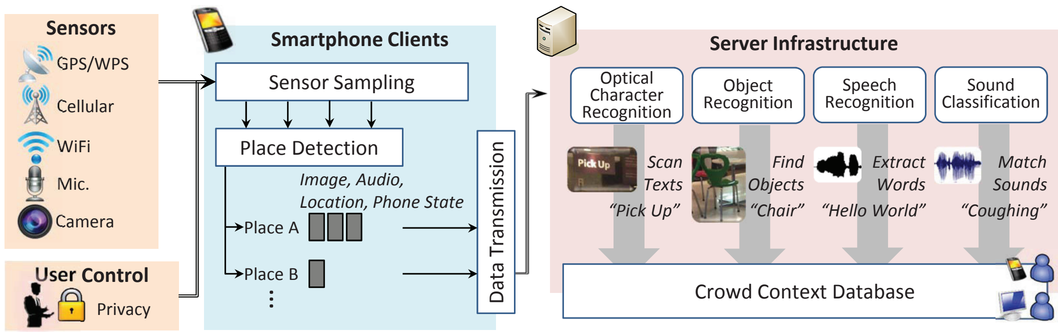


Figure 1. Architecture and data flow of the CrowdSense@Place system used in the study.

ployed system is representative of crowdsensing systems in the literature. As such, we make no claims of novelty for this system; our work's contribution is based on the execution of a large-scale field trial and subsequent analysis of the collected data and experiences gained.

Overview

Figure 1 shows both the architecture and data flow of the crowdsensing system. This system comprises two components: (1) Smartphone App and (2) Server Infrastructure.

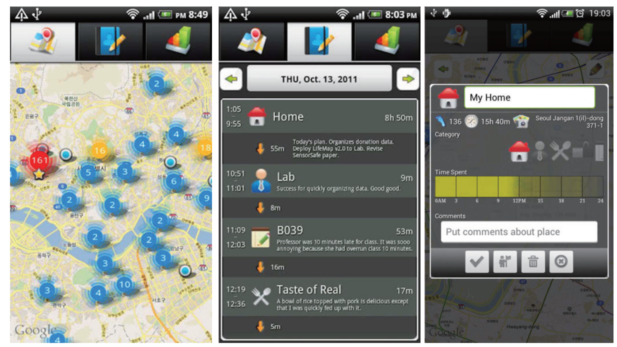
Smartphone App. Smartphone App samples low-level sensor data (viz. image, audio, WiFi, or GPS) that is later transferred to Server Infrastructure. All sensor data is collected opportunistically (that is, via automated sampling), with the exception of image data that is sampled with a participatory approach (user-driven sampling). Study subjects receive an extensive set of privacy controls, such as customizable data-filtering policies (for example, "never sample data at this location") and the ability to manually review and remove data before it is transferred from the phone.

Server Infrastructure. Server Infrastructure collects, stores, and analyzes data received from Smartphone App. All uploaded data is processed by multiple sensor-data classifiers that attempt to extract additional high-level information, such as objects (cars, buildings, and so on) recognized in images or spoken words recognized in audio. The results of classifiers along with the raw data are stored for later analysis and use within crowdsensing applications.

Smartphone App

Figure 2 presents screenshots of Smartphone App that include the following components: (1) sensor sampling; (2) privacy controls; (3) place detection; and (4) data transmission.

Sensor Sampling. Sensor data collection occurs primarily through an Android background service that samples data from all sensors except the camera. Sensor sampling is not periodic but is based on a series of techniques adopted from the literature in an effort to keep Smartphone App's overall energy consumption to $\approx 10\%$ of the phone battery. We adopt the WiFi sampling policy proposed in [8]; this prediction-based scheduling approach is based on the users everyday routine mobility patterns – WiFi scanning occurs when a user is likely to change a location, and re-uses the stored location information to minimize GPS sensing. GPS is sampled



(a) (b) (c)

Figure 2. Screenshots of Smartphone App. (a) Collected data is visualized using a map. (b) Alternatively, data is shown by list view. (c) From either view, users can access data collected from a specific place and edit privacy policies regarding this place.

only when WiFi scans detect strong signs that the user is in motion, such as being in a car (details of this process are provided in the upcoming Place Detection description). Along with these sample strategies, a simple daily budget helps to limit the number of GPS and audio samples and maintain our energy consumption goal.

Image capture is only manually triggered by users through an explicit action within Smartphone App. The audio sampling is basically disabled in all places, but the image capture enables audio sampling on a place-by-place basis. Audio is sampled based on a simple heuristic to improve audio clip quality: the microphone records once users turn off the smartphone screen after using an application or making a phone call at a place where they have taken a picture. In such cases, the user is no longer speaking into the microphone, and the phone is exposed to the environment.

Place Detection. Recognition of logical places (such as a store or a user's home) is performed by WiFi fingerprint matching. This approach is widely used throughout the literature, and we integrate the specific techniques detailed in [8, 18]. Smartphone App utilizes place detection to (1) associate collected data with a place (where applicable), and (2) allow users to annotate a logical place as a particular place category. Based on the sensing schedule generated by the sampling policy (described above), Smartphone App adaptively performs WiFi scans to identify nearby WiFi access points. Whenever a WiFi fingerprint is encountered that is unlike those previ-

Table 1. Summary of classifiers.

