**Multimodal Indoor Localization**

COMP4971C Report

LEUNG, Ko Tsun

20516287

ktleungai@connect.ust.hk

**1. Background**

Localization technology has wide applications in navigation, location-based marketing, geofencing, etc. GPS typically offers an effective solution for open spaces; however, its performance deteriorates in urban-canyon, semi-indoor, or deep indoor settings due to weak or unavailable signals. Fingerprinting emerges as a promising technique for complex indoor cases. Given the fingerprints, user location can be estimated based on the signals they samples. WiFi, Bluetooth, magnetic fields are various fingerprint examples. An Inertial navigation system (INS) may be added to enhance localization accuracy and user experience further.

Our framework supports a dynamic set of signals in terms of type and number, enabling seamless roaming over a wide range of signal environments. Consider a user walking from place to place, an arbitrary combination of WiFi, Bluetooth, magnetic fields, and GPS may be involved. SiFu overcomes this problem by supporting signal addition and removal at any time resulting in an incrementally extensible to new signals without the need for retraining the whole system.

I was responsible for incorporating BLE signals into the current module to verify and improve the siFu framework. In the remaining parts of the report, I will discuss the results of researches that I have conducted, my implementation of BLE modules, deployment of other machine learning models, and what I have learned from the project and suggest some future work in multimodal localization for the project.

**2. Literature Review**

I took much of the time to perform research on relevant topics and study the project code. To overcome the lack of generalization ability of the current single-modal system, more researchers are focusing on the signal fusion framework [1,2]. The fusion approaches can be classified into probabilistic and non-probabilistic in general.

**Probabilistic**

1. WiFi-Assisted Particle filter(WaP)

WaP contains three major components: Pedestrian Dead Reckoning (PDR), WiFi, and Particle Filter (PF). Based on the accelerometer readings, the PDR component counts the user's steps and calculates each step's stride length. All the WaP needs are the floor plan and the coarse locations on which the APS reside. WaP takes the WiFi sensor, inertial sensor, and magnetometer readings embedded as an input on smartphones, outputting the user's position at each stage. However, it can not be applied to other signals beyond WiFi and inertial sensors [3].

2. Magnetic Field-Based indoor localization (Magicol)

The indoor geomagnetic field anomalies are omnipresent, location-specific, and temporally stable; Magicol leverages the locally disrupted magnetic signals. It uses the magnetometer commonly found on smartphones, without resorting to special hardware. However, the strength of the magnetic field is generally weak, usually within a few tens of uT. Single magnetic signals, therefore, have very little discernibility [4].

3. Unified Mobile Localization Framework (UniLoc)

UniLoc is a unified framework that exploits existing localization schemes' diversity to achieve accurate and robust positioning across variant environments. No particular localization schemes or sensor data are restricted to UniLoc, while it can automatically adjust to the spatiotemporal variation of sensor data at any location. It can be used without pre-training in new unknown locations [2].

**Non-probabilistic**

1. WiFi-based Indoor Localization System

Many indoor localization methods are currently based on the Wi-Fi signals obtained by the devices (smartphones) and create these signals' fingerprint maps. Then it can measure the strength of these Wi-Fi signals and predict position points. The precision of these approaches depends on how many reference points the database has reached [5].

However, Wi-Fi signals are unstable and sometimes even inaccessible in most indoor environments, particularly in commercial buildings. This approach does not return an accurate result if Wi-Fi coverage is low or mobile phones are not linked to the malls' Wi-Fi. Also, WiFi-based IPS can record an object's location, but it cannot detect and predict the object's orientation and direction. Nevertheless, current WiFi-based IPSs involve installing dedicated software on smartphones to constantly scan nearby Wi-Fi routers for data acquisition, which consumes a large battery. Hence, no sole sensing technology could provide reliable, stable, and consistent localization and monitoring services [6].

