Mobicom Poster: SurroundSense: Mobile Phone Localization Using Ambient Sound and Light

Martin Azizyan
Dept. of Computer Science,
Duke University
ma45@duke.edu

Romit Roy Choudhury
Dept. of Electrical and Computer Engineering,
Duke University
romit@ee.duke.edu

ABSTRACT

Proliferating mobile phones provide a foundation for revolutionary innovations in people-centric computing. Numerous applications are on the rise, many of which exploit the phone's location as the primary indicator of *context*. Existing physical localization schemes based on GPS/WiFi/GSM have been shown to achieve, at best, localization accuracy of several meters. This paper argues that such accuracies may not be sufficient for several location based applications. For example, while transmitting location based advertisements to mobile phones, it is important to learn whether a mobile phone is in a grocery store, or in the adjacent coffee shop. Small errors can place the phone on the wrong side of the partition, affecting the application. This poster proposes to sense the surrounding in which a phone is located, and use this ambient information to classify its location. Put differently, we postulate that different surroundings have photo-acoustic fingerprints, that can be sensed and used for localization. We demonstrate the feasibility using Tmote Invent motes that have light and sound sensors. Our ongoing work is extending *SurroundSense* to the mobile phone platform, and exploiting additional sensors (such as accelerometers and compasses) towards even better localization.

1. INTRODUCTION

Mobile phones are becoming a powerful platform for people-centric applications. Emerging ideas include participatory sensing [1, 2], pervasive computing [3, 4], and social networks [5]. Many of these applications rely on the location of the phone as a primary indicator of the user's context. While GPS services are the popular choice for localization, new investigations are showing that the battery lifetime with GPS is less than an unacceptable 10 hours [1]. Moreover, GPS signals do not work indoors, and hence, are unsuitable for many location-specific applications. Technologies such as Skyhook and Place Lab have proposed WiFi and GSM based localization schemes that are less energy hungry [6, 7]. However, the accuracy of such schemes range from 40 to 400 meters, making them excessive for many applications. We believe that publicly available localization technology will not achieve sub-meter accuracy in the near future. Hence, a user sitting near the wall of a coffee shop, may often be placed on the other side of the wall, in an adjacent grocery store. This will disrupt services that provide, for example, location-specific advertisements. A coffee drinker may get annoyed if she receives a coupon for toothpaste.

Towards developing a practical localization technology, we observe that *accurate physical coordinates may not be essential to many applications*. Put differently, an application may only require the context in which a phone is situated, e.g., a coffee shop, a grocery store, a restaurant. Reliably identifying this context can be valuable, and may even offer the physical coordinates by looking up a geocode database. This poster attempts to use the variety of sensors embedded in mobile phones to sense the context of a phone. In particular, we combine ambient light and sound sensors to find a *fingerprint* of the surrounding. We postulate that different contexts have a reasonably dissimilar fingerprint. For example, ambient sound in Starbucks may include specific noise signatures from coffee machines and microwaves, that are different from forks and spoons clicking in restaurants. Even lighting styles may be different in order to match with the type of service a place may provide – bars with dim yellow lights versus BlockBusters with bright white light (Figure 1). Thus, if the macro-location of a phone can be obtained through GPS, WiFi, or GSM, ambience-based fingerprinting can be used to pinpoint the phone's precise context within the macro-location. Put differently, knowing that a phone is in the shopping mall, it might be possible to pinpoint the store in which it is located. We demonstrate the feasibility of this postulation through this project, *SurroundSense*. Our main intuition is that, although each of the individual sensors may not offer concrete information about the phone's context, together they may produce a fingerprint. Our measurements on Tmote Invent sensor motes, along with simple classification algorithms, show encouraging results. Ongoing work is extending *SurroundSense* to Nokia N95 mobile phones, enriching the fingerprint with additional on-phone sensors (cameras, accelerometers, compasses), and drawing on machine learning algorithms to classify sensed data.