Classifiers	Classification Model	Recognized Classes	Sources of Training data
Optical character recognition	Microsoft API [12]	words in images	-
Object recognition	exemplar-SVM [23]	bicycle, sofa, bottle, bus, car, chair, dining table, person, potted plant, monitor	PASCAL VOC 2011 dataset [14]
Speech recognition	Google API [16]	words in audios	-
Sound category classification	Gaussian mixture model [31], Random forest (cough detector) [28]	coughing (domain specialized), chattering, typing, horns of car, toilet flush, vacuum cleaner	Freesound [3], self collected cough dataset

ously seen, a new place is assumed to have been discovered. Similarly, previously visited places are recognized based on the WiFi fingerprint being sufficiently similar to fingerprints that have been observed earlier. We adopted the Tanimoto Coefficient as the distance function in this process and set the threshold to 0.7, as suggested by [18].

Privacy Controls. Users have three forms of privacy controls at their disposal. First, they can disable the sampling of certain sensors at specific locations/places. This option is available as soon as Smartphone App has learned the place's WiFi fingerprint or gained a GPS lock. Second, users can review all data collected and delete it if desired. Data uploading is delayed at least one day to allow the user to perform a manual review. Image and audio data can be viewed and played back, respectively. All other data modalities are visualized as list entries at a certain time and place/location. Data navigation is provided based on either (1) the time and date using a list interface, or (2) a map interface that shows places the user has visited. Third, the user can temporarily disable sensor sampling for a window of time (e.g., 3 hours).

Data Transmission. To minimize data collection's energy overhead to the user, we utilize a simple approach previously described in the literature (e.g., [7, 22].) Data transmission from Smartphone App to Server Infrastructure occurs only when (1) a WiFi connection is available and (2) the phone is line-powered. As a result, the impact to battery life and mobile bandwidth is reduced to negligible levels.

Server Infrastructure

As illustrated in Figure 1, Server Infrastructure comprises two components: (1) sensor-data classifiers and (2) application support. The classifiers extract high-level context from image and audio data, such as written texts or spoken words. The analysis support manages recognized contexts, raw data, location/places, and any auxiliary information.

Sensor-data Classifiers. We integrate four sensor-data classifiers into our system: (1) optical character recognition, (2) object recognition, (3) speech recognition, and (4) sound classification (embedding a specialized cough classifier). Prior crowdsourcing systems (e.g., [6, 7, 11]) have also closely incorporated classifiers to extract information from urban regions. In addition, we include a variety of classifiers used in examples of mobile sensing systems (e.g., [21]). Our intent is to incorporate a variety of often-used classifiers, so we can examine them at scale as part of our study.

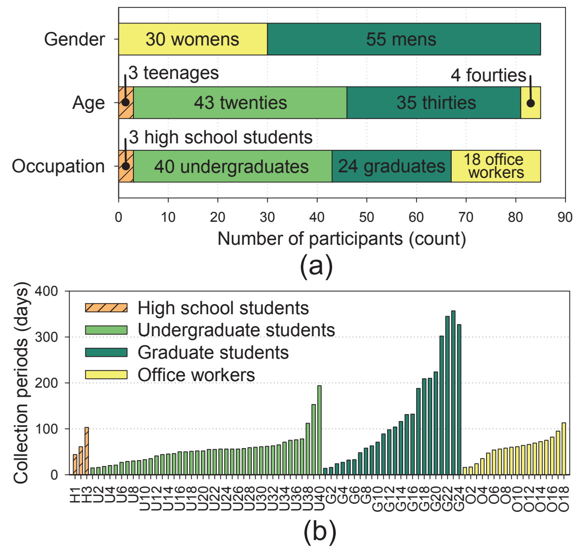


Figure 3. Composition of our study population in terms of (a) age, gender, occupation, and (b) collection periods.

Classifier Design. Each classifier uses techniques that are either commercial-quality or are recently published in the literature. Table 1 provides the specific details regarding each classifier, including features and classification model used, source of training dataset, and classes recognized.

Classifier Confidence. For each classifier, we also want to have an estimate of the algorithm's certainty of the classification. In the case of the two classifiers based on commercial solutions (OCR and speech recognition), a confidence algorithm is built-in to each and can be accessed via an API. For all others, we implement well-known classifier confidence techniques based on the classification model used. Object recognition is based on an exemplar-SVM approach; we implement a technique called binning that allows well-calibrated class probabilities to be estimated from a model [32]. Similarly, our primary sound classifier relies on Bayesian inference applied to Gaussian mixture models – we adopt the techniques described by Paalanen et al. in [34]. Finally, we also implement a specialized cough classifier that uses a random forest classifier to provide class probabilities. For this classifier, we adopt the technique proposed in [33].

STUDY FINDINGS AND IMPLICATIONS

In the following, we present key experiment results and implications from our study that touch on the four major themes: (1) place-temporal coverage, (2) the relationship between the user population and coverage, (3) privacy concerns and (4) characterizing the collected data.

Data Collection

We begin by describing our dataset and experiences during its collection.

CrowdSensing Participants. A total of 85 participants were selected out of approximately 150 applicants. Figure 3(a) shows the composition of our study population in terms of age, sex and occupation. The age of subjects ranged from 17 to 44, with 65% of subjects being male (35% female). We

Table 2. Description of collected dataset.

Type		Values
Locations	Covered Areas	230.2 km^2
	Unique Places	13,447
	Unique Paths	483,379
	Visit Counts	48,068
	Total Size of Database	813MB
Photos	Unique Places	1,580
	Number of Photos	6,242
	Total Size	9.1GB
Audios	Unique Places	1,517
	Number of Audios	22,604
	Total Length	192 hours
	Total Size	1.3GB

find participants have one of the following occupations: office workers, graduate students, undergraduate students, and high school students. All users live in Seoul, South Korea. During our study, subjects drop out of the experiment due to personal reasons (i.e., taking time off from school or changing their phone without notifying us). We pay 10 USD into drop-out users and removed the data they collect.