Some works fuse multiple sensors' raw data to provide more stable outcomes, e.g., Wi-Fi is used by particle filtering to enhance motion-based Pedestrian Dead Reckoning(PDR) findings. Therefore, it is difficult for a fixed fusion algorithm to automatically adapt to all possible environmental conditions with predefined sensor types. For example, in some regions with poor Wi-Fi signals (due to high wireless interference or sparse distribution of access points), Wi-Fi may not support the default motion-based PDR or even make the estimated position deviate from the user's true place. As a result, there is no one-size-fits-all mobile localization scheme that can cover all the places in people's daily lives [7]-[9].

2. Bluetooth Corrections Approach

Bluetooth technology detects the sensor device's proximity to the smartphone, which tackles the weak Wi-Fi signal's drift problems, but Bluetooth sensors need to be installed [10].

3. Pedestrian Dead Reckoning (PDR)

By the user's previous position, walking length, and direction, PDR estimates its current position. Acceleration sensors can calculate the walking displacement, while magnetic field sensors and gyroscope sensors can measure direction angles [10].

However, this approach is only applicable to the prediction of short-distance position points. Drifting problem arises when predicting long-distance position due to calculation error.

4. Visible light positioning (VLP)

VLP employs “smart LEDs'' as location land-marks and photodiodes (PDs) or cameras as sensors. The dense overhead deployment of light fixtures and multipath-free prole propagation allows for the high spatial resolution and sensitivity of VLP to environmental dynamics. When using PD and camera, current VLP technologies have achieved meter or decimeter accuracy, respectively. The adoption of VLP has, however, been impeded by two fundamental limitations. The main limitation is that specially designed smart LED bulbs are required, which modify the light signals to generate specific beacon IDs. These bulbs are more expensive and bulkier than commercial LEDs or fluorescent lights, hindering mass deployment in the near term. On the other hand, it depends on the regular light shape and angular response. VLP systems that employ RSS-based light propagation models commonly assume Lambertian radiation patterns, which apply only to round-shaped LED bulbs. The camera’s narrow field-of-view, high power consumption [11].

**2.2 Comparison between different signal**

Every signal has its own strengths and weaknesses. For example, GPS works well outdoors but not so indoors. Radio-frequency (RF) signals such as Wi-Fi or Bluetooth are pervasive and differentiable over a long-range. However, it suffers from relatively high noise due to multipath and fading effects. The localization accuracy also depends heavily on the strength and density of the signals. Geomagnetism is also omnipresent with a fast sampling rate and low noise. However, a certain geomagnetic sequence may be matched to multiple places due to its global ambiguity. Inertial sensors provide user movement information and are highly available on mobile phones. However, it only provides relative location information, and error may accumulate and diverge over time. Besides, light(computer vision) is highly accurate but overhead in deployment. As a result, INS is often combined with other signals to enhance localization accuracy.

Moreover, I was responsible for incorporating BLE signals into the siFu framework. Compared to the Wi-Fi signal, BLE has a low energy consumption and configuration options, and it is more accurate than the Wi-Fi signal. Moreover, Android limited the frequency of Wi-Fi scans for devices later than Android 9 to four times in a 2-minute period. [12] Therefore, the BLE signal is a better supplement to the Wi-Fi signal.

**3. Implementation**

I was given an android application to collect data from different sensors, and I was responsible for incorporating BLE signals in the siFu network, verifying and improving the siFu framework. Moreover, I spent most of the time understanding engineering aspects of the project, for example, learning some important packages in machine learning such as TensorFlow and Keras. Having a practical project experience enables me to train my engineering ability to make things into reality. After I understood the codes and resolved issues during the development of the models, I was able to try different machine learning models to replace the autoencoder network, minimizing localization error in prediction.

