Simultaneous Fingerprinting and Mapping for Multimodal Image and WiFi Indoor Positioning

Plamen Levchev, Michael N. Krishnan, Chaoran Yu, Joseph Menke, and Avideh Zakhor
Department of EECS, U.C. Berkeley

***Abstract*—In this paper, we propose an end-to-end system which can be used to simultaneously generate (a) 3D models and associated 2D floor plans and (b) multiple sensor e.g. WiFi and imagery signature databases for the large scale indoor environments in a fast, automated, scalable way. We demonstrate ways of recovering the position of a user carrying a mobile device equipped with a camera and WiFi sensor in an indoor environment. The acquisition system consists of a man portable backpack of sensors carried by an operator inside buildings walking at normal speeds. The sensor suite consists of laser scanners, cameras and an IMU. Particle filtering algorithms are used to recover 2D and 3D path of the operator, a 3D point cloud, the 2D floor plan, and 3D models of the environment. The same walkthrough that produces 2D maps also generates multi-modal sensor databases, in our case WiFi and imagery. The resulting WiFi database is generated much more rapidly than existing systems due to continuous, rather than stop-and-go or crowd-sourced WiFi signature acquisition. We also use particle filtering algorithms in an Android application to combine inertial sensors on the mobile device, with 2D maps and WiFi and image sensor databases to localize the user. Experimental for the second floor of the electrical engineering building at UC Berkeley campus show that our system achieves an average localization error of under 2m.**

I. Introduction and Related Work

In recent years, indoor localization has received a great deal of attention among researchers. On one hand, it has a number of important applications such as location-aware intelligent shopping assistant and indoor real-time navigation. On the other hand, it is a technically challenging problem due to the fact that most buildings virtually block GPS signals. Hence alternate localization approaches are needed for the indoor environment.

The prevalence of WiFi infrastructure inside most buildings provides a natural starting point for this problem. A well-known approach is to construct a database of WiFi Received Signal Strength Indicator (RSSI) fingerprints for the building. The RSSI fingerprint for each location is a vector of decibel values where each entry corresponds to the WiFi signal strength of a particular access point detected at that location. As the client-side application queries the database with an RSSI measurement, algorithms such as Redpin [2] or variants of $k-$NN [9, 22] are used to retrieve the location whose fingerprint is closest to the querying fingerprint. A major advantage of this method is its cost-effectiveness, since hardware infrastructure is already in place in most indoor commercial and residential buildings. Furthermore, practically all mobile phones and consumer electronic devices have WiFi scanning capability. One disadvantage of this approach is that the fingerprinting process is slow and cumbersome and requires a map. Another disadvantage is that the location dependency of RSSI is not reliable and is subject to interference. Even though room-level accuracy e.g. approximately 5 to 10m has been demonstrated, the method does not achieve meter or sub-meter level accuracy [12].

Another approach that could be used for localization is inertial dead reckoning using a smartphone's on-board accelerometer, gyroscope and magnetometer. Nowadays consumer electronic devices are equipped with increasingly more accurate sensors that are capable of sampling at fast data rates. Utilizing these sensors, users' speed and orientation of movement can be estimated and their path can therefore be tracked in real time. Magnetometer can be used to get users' orientation, but it suffers from interference from various indoor magnetic anomalies such as steel cabinets. Integrating raw accelerometer and gyroscope readings provides displacement information but integration introduces significant drift error [25]. To estimate users' movement, a commonly adopted method is to first detect their steps, and then estimate the corresponding step lengths [8, 10, 21]. The inertial sensor-based approach is different from the WiFi-based one in that it only estimates relative change in position, whereas WiFi measurements provide position estimates in the global coordinate frame. In an attempt to combine strengths of above two approaches and combat their respective weaknesses, it is possible to use a probabilistic technique, e.g. particle filtering, to derive an integrated location estimate [13].

Recently an image-based indoor localization scheme [7, 14, 15] has been proposed for mobile devices with cameras. Firstly, a database of images is constructed via a man portable ambulatory backpack of sensors, and then images taken by the client mobile device are used to retrieve those images with most number of matching features in the database. The Shift Invariant Feature Transform (SIFT) [18] finds distinct features contained in images that are invariant to uniform scaling and partially to affine distortion. Even though this method generally achieves higher accuracy than WiFi RSSI matching, its performance is degraded when the query image has few distinguishing features, or when the pictures are of low quality due to out-of-focus and/or motion blur. In the first case, lack of distinct features would adversely affect the ranking of retrieved database images. In the second case, it is difficult to detect and extract features from blurry images. Liang et al.

[14, 15] have shown that in the static case, where blur-free images are captured with a stationary camera, image-based localization could achieve an accuracy of 2 meters over 80% of the time, and 4 meters over 90% of the time.

In this paper, we propose an end-to-end system which can be used to (a) construct 3D models of large-scale indoor environments in a fast, automated, scalable way; (b) construct multiple sensor, e.g. image and WiFi, signature databases for the same environment; (c) use the reference databases in part (b) to recover position of users carrying a mobile device equipped with a camera and WiFi sensor. The system consists of a man-portable backpack equipped with sensors that is carried by an operator inside buildings walking at normal speeds. The sensor suite consists of laser scanners, cameras and an IMU [3, 4, 11, 17, 20]. Particle filtering algorithms are used to recover 2D and 3D path of the operator, the 3D point cloud, the 2D floor plan, and 3D models of the environment. The same walkthrough that results in 2D maps, is used to generate multi-modal sensor databases, in our case WiFi and imagery. We also use particle filtering algorithms in an Android application to combine inertial sensors on the mobile device, WiFi and image sensor databases, and a 2D map of the environment to localize the user.