CrowdSensing Dataset. We collected the dataset from March 2011 to September 2012. Figure 3(b) presents the collection period across all participants. The average collection period was 79 days and the median was 58 days. Within the 6,700 user-days of collected data, subjects visit 48,000 places and make 13,500 distinct places – as presented in Table 2 that provides key statistics characterizing the dataset. The dataset contains more than 11 GBs, along with 22,000 audio clips (190 hours of audio) and 6,200 photos. Figure 4 shows the locations where data is collected in the city, with color indicating the volume of data collected at each particular location.

Some experiments require ground truth (e.g., labeling objects within images to act as training data) or data categorization. To make this possible we employ 20 undergraduates to hand label/categorize image and audio files, they are paid five cents (USD) per file. Due to the overwhelming size of the data, the portion of our dataset which is manually inspected is $\approx$ 15%. Data selection is performed by random sampling.

Social Media Dataset. During the same period as our study, we gathered a large social media dataset from FourSquare [2]. We restricted spatial regions within Seoul, Korea due to the huge amount of data in FourSquare. We collect 1,078,100 checkins, 99,000 POIs, 9,200 pictures by 31,000 unique users. Note that we collected only publicly available data.

For all experiments where place categories are required we adopt the category assignments determined by FourSquare. Unless otherwise stated, only the top-level of the FourSquare place category hierarchy is used. In some cases, FourSquare does not have a record of a place visited during our study. For these situations we apply the methodology described in [7], and performing manual coding using a group of five people.

Place-Temporal Coverage Properties

Our first set of results investigate the following coverage-related issues: (1) How extensively are places and place categories covered? (2) How does this coverage compare to data available from a social media (FourSquare) dataset?

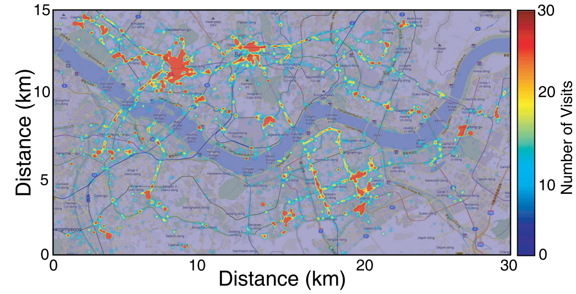


Figure 4. CrowdSensing study was performed in Seoul, Korea.

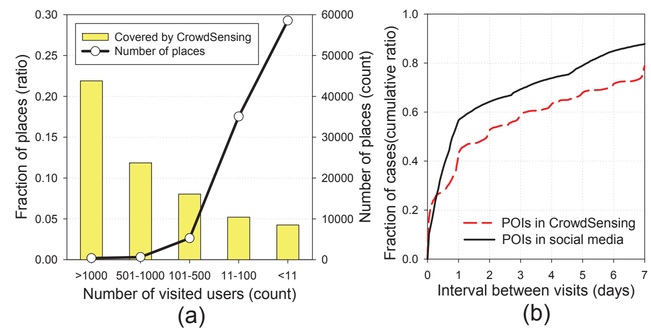


Figure 5. (a) Coverage of CrowdSensing and (b) interval between visits.

To estimate place coverage we rely on FourSquare that lists 98,899 places in Seoul. We find our users visit a total of 5,999 places during the study – 6% of this total. However, higher coverage is found for select place categories, such as: colleges (14%) or arts and entertainment (9%). Higher coverage levels are also found when the popularity of places is taken into consideration. Figure 5(a) groups places based on their popularity (as measured by FourSquare check-ins) and shows the coverage our dataset achieves at each level. For example, this figure shows our 85 users visited 22% of the top 0.4% percentile of places (that are visited by 1,000+ users).

Figure 6 compares the place visits made in our dataset to FourSquare check-ins made during the study period. Obviously our dataset contains far fewer place visits due to a much lower number of participants. For example, FourSquare check-ins dominate in many place categories, such as movie theaters, restaurants, or shopping stores. But interestingly in certain place categories – such as the examples provided in Figure 6, namely: residences, workplaces, internet cafes, or karaoke bars – crowdsensing (85 users) collects an appreciably higher number of visits than FourSquare ($\approx$ 31,000 users). When information is required about such places crowdsourcing may be a more effective method to use. We believe this is because the social motivations that underpin check-in activities (e.g., sharing experiences, receiving rewards) do not apply equally to all place categories.

In Figure 7(a), we present the time-of-day when crowdsensing data collection occurs for a variety of place categories. Because crowdsensing semi-continuously gathers data certain locations – for example, places visited later at night – are collected much more heavily than in the social media dataset.

Figure 7(b) examines the speed at which unique place visits

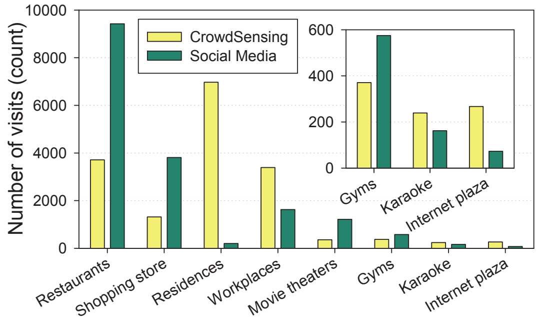


Figure 6. Visit counts of places in crowdsensing and social media according to categories.

