Enabling Wide Deployment of GSM Localization over Heterogeneous Phones

Mohamed Ibrahim
Wireless Research Center
EJUST, Egypt
Email: mohamed.ahmed@ejust.edu.eg

Moustafa Youssef
Wireless Research Center
Alexandria University and E-JUST, Egypt
Email: moustafa.youssef@ejust.edu.eg

***Abstract*—Wide deployment of GSM based location determination systems is a critical step towards moving existing systems to the real world. The main barrier towards this critical step is the heterogeneity of existing types of cell phones which results in different readings of received signal strength. Specially, in the context of fingerprinting localization where offline phases are needed for system training and different types of phones may be used in the offline and the online phases. Therefore, a mapping function, that maps the RSSI values between different types of cell phones, is inevitably needed. A trivial solution is to build a radio map for each type of phone. Obviously, this solution can neither scale in terms of number of phone types nor fingerprint size. In this paper, we address this problem by proposing the following two-way approach: A mathematical approach that maps RSSI values of different types of phones using linear transformation with regression, or logging ratios of readings instead of absolute values. We have empirically evaluated the proposed approach on Android-based phones. Our experimental results show that applying our approach can improve location accuracy with at least 127.84% in multiple cell tower configuration and at least 22.11% in the single cell tower configuration compared to the state-of-the-art GSM localization systems.**

I. Introduction

The demand for context-aware applications has been increased significantly in the last few years [1], [2]. A wide set of location-based services including location-aware social networking, navigation, and security applications are enabled by knowing cell phones' locations. Many systems have been proposed to address the localization problem including the GPS system, cellular-based system, and WiFi-based systems. However, GPS [3] is considered one of the most well known localization techniques, it is not available in many cell phones, requires direct line of sight to the satellites, and consumes a lot of energy. Therefore, research avenues directed towards other techniques for obtaining cell phones' location has gained momentum recently. This was motivated by both the user needs for location-aware applications and government requirements, e.g. FCC [4]. WiFi-based localization has been proposed in [5]–[10] and commercial products are currently available [11]. However, sufficient WiFi coverage is not available in all the cities in the world to obtain ubiquitous localization and not all the cell phones support WiFi networks. Similarly, the same availability issues exist for the localization systems that use augmented sensors in cell phones, e.g. accelerometers and compasses [12]–[17]. As these sensors are only available in smart phones. On the other hand, GSM-based localization, by definition, is available on all GSM-based cell phones, which represents 80-85% of today's cell phones [18], works all over the world, and consumes minimal energy in addition to the standard cell phone operation. Many research work have addressed the problem of GSM localization [4], [6], [19]–[23], including time-based systems, angle-of-arrival based systems, and received signal strength indicator (RSSI) based systems. GSM-based localization systems have been implemented [6], [19]–[23], Only recently, with the advances in cell phones. The RSSI information is mainly used in these systems as it is easily available to the user applications.

Implementing GSM-based localization on different types of phones is not a straightforward task, as the RSSI readings are not the same for different types of phones. A trivial, but not scalable, solution is to build a radio map for each type of phone. Therefore, a mapping function that maps the RSSI values between different types of cell phones is needed. This brings up the challenge of providing accurate GSM localization using different types of cell phones. Therefore, new techniques to address this new problem are inevitably needed.

In this paper, we address this problem by proposing the following two-way approach:

1) Relative power
2) Linear transformation with Gaussian kernel

The relative power technique is simply logging ratios of RSSI readings instead of absolute values. While the other technique is mapping RSSI values of different types of phones using linear transformation with regression. Similar approaches have been used before only in the context of indoor WiFi localization [24]–[26].

To evaluate our approach, we implemented it on Android-enabled cell phones, and compare their performance to probabilistic [21] fingerprinting technique, and to Google's MyLocation service [27] under a testbed representing urban environment. The obtained results show that our approach provide at least 127% enhancement in median error in multiple cell tower configuration and at least 22% enhancement in median error in single cell tower configuration compared to other systems.

The organization of the rest of the paper is as follows: Section II presents a brief background on the current RSSI-based localization techniques used in GSM networks. In Section III we discuss our solutions to the heterogeneity problem for a single and multiple cell tower configurations. Section IV presents the experimental evaluation of our system. Finally, Section V gives the conclusion of the paper.

II. Background

A brief background is presented here on the current RSSI-based localization techniques for GSM networks. We use them for comparison with our techniques, namely: cell-ID based techniques and fingerprinting techniques.