Our proposed simultaneous fingerprinting and mapping system has two main advantages over existing indoor fingerprinting schemes. First, it does not require the knowledge of the map of the environment a priori in order to collect the fingerprints. Second, it requires significantly smaller set up time compared to existing schemes that fall into two categories: The original idea of [1] require tens of scans at each location with 1-3 meters between points; other techniques, starting with [2] obtain the training data from crowd-sourcing. Both of these methods require a significant amount of time to be spent on data collection and can be problematic if the environment changes. In [23], Quigley et. al. map the building interior with a robot, then make a second pass with an ambulatory backpack to collect fingerprint data with the specific sensors to be used by the client. Our method generates both the map and the fingerprint map simulatneously so that they can be used by any client.

The remainder of the paper is organized as follows: in Section II we present the data collection system, in Section III we discuss the client system, experimental results are presented in Section IV, and the paper is concluded in Section V.

II. DATA COLLECTION SYSTEM OVERVIEW

In previous work, we have demonstrated that an ambulatory backpack, shown in Figure 1, equipped with two cameras, one orientation sensor, and five 2D laser range sensors can be used to generate a 2D floor plan and 3D model of an indoor environment with a single walkthrough of the building with 10cm average 3D position accuracy [3, 4, 11, 17, 20]. An example of the recovered 2D floor plan for the second floor of the Electrical Engineering building, Cory Hall, at

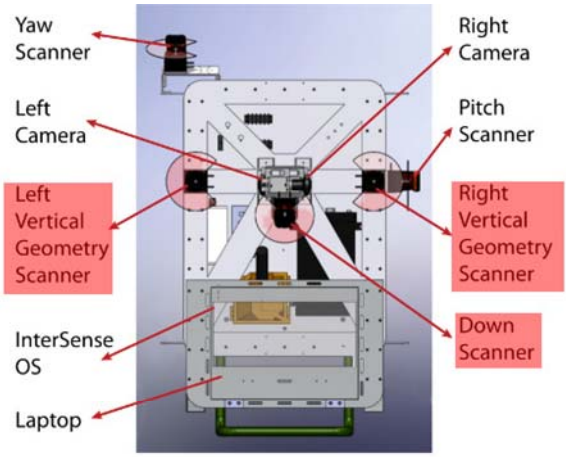
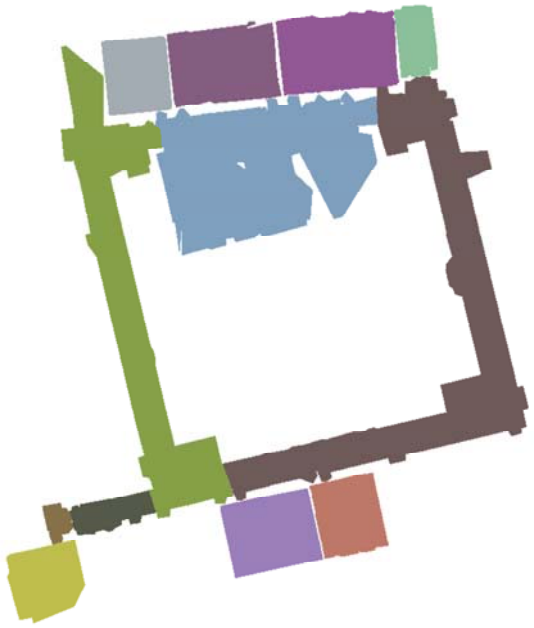


Fig. 1. Ambulatory backpack equipped with various sensors.

U.C. Berkeley is shown in Figure 2. Our system builds on top of that work by adding WiFi scanning capability to the backpack. Specifically, we use three USB AirPcap NX cards from Riverbed Technology to capture 802.11 beacons. As the human operator walks with the backpack, the Multi-Channel Aggregator captures WiFi beacons on all supported channels in the 2.4GHz and 5GHz bands, recording the timestamp and RSSI of each beacon. The operator walks at normal speeds, i.e. at about 0.7 m/s, to generate a dense WiFi signature and image database. Thus, with one walkthrough it is possible to generate both a 2D map of the environment and the associated WiFi and image databases for indoor localization.



(c)

Fig. 2. 2D floor plan and room labels for second floor of Cory Hall.

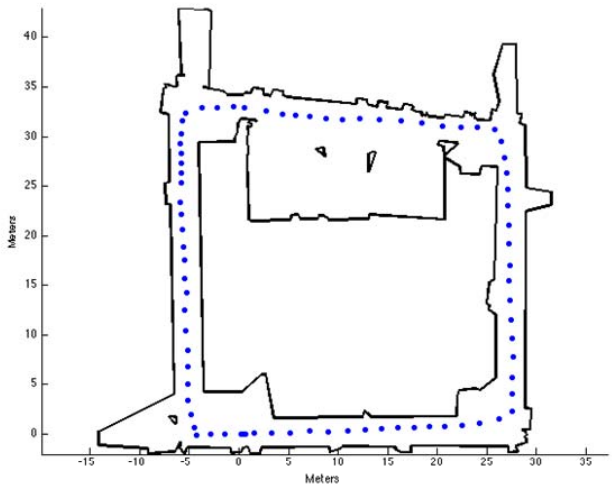


Fig. 3. Spatial distribution of data points in our WiFi database, where each circle represents a WiFi signature at the corresponding location.