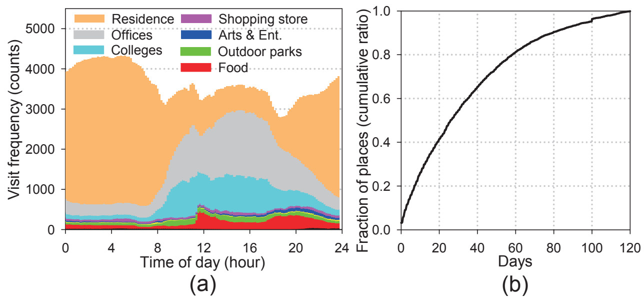


Figure 7. (a) Stacked plot of users' place visits by category over weekdays and (b) accumulation of place visits according to days.

accumulate during our study. We find that the rate at which our crowdsensing users visit new (i.e., previously unvisited) places quickly declines as time passes. For example, after 35 days at least one visit has already occurred for $\approx 60\%$ of the complete set of places eventually seen during the study. Our data suggests simply waiting for coverage to increase over time is of limited effectiveness. Instead, as examined in the following subsection, to reach new areas the crowdsensing user population must altered. Because place coverage saturates quickly one viable approach – especially when designers wish to deliberately guide coverage changes – appears to be incrementally adding users and waiting for the new system coverage to converge. After convergence, the next batch of recruits can be added and redundant users removed.

Finally to examine temporal place coverage we compute the interval between visits of users to each unique place. Figure 5(b) presents the CDF of visit intervals for both crowdsensing and social media datasets. From this figure we find that 60% of places are visited every ≈ 3.1 days. This longer interval for places is understandable given their small size but highlights the difficulty in providing up-to-date information regarding places as often proposed by crowdsourcing systems (e.g., [24]). In comparison, 60% of FourSquare places receive a check-in every ≈ 1.4 days.

Predicting Coverage at Scale

The next set of experiments consider: How many users are likely needed to scale-up place coverage? Accurately estimating this relationship must be done carefully. Simply by naively extrapolating from the coverage achieved by small numbers of people can potentially underestimate the number of participants required. One reason for this is that we find

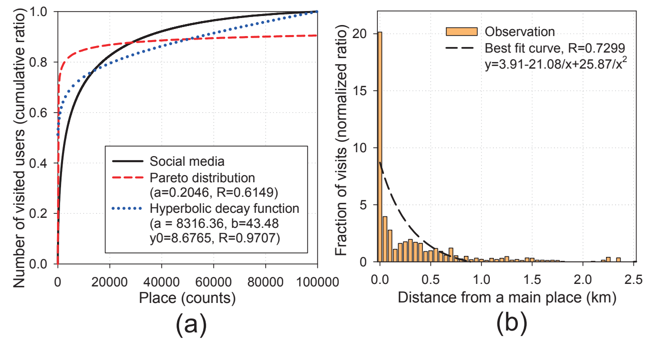


Figure 8. (a) Distribution of number of visited users to places and (b) visit probability to distances from top two places.

new users are more likely to share place routines and preferences with existing users as the population increases; as a result, these new users contribute fewer visits to previously unseen places than would be otherwise expected.

We propose a data-driven estimation model for predicting coverage relative to user population size. This model is based on two fundamental observations made from our dataset. First, we find in our study the frequency of place visits follows a power-law distribution (this result is supported by the findings of [26]). Figure 8(a) illustrates this observation using a CDF of FourSquare check-ins across all places in Seoul during our study. This CDF is closely fit by a hyperbolic decay function ($R = 0.97$) or more loosely by a pareto distribution ($R = 0.61$). The implication of this result is that many places will exist in the 'long tail' of this distribution, so will be infrequently visited – and so difficult to cover. Second, we find that the probability of a user place visit falls sharply as the distance between the place and the user's own significant places (i.e., often visited locations) grow. Figure 8(b) shows a histogram of place visits relative to the shortest distance of the top two significant places of the user (based on the significant places algorithm in [5]). This distribution is fit by an second order inverse function ($R = 0.72$). This result is important as it indicates the majority of user place visits are relatively localized; places are most likely to be visited by those who live or work near the place. Collectively, these two findings are useful when selecting a final crowdsensing user population from a cross-section of willing participants.

Our approach is a generative model based on the conditional probabilities representing the two observations, as just described, with functions used to the model data. To validate the model we assume the actual place locations (from FourSquare) and physical dimensions of Seoul. For each user in the simulation we must assume a location where they live. In addition, we grouped the places into four categories based on the number of visited users (i.e., 1000+, 100-999, 10-99, 1-9) as observed in the social media dataset. We then estimate the parameters in the model according to each categories. This approach incorporates an important behavioral tendency of users: users likely visit popular places rather than unpopular ones. For example, a user may visit a supermarket very frequently, but he/she may rarely visit a travel agency (unless a trip is planned) even if it is nearby to his/her home.

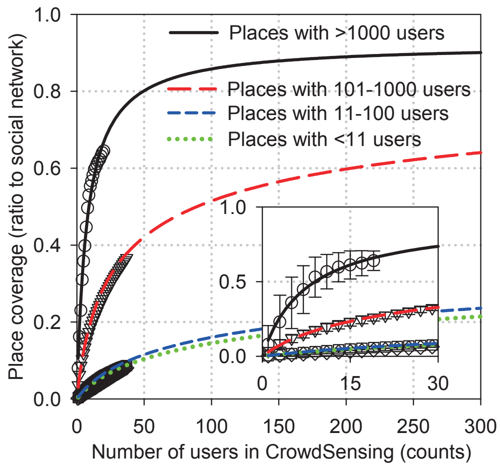


Figure 9. Estimation model about number of users and place coverage.

