
Automated Wireless Localization Data Acquisition and Calibration with 6DOF Image Localization

Jonathan Fürst

NEC Labs Europe
69115 Heidelberg, Germany
jonathan.fuerst@neclab.eu

Gürkan Solmaz

NEC Labs Europe
69115 Heidelberg, Germany
gurkan.solmaz@neclab.eu

Kaifei Chen

UC Berkeley
Berkeley, CA 94720, USA
kaifei@cs.berkeley.edu

Ernö Kovacs

NEC Labs Europe
69115 Heidelberg, Germany
ernoe.kovacs@neclab.eu

Abstract

Radio frequency (RF) signals have been used extensively to enable (indoor) localization and proximity detection based on Received Signal Strength Indication (RSSI). However, localization systems often suffer from large data collection and calibration overhead, especially when being deployed in a new environment. RSSI fingerprinting based localization systems require the construction of a fingerprinting database. This localization data acquisition is a hindrance for the proliferation of localization systems in practice. Similarly, RSSI proximity applications require an RSSI calibration for the receiver hardware and the deployment environment. To overcome these problems, we propose the usage of visual 3D models which enable 6DOF localization and distance measurement with high accuracy. We then fuse this physical knowledge with RF data: (1) for automated acquisition of fingerprinting data and (2) easy calibration of a RF propagation model for proximity estimation.

Author Keywords

Localization; CV; RSSI Fingerprinting; Bluetooth

ACM Classification Keywords

500 [Human-centered computing]: Ubiquitous and mobile computing systems and tools

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Copyright held by the owner/author(s). Publication rights licensed to ACM.
UbiComp/ISWC'18 Adjunct, October 8–12, 2018, Singapore, Singapore
ACM 978-1-4503-5966-5/18/10.

<https://doi.org/10.1145/3267305.3274186>

Introduction

In mobile computing, radio frequency (RF) signals have been used extensively to enable (indoor) localization and proximity detection. RF signals fade in space, while the extend of signal fading is influenced by the physical characteristics of the space. As such, RSSI measured at the receiver is a proxy for transmitter-receiver distance. Specifically, two common approaches make use of this RF signal fading behavior: (1) RSSI fingerprinting based localization and (2) RSSI based proximity estimation [10].

RSSI fingerprinting faces the problem that RSSI values need to be mapped to specific locations. Because the RF signal characteristics heavily depend on environment and transmitter/receiver characteristics, an explicit/implicit data acquisition, site-survey is required to build the initial fingerprinting database. Likewise, for RSSI based proximity estimation, a calibration to the deployment environment and transmitter/receiver characteristics is required to obtain accurate distance estimations (details in the following section).

In this work, we propose to apply physical knowledge obtained from a 3D model (i.e., user location and transmitter-receiver distance) to address such problems. Specifically, we **apply image localization in a visual 3D model to automate wireless localization data acquisition and calibration**. Towards this goal, we first combine vision based localization with RSSI fingerprinting, by proposing a new technique for simple, explicit crowd-sourcing that (1) reduces the user effort, while (2) increasing the quality of the site survey data. We achieve this by combining the RSSI data collection with 6DOF (Degree of Freedom) image localization against a 3D model. Our image localization removes the need for a manual location tagging by users and provides high localization accuracy. Second, we show how distance metrics obtained from a 3D model can also be

used to automatically calibrate RSSI based proximity applications.

We perform an experimental evaluations for both application scenarios: (a) with an indoor localization system that uses Bluetooth Low Energy (BLE) fingerprinting and (b) with a BLE proximity application that determines beacon-phone distances based on RSSIs of BLE advertisements.

Note that pure image localization is accurate, but not as ubiquitous as our approach of vision supported RSSI based localization. Image localization is resource intensive [7], it can violate people's privacy requirements, and it requires explicit user involvement (i.e., camera actuation). On the other hand, RSSI based localization (proximity estimation) has advantages for privacy, latency, computational requirements and mobile battery consumption [14].

Our preliminary results show that image localization provides precise results (mean standard deviation at all locations: 5 cm) and accurate distance estimation (mean error: 33 cm) to automate localization data acquisition and calibration. For RSSI fingerprinting, we show how image localization can be used to construct a fingerprinting database that can exclusively be used for lightweight localization. Further, we show that we can improve RSSI based proximity estimation error from 3.3 m to 1.7 m by using image localization to calibrate an RF signal propagation model to the specific phone and environment. These results open new avenues for crowd-sourced localization data acquisition and calibration approaches that now may build on Augmented Reality (AR) based interactions, such as smart appliance control [7] to recruit participants.