The WiFi database consists of points in our local coordinate system. Each point has an associated fingerprint. A fingerprint consists of a variable-sized set of pairs of the form (AP MAC address, RSSI). This fingerprint is associated with a single location or point in the coordinate system. The MAC addresses and corresponding RSSI values are collected by the WiFi cards over time as they sweep through all the channels. As a complete sweep of all channels may take several seconds, depending on channel-switching delay, the set of MAC addresses and corresponding RSSI values does not precisely correspond to a single location. Since the standard 802.11 beacon period is 102 milliseconds, it is possible to use 3 AirPcap cards to scan each channel for 102 ms before switching to a new channel. Assuming there are a total of 30 channels to scan in the 2.4 and 5 GHz bands, this results in 1 second end-to-end scanning of all 30 channels; hence at walking speed of 0.7m/s, the spacing between fingerprint data points in our WiFi database is approximately 0.7m. It is also possible to scan each channel for a period of 204 or even 306 ms so as to obtain 2 or 3 beacons from each AP on the current channel before switching to the next one in order to both minimize the overhead associated with switching between channels and reduce the impact of single noisy observations. We can use the median signal strength of the beacons for each entry in our WiFi signature database to reduce the noise. However, this comes at the expense of lower spatial density of the fingerprints in our database. Specifically, assuming the same walking distance, the spatial density of the 204 (306) ms database is half (a third) of that of the 102 ms database. An example of 306 ms database superimposed on the 2D floor plan of the 2nd floor of Cory Hall is shown in Fig 3. As shown, for a donut shaped square corridor of dimension 35 meters, there are 76 fingerprints in our database. We use this database for the experiments in this paper.

We have experimentally verified that a normal walking speed of 0.7 m/sec does not result in significant blur in database images which would adversely affect the performance of our image-based localization method [14–16, 18]. Images taken with two fisheye cameras on the backpack at 4 frames per second are post-processed using algorithm in [14–16] to construct a database containing processed images, their corresponding SIFT features, pose information and depth maps.

To summarize, with one walkthrough at normal walking speed with the backpack shown in Figure 1 we generate 2D maps and multiple sensor modality e.g. WiFi and image databases for an indoor environment in order to enable indoor positioning.

III. MULTIMODAL CLIENT-SIDE POSITIONING

A. System design

We use particle filtering to fuse inertial sensors, WiFi readings and images to localize the mobile device or the AR glass of the user. Particle filtering is particularly attractive for our application because it can incorporate observations from WiFi and image localization estimates arriving at different rates. Since the localization is done in 2 dimensions, each particle is a 3-dimensional vector consisting of the x and y position and an orientation.

The client device continually scans the WiFi channels and performs step detection using the accelerometer. Images are taken in one of two ways: “on-demand”, whereby the user takes pictures of feature rich areas, and continuously, whereby every time a positioning result is returned from the image server, another image is captured. The first method requires more user intervention, but yields a greater localization success rate. The second method is more automatic, but can result in many featureless images being sent to the server. In order to minimize the impact of these featureless images, the server first checks if the number of SIFT features is above 150, and if it not, it returns immediately.

The particle filter to combine these observations with the step detection is performed on the device with 500 particles. The observations and step estimates are also timestamped and recorded to allow for offline processing and thus controlled comparison between various methods. In the following subsections, components of the particle filtering are discussed.

B. Step Model

Steps are detected using the accelerometer on the mobile device. We use the algorithms in [24, 25] to detect steps and estimate their lengths, and estimate the direction based on change in orientation reported by the device since the previous step. It is assumed that the device stays oriented at a fixed known offset from the direction of motion. For each particle, we add a random noise to this step estimation and apply the appropriate rotation and translation to update the particle. The noise model comes from empirical observations

of the variability in the estimated step lengths and orientations. If the step results in a particle crossing a wall, the particle cannot represent the true location and is thus eliminated. If more than 90% of particles are eliminated, the particles will be resampled, and if all particles are eliminated, a new set of particles is generated in a wide Gaussian distribution around the last estimated location.

C. Observation Update: WiFi and Images

The observations consist of two types: WiFi and image. The WiFi observations occur periodically as the device completes scans of the WiFi channels. The image observations occur less regularly, when the device is able to capture an image with a sufficient amount of features to perform the localization described in [13]. At the time an observation is available, importance resampling is performed.

For WiFi observations, importance factor is computed via the Redpin algorithm [2]. For each particle, the fingerprint corresponding to the closest point in the database is compared to the current observed fingerprint. The weight is proportional to the Redpin score, which compares the similarity of 2 fingerprints by considering the set of command and distinct access points observed as well as the signal strengths.

For image observations, the importance factor is a function of the distance between the particle and the result of the image localization. The probability distribution of the location is assumed to be Gaussian with variance depending on the confidence returned by the image localization [15]. The confidence, c, is a number between 0 and 1, where 1 represents 100% confidence, and the standard deviation of the Gaussian is taken to be $1 - c$.

Since the observations require processing on a server, the particles will have moved between the time of the actual observation and the time the result is available. In order to combat this delay, each particle maintains an additional state which records the position at the time of the outstanding observation. The importance factor is then computed based on the stored location rather than the current location.

D. Location Report

After fusing multiple sources of information, the particle filter returns a single location estimate, which is the centroid of all the living particles. In computing the average position of the particle cloud, it is possible for the centroid to lie outside of the valid region e.g. outside of the walls of floor plan. In that case, we use binary search along the past trajectory to find the closest valid point as the answer.