Figure 9 presents results that validate the estimation model. We again assume use the number of places in the social media dataset as the total number of places in the entire city. For low numbers of users we can compare the model predictions to real values from our dataset. Based on the 100 round simulation using real traces, our estimation model accurately predicts the coverage with an error of 11.3% (see inset in Figure 9). Next, we use the model to predict the increase in coverage assuming users are systematically selected from one of the 25 districts in Seoul with their living location set to a random location within the district. Under this scenario, the model predicts in the extreme ≈ 25,000 users will necessary to cover 60% of popular places (visited by 100+ users) in Seoul over a two month span – however, this is only 0.2% of entire population. Smartphone app user populations of this size are not that uncommon. For example, *SeoulBus* [36] used for checking bus arrival information has 268,000+ users living in Seoul today. This same model can be applied to estimate the required user population necessary for various targeted groups of places. One key limitation of this result however is the majority of our participants are connected to Yonsei University and may be more mobile and socially active than typical members of the city population.

Influence of User Privacy Concerns

During this study three highly sensitive data types are collected – images, audio clips and location (place visits and explicit location). In the following set of experiments we investigate: (1) Are users willing to collect this type of data? (2) If so, in which places do they collect this data? (3) How does these aspects of user behavior vary within the demographic groups represented in the study?

Feasibility. We find the majority of users are willing to collect and share all three data types, at least under certain conditions, during the study. As suggested in prior work (e.g., [19]), audio is the most sensitive modality with 29% of users considering too sensitive to share and choose to disable it for the entire study. Nevertheless, on average for each place visit 15 audio clips are collected along with 4 photos while location is also enabled 95% of the time. This overall result is important in that it shows despite natural concerns there are

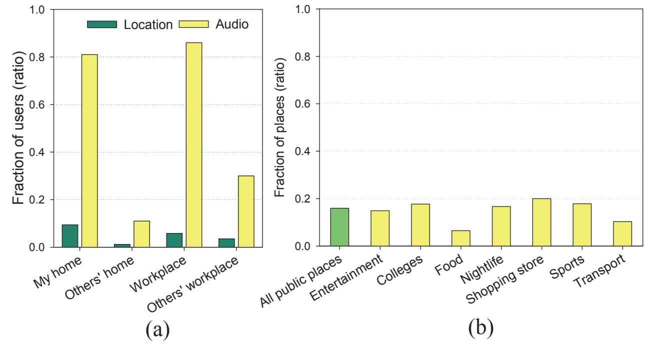


Figure 10. (a) Ratio of users who prohibit the collection of location and audio sensing. (b) Ratio of places where users prohibited the collection of audio data.

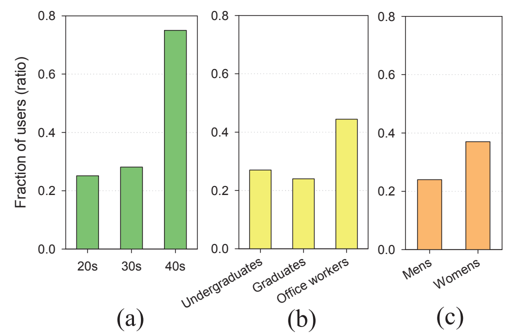


Figure 11. Ratio of users who prohibit the collection of audio according to (a) ages, (b) occupation, and (c) sex.

still enough willing users to crowdsource all three of these rich, yet sensitive, data types. As discussed in the following subsection, the availability of such data can act as valuable building blocks for a variety of crowdsensing applications.

Place Categories. Figure 10 shows how often users disable location and audio collection across a range of place types across all users. Figure 10(a) highlights place categories that are personal to each user that we determine by interviewing each study participant. Not surprisingly users are cautious when sampling in their own homes (81%) and workplace (86%) but much more open to homes of friends (11%) and the workplaces of others (30%) – likely because of it being considered a semi-public place. However, most users are willing to share location data (unlike audio) within private places: only 7% of users are cautious about sharing location data in personal places. Figure 10(b) presents the places where a user prohibits the audio collection for all other place categories based on the FourSquare category assignment. We find users are much more willing to capture data in public places. Users prohibit audio sampling in 16% of places. In contrast, users enable audio collection in 93% of food places (e.g., coffee shops/restaurants). These results have the following implications for building crowdsensing applications: (1) privacy concerns will not effect the coverage of public place types; but, (2) applications that require sampling in personal spaces (e.g., to understand user or community behavior) will be more challenging to develop and require larger user populations – although potentially still feasible as data is still collected.

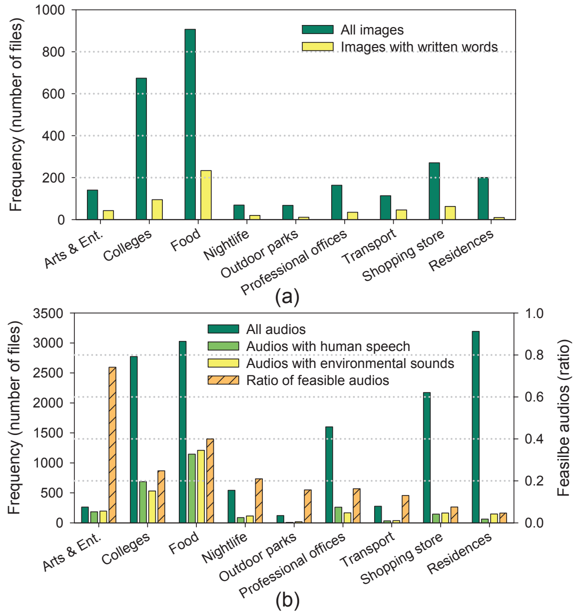


Figure 12. Amount of (a) image and (b) audio data according to place category.