Challenges of RF Based Localization

This section discusses the challenges that RF based localization, such as RSSI fingerprinting and proximity estimation face to motivate our approach.

RSSI Fingerprinting

RSSI fingerprinting usually follows two phases: (1) In the offline site survey phase, RSSI values of transmitters (i.e., WiFi access points, Bluetooth Beacons) are densely surveyed at known locations and stored in a database; (2) In the online phase, users of the localization system sample RSSI values with their devices and the system compares user-sampled RSSI values with the values in the survey database according to some similarity metric (e.g., Euclidean distance [4]).

It becomes clear that the site survey phase is a hindrance to a fast adoption of such systems. As such, several works try to reduce the involved efforts, mainly through explicit and implicit crowdsourcing. In explicit crowdsourcing, users are requested to collect RSSI data at unexplored locations, e.g., using their mobile devices [20]. Explicit crowdsourcing often suffers from poor quality of user input and difficulties to recruit participants [10]. In implicit crowdsourcing, RSSI data is collected on user's devices without manual user intervention, e.g., in [25, 21] the authors propose effortless RSSI fingerprinting based indoor localization by combining an indoor floor map with RSSI and inertial data sampling with a particle filter. However, Yang et al. [25] note that localization accuracy for implicit crowdsourcing systems are lower than a traditional site survey based systems, while system complexity increases because of the required signal processing (e.g., noise reduction with filtering). Summarized, RSSI fingerprinting faces the following challenges:

C1 Slow adoption due to offline phase.

C2 Inaccurate user input for explicit crowdsourcing.

Proximity Estimation

RSSI based proximity estimation relies on signal fading properties to estimate proximity based on received signal strength. However, this proximity estimation is heavily influenced by environmental factors, like obstacles, multipath effects, channel fading or wireless interference. Thus, formulae for proximity estimation are usually based on both physical RF signal propagation (e.g., Friis transmission equation) and heuristics obtained experimentally to fit different physical environments and hardware.

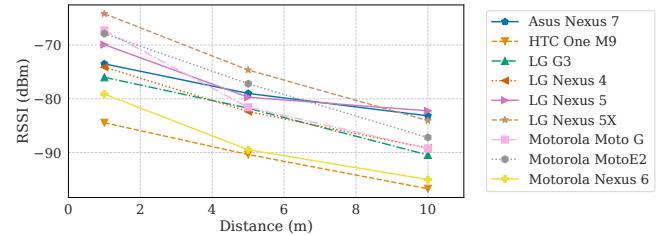


Figure 1: Hardware heterogeneity problem. Different smartphone models exhibit greatly different RSSI measurements.

Figure 1 shows the variances in the RSSI values in the lab environment for different distances and from different phone models. RSSI varies by 20 dBm between models at the same distance. As such, RSSI based proximity estimation faces the following main challenge:

C3 The high hardware and environment dependency requires a calibration of the signal propagation model to the deployment environment.

In the following, we show how 6DOF image localization can help localization systems to overcome these challenges.

Localization Data Acquisition and Calibration with 6DOF Image Localization

Figure 2 depicts the general working of our approach. During an offline phase, a visual 3D model is constructed using a RGB-D camera. Note, that such model might also be constructed through crowd-sourcing as in [9] to reduce efforts. Mobile phone users collect RSSI values from the surrounding wireless transmitters (e.g., WiFi access points or Bluetooth beacons). The users provide RSSI values and the images taken from their mobile phone camera. We then localize the images in the visual 3D model, which gives us the mobile phone location of the user in the physical space. For fingerprinting, a measurement contains the RSSI values of the transmitters by the mobile device, the mobile device model, and the location of the user. The measurement is then saved to the database. After a measurement has been stored, subsequent users of our system can use a purely RSSI fingerprinting based localization approach that does not require any more images to be taken. For RSSI proximity estimation, we obtain transmitter-receiver distances from the 3D model to calibrate the RF propagation model to the phone hardware and environment.

6DOF Image Localization

Vision based localization localizes camera images in an existing 3D model. To infer the location of a query image, it localizes the query image in the 3D model using the most likely location (3DOF) and angle (3DOF) [24].