2. SYSTEM SETTING

We use Tmote Invent sensor motes on behalf of mobile phones. Our aim was to validate the feasibility of photo-acoustic fingerprinting, and upon success, extend our system to the mobile phone platform. Each Tmote Invent is equipped with a light sensor, a microphone, and other sensors we did not use. One mote was configured to send acoustic samples from its microphone every 2 milliseconds (a sample rate of 500, translating to a frequency range from 20 to $250Hz$). The other mote was programmed to send light intensity readings every 50 milliseconds. A laptop captured and saved the data as shown in Figure 2. The subsequent process is described next, with a flowchart representation in Figure 3.

3. SURROUND SENSE: DESCRIPTION

After audio and light data is recorded from one location, it

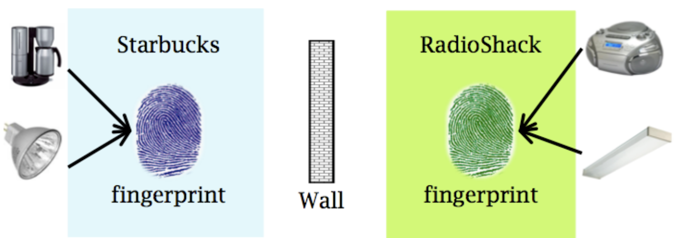

Figure 1: Ambiances in different locations may exhibit diversity, which can in turn be sensed towards localization.

needs to be converted to a fingerprint. The audio and light signatures may vary over time scales of seconds or minutes. In the case of light, a change can occur if the light sensor is directly exposed to a light source, or if it is pointed towards a shadow. In audio, change can be caused by a large group of people entering or leaving the room, starting and ending conversations, or the person holding the sensor moving closer to some device with a unique audio signature (microwaves, vacuum cleaners, refrigerator, etc.). Hence, fingerprints need to be some type of an average (over multiple samples in time) as opposed to an instantaneous reading of the sensor.

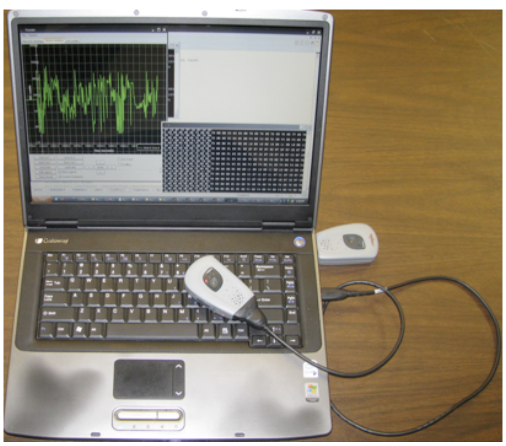

Figure 2: Tmote Invent with light and sound sensors sense the ambience. The sensed data is downloaded onto a laptop and fingerprints generated from them.

3.1 Fingerprinting

In the interest of space, we report our fingerprinting scheme without detailed justification. To generate an acoustic fingerprint, 10 second blocks are extracted from the recorded clip, at a rate of one block every 5 seconds. A spectral representation of a block is computed by performing a Fourier Transform. The frequency range from 20 to 250 Hz is divided into 23 bands, each 10 Hz wide. For each band, the *average* power of the signal falling within the corresponding frequency range is computed. This results in an array of 23 numbers describing each block. The average value and variance of each number over all blocks is computed. These 46 values, combined with the average and variance of the light intensity readings, yield a 48-dimensional fingerprint for a location. Fingerprints are gathered from different locations and stored in a repository, which we call the Fingerprint Database.

3.2 Localization

We need to accept a <light, sound> tuple from a sensor (call it a test fingerprint), and match it to one of the fingerprints in our repository. Numerous matching algorithms exist, and the selection of the best candidate is a topic of ongoing work. In this poster, we evaluate the results from a simple matching algorithm. The test fingerprint, T, is compared with each fingerprint in the repository, R_1 to R_n. Each comparison outputs a value of *similarity*, defined as the inverse of the Euclidean distance between the fingerprints (essentially points in 48 dimensional space). Smaller the Euclidian distance, higher is the similarity. Mathematically, if S_i is the similarity between T and R_i, then S_i can be expressed as

$$S_i = \frac{1}{\sum_{j=1}^{48}(T_j - R_{ij})^2}$$

where T_j and R_{ij} are the values in position j in the test (T) and repository fingerprint (R_i), respectively. The location of fingerprint R_k, for which S_k is largest (most similar to test fingerprint) is assigned as the location of the phone.
This comparison method uses a combined photo-acoustic fingerprint. Our hope is that even if two locations are similar with respect to sound or light alone, they may prove to be discernible through a combined fingerprint. Our results confirm this in the following subsection.