Demographic Differences. Attitudes to privacy are well-known to vary based on personal beliefs and characteristics (e.g., [15]). Figure 11 presents the same privacy sharing information discussed earlier in this section but separated by subject age, occupation and sex. We omit the results of teenagers (i.e., high school students) as the population size is too small (3 users) to be meaningful. It appears office workers and women tend to prohibit the sampling in more places relative to students and men. Interestingly, this pattern does not hold in terms of the amount of data collection. On average, women collected 71% more images than men (the top-5 individual data contributors in the study are all women), and students take 24% more images than office workers. The result suggests women (at least in our study) are inclined to perform crowdsourcing tasks (i.e., picture taking), but they are sensitive to privacy and carefully use our data collection tool when doing so. These findings are useful in understanding the type of user population needed to capture certain types of data. When deploying a crowdsensing system this should jointly be considered along with user mobility patterns as both factors have considerable influence on the data that is collected.

Characterizing CrowdSensed Place Data

In our final set of experiments, we examine the data collected during the study and attempt to mine place-related contexts using off-the-shelf image and audio classifiers. Specifically, we consider: (1) What type of images and audio are collected in each place category? (2) How can the problem of low-quality data collected by untrained users be overcome? (3) Which contexts are contained in the data collected?

Place Categories. As already shown in Table 2, a large volume of image and audio data is collected during the study. Figure 12 further categorizes collected images and audio clips based on content, while also showing the place category where the data is captured. From Figure 12(a) we observe food and shopping related places are where the majority of images are captured; in such places, a great fraction of images also contain text detected by OCR. In contrast, place categories where users spend the majority of their time (e.g., residences, colleges, office) few images are collected. These results indicate such places are not well suited to exploit image-based contexts. This may be significant for systems seeking to monitor user place visits more continuously throughout the day. Figure 12(b) groups audio clips into three categories, namely audio containing: human voice, environmental sounds and one "other" category for the remainder. We find from this figure large amounts of audio is collected in place categories including: residences, food places, colleges, shopping places, and offices. Potentially audio may be able to compensate, under some scenarios, for the lack of images available in certain place categories (e.g., offices). However, we also find overall only 8% of collected audio contains recognizable sounds (viz. speech or environmental categories). Such inefficiency highlights the need for more intelligent sampling strategies than the simple scheme used in our system.

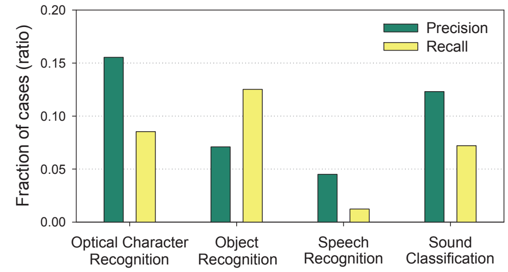


Figure 13. Precision and recall of primary classifiers.

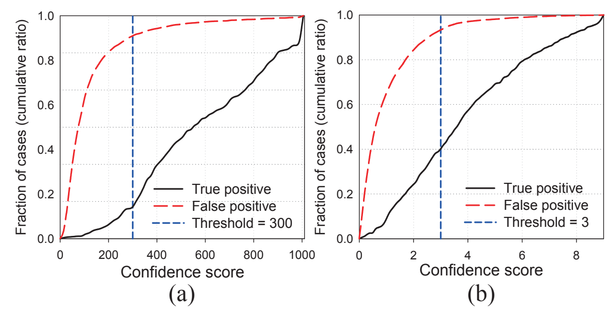


Figure 14. (a) Distribution of confidence scores in OCR and (b) object recognition.

Coping with Noise. Perhaps the most significant problem of crowdsensing is the low quality of sensing even though the system collects a huge volume of data. Users are not precise when capturing images, and audio clips are automatically sampled regardless of the ambient conditions. As a result, the automatic extraction of contexts from this dataset is challenging. Figure 13 quantifies this difficulty by presenting the accuracy of the four primary classifiers used in our study: OCR, object recognition, speech recognition, and sound classification. Despite our use of state-of-the-art techniques, the baseline precision and recall results for all classifiers are extremely low, ≈10% and ≈7% respectively.

Table 3. Example Classifiers Output.

Data Source	Semantic Contexts	Frequencies	Locations
Image	Chair/Table	598	143
	TV/Monitor	160	34
	ATM Machine	17	4
	Flowerpot	21	7
Audio	Conversation	1,716	470
	Horn of Cars	419	212
	Typing	78	18
	Cough	248	69
Image/Audio	14,081 Unique Words	56,867	913

Table 4. Frequently occurring Sounds, Words, and Objects.

Type	Frequency	Top Results (ordered by high frequency)
Sounds	> 500	music, laugh, foot stepping, car & street
	> 100	put down objects, cough, type keyboard, click mouse, *beep* (card, button), writing, crashing, game
	> 50	printing, paper passing, door open & close, bell ringing, moving objects, clapping, zipper
	> 20	glass hitting, humming, screaming, knocking, vibration, water flowing, sign, crying
	< 20	running, vacuum, dryer, brushing, sign, *animal sounds* (cat, dog, bug)
Words	> 100	*store name* (McDonald, Starbucks), *food name* (rice, noodle), *province name* (Seoul, Shinchon)
	> 50	cafe, station, pc, information, *date info.* (June, Sunday), *floor info.* (1F, B1F), *brand name* (Samsung, Hyundai)
	< 50	here, cashier, open, transfer, *URL info.* (www, kr), toilet, pick, service, new, room, best, fresh, atm
Objects	> 500	person, chair, table, light
	> 100	sign, plate, flowerpot, monitor, board, window, tv, bag, menu board, picture frame
	> 50	display stand, desktop, paper, box, clock, sofa, trash box, refrigerator, bottle, book shelf
	< 50	cup, speaker, curtain, price tag, keyboard, door, camera, laptop, shoes, vending machine, calendar, cosmetics, atm

Figure 14 highlights a potential solution. We find the classifier confidence algorithms we adopt in our crowdsensing system (see earlier system design section) are reasonably accurate even with low quality data. As an example, Figure 14 shows how false positives decrease if classifier results are ignored unless they also have a high confidence score. This is only possible if the confidence algorithm is effective. In Figure 14(a) 91% of OCR false positive results are filtered if we set a threshold of 300. Similarly, Figure 14(b) shows for object recognition we can eliminate 92% of false positives by using a threshold of 3. We find similar results across all classifiers in our system.