To determine whether a point is in valid regions we use ray tracing as shown in Figure 4; specifically we draw a line from an arbitrary point outside of the polygon to the point in question, check whether the number of edges crossed is odd, and make sure that if the crossing is a vertex, it is counted only once. Odd number of crossed edges/vertices would indicate a valid point. This result is known in topology as Jordan curve theorem [6].

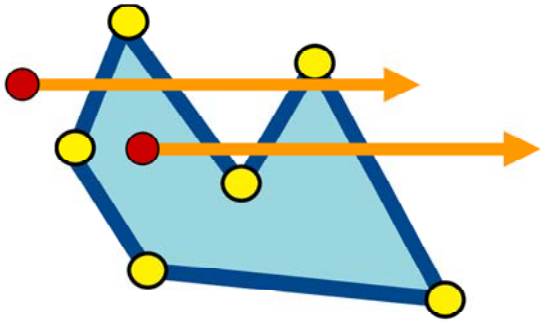


Fig. 4. Ray tracing used to determine validity of a given coordinate.

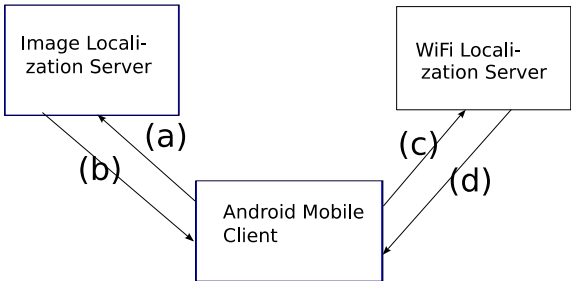


Fig. 5. Client-side architecture and communication paths. The Android mobile client is either a smartphone or an AR glass. Arrows indicate direction of message transfer. Message content for each arrow is as follows. (a) Picture taken by smartphone camera. (b) Image-based location estimate (x_i, y_i). (c) WiFi measurements. (d) WiFi-based location estimate (x_w, y_w).

IV. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of our system consisting of step detection/step length estimation algorithm, and the tracking/localization system for the second floor of the U.C. Berkeley Electrical Engineering building, Cory Hall. The map and fingerprint database is generated from a single walk-through with the backpack, using a 306 ms channel dwell time for WiFi, resulting in a total of 76 data points in the database as shown in Figure 3. The entire path is about 130m, which on average results in about one WiFi fingerprint per 1.71m in the database.

The tracking system on the client side is implemented and tested on Android platform with the Samsung Galaxy S4 smartphone running Android 4.3, which supports required sensor APIs. The system architecture is shown in Figure 5. The users walks around the second floor of Cory Hall orienting the phone such that the camera is facing the outer wall, i.e. 90 degrees to the right of the user.

While the particle filter is executed on the phone, the observations and step estimations are stored in a text file to

TABLE I

	image	wifi	combined
mean error	2.37	2.53	1.81
max error	10.53	11.56	10.96
avg std dev	1.77	2.08	1.66

Average performance over 10 runs with on-demand image capture

be processed offline. This data can then be used to simulate the performance of the particle filter with only image or only WiFi observations in order to allow for a controlled comparison between the three methods.[1] For both the on-demand and continuous image capture, we performed 10 runs at 3 different times of day in order to capture a variety of levels of interference, both in terms of WiFi and image obfuscation from the presence of other humans in the hallways. In all runs, the operator walked an identical path counter-clockwise around the hallway. We recorded the time taken to reach each corner of the path and assume the user walked at a constant velocity over each straight section in order to create a ground-truth path. We measure the error at the time of each step as the difference between the estimated location and the ground-truth location at the step with the closest timestamp.

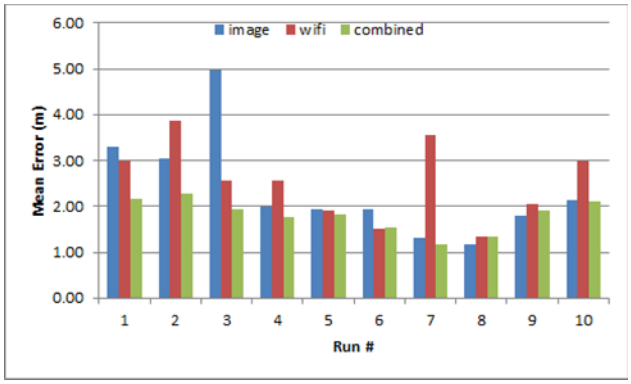
A. *On-demand Image Capture*

The first set of experiments was performed with on-demand image capture. The user selectively took pictures of areas with sufficient SIFT features for image localization, i.e. he avoided taking pictures of blank walls. Figure 6 compares the localization performance for the 10 runs in terms of mean error, maximum error, and the average spread of the particle cloud. Combined stats for all 10 runs are shown in Table I.

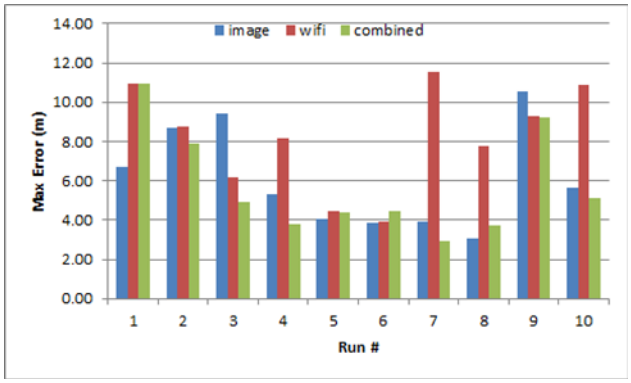
It can be seen from the table that the combined method has the best overall performance in terms of mean error and average standard deviation. Additionally, it can be seen from Figure 6(a) that in all 10 cases, mean error of the combined method is either on par with or significantly lower than both single modality methods. While the relative performance of the single modality methods varies greatly on a run-by run basis, it can be seen from Table I that the average error for image-based localization is slightly lower than for WiFi. It should be noted that there are an average of 41.7 WiFi observations and only 9.8 image observations, as the operator was very cautious about taking images, yet the addition of these few reliable observations allows for a noticeable improvement in performance.