3.1 Implementing BLE module in siFu framework

Similar to the Wi-Fi module, after the BLE signal is preprocessed, signal data(including UUID, RSSI signal strength, transmission power, and distance) is passed to the autoencoder module to learn the latent representation and compute similarity measure. For the preprocessor module, there are four steps involved: beacon filtering, feature mapping, feature scaling, feature selection, and noise injection. Site surveys are conducted at least 3 times at different hours, ensuring we can get the static beacon list but not any other portable Bluetooth devices at the location. Then, the filtered beacon list is mapped to the same space as the inputted signal. After normalizing RSSI signal strength and transmission power, feature selection is processed, which we only select the top k percentile strong of BLE signals. Finally, noise is injected based on a hyperparameter noise ratio, where injected signal noise and missing measurements will be implicitly considered in the latent representation generation.

After the processor module, extracted features are passed to pretrain and training modules of the autoencoder with 2 fully connected layers with an ADAM optimizer.

3.2 Comparison between different machine learning model

I also tried different neural network modules to replace the autoencoder proposed in siFu to verify the framework's general prediction ability. The following table illustrates the experimental results. We define accuracy as the probability of the input signal matching the corresponding ground truth reference point at a specific spot. In the siFu paper, it is proposed to use autoencoder with fully connected layers, where I also tried autoencoder with Long Short-Term Memory(LSTM), Recurrent Neural Network(RNN) networks, and some common machine learning models such as 1 or 3 layer(s) Bi-directional LSTM, RNN model, Gated Recurrent Unit(GRU) model, Support Vector Machine(SVM).

Table 1.1 Accuracy of different machine learning models

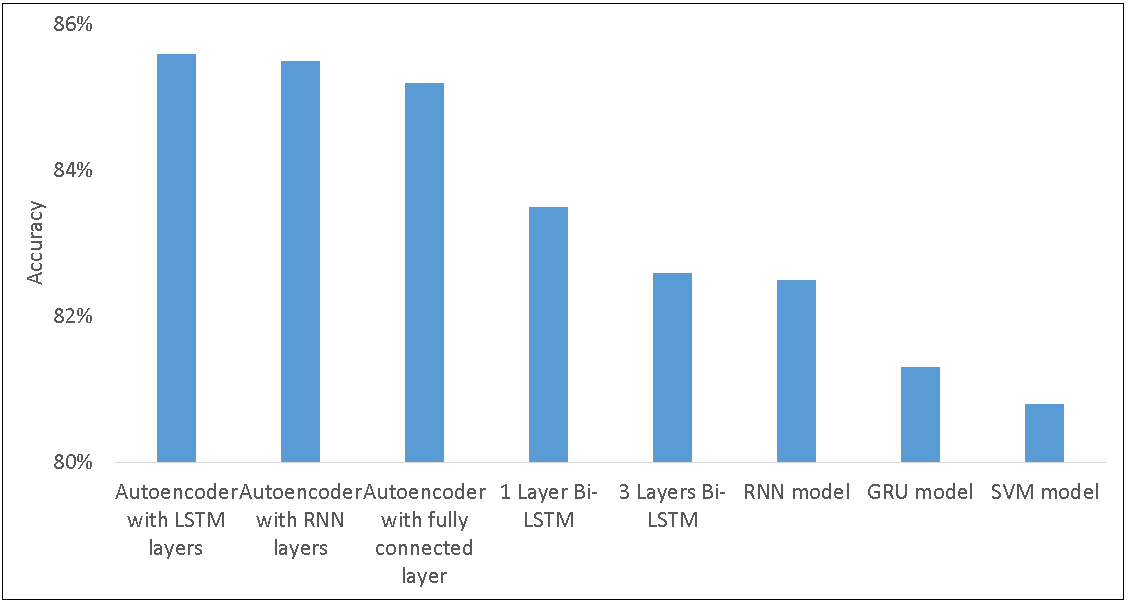
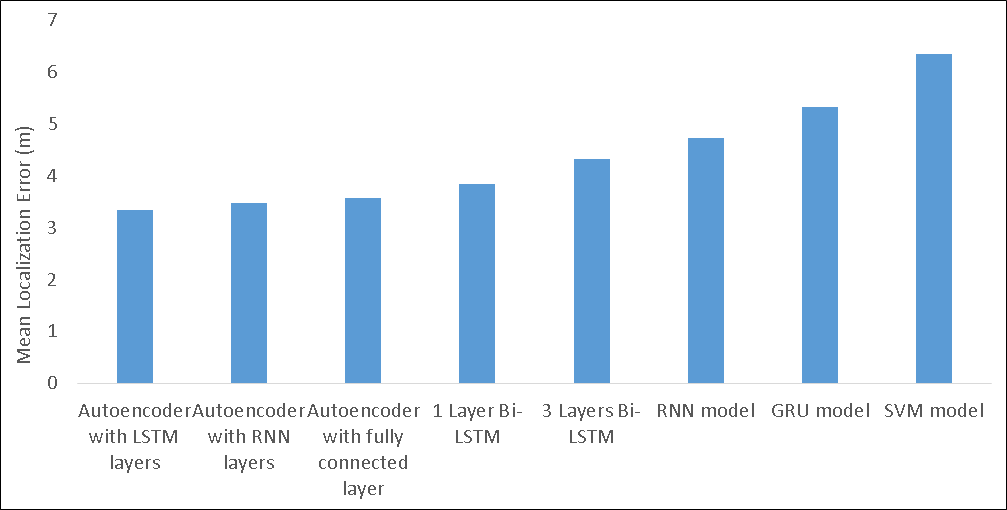