To localize an image, we build a 3D model, which is essentially a set of 3D points (a.k.a., a point cloud). Some points in the 3D model are associated with a visual feature (e.g., SIFT [16], SURF [5]) that describes the appearance around it. We find the most similar visual features in the 3D model from those extracted in the image, and compute the geometry relation among them to get the 6DOF the image [24].

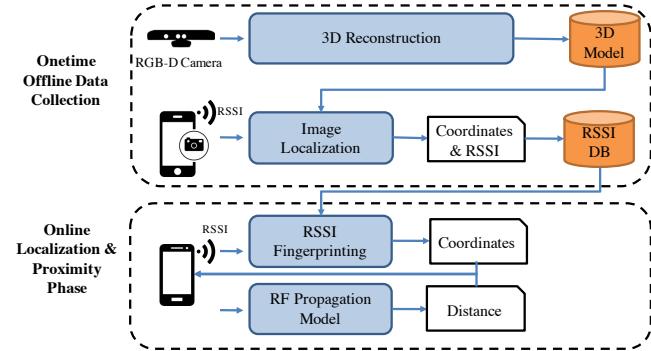


Figure 2: 3D Model Supported RSSI Fingerprinting and Proximity Estimation. We construct a 3D model using a RGB-D camera and subsequently use this model for RF localization data acquisition and calibration by combining image localization with RSSI sampling on end-user devices.

In this paper, we use RTAB-Map [12] to build the 3D model, and use SnapLink [7] for image localization. As an example, Figure 3 shows a 3D model and two images localized at the location and orientation where they were taken in the real world.

Interpretation of 3D Model and 6DOF Locations

Image localization allows us to estimate location in the form of 3D coordinates in a 3D model. However, applications that use the localization services still need to know how to interpret these coordinates in the context of the corresponding 3D model. Fortunately, 3D point clouds are easy to label by humans [11], and many algorithms can be used to understand semantics in 3D point cloud automatically. For example, we can perform scene understanding on the point cloud [3], such as detecting doors, obstacles, and pathway for indoor navigation. We can also perform 3D registration [23] to combine room 3D models together to form a 3D model of the building, which allows localization applications to operate at the scale of a building.

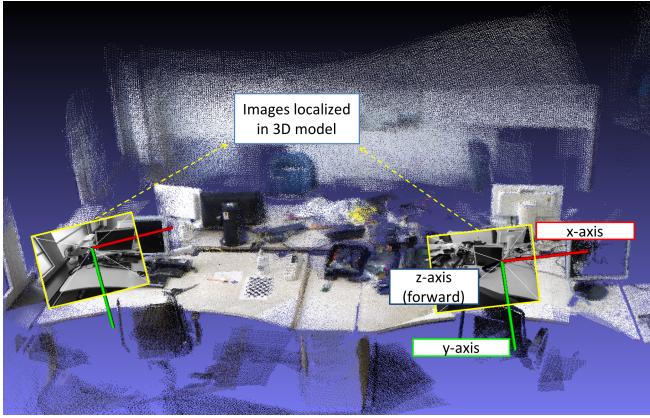


Figure 3: An example from our deployment, showing two images localized in a 3D model. Each image has its own coordinate system, which is annotated in the figure.

Prototype and Experiment Design

We implement a prototype of our system and deploy it in a shared office space with two rooms (see Figure 4). In total we deploy 5 Estimote Bluetooth beacons [1] in the corners of the rooms at around 1.5 m height (B1-B5). We then collect a visual 3D model with a Microsoft Kinect, using RTAB-Map [12]. We use a Nexus 5x Android phone to collect pictures and RSSI samples at six locations in the rooms (L1-L6). Further, we re-collect RSSI samples for 7 days to evaluate the stability of our approach.

For RSSI fingerprinting, we implement a simple Euclidean distance based approach following the system outlined in [6]. Specifically, to localize a device, we match the sampled RSSI values to the closest set of RSSI values stored in the fingerprinting database. We use the median of sampled RSSI values to remove outliers. Note, because our system is intended to be used with off-the-shelf mobile devices, we use RSSI instead of more stable channel state information

