On Wi-Fi Fingerprint Datasets

Kyeong Soo (Joseph) Kim
Department of Electrical and Electronic Engineering
Centre of Smart Grid and Information Convergence
Xi'an Jiaotong-Liverpool University (XJTLU)
07/31/2018

Outline

- Review of "Wi-Fi crowdsourced fingerprinting dataset for indoor positioning"
- On Wi-Fi Fingerprint Datasets for Indoor Localization &Navigation

2

Review of "Wi-Fi crowdsourced fingerprinting dataset for indoor positioning"*

* *E. S. Lohan et al., "Wi-Fi crowdsourced fingerprinting dataset for indoor positioning," Data, vol. 2, no. 4, article no. 32, pp. 1-16, 2017.*

Measurement

- Period: January–August 2017
- Place: A 5-floor building at Tampere University of Technology, Finland
  - The basement was not used.
- Client: Android app coded in Java using
  - Android Studio 2.2.3
  - Google cloud server-based application
- Server: Written in Python 2.7 with
  - REST API based on Flask.
  - Google services based on App Engine SDK for Python.
- The server stored the following information reported by users:
  - 3D Location (local coordinates in meter)
  - Time stamp
  - Device model (total 21)
  - MAC address
  - RSS (dBm in 2.4- and 5-GHz bands; +100 for non-heard APs)

Database

- Total number of fingerprints: 4648
- Randomly split the measurements into non-overlapping subsets:
  - Training: 697 (15%)
  - Test: 3951 (85%)
  - ***No problem of mismatched training and dev/test sets.***
    - Note that UJIIndoorLoc DB has this problem.

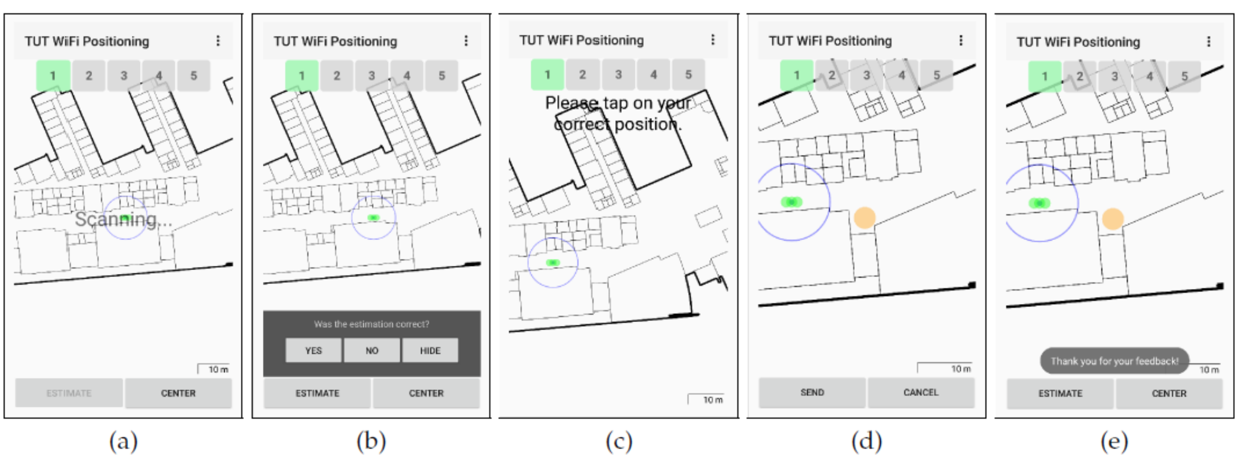

Figure 1. Five snapshots, in chronological order, of the interface of the Android application "*TUT WiFi Positioning*" used to collect the data. (**a**) initial position estimate; (**b**) aksing for estimation feedback; (**c**) selecting the correct floor; (**d**) selecting the correct location; and (**e**) notification of received feedback.

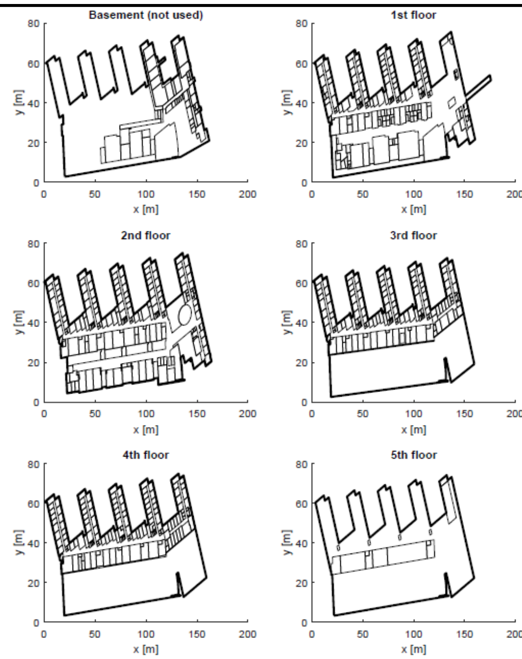

Figure 2. Floor maps of the 6-floor university building in which the measurements were recorded over five floors.