Table 1.2 Mean Localization Error of different machine learning models



We can observe that the Autoencoder with LSTM layers scores the lowest localization error among all models from the table. LSTM can process single data points(such as spot signal input from RSSI vectors) and sequential data from magnetic signals or INS data. It can perform better than fully connected or RNN networks by solving vanishing or exploding gradient problems in the training process. Besides, we can also process radio fingerprint signals with LSTM. This is because a fingerprint can be viewed as a sequence of ordered received strength values.

3.3 Fine-tuning Parameters and Explanation

Fig. A: Illustration of LSTM layer with spot based signal (x) and sequence-based signal(s)

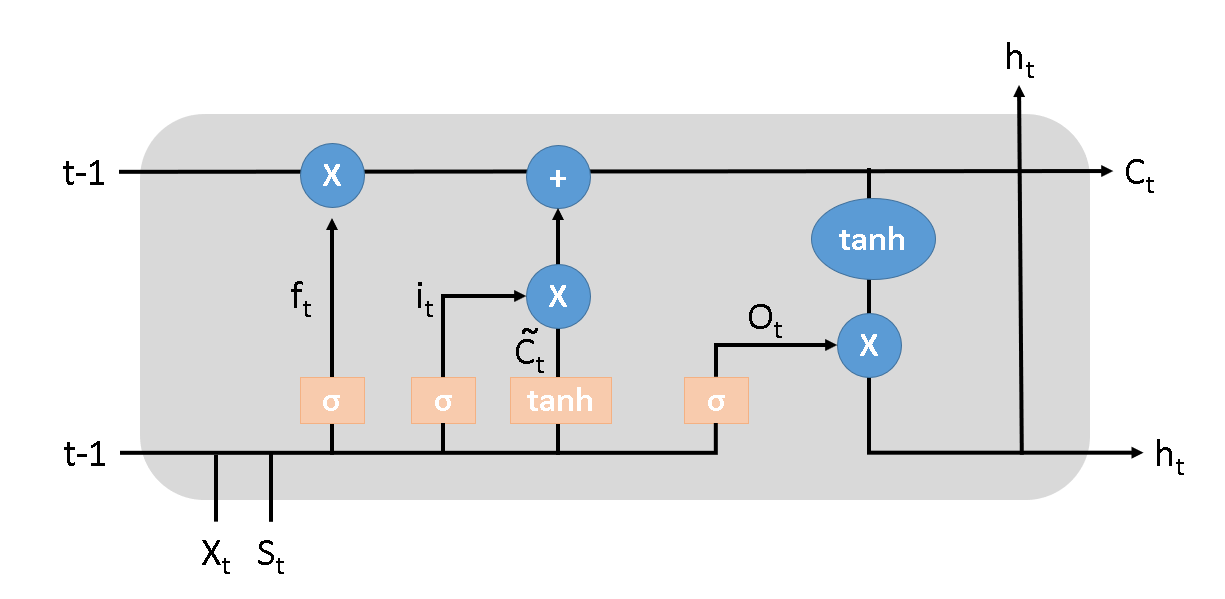
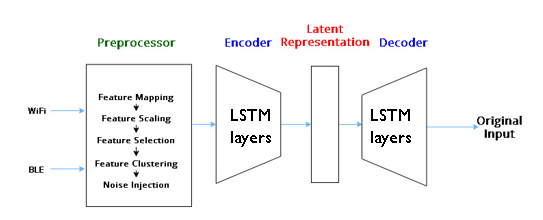