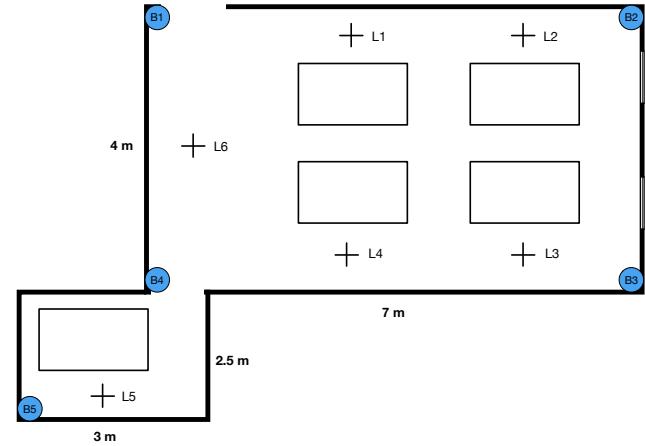


Figure 4: Experimental setup for fingerprinting experiments. We deploy 5 beacons (B1-B5) and collect RSSI fingerprints for 6 locations (L1-L6) in two rooms.

(CSI) [26] for fingerprinting.

For RSSI proximity estimation, we use the commonly used log-distance path loss model [8] given by the following equation,

$$d = 10^{\left(\frac{P_{tx} - P_{rx}}{10 * \gamma}\right)} \quad (1)$$

where d is the estimated distance between phone and transmitter, γ is the path loss component, which determines the rate of decay of the RSSI values when moving away from the transmitter, P_{rx} is the received signal strength, and P_{tx} is the signal strength at 1 m. As such, γ captures the effects of the environment (e.g., obstacles in line of sight result in faster decay), whereas P_{tx} captures hardware effects on transmitter and receiver side (SoC implementation, amplifier and antenna). We then calibrate γ and P_{tx} through image localization in the 3D model. I.e., we obtain two distances d_0 and d_1 from the 3D model and

solve Equation 1 for both unknown variables. Note that due to variations in signal propagation in indoor environments (multipath, obstacles), there exists no single solution that perfectly solves Equation 1 for multiple distances. We therefore minimize the mean square error for multiple distances as shown in Equation 2 and 3.

$$err_i = 10^{\left(\frac{P_{tx} - P_{rx}}{10 \gamma_i}\right)} - d_i \quad (2)$$

$$err_{total} = \sum_{i=0}^n err_i^2 \quad (3)$$

Evaluation

To evaluate our approach, we first perform experiments to gain an understanding for the localization accuracy that we can obtain with the visual 3D model (offline phase). We then separately evaluate the resulting RSSI fingerprinting based localization and RSSI based proximity estimation (online phase).

Visual Image Localization

To estimate the localization error using the visual 3D model, we collect 50 images at the same location but having different angles. We do this for all six locations (L1-L6) marked in Figure 4. We then use SnapLink [7] to localize the images in the 3D model. Figure 5 shows the result of this experiment by depicting x , y and z coordinates for each location. Note that this data is based on multiple 3D models. Locations with similar coordinates are not necessarily physically close to each other. Out of 313 images, only 16 (5%) fail to be localized due to noise in the images and 3D model and the thresholds we set in the system, which is not uncommon in state-of-the-art image localization systems [24].

For 95% of the images, image localization achieves highly accurate results (we empirically verify the results in the 3D model as shown in Figure 3). These successfully localized

images exhibit only a small standard deviation (SD): $x_{SD} = 4.3$ cm, $y_{SD} = 4.7$ cm, $z_{SD} = 4.6$ cm. Overall, our results show that image localization provides high accuracy and it can be used as the near-ground-truth input for RF based localization systems.

RSSI Fingerprinting Localization

We collect RSSI samples at locations (L1-L6) over the time frame of a week (10 different dates and times). We then use the initial collection of RSSI data as training data to build a fingerprinting database where we store the median RSSI value for each location. Then, we test our system with the remaining 9 data sets. Figure 6 depicts the cumulative distribution function (CDF) of the localization errors for all locations together, while Figure 7 shows results for each location separately. We can localize most collected samples correctly (75%). As expected, due to the close proximity of some locations (e.g., < 2 m for L3 and L4) and changes in time in RF signal propagation (e.g., due to human activity, WiFi traffic), some localization attempts result in wrong locations. Likewise, due to specifics of the environment (see Figure 4), L1 and L5 provide near perfect localization results.