7

Figure 3. Environment of the measurements: example images at different floors of the building. (a) Main hallway seen from the second floor; (b) corridor partitioned by glass on the 4th floor; (c) restaurant area on the first floor; (d) corridor connecting different office spaces on the 3rd floor.

8

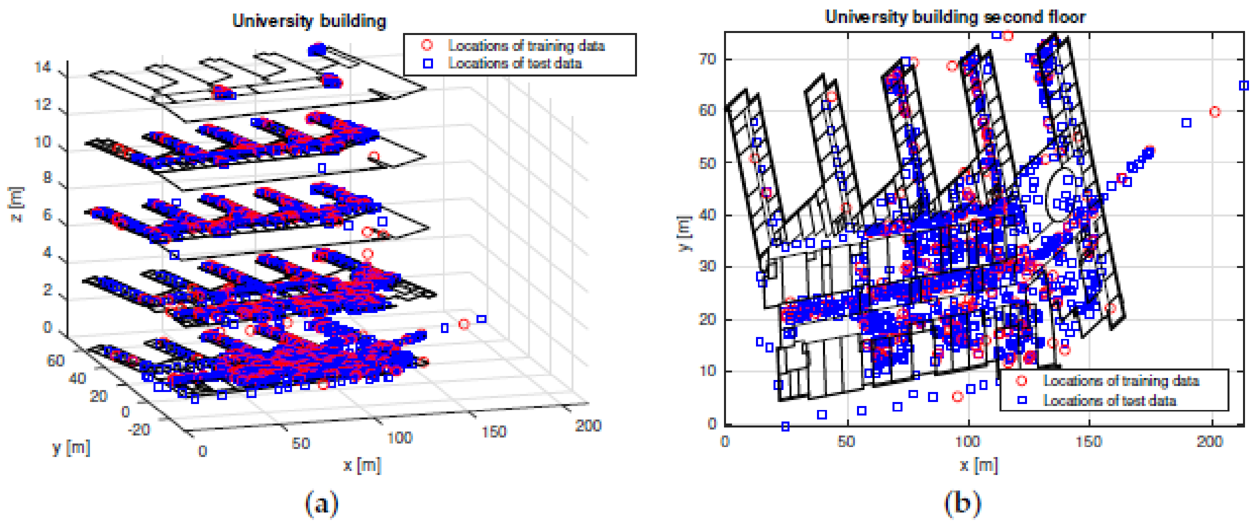

Figure 4. Collected fingerprints. (a) 3D view, whole building; (b) 2D view, second floor.

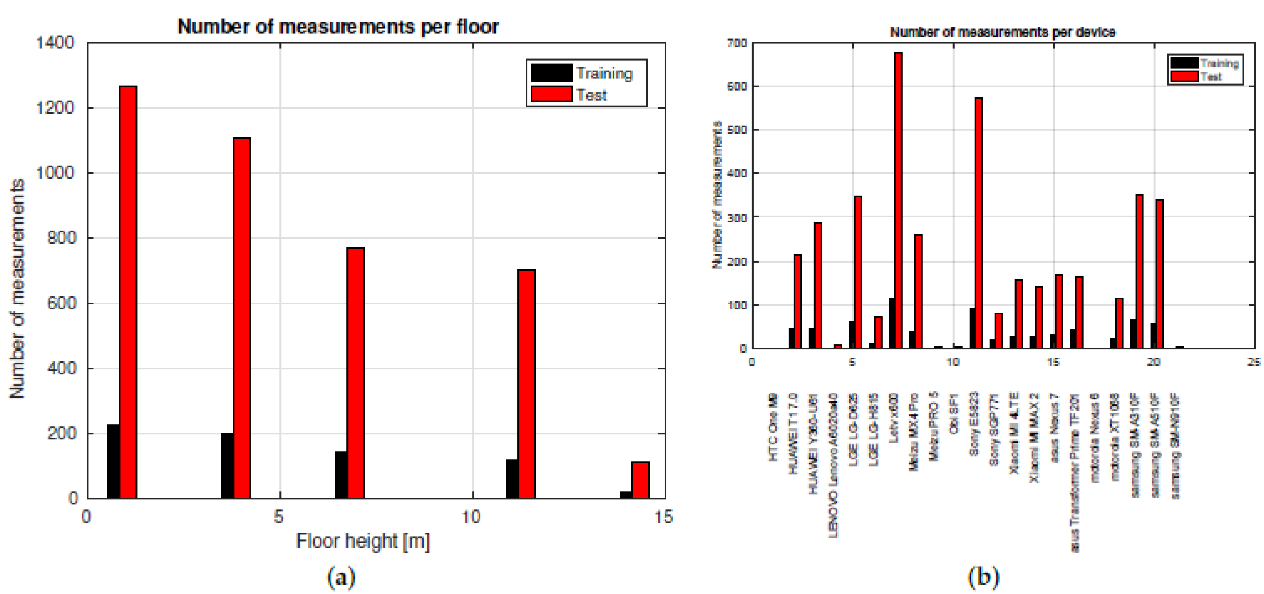

Figure 5. Number of measurements per floor (a) and per device type (b).