A. Cell-ID based Techniques

Cell-ID based techniques, do not use RSSI explicitly, e.g. Google's MyLocation [27], but rather estimate the cell phone location as the location of the cell tower with which the phone is currently associated. A database of cell towers' locations is required for these techniques to provide an efficient, though coarse grained localization method.

B. Fingerprinting Techniques

During an offline phase, fingerprinting techniques store in a database the RSSI signature (fingerprint) of cell towers at different locations in the area of interest. Then, during the tracking phase, this database is searched for the closest location in the RSSI space to the unknown location. Fingerprints are usually constructed by war driving, where a car drives through the area of interest while continuously scanning for cell towers and recording the cell tower ID, RSSI, and GPS location for the associated and neighboring cell towers.

Current fingerprinting techniques for GSM localization are either deterministic [19], [20] or probabilistic [21] techniques. Deterministic techniques do not take the signal strength distribution into account. For example, each location in the fingerprint of [19] stores a vector representing the RSSI value from each cell tower heard at this location. The K-Nearest Neighbors (KNN) classification algorithm is used, during the tracking phase, where at an unknown location the RSSI vector is compared to the vectors stored in the fingerprint. Then, in terms of Euclidian distance in the RSSI space, the K-closest fingerprint locations to the unknown vector are averaged as the estimated location. On the other hand, probabilistic techniques store information about the RSSI distribution in the fingerprint. Then, during the tracking phase, they try to estimate the most probable user location using these information. For example, the CellSense system [21] stores the RSSI histogram for each cell tower at a particular location and estimate the user location using Bayesian-based inference.

Fingerprinting techniques provide higher accuracy than cell-ID based techniques but require searching a larger database. However, both of them require war driving for constructing their databases.

III. Heterogeneity Solutions

In this section we discuss our solutions to the heterogeneity problem.

A. Multiple Cell Tower

First, we tackle the heterogeneity problem using the readings of multiple cell towers. The problem, as stated earlier, is seeing different readings from the same cell towers using different cell phones at the same locations. To address this problem, we use the relative power approach which implies saving the ratio of readings from each pair of cell towers instead saving the readings for each cell tower. The main idea here is that this ratio remains the same even if we change the type of the cell phone.

More formally, given an RSSI vector $s = (s_1, ..., s_q)$ we store it as $r_s = (r_{s_{12}}, r_{s_{13}}, ..., r_{s_{q-1q}})$ where s_i is the RSSI value received from the cell tower number i, $r_{s_{ij}} = \frac{s_i}{s_j}$, and $|r_s| = \binom{q}{2}$. Then for each pair of cell towers, we build the histogram of the ratio of the RSSI readings. However, we also save the histogram for each individual cell tower as the standard *CellSense* approach [21]. This will help during the online phase when only one cell tower is available from the pair of cell towers seen in the training phase. In this case, we compute the probability from the histograms of the individual cell tower. It should be noted that this approach can not be used in the case of the single cell tower configuration([22]) because for the same cell towers the successive samples are almost the same.

B. Single Cell Tower

In this case, we have only the associated cell tower information. Therefore, we introduce here another approach, namely, the linear transformation with Gaussian kernel. The key feature in this approach, is to map the histograms of the different cell towers using a linear transformation (offset) and smoothing (scaling). In order to do that, we collect readings using different types of phones at the same locations and compute the least squares fit between them. Then, we use the Gaussian kernels in order to smooth the histograms of the cell towers.

The Gaussian kernel function using mean $= 0$ and variance $= 1$ is:

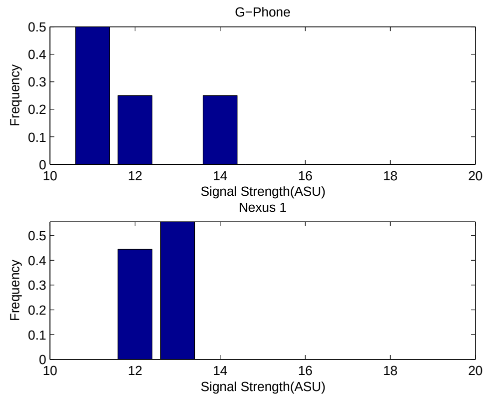
$$F(X) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{1}{2}(X^T X)} \quad (1)$$