Fig. B: Illustration of autoencoder with LSTM network



From the discussion above, we conclude that the autoencoder with LSTM has the highest accuracy than other networks using the same set of parameters. Thus, an autoencoder with an LSTM network is selected to fine-tune the parameters to make it more robust. The following table illustrates different changes made to the autoencoder with the LSTM network.

Table 1.3 Fine-tuning parameters for autoencoder with LSTM network

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ID | Modification | Value | Test Accuracy | Computation Time |
| 1 | embedding\_size | 32 → 64 | 85.3% | 450s |
| 2 | embedding\_size | 64 → 128 | 85.0% | 436s |
| 3 | activation function | tanh → relu | 85.5% | 363s |
| 4 | learning\_rate  activation function | 0.01→0.001  tanh → relu | 85.8% | 336s |
| 5 | learning\_rate  noise\_ratio  activation\_function | 0.01 → 0.001  0.7 → 0.75  tanh → relu | 86.2% | 323s |
| 6 | learning\_rate  noise\_ratio  activation\_function  noise\_stdev | 0.01 → 0.001  0.7 → 0.75  tanh → relu  0.15 → 0.2 | 85.8% | 337s |
| 7 | learning\_rate  noise\_ratio  activation\_function  num\_layers | 0.01 → 0.001  0.7 → 0.75  tanh → relu  2 → 3 | 84.3% | 913s |

First, I tried to increase the size of embedding from 32 to 64 or 128, but overall, it does not improve. Then, I used relu instead of tanh, bringing improvement to localization accuracy and training time. Since the computation and derivatives of relu are less complex than hyperbolic tangents. Furthermore, using relu can solve the vanishing gradient problem of the hyperbolic tangent. Since when the input x using hyperbolic tangent as activation function is smaller than -2 or larger than 2, the network's optimization step may not converge, and training time will be long.

Also, from ID number 4 to 7 records in the table, we can see that adjusting noise ratio, noise variance, and learning rate are beneficial to accuracy and computation time. Having a larger noise ratio during training can reflect the real environment.

Finally, increasing the number of layers in the encoder and decoder does not improve the model’s accuracy since a complex model may overfit easily, lowering the testing accuracy. Moreover, the training time dramatically increases by over 100% since a deeper encoder and decoder requires more time to minimize the loss function.