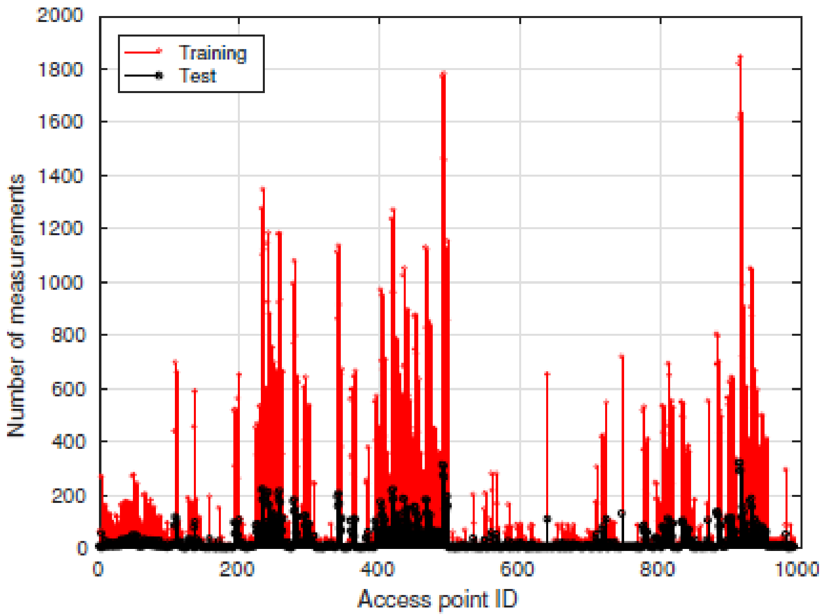

Figure 6. Number of measurements per access point (or MAC address). 11

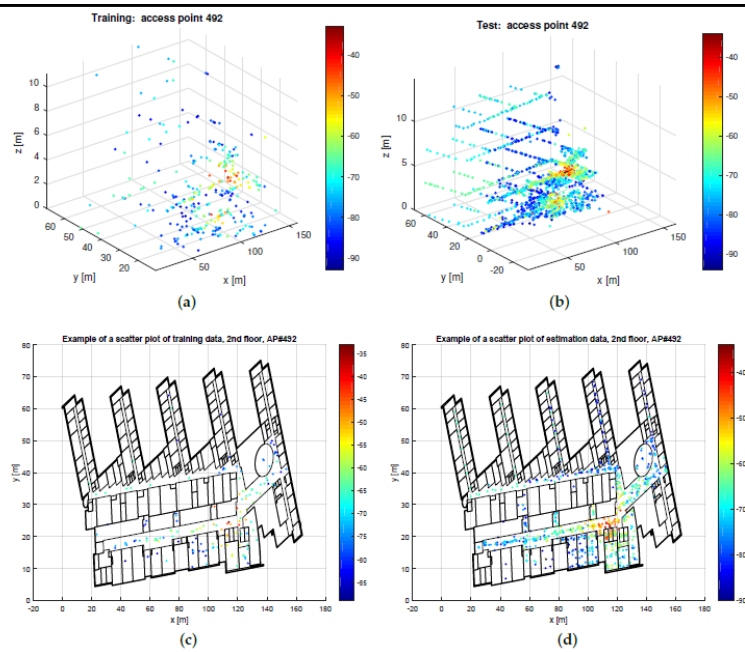

Figure 7. Examples of the 3D (up) and 2D (down) scatter diagrams (non-interpolated power maps) of AP 492. (a) 3D training data; (b) 3D test data; (c) 2D training data; (d) 2D test data. 12