Where d is the length of the vector X. In our system, we build the histogram for the RSSI values so X is in one dimensional space. In order to smooth the histogram for a given value x:

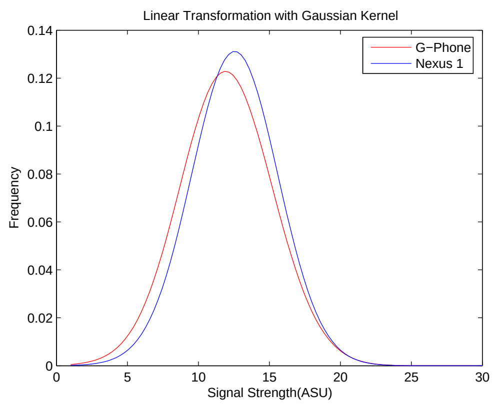
$$p(x) = \frac{1}{Nh} \sum_{i=1}^{N} F(\frac{x - x_i}{h}) \quad (2)$$

Where h determines the width of the Gaussian kernel and N is the total number of samples used to build the histogram.

Figure 1 shows the effect of applying our approach on the histograms of different types of phones. After applying the linear transformation to offset the histograms and after smoothing them using Gaussian kernels, they become almost the same. Therefore, our approach minimizes the effect of the heterogeneity problem.



(a) Before applying linear transformation with Gaussian kernel approach.



(b) After applying linear transformation with Gaussian kernel approach

Fig. 1. Effect of applying the linear transformation with Gaussian kernel approach on the histograms.

IV. Performance Evaluation

We study in this section the effect of grid cell size on the two approaches mentioned before and compare their performance to other GSM-based localization systems in terms of localization accuracy.

A. Data Collection

We collected data for a testbed that covers a $5.45\ Km^2$ in Alexandria, Egypt representing a typical urban area. Data was collected using three phones: T-Mobile G1, Nexus1 and Samsung Galaxy S. These three phones have GPS receivers (used as the ground truth for location) and running the Android 1.6 and 2.1 operating system respectively. The experiment was performed using three phones, each with a SIM card belonging to the same service provider in Egypt.

We implemented the scanning program using the Android SDK with one per second scanning rate. The program records the (cell-ID, signal strength, GPS location, timestamp) for the cell tower to which the mobile is connected as well as the six neighboring cell towers information as dictated by the GSM specifications. Two independent data sets were collected for the testbed: one for training and the other for testing.

It should be noted that Samsung phones provide only the associated cell tower information so they are used only in the single cell tower experiment.

B. Grid Cell Size Effect

The effect of grid cell size is studied here on the accuracy of our system for both configurations (single and multiple cell towers).

1) Single cell Tower: Figure 2 shows the effect of changing the grid cell length on the median localization error of the linear transformation with Gaussian kernel technique. We show here the effect of using each of the three types of phones for training the system while using another type for testing the system. From the figures we can notice that Nexus phone always provides better accuracies than the G-Phone and Samsung phone. Because Nexus phone, which may has more sensitive antenna, was always seeing more cell towers compared to the other types in our experiment. Therefore, using Nexus phones for training will always results in better accuracies than any of the other two types of phones. Also the effect of the grid cell size parameter on performance has different behavior than the standard behavior of *CellSense* [21]. This is because the dominant factor here is as the cell size increases, the number of samples increases and therefore, the accuracy increases. Figure 3 shows that the linear transformation with Gaussian kernel approach improves the accuracy compared to the ordinary *CellSense* approach.

2) Multiple cell Tower: Figure 4 shows the effect of changing the grid cell length on the median localization error of the ratio technique.

The effect of changing the grid cell length is almost the same as the effect on *CellSense*, two opposing factors affect the accuracy:

1) As the cell size increases, the number of samples increases
2) As the cell size increases, the centroid of the cell may be far from the actual location

For the relative power approach, the accuracy decreases as the cell size increases. However for *CellSense* the accuracy increases as the cell size increases. As the accuracy of *CellSense* is too low, the effect of improving the histogram here is more dominant than the effect of the estimated location being far from the centroid of the cell.

The figure also shows that the relative power approach improves the accuracy compared to the *CellSense* approach.