It can be seen in Figure 6(b) that maximum error of the combined method is on par with or lower than that of the single modality methods except for run 1, in which the image-based localization has a significantly lower maximum error

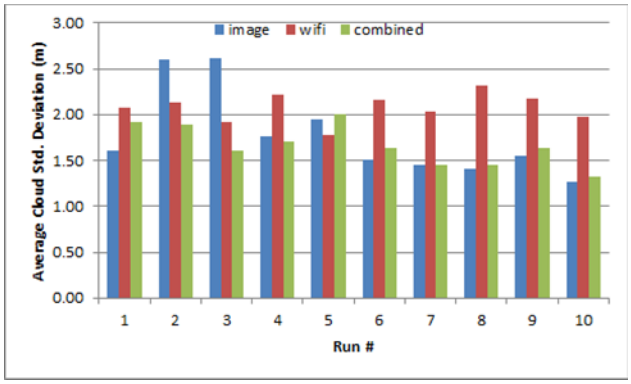
[1]We have implemented the entire particle filter on the phone to ensure our approach can run real time on a Galaxy S4.



(a)



(b)



(c)

Fig. 6. (a) Mean error, (b) max error, and (c) average cloud standard deviation for image only, wifi only, and combined image and wifi for each of 10 runs.

than either of the other methods. Similarly, Figure 6(c) shows that the average standard deviation of the combined method is on par with or lower than that of the single modalities except for run 1.

Figure 7 shows the localization error as a function of time for the fourth run, which is the run with mean errors closest to the average of all ten runs: 2.00, 2.55, and 1.77 for image, WiFi, and combined, respectively. The dashed blue curve corresponds to image only, the dotted red curve corresponds to WiFi only, and the solid green curve corresponds to the

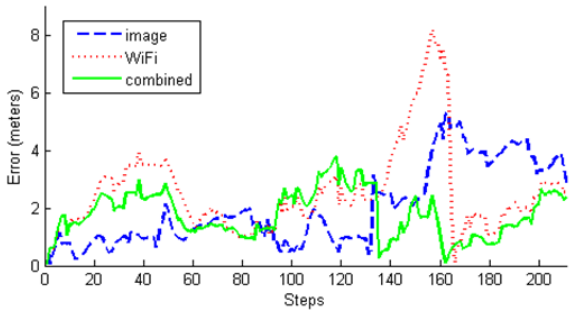


Fig. 7. Error vs time

combination. It can be seen that for much of the time, the performance of the combined method is similar to WiFi only, since the WiFi observations are more frequent than image observations; however, at the time the WiFi only method starts to perform poorly, around 135-165ms, the combined method is able to maintain a much lower error.

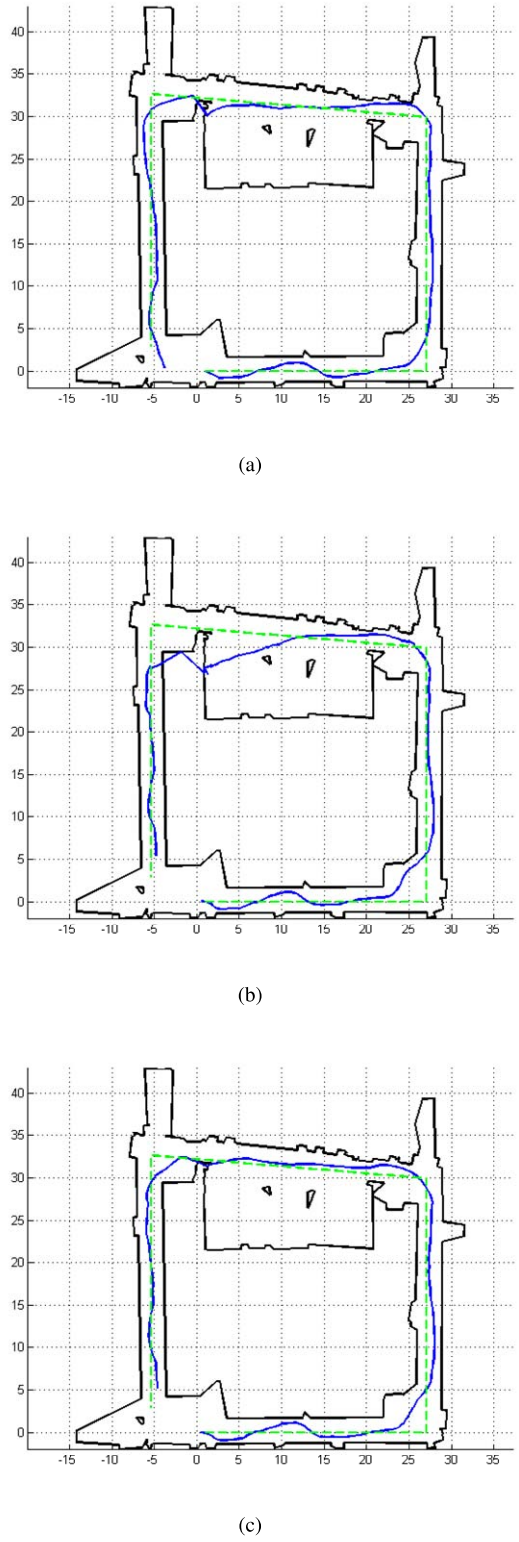
Figure 8 shows the resulting paths for each of the three methods for the same run as in Figure 7. The dotted green curve represents the ground truth path and the solid blue curve represents the centroid of the particle cloud. It can be seen that the WiFi only path in Figure 8(b) starts to drift around the location (12, 32), hits the wall at (1, 28), and is forced to jump around the corner. However, the combined path in 8(c), the estimated curve tracks cleanly around the corner. A video of the particles corresponding to Figure 8 can be found at [26].