RSSI based Proximity Estimation

To evaluate vision based RSSI calibration, we place a single BLE beacon in a larger conference room at the height of 1.5 m and then build a 3D model using RTAB-Map [12]. We manually select the beacon location in the 3D model using the labeling tool developed in [7]. We collect RSSI data at five transmitter-receiver distances (1 m, 3 m, 5 m, 7 m and 10 m) and with multiple phone models, together with images captured by the smartphone camera. We use the 3D model to localize the captured images and calculate the beacon-camera distances. Figure 9 depicts the visual distance error for different camera-beacon distances based on 20 images

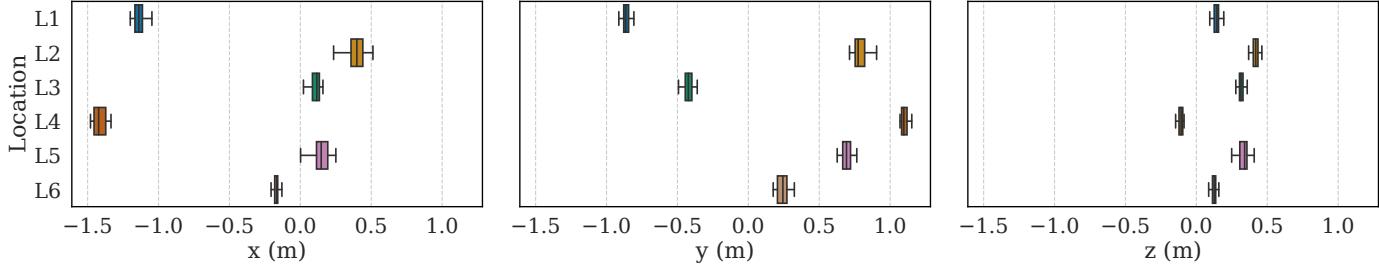


Figure 5: Visual localization results for the successfully localized images at different locations (L1-L5). x , y and z are coordinates in 3D space.

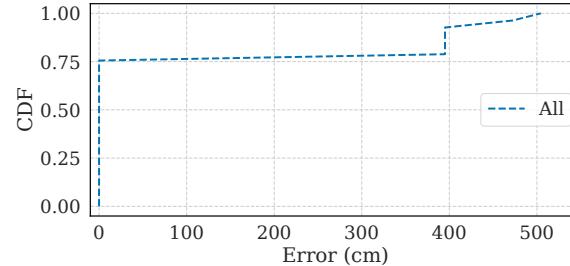


Figure 6: CDF of Overall Localization Errors.

per location. Note that we omit results for 1 m, because image localization does not work reliable due to our specific deployment environment (the beacon is placed on the wall and images captured at 1 m distance do not contain enough SURF [5] features for reliable localization). Overall, we derive accurate distance estimations from the 3D model for the remaining locations (mean error: 33 cm).

We then use the collected RSSI data at two locations ($P_{0_{rx}}$, $P_{1_{rx}}$) and the distances d_0 and d_1 obtained from the 3D model for these locations to minimize the overall error in Equation 3 for a (γ, P_{tx}) pair. We use the Nelder-Mead algorithm [17] to solve this multidimensional unconstrained optimization problem. We use a single set of RSSI val-

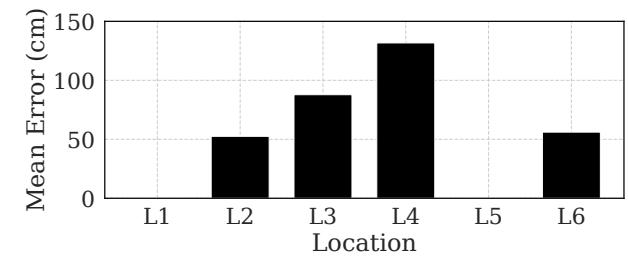


Figure 7: Mean Localization Error for each location.

ues for this calibration phase and test our approach with solely RSSI data collected in a consecutive sampling iteration. Figure 8 depicts (top), the generic proximity estimation results (using the P_{tx} , as advertised by the beacon (-74 dBm), and an environment factor $\gamma = 1.7$, based on empirical data for non-residential buildings published in [22]) and (bottom), the results after calibration to phone model and environment through image localization in the 3D model using $d_0 = 7 \text{ m}$ and $d_1 = 10 \text{ m}$. The generic model fails to capture phone model and environment characteristics (root mean square error: 3.3 m), while the calibrated model captures device and environment characteristics better (root mean square error: 1.7 m). However, 1.7 m