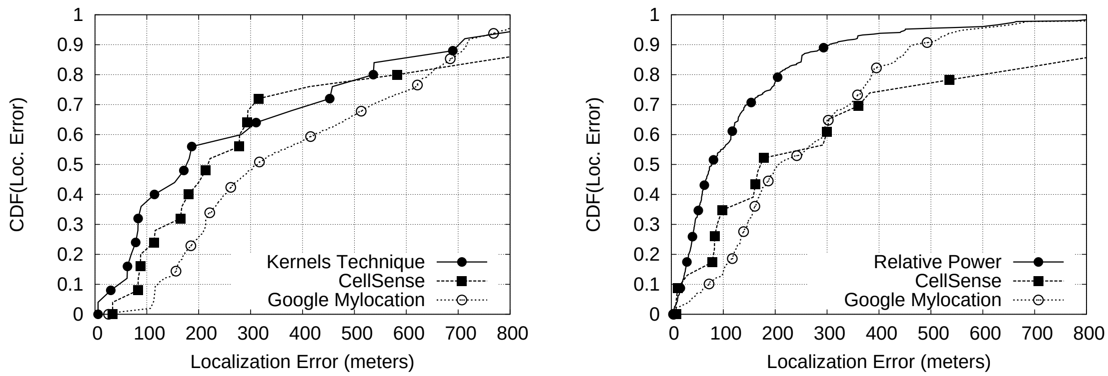
Figure 5 provides a comparison between the two approaches under the multiple cell tower configuration. The figure shows that comparable accuracies are provided from the both approaches. However for small grid cell lengths, the relative power approach provides better accuracy compared to the linear transformation with Gaussian kernel approach. As for small grid cells, there is no enough samples to compute the relation and smooth the histograms.

C. Comparison with Other Techniques

In this section, we evaluate our system's performance, in terms of localization error, compared to other GSM-based

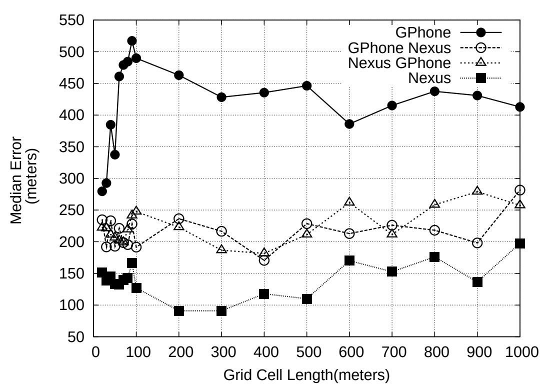
Algorithms	Google's MyLocation	*CellSense*	Relative Power	Linear Trans. with Gauss. kernel
Single cell tower configuration's median error(meters)	314.84 (73.35%)	221.78 (22.11%)	N/A	181.61 (**Reference**)
Mult. cell towers configuration's median error(meters)	208.28 (167.09%)	177.66 (127.84%)	77.97 (**Reference**)	N/A

TABLE I
COMPARISON BETWEEN DIFFERENT TECHNIQUES USING THE COLLECTED TESTBED. NUMBERS BETWEEN PARENTHESIS REPRESENT PERCENTAGE DEGRADATION COMPARED TO THE REFERENCE TECHNIQUE(THE BEST TECHNIQUE).

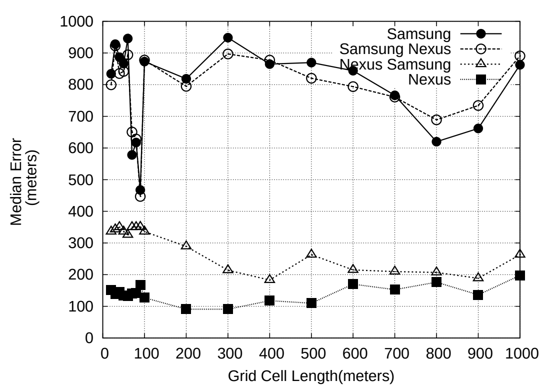


(a) Single cell tower configuration (b) Multiple cell tower configuration

Fig. 6. CDF's of distance error for different techniques under the two configurations. The tails of the CDF's are truncated for clarity of presentation.



(a) Nexus GPhone- median error



(b) Nexus Samsung- median error

Fig. 2. Effect of changing the grid cell length on the linear transformation with Gaussian Kernel approach median error.