Although simply requiring a high confidence score is effective in lowering false positives, and thus increasing overall accuracy, it often causes many true positives (i.e., correct classifier results) to also be ignored. We find this occurs for all classifiers we test, with the exact quantity depending on confidence threshold used. However, just as noise is an inherent property of crowdsourcing so is redundancy. For example, Figure 5(b) indicates during the study users in social media and crowdsensing visit 60% of places every 1.4 days and 3.1 days respectively. Such redundancy allows us to set confidence thresholds at levels that would otherwise (in non-crowdsourcing scenarios) be impractical because of a limited number of opportunities to classify a particular context.

Frequently Extracted Contexts. To better understand potential crowdsensing application scenarios we present a selection of classifier results, as seen in Table 3. In all cases we tune each classifier until false positives fall to *2%* even though we understand this will filter out a large number of correct classification results. Table 4 provides examples of contexts extracted by manual labeling.

From Table 3, we find words extracted from conversation and writing in the environment are particularly promising sources of context. A total of 14,081 unique words (among 56,867 total recognized words) are extracted using both OCR and speech classifiers. One key reason we believe the number of classification results can remain high (such as, in the case of written words) – even when applying aggressive confidence thresholding – is because many sources of context (signs etc.) are static in the environment. As a result, these sources provide multiple opportunities for classification to occur correctly each time a user visits a place; this appears to be a general principle designers of these systems can leverage. In Table 4, we find common words extracted are tied, for example, to entities, brands and locations. Such words describe different attributes of a place. This suggests that a number of opportunities exist for various types of automated place understanding. Potential scenarios include: inferring store open hours, performing price comparisons, and measuring customer service – for instance, queuing wait times.

In Table 4, we also observe contexts commonly related to activities or actions. For example, contexts such as: music, clapping, laughing, coughing and typing. These contexts can be informative regarding routine user behavior during a place visit. Possible applications include: building an understanding of what occurs during place visits (e.g., how entertained is an audience based on overheard laughter or clapping?); or, even for performing place comparisons based on activity. For example, comparing two coffee shops based on if people tend to quietly work there or if people often talk and listen to live music. However, because context inferences tend to accumulate over multiple place visits – as discussed above – it appears crowdsensing is not well suited to making assessments on a visit by visit basis (i.e., a form of activity recognition) but instead making assessments incrementally as aggregate trends slowly build up.

LIMITATIONS

Despite the scale of our study, the participant population is not particularly diverse. Our data was collected only in a single city in South Korea, with 65% of subjects being university students (both graduate and undergraduate). Moreover, smartphone users in South Korea are some of the most active phone users in the world [35]. Similarly, user behavior with respect to privacy is known to be sensitive to characteristics like culture and occupation. As a result, we accept some of our findings still remain to be verified by performing similar experiments elsewhere in the world.

This study seeks to investigate a broad range of coverage and scalability related issues encountered in place-centric crowdsensing. Many of these issues (e.g., modeling the relationship between coverage and the user population) warrant additional experiments and investigation beyond the results presented in this paper. We anticipate revisiting these topics as part of future focused follow-up investigations, in addition to inviting other researchers to examine these topics using the study dataset that we have released publicly [30].

RELATED WORK

In this section, we describe closely related prior work while also highlighting the novel contributions of our research.

Mobile Crowdsourcing. Crowdsourcing is an active area of interest, and has been applied to a variety of different domains, ranging from translation [13] to image search [29]. Recently, a number of projects have examined the architectural requirements of these systems at scale [9, 27]. We complement this growing body of work by providing further insights that aid the evolution of crowdsourcing architecture and application scenarios.

[6, 24, 20, 25, 11, 7] are recent studies of place-centric crowdsensing. For example, SurroundSense [6] used smartphone sensors to build sensor fingerprints for place recognition. Similarly, VibN [24] leveraged crowds to enhance local search based on data from audio clips and user behavior. Again, our study fulfills a complementary role to this research by providing analysis of use to designers of such systems.

Although we perform our study using the same crowdsensing system proposed and evaluated in our own prior work [7] – this paper and [7] make distinct contributions. Specifically, all results presented in [7] are tied to a single place-centric application (place category classification); furthermore, [7] performs no analysis of coverage, scalability or privacy and only indirectly considers collected data in terms of how it impacts the targeted place-centric application.

Participatory and Opportunistic Sensing. Mobile crowdsourcing and the results of our study are tightly linked with other sensing system architectures, namely participatory and opportunistic sensing. [11, 28] have a particular city focus: LiveCompare [11] leverages the camera to explore the discovery of grocery bargains; EarPhone [28] constructs a noise map by opportunistically collecting audio samples. These systems motivated our study and help guide our study aims.

Privacy. [17, 19, 10] have investigated privacy concerns of sensor usage; [19] in particular provided detailed insights with respect to the usage of the microphone. Our findings are fairly consistent with both these studies. In addition, we contribute by providing further results under the specific domain of crowdsensing, and in a different culture and city.