B. Continuous Image Capture

The second set of experiments was performed with continuous image capture. Again the phone is oriented in the same direction, but this time the pictures are taken whenever the server is available, and not necessarily in areas with a significant number of features. Counts of images captured, used, and successfully localized, for each run are shown in Figure 9(a). As seen, on average 73.9 images are taken, 24.5 or 33% are deemed to have sufficient features to attempt retrieval, and 16.4 are correctly localized. Retrieval success rate for the images with a sufficient number of features for both on-demand and continuous captures are shown in Figure 9(b). As expected, the average success rate for on-demand, 86.3%, is higher than that of continuous capture, 66.9%.

A few representative example images are shown in Figure 10. Figure 10(a) is an example of a correctly retrieved image. The image in Figure 10(c) is primarily of a blank wall, so no retrieval is attempted. The image in Figure 10(b) appears as a blurry door, which has sufficient features to attempt retrieval but fails to return a correct result. On average, there are 8.1 incorrect retrievals per continuous image capture run, as compared to 1.2 for the on-demand image capture.

Figure 11 compares the localization performance for the 10 runs in terms of mean error, maximum error, and the average

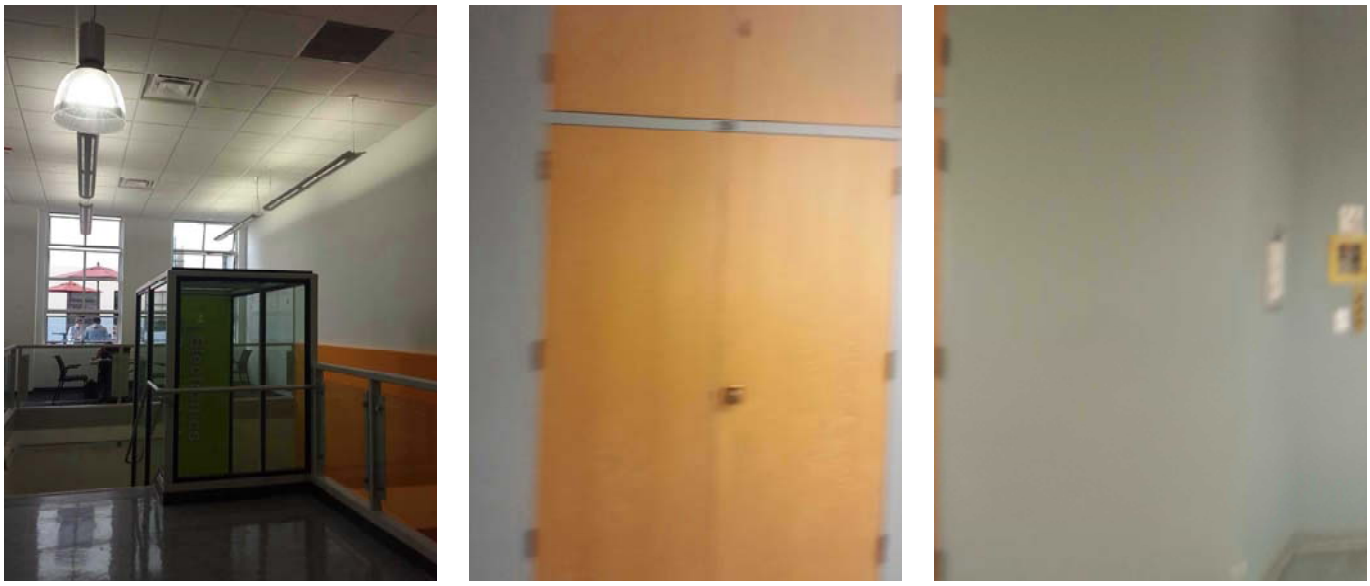


(a)

(b)

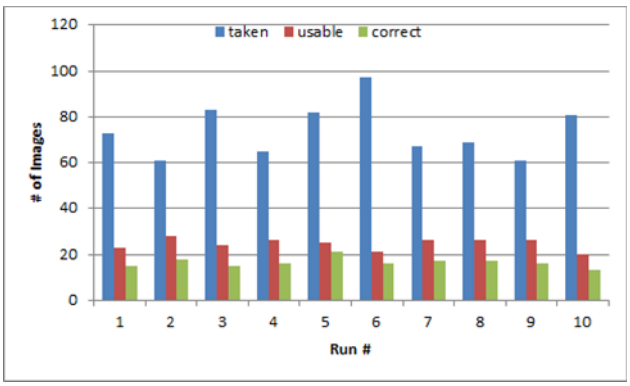
(c)

Fig. 8. Example paths for (a) image, (b) WiFi, and (c) combined.