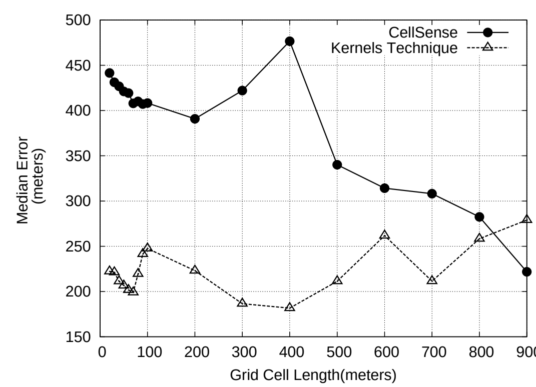


Fig. 3. Effect of changing the grid cell length on the linear transformation with Gaussian kernel approach median error compared to *CellSense*.

localization techniques described in Section II. The CDF of distance error for the different techniques is showed in Figure 6 under the two configurations. Table I summarizes the results. The table shows that our proposed techniques are better than any other technique with at least 127.84% in multiple cell tower configuration and at least 22.11% in single cell tower configuration. All techniques perform better in the multiple cell tower configuration than single cell tower configuration due to the more information used in the multiple cell tower configuration.

V. CONCLUSION

We discussed the cell phones heterogeneity problem and addressed the problem using the relative power approach and the linear transformation with Gaussian kernel approach. Also

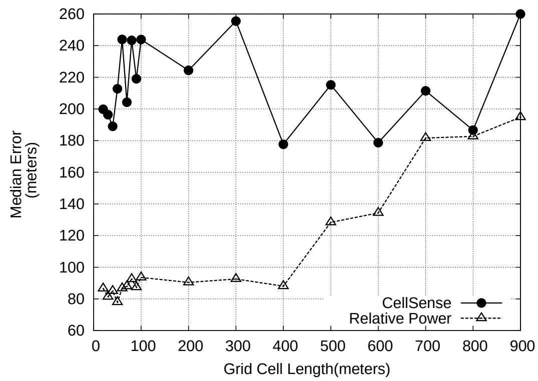


Fig. 4. Effect of changing the grid cell length on the relative power approach median error.

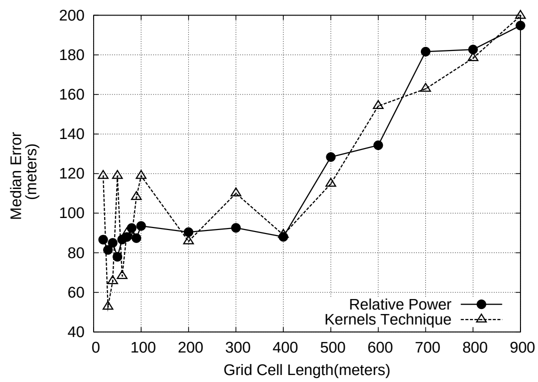


Fig. 5. Comparison between the relative power approach and the linear transformation with Gaussian kernel approach.

we presented the evaluation of the two approaches. The results show that using the relative power approach for the multiple cell towers problem will improve the accuracy compared to the current GSM localization techniques with at least 127.84% as the ratio of the readings between the same cell towers is almost the same for different cell phones. Also for single cell tower problem, using the linear transformation with Gaussian kernels also improves the accuracy with at least 22.11% in median error. The results also show that training the system with Nexus phones will be better than using the other phones.

ACKNOWLEDGMENT

This work is supported in part by a Google Research Award.

REFERENCES

[1] N. Kassem, A. E. Kosba, and M. Youssef, "RF-based Dehicle Detection and Speed Estimation," in *Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th*, 2012.

[2] A. Al-Husseiny and M. Youssef, "RF-based Traffic Detection and Identification," in *Vehicular Technology Conference (VTC Fall), 2012 IEEE 76th*, 2012.

[3] P. Enge and P. Misra, "Special issue on GPS: The Global Positioning System," *Proceedings of the IEEE*, 1999.

[4] S. Tekinay, "Special issue on Wireless Geolocation Systems and Services," *IEEE Communications Magazine*, 1998.

[5] Y.-C. Cheng, Y. Chawathe, A. LaMarca, and J. Krumm, "Accuracy characterization for metropolitan-scale Wi-Fi localization," in *MobiSys '05: Proceedings of the 3rd international conference on Mobile systems, applications, and services*, 2005.

[6] I. Smith, J. Tabert, A. Lamarca, Y. Chawathe, S. Consolvo, J. Hightower, J. Scott, T. Sohn, J. Howard, J. Hughes, F. Potter, P. Powledge, G. Borriello, and B. Schilit, "Place lab: Device positioning using radio beacons in the wild," in *Proceedings of the Third International Conference on Pervasive Computing*, 2005.