CONCLUSION

In this paper, we presented a detailed study based on one of the largest examples of urban crowdsourcing using smartphones performed to date. We collected two-months of smartphone sensor data from 85 study subjects who made a total of $\approx$ 48,000 place visits in Seoul, Korea. Our investigation examined key issues for place-centric crowdsensing – specifically, we studied place-temporal coverage properties, coverage scalability, participant privacy concerns and performed a characterization of the collected data. We believe the analysis and findings we have presented provide valuable insights useful not only for builders of crowdsensing systems; but, also apply to other closely related sensing systems that rely on having a close engagement with the user.

ACKNOWLEDGEMENTS

This work was supported by the NRF of Korea – grant funded by the Korean government, MEST (No.2011-0006464, No.2013-027363).

REFERENCES

1. FieldAgent. `http://www.fieldagent.net/`.
2. FourSquare. `http://foursquare.com/`.
3. Freesound. `http://freesound.org/`.
4. Gigwalk. `http://gigwalk.com/`.
5. D. Ashbrook and T. Starner. Using gps to learn significant locations and predict movement across multiple users. *Personal Ubiquitous Comput.*, 7(5):275–286.
6. M. Azizyan et al. Surroundsense: mobile phone localization via ambience fingerprinting. In MobiCom '09, pp.261–272, 2009. ACM.
7. Y. Chon et al. Automatically characterizing places with opportunistic crowdsensing using smartphones. In UbiComp '12, pp.481–490, 2012. ACM.
8. Y. Chon et al. Mobility prediction-based smartphone energy optimization for everyday location monitoring. In SenSys'11, pp.82–95. ACM, 2011.
9. E. Cuervo et al. Crowdlab: An architecture for volunteer mobile testbeds. In *COMSNETS*, pp.1–10. IEEE, 2011.
10. A. Raij et al. Privacy Risks Emerging from the Adoption of Innocuous Wearable Sensors in the Mobile Environment. In *CHI '11*
11. L. Deng and L. P. Cox. Livecompare: grocery bargain hunting through participatory sensing. In HotMobile '09, pp.4:1–4:6, 2009. ACM.
12. J. Du, Q. Huo, L. Sun, and J. Sun. Snap and translate using windows phone. In ICDAR'11, pp.809 –813, IEEE. 2011.
13. N. Eagle. txteagle: Mobile crowdsourcing. In IDGD '09, pp.447–456, 2009. Springer-Verlag.
14. M. Everingham et al. The pascal visual object classes challenge 2011 (voc 2011) results. `http://www.pascal-network.org/challenges/VOC/voc2011/workshop/index.htm`.
15. J. Fogel and E. Nehmad Internet social network communities: Risk taking, trust, and privacy concerns. *Computers in Human Behavior*, 25:1 153–160, 2009.
16. G. Hinton et al. Deep neural networks for acoustic modeling in speech recognition. *Signal Processing Magazine*, 2012.
17. G. Iachello et al. Prototyping and sampling experience to evaluate ubiquitous computing privacy in the real world. In CHI '06, pp.1009–1018, 2006. ACM.
18. D. H. Kim et al. Sensloc: sensing everyday places and paths using less energy. In SenSys '10, pp.43–56. ACM, 2010.
19. P. Klasnja et al. Exploring privacy concerns about personal sensing. In Pervasive '09, pp.176–183, 2009. Springer-Verlag.
20. J. Weppner et al. Bluetooth based Collaborative Crowd Density Estimation with Mobile Phones. In Percom '13, pp.193–200, 2013.
21. E. C. Larson et al. Accurate and privacy preserving cough sensing using a low-cost microphone. In UbiComp '11, pp.375–384, 2011. ACM.
22. K. Lee et al. Mobile data offloading: how much can wifi deliver? In Co-NEXT '10, pp.26:1–26:12, 2010. ACM.
23. T. Malisiewicz et al. Ensemble of exemplar-svms for object detection and beyond. In *ICCV '11*, pp.89–96, 2011.
24. E. Miluzzo et al. Tapping into the vibe of the city using vibn, a continuous sensing application for smartphones. In SCI '11, pp.13–18, 2011. ACM.
25. P. Kannan et al. Low cost crowd counting using audio tones smartphones. In SenSys '12, pp.155–168, 2012.
26. A. Noulas et al. A tale of many cities: universal patterns in human urban mobility. *PloS one* 7:5, 2012.
27. M.-R. Ra et al. Medusa: a programming framework for crowd-sensing applications. In MobiSys '12, pp.337–350, 2012. ACM.
28. R. K. Rana et al. Ear-phone: an end-to-end participatory urban noise mapping system. In IPSN '10, pp.105–116, 2010. ACM.
29. T. Yan, V. Kumar, and D. Ganesan. Crowdsearch: exploiting crowds for accurate real-time image search on mobile phones. In MobiSys '10, pp.77–90, 2010.
30. CrowdSense Study Dataset. `http://lifemap.yonsei.ac.kr/ubicomp13.htm`.
31. H. Lu et al. SoundSense: Scalable Sound Sensing for People-Centric Applications on Mobile Phones. In Mobisys '09
32. J. Drish et al. Moderating the outputs of support vector machine classifiers IEEE Transactions on Neural Networks
33. H. Bostrom et al. Estimating Class Probabilities in Random Forests In ICMLA '07
34. P. Paalanen et al. Feature representation and discrimination based on Gaussian mixture model probability densitiesPractices and algorithms. Pattern Recognition 39(7) 1346–1358
35. Nielsen Research The Mobile Consumer Report 2013 in Mobile World Congress '13
36. SeoulBus. `https://play.google.com/store/apps/details?id=com.astroframe.seoulbus`.