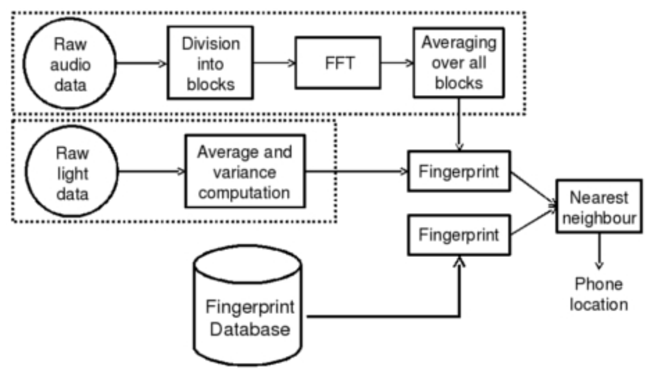

Figure 3: The flow of operations in *SurroundSense*. A Test Fingerprint from the phone is compared with fingerprints from the Fingerprint Database created *a priori*. The fingerprint with highest similarity indicates the location of the phone.

3.3 Evaluation

Towards a proof of concept, we collected light and sound data from six nearby businesses at the Ninth Street area in the Duke University campus. The locations were Dale's, Twinnie's, Blue Express restaurant, Francesca's coffee shop, Whole Foods grocery, and the street itself. The recordings were divided equally to compose the training set and the test set. Figure 4 shows the pairwise similarity between training and test data for all locations. The X axis of the figure has 6 clusters of bars, cluster i reflecting the comparison of store i with all other stores[1]. Put differently, a bar in the j^{th} position in the i_{th} cluster represents the likelihood of estimating store i as store j. It can be seen, that in the first cluster, the first bar is the tallest. In the second cluster, the second bar

[1]The numbering on the X axis is in the same order as that in the legend.

is the tallest, and so on. Essentially for all stores, the estimation algorithm achieved the best match (highest similarity) for the correct store. This accuracy is quite encouraging and motivates further research in the design of SurroundSense.

Interestingly, although the algorithms we used are simple, they were able to discern between nearby shops in our campus. Observe that in the case of Francesca's and the street, the match is correct by a relatively narrow margin. We believe more sophisticated fingerprinting and classification will decrease the occurrence of such cases.

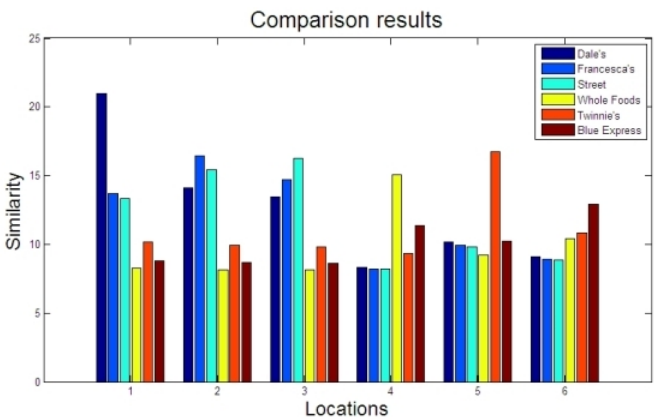

Figure 4: Localization of different contexts in the Duke University campus. A coffee shop, restaurants, and a grocery store are well localized.

4. LIMITATIONS AND ONGOING WORK

Strengthening SurroundSense: While results reported in this paper are encouraging, they may not be treated as conclusive. Our ongoing work is augmenting *SurroundSense* across multiple directions, including (1) exhaustive training sets across different times and locations, (2) more sophisticated fingerprint generation and localization, and (3) inclusion of other sensors to increase the richness of sensed data. The performance evaluation also needs to be more exhaustive.