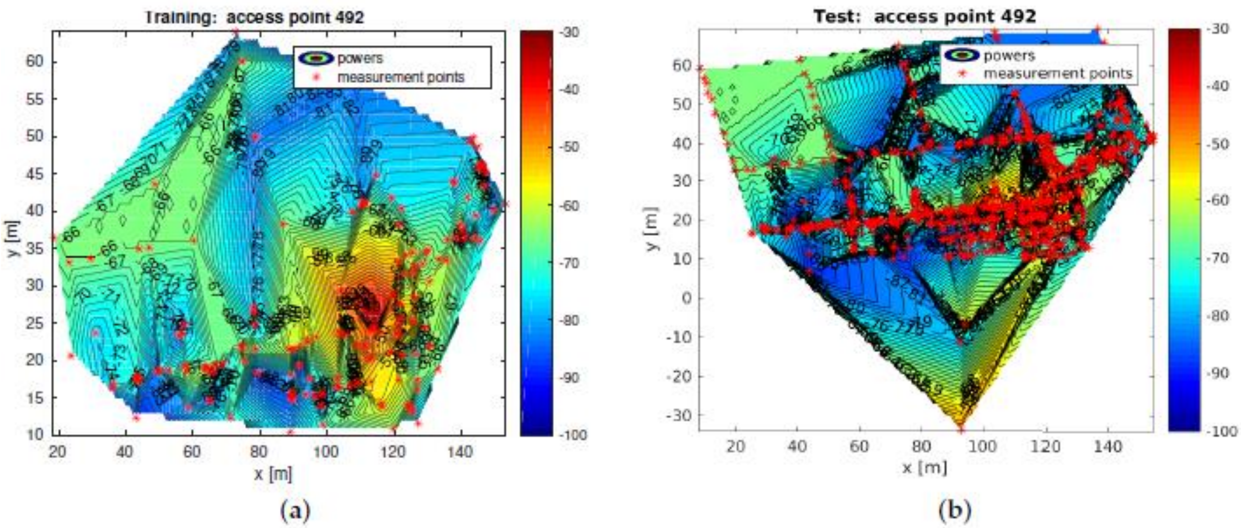

Figure 8. Example of interpolated power maps for Access Point 492 at second floor. (**a**) Training data; (**b**) test data.

13

Table 2. Benchmark positioning results based on our dataset. RBF, Rank-Based Fingerprinting.

Algorithm	Mean 2D Error (m)	Mean 3D Error (m)	Floor Detection (%)	Reference
Weighted centroid	10.64	11.57	83.19	[18]
Log-Gaussian probability ($\sigma = 10, N_{nn} = 3$)	10.18	11.19	82.92	[20,23]
Log-Gaussian probability ($\sigma = 7, N_{nn} = 1$)	9.78	11.03	85.29	[20,23]
RSS clustering (affinity propagation)	$\boxed{8.09}$	$\boxed{8.70}$	90.81	[20]
3D clustering (k-means)	17.35	24.73	72.90	[20]
UJI kNN algorithm (data=positive,dist=sorensen, $N_{nn} = 1, Not_{heard} = -103$)	8.45	8.73	92.26	[21]
UJI kNN algorithm (data=exponential, dist=neyman, $N_{nn} = 1, Not_{heard} = -103$)	8.60	9.02	91.98	[21]
UJI kNN algorithm (data=powed, dist=sorensen, $N_{nn} = 1, Not_{heard} = -103$)	8.65	8.92	$\boxed{92.99}$	[21]
RTLS@UM ($approach = 1, variant = 1, n = k1 = 5, k2 = 3$)	9.18	10.29	86.99	[31,32]
RTLS@UM ($approach = 1, variant = 3, n = 5, k1 = 1, k2 = 3$)	9.18	9.92	90.05	[31,32]
RBF ($N_{nn} = 1$), $distance = spearman$	9.77	10.32	86.51	[33]
Coverage area, pointwise defined (probability of AP match = 0.9)	10.03	9.44	86.64	[34]
Coverage area, distribution based (Gaussian distribution)	13.01	11.68	69.07	[36,37]

14

Diversity of The New Database

- A new environment where the AP deployment might highly differ from other available databases.
- A new building where its geometry, building materials, structural elements and obstacles might highly differ from the buildings in other available databases.
- Different conditions (e.g., density of people, and weather, among others).
- A higher number of APs (≈ 1000 MAC addresses).
- Benchmark results with the available dataset.

Features of The New Database

- Samples are collected at random positions and orientations decided by the user
  - i.e., no grid-based or pre-established mapping.
- Just one sample per reference point.
- Different devices used to generate the database.
- Database division is more challenging
  - 15% of samples for training/reference and 85% of samples for evaluation.

On Wi-Fi Fingerprint Datasets for Indoor Localization &Navigation

What To Measure?

- e.g., UJIIndoorLoc database

Columns	Data
001-520	RSSI levels
521-523	Real world coordinates of the sample points • Longitude, latitude, floor
524	BuildingID
525	SpaceID
526	Relative position with respect to SpaceID
527	UserID
528	PhoneID
529	Timestamp

How To Measure?

- # of people
- # of devices
- Measurement period

How To Organize Database?

- One common dataset vs. separate ones for training/validation/testing.