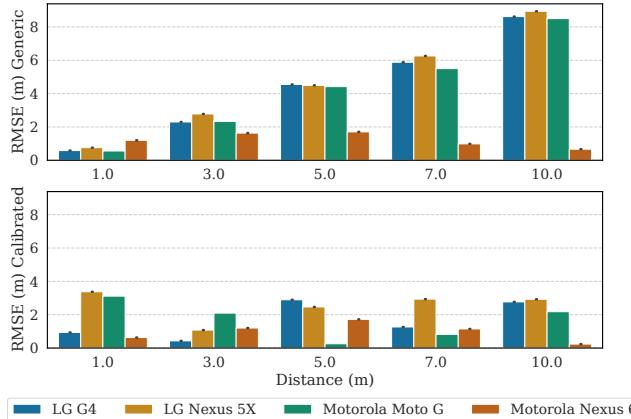


Figure 8: RSSI Based Proximity Estimation using a Generic (top) and a Calibrated (bottom) RF Signal Propagation Model.

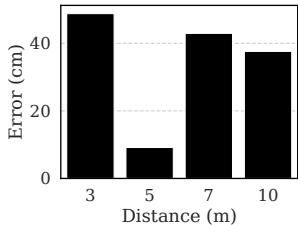


Figure 9: Distance Estimation Error from Image Localization. Images captured at 5 m exhibit the smallest error, because our deployment environment contains a projector-table at this location, which provided many additional visual features.

Discussion and Future Work

Our evaluation show the potential of using image localization to support wireless localization data acquisition and calibration. We can explore this combination in several directions. Because the 3D model contains the full geometry (i.e., room sizes, walls, furniture etc.), it is possible to utilize this information to build a more accurate RF propagation model for localization applications. We consider employing crowdsourcing, together with the structure from motion (SfM) to build the model from pictures without depth information as in [2]. This would remove the current one-

time effort of 3D reconstruction using a RGB-D camera and allow us to update the model to environment changes as in [9]. However, it would also result in a sparser point cloud. Another option is to deploy fiducial markers (e.g., AprilTag [19]) that allow a camera to directly localize itself in space without requiring a 3D model.

Related Work

Liu et al. [15] present a visual approach to fingerprint map creation, combining WiFi data with video frames and inertial readings. Levchev et al. [13] design an end-to-end system for simultaneous WiFi fingerprinting and mapping using a customized sensor selection (camera, laser scanners, IMU). They show that fusing camera with RF data improves localization accuracy in some cases. Noreikis et al. [18] show that combining 3D model based image localization with inertial smartphone sensors greatly improves energy consumption for context-aware localization task scheduling. Our approach assumes the existence of an indoor 3D model (or fiducial markers) to be used as reference points to calibrate RSSI measurements and the resulting localization applications.

Conclusion

In this paper, we proposed to use image localization in 3D model to automate data collection and calibration in RSSI-based indoor localization systems. Image localization cannot be used in daily activities in shared environment because of privacy concern and its heavy energy overhead. However, using it can eliminate human efforts in the data acquisition and calibration phases of other indoor localization systems. Our evaluation showed that image localization can be performed fully automatically and can provide accurate ground truth locations. We believe our work will be another stepping stone for large deployment of various indoor localization algorithms in the future.