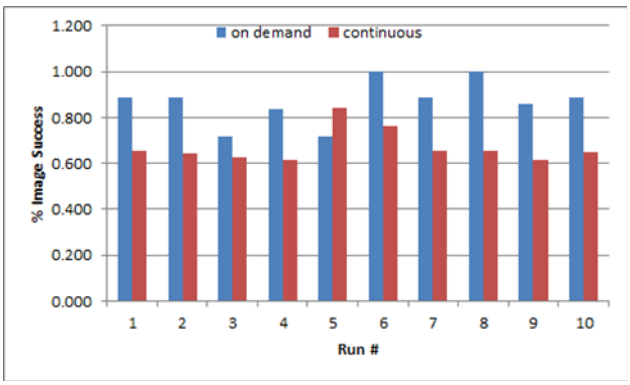


(a) (b) (c)

Fig. 10. Representative images for which (a) localization is successful, (b) localization fails, and (c) there are insufficient SIFT features to attempt localization.



(a)

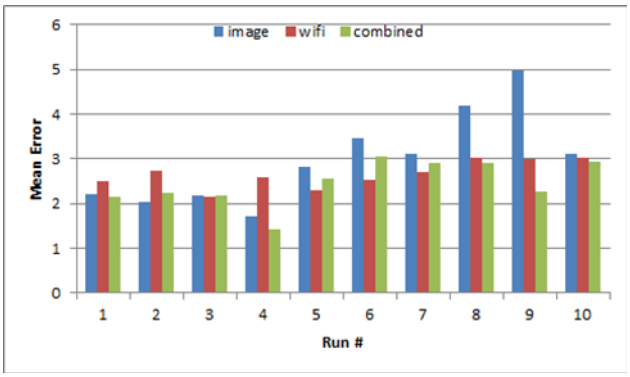


(b)

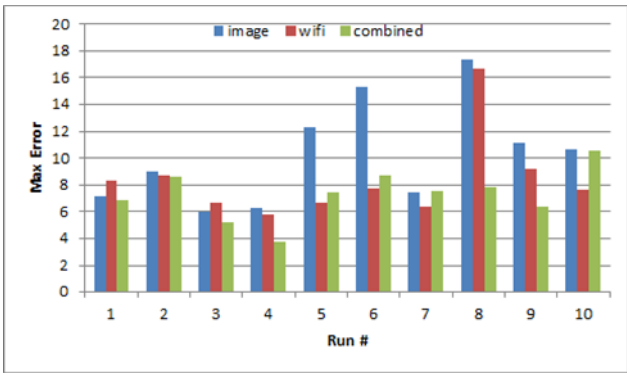
Fig. 9. (a)Number of images captured, used, and correctly localized for the 10 runs and (b) percent s uccess for on-demand and continuous capture for all 10 runs.

spread of the particle cloud. From Figure 11(a), the mean error for the combined method is on par with or lower than the single modality methods except for run 6. Similarly, from Figure 11(b), the maximum error of the combined method is on par with or lower than the single modality methods except for run 10. While the maximum error for image only or WiFi only can exceed 16m, the combined method never exceeds 11m. From Figure 11(c), the average cloud standard deviation is on par with or lower than the single modality methods except for run 7. Notably, it is significantly lower than WiFi only in all 10. This is due to the greatly increased number of observations and the fact that correct image observations localize to greater certainty than is possible for WiFi observations.

Aggregate statistics for all 10 runs in Figure 11 are shown in Table II. As seen, the combined method has lower mean error and maximum error than either of the single modality methods. Comparing Tables I and II, we see that the mean error of both image and combined are significantly higher than with on-demand image capture, owing to the lower retrieval success rate. With on-demand image capture, image-based localization has slightly lower error than WiFi-based, and with continuous capture, it has slightly higher error. The lowest overall mean error is for the combined method with on-demand image capture at 1.81m. Even though the mean error for the combined method with continuous capture, 2.46m, is higher than for combined with on-demand, 1.81m, it is still lower than for WiFi only. This indicates that even in an environment with primarily blank walls and non-distinctive features, the use of random images can still provide some improvement over WiFi only.



(a)



(b)

(c)

Fig. 11. (a) Mean error, (b) max error, and (c) average cloud standard deviation for image only, wifi only, and combined image and wifi for each of 10 runs.

TABLE II

	image	wifi	combined
mean error	2.99	2.66	2.46
max error	17.34	16.72	10.55
avg std dev	1.35	1.98	1.35

Average performance over 10 runs with continuous image capture

An example video of particles for a run with continuous image capture can be found at [27].

V. Conclusions and Future Work

We have presented a complete end-to-end localization system including map and database generation and client localization which is faster than any other current solution while maintaining good localization accuracy, with an average error of under 2m. We have demonstrated that by combining WiFi and image observations, it is possible to achieve accurate localization even when neither system can do so on its own. Additionally, we have shown that image-based localization can perform similarly to WiFi-based localization, and can thus be a viable alternative for indoor positioning.

It may be possible to improve the performance of continuous image capture by doing coarse image processing on the phone to avoid sending featureless images to the server, which wastes transmission time and may result in missed opportunities to capture images with more features.

In addition to the Galaxy S4, we have also implemented the client system on the Google Glass. The performance has similar trends, but is worse overall due to the lack a 5 GHz radio. It has been observed in [5, 19] that localization performance using 5 GHz is superior to using 2.4 GHz. The Google Glass does, however, have some inherent advantages due to the fact that it is attached to the user's face: since the face is more rigidly attached to the torso than the hand, it results in fewer erroneous step detections as compared to a phone held in the user's hand; and additionally, since the camera faces out from the user's face, it will more often be in a position to capture images than a hand-held phone whose camera will often be facing the ground while the user view the screen.