Energy Implications: With increasing number of applications running on mobile phones, battery life will be a prime concern. We will study the energy overheads of *SurroundSense*, and compare it with alternate forms of GPS/WiFi/GSM localization. Early measurements on Nokia N95 phones showed that, except for the camera flash, other sensors consume less power than GPS and WiFi. We will characterize and address these tradeoffs in detail.

Pictures from Cameras: We are investigating the applicability of pictures taken from phone cameras, as an indicator of context. Our basic idea is to extract color features from pictures taken by the phone, and use them to classify location. Since some places have unique color signatures in the surrounding (say, a red theme in wall colors in Target versus a dark orange in Panera Breads), they may prove useful in some cases. Moreover, if the orientation of the phone is known via a compass, it may be possible to deduce if the picture is of the wall or of the ceiling.

Accelerometers: We also plan to add accelerometer-based activity recognition, and use it in conjunction with *SurroundSense*. In a separate work, we have used accelerometers to classify human activity, and relate it to location [8]. If accelerometer readings from a user's phone indicates that the user is moving (as opposed to say, sitting), then the user may be positioned in a grocery store (as opposed to a coffee shop). This project will unify accelerometers into its fingerprinting/localization framework to further improve accuracy.

5. CONCLUSION

Various mobile computing applications require a user's context to operate correctly. Rough location information may not be a perfect indicator of context because the spatial separation between two adjacent contexts may be arbitrarily small. Consequently, slight error in physical coordinates may translate into a completely erroneous context. We present *SurroundSense*, a system that exploits ambient light and sound for context identification/localization. The key idea is to find a fingerprint for different contexts through a (one-time) training phase. Subsequently, each phone can export its sensed information to a remote server, which can then localize it accurately. Of course, a photo-acoustic fingerprint may not qualify as a stand-alone localization technique. However, given a macro location (via GPS/WiFi/GSM), *SurroundSense* can perform micro-localization based on the properties of the environment. In conjunction with other phone sensors (camera, accelerometer, and compass), the confidence on micro-localization can be even higher.

6. REFERENCES

[1] Shravan Gaonkar, Jack Li, Romit Roy Choudhury, Landon Cox, and Al Schmidt, "Micro-blog: Sharing and querying content through mobile phones and social participation," in *ACM MobiSys*, 2008.

[2] Shane B. Eisenman, Nicholas D. Lane, Emiliano Miluzzo, Ronald A. Peterson, Gahng-Seop Ahn, and Andrew T. Campbell, "Metrosense project: People-centric sensing at scale," in *First Workshop on World-Sensor-Web (WSW'2006)*, Oct, 2006.

[3] Timothy Sohn, Kevin A. Li, Gunny Lee, Ian E. Smith, James Scott, and William G. Griswold, "Place-its: A study of location-based reminders on mobile phones," in *Ubicomp*, 2005, pp. 232–250.

[4] Jon Froehlich, Mike Y. Chen, Sunny Consolvo, Beverly Harrison, and James A. Landay, "Myexperience: a system for in situ tracing and capturing of user feedback on mobile phones," in *ACM MobiSys*, 2007.

[5] Mehul Motani, Vikram Srinivasan, and Pavan S. Nuggehalli, "Peoplenet: engineering a wireless virtual social network," in *MobiCom*, 2005.

[6] Anthony LaMarca, Yatin Chawathe, Sunny Consolvo, Jeffrey Hightower, Ian E. Smith, James Scott, Timothy Sohn, James Howard, Jeff Hughes, Fred Potter, Jason Tabert, Pauline Powledge, Gaetano Borriello, and Bill N. Schilit, "Place lab: Device positioning using radio beacons in the wild," in *Pervasive*, 2005, pp. 116–133.

[7] Y. Chen, Y. Chawathe, A. LaMarca, and J. Krumm, "Accuracy characterization for metropolitan-scale wi-fi localization," in *ACM Mobisys*, 2005.

[8] Andrew Ofstad, Emmett Nicholas, Rick Szcodronski, and Romit Roy Choudhury, "Aampl: Accelerometer augmented mobile phone localization," in *ACM MELT Workshop (in conjunction with Mobicom)*, 2008.