REFERENCES

1. Estimote. <https://estimote.com/>. (2018 ????).
2. Sameer Agarwal, Noah Snavely, Ian Simon, Steven M Seitz, and Richard Szeliski. 2009. Building rome in a day. In *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 72–79.
3. Iro Armeni, Ozan Sener, Amir R Zamir, Helen Jiang, Ioannis Brilakis, Martin Fischer, and Silvio Savarese. 2016. 3d semantic parsing of large-scale indoor spaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1534–1543.
4. Paramvir Bahl, Venkata N Padmanabhan, and Anand Balachandran. 2000. Enhancements to the RADAR user location and tracking system. *Microsoft Research* 2, MSR-TR-2000-12 (2000), 775–784.
5. Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. 2006. Surf: Speeded up robust features. In *European conference on computer vision*. Springer, 404–417.
6. Philipp Bolliger. 2008. Redpin-adaptive, zero-configuration indoor localization through user collaboration. In *Proceedings of the first ACM international workshop on Mobile entity localization and tracking in GPS-less environments*. ACM, 55–60.
7. Kaifei Chen, Jonathan Fürst, John Kolb, Hyung-Sin Kim, Xin Jin, David E Culler, and Randy H Katz. 2018. SnapLink: Fast and Accurate Vision-Based Appliance Control in Large Commercial Buildings. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 129.
8. Krishna Chintalapudi, Anand Padmanabha Iyer, and Venkata N Padmanabhan. 2010. Indoor localization without the pain. In *Proceedings of the sixteenth annual international conference on Mobile computing and networking*. ACM, 173–184.
9. Jiang Dong, Yu Xiao, Marius Noreikis, Zhonghong Ou, and Antti Ylä-Jääski. 2015. imoon: Using smartphones for image-based indoor navigation. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*. ACM, 85–97.
10. Suining He and S-H Gary Chan. 2016. Wi-Fi fingerprint-based indoor positioning: Recent advances and comparisons. *IEEE Communications Surveys & Tutorials* 18, 1 (2016), 466–490.
11. Hema S Koppula, Abhishek Anand, Thorsten Joachims, and Ashutosh Saxena. 2011. Semantic labeling of 3d point clouds for indoor scenes. In *Advances in neural information processing systems*. 244–252.
12. Mathieu Labbe and Francois Michaud. 2013. Appearance-based loop closure detection for online large-scale and long-term operation. *Robotics, IEEE Transactions on* 29, 3 (2013), 734–745.
13. Plamen Levchev, Michael N Krishnan, Chaoran Yu, Joseph Menke, and Avideh Zakhor. 2014. Simultaneous fingerprinting and mapping for multimodal image and WiFi indoor positioning. In *Indoor Positioning and Indoor Navigation (IPIN), 2014 International Conference on*. IEEE, 442–450.
14. Kaisen Lin, Aman Kansal, Dimitrios Lymberopoulos, and Feng Zhao. 2010. Energy-accuracy trade-off for continuous mobile device location. In *Proceedings of the 8th international conference on Mobile systems, applications, and services*. ACM, 285–298.
15. Tao Liu, Xing Zhang, Qingquan Li, and Zhixiang Fang. 2017. A Visual-Based Approach for Indoor Radio Map

- Construction Using Smartphones. *Sensors* 17, 8 (2017), 1790.
16. David G Lowe. 2004. Distinctive image features from scale-invariant keypoints. *International journal of computer vision* 60, 2 (2004), 91–110.
 17. John A Nelder and Roger Mead. 1965. A simplex method for function minimization. *The computer journal* 7, 4 (1965), 308–313.
 18. Marius Noreikis, Yu Xiao, and Antti Ylä-Jääski. 2017. SeeNav: Seamless and Energy-Efficient Indoor Navigation using Augmented Reality. In *Proceedings of the on Thematic Workshops of ACM Multimedia 2017*. ACM, 186–193.
 19. Edwin Olson. 2011. AprilTag: A robust and flexible visual fiducial system. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 3400–3407.
 20. Jun-geun Park, Ben Charrow, Dorothy Curtis, Jonathan Battat, Einat Minkov, Jamey Hicks, Seth Teller, and Jonathan Ledlie. 2010. Growing an organic indoor location system. In *Proceedings of the 8th international conference on Mobile systems, applications, and services*. ACM, 271–284.
 21. A. Rai, K.K. Chintalapudi, V.N. Padmanabhan, and R. Sen. 2012. Zee: Zero-effort crowdsourcing for indoor localization. In *Proceedings of the 18th annual international conference on Mobile computing and networking*. ACM, 293–304.
 22. Theodore S Rappaport and others. 1996. *Wireless communications: principles and practice*. Vol. 2. prentice hall PTR New Jersey.
 23. Radu Bogdan Rusu, Nico Blodow, and Michael Beetz. 2009. Fast point feature histograms (FPFH) for 3D registration. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*. Citeseer, 3212–3217.
 24. Torsten Sattler, Bastian Leibe, and Leif Kobbelt. 2011. Fast image-based localization using direct 2d-to-3d matching. In *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 667–674.
 25. Zheng Yang, Chenshu Wu, and Yunhao Liu. 2012. Locating in fingerprint space: wireless indoor localization with little human intervention. In *Proceedings of the 18th annual international conference on Mobile computing and networking*. ACM, 269–280.
 26. Zheng Yang, Zimu Zhou, and Yunhao Liu. 2013. From RSSI to CSI: Indoor localization via channel response. *ACM Computing Surveys (CSUR)* 46, 2 (2013), 25.