We are also in the process of developing a simplified system for the database generation, consisting only of a foot-mounted IMU and a mobile phone. Simplifying the map and database generation will be beneficial for frequently changing environments where 3-D models are not required.

References

[1] P. Bahl and V. Padmanbhan, "Radar: An in-building rf-based user location and tracking system," in *IEEE INFOCOM*, 2000.

[2] P. Boliger, "Redpin - adaptive, zero-configuration indoor localization through user collaboration," in *ACM MELT*, 2008.

[3] G. Chen, J. Kua, S. Shum, N. Naikal, M. Carlberg, and A. Zakhor, "Indoor localization algorithms for a human-operated backpack system," *3D Data Processing, Visualization, and Transmission, Paris, France*, 2010.

[4] N. Corso and A. Zakhor, "Indoor localization algorithms for an ambulatory human operated 3d mobile mapping system," *Remote Sensing*, vol. 5, pp. 6611–6646, 2013.

[5] A. Farshad, J. Li, M. K. Marina, , and F. J. Garcia, "A microscopic look at wifi fingerprinting for indoor mobile phone localization in diverse environments," in *International Conference on Indoor Positioning and Indoor Navigation*, 2013.

[6] T. C. Hales, "The jordan curve theorem, formally and informally," *The Mathematical Association of America Monthly*, vol. 114, 2007.

[7] A. Hallquist and A. Zakhor, "Single view pose estimation of mobile devices in urban environments," *IEEE Workshop on the Applications of Computer Vision (WACV), Clearwater, FL, USA*, 2013.

[8] M. Holcik, "Indoor navigation for android," *Masters thesis, Faculty of Informatics, Masaryk University*, 2012.

[9] C. Y. J. So, J. Lee and H. Park, "An improved location estimation method for wifi fingerprint-based indoor localization," *International Journal of Software Engineering and Its Applications*, 2013.

[10] Y. Jin, H. Toh, W. Soh, and W. Wong, "A robust dead-reckoning pedestrian tracking system with low cost sensors," *IEEE International Conference on Pervasive Computing and Communications (PerCom), Seattle, WA, USA*, 2011.

[11] J. Kua, N. Corso, and A. Zakhor, "Automatic loop closure detection using multiple cameras for 3d indoor localization," *IS&T/SPIE Electronic Imaging*, 2012.

[12] D. M. Lambeth, "Design considerations for an indoor location service using 802.11 wireless signal strength," Master's thesis, Department of Aeronautics and Astronautics, MIT, 2009.

[13] F. Li, C. Zhao, G. Ding, C. Liu, and F. Zhao, "A reliable and accurate indoor localization method using phone inertial sensors," in *UbiComp '12, Pittsburgh, PA, USA*. IEEE, 2012, pp. 421–430.

[14] J. Z. Liang, N. Corso, E. Turner, and A. Zakhor, "Image based localization in indoor environments," *International Conference on Computing for Geospatial Research and Application (COM.Geo) '13, San Jose, CA, USA*, 2013.

[15] ——, "Reduced-complexity data acquisition system for image-based localization in indoor environments," *International Conference on Indoor Positioning and Indoor Navigation (IPIN '13), Montbeliard-Belfort, France*, 2013.

[16] ——, "Image-based positioning of mobile devices in indoor environments," in *Multimodal Location Estimation*, S. Gerald Friedland, Ed., 2014.

[17] T. Liu, M. Carlberg, G. Chen, J. Chen, J. Kua, and A. Zakhor, "Indoor localization and visualization using a human-operated backpack system," *International Conference on Indoor Positioning and Indoor Navigation (IPIN), Busan, Korea*, 2010.

[18] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, pp. 91–110, 2004.

[19] G. Lui, T. Gallagher, B. Li, A. G. Dempster, and C.Rizos, "Differences in rssi readings made by different wi-fi chipsets: A limitation of wlan localization," in *International Conference on Localization and GNSS*, 2011.

[20] N. Naikal, J. Kua, and A. Zakhor, "Image augmented laser scan matching for indoor dead reckoning," *International Conference on Intelligent RObots and Systems (IROS), St. Louis, MO, USA*, 2009.

[21] F. H. T. Pinto, "An indoor localization solution for mobile devices," Master's thesis, FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO, 2011.

[22] J. T. Quah and L. Lim, "Location cluster with nearest neighbors in signal space: An implementation in mobile service discovery and tracking," *International Conference on Wireless Networks (ICWN '11), Las Vegas, NV, USA*, 2011.

[23] M. Quigley, D. Stavens, A. Coates, and S. Thrun, "Sub-meter indoor localization in unmodified environments with inexpensive sensors," in *IEEE/RSJ International Conference on Intelligent Robots and Systems, Taipei, Taiwan*, 2010.

[24] A. Serra, T. Dessi, D. Carboni, V. Popescu, and L. Atzori, "Inertial navigation systems for user-centric indoor applications," *Networked & Electronic Media (NEM '10) Summit, Barcelona, Spain*, 2010.

[25] O. J. Woodman, "An introduction to inertial navigation," University of Cambridge, Computer Laboratory, Tech. Rep. UCAM-CL-TR-696, Aug. 2007. [Online]. Available: http://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-696.pdf

[26] http://www-video.eecs.berkeley.edu/~avz/ipin2014-particles-ondemand.mp4.

[27] http://www-video.eecs.berkeley.edu/~avz/ipin2014-particles-continuous.mp4.