[7] M. Youssef and A. Agrawala, "On the optimality of WLAN location determination systems," in *In Communication Networks and Distributed Systems Modeling and Simulation Conference*, 2003.

[8] M. Youssef, M. Abdallah, and A. Agrawala, "Multivariate analysis for probabilistic WLAN location determination systems," in *Mobile and Ubiquitous Systems: Networking and Services, 2005. MobiQuitous 2005. The Second Annual International Conference on*. IEEE, 2005.

[9] A. E. Kosba, A. Saeed, and M. Youssef, "Rasid: A robust wlan device-free passive motion detection system," in *Pervasive Computing and Communications (PerCom), 2012 IEEE International Conference on*, 2012.

[10] I. Sabek and M. Youssef, "Multi-entity Device-Free WLAN Localization," in *IEEE Globecom*, 2012.

[11] Skyhook wireless, "http://www.skyhookwireless.com."

[12] R. R. C. Ionut Constandache and I. Rhee, "Towards mobile phone localization without war-driving," in *IEEE Infocom*, 2010.

[13] R. S. Andrew Offstad, Emmett Nicholas and R. R. Choudhury, "AAMPL: Accelerometer Augmented Mobile Phone Localization," in *ACM MELT Workshop (with Mobicom 2008)*, 2008.

[14] I. C. Martin Azizyan and R. R. Choudhury, "SurroundSense: Mobile Phone Localization via Ambience Fingerprinting," in *ACM MobiCom*, 2009.

[15] M. Youssef, M. A. Yosef, and M. N. El-Derini, "GAC: Energy-Efficient Hybrid GPS-Accelerometer-Compass GSM Localization," in *IEEE GLOBECOM*, 2010.

[16] M. Alzantot and M. Youssef, "Uptime: Ubiquitous pedestrian tracking using mobile phones," in *Wireless Communications and Networking Conference (WCNC), 2012 IEEE*, 2012.

[17] H. Wang, S. Sen, A. Elgohary, M. Farid, M. Youssef, and R. Choudhury, "No need to war-drive: Unsupervised indoor localization," in *Proceedings of the 10th international conference on Mobile systems, applications, and services*, 2012.

[18] Wikipedia, "Comparison of mobile phone standards — Wikipedia, the free encyclopedia," 2010. [Online]. Available: \url{http://en.wikipedia.org/wiki/Comparison_of_mobile_phone_standards}

[19] M. Y. Chen, T. Sohn, D. Chmelev, D. Haehnel, J. Hightower, J. Hughes, A. Lamarca, F. Potter, I. Smith, and A. Varshavsky, "Practical metropolitan-scale positioning for GSM phones," in *Proceedings of the Eighth International Conference on Ubiquitous Computing (UbiComp)*, 2006.

[20] A. Varshavsky, M. Y. Chen, E. de Lara, J. Froehlich, D. Haehnel, J. Hightower, A. LaMarca, F. Potter, T. Sohn, K. Tang, and I. Smith, "Are GSM phones THE solution for localization?" in *WMCSA '06: Proceedings of the Seventh IEEE Workshop on Mobile Computing Systems & Applications*, 2006.

[21] M. Ibrahim and M. Youssef, "CellSense: A Probabilistic RSSI-based GSM Positioning System," in *IEEE Globecom*, 2010.

[22] ——, "A Hidden Markov Model for Localization Using Low-End GSM Cell Phones," in *IEEE ICC*, 2011.

[23] ——, "CellSense: An Accurate Energy-Efficient GSM Positioning System," *Vehicular Technology, IEEE Transactions on*, 2012.

[24] M. B. Kjrgaard, "Indoor location fingerprinting with heterogeneous clients," *Pervasive and Mobile Computing*, 2011.

[25] J. geun Park, D. Curtis, S. Teller, and J. Ledlie, "Implications of device diversity for organic localization," in *IEEE INFOCOM*, 2011.

[26] A. Haeberlen, A. Rudys, E. Flannery, D. S. Wallach, A. M. Ladd, and L. E. Kavraki, "Practical robust localization over large-scale 802.11 wireless networks," in *in Proceedings of the 10th Annual International Conference on Mobile Computing and Networking (MOBICOM)*, 2004.

[27] Google Maps for Mobile, "http://www.google.com/mobile/maps/."