3.4 Obstacles

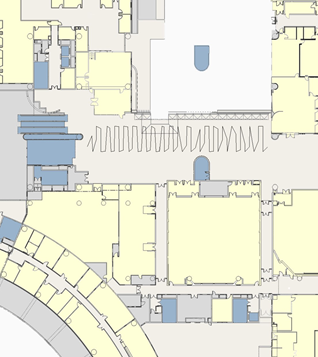
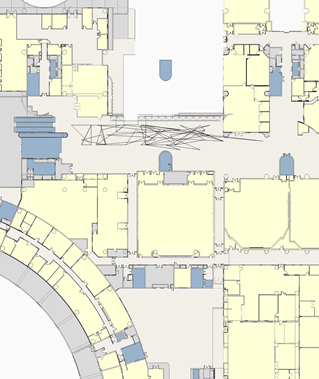
The research faced some difficulties, and I learned so much from self-learning and communicating with seniors.

The BLE signal's accuracy is not satisfactory since the collected BLE signals are sparse. Although the BLE signal's sampling rate is high, about 1 sample per second, BLE vectors have a severe missing value problem. So the signal needs to be mapped to a lower-dimensional space or deploy other techniques to extract relevant features. However, at the same time, caution needs to be drawn that the siFu framework needs to retain the generalization ability to extend to other signals. So, in the preprocessor module, I deployed the singular value decomposition(SVD) algorithm to extract features. Given a sparse matrix M, it can be decomposed into UΣV\*. If singular values in Σ are sorted descendingly, we could extract r features from the sparse matrix, where r >= number of singular values such that the sum of their squares is 90% of the total sum squared singular values.

**4. Experimental Result**

I performed site visits in Atrium(semi-outdoor environment), LTA to LTB(indoor open space), and 2F corridors at HKUST using the given android app to collect sensor signals. Autoencoder with LSTM layers is used to handle BLE and WiFi signals. There are 103 APs in the atrium area, 93 APs in the indoor open space, and 91 APs in the 2F corridors area. A mixture of different signals is handled using the siFu framework in the indoor open area, including Google API, WiFi, magnetic, and BLE signals.

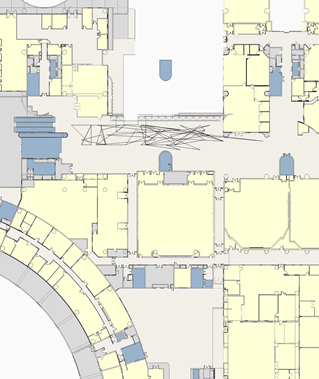
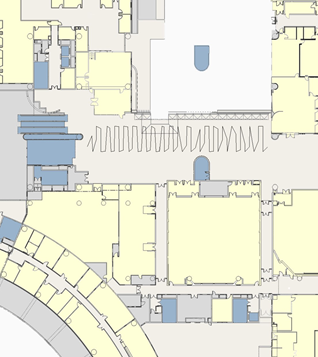
Fig. A: Designed Path for #1 experiment Fig. B: Predicted Path for #1 experiment

#1 Experimental result

In the first experiment, the result was not as satisfactory as expected. The environment was noisy as many signals overlapped, so it was difficult to do reference point fingerprinting. The mean error of prediction was 6.57m. Also, BLE signals were sparser than expected, with only approximately 3.92 beacon signals were received per second, but more than 25 beacons were installed in the region.

Fig. C: Designed Path for #2 experiment Fig. D: Predicted Path for #2 experiment



#2 Experiment Result

For the second training and prediction processes, I increased noise\_ratio and noise\_stdev parameters, as discussed before. Also, I deployed SVD algorithms to deal with sparse vectors for BLE signals. Furthermore, after these changes, the result was satisfactory, with a mean localization error smaller than 5m.

Overall, after incorporating the BLE signal module in the siFu framework, the localization error decreases. Therefore, it further proved that the siFu framework worked in an arbitrary number of available signals.

**5. Future Work**

5.1 Interactive Data Visualization

Interactive visualization can be implemented for real-time monitoring of the prediction result. For example, Augmented Reality(AR) technology can be deployed to show relative location indicators to show users the direction. In this setting, visible light position(VLP) is available, and the signal can be incorporated in the siFu framework to increase the accuracy.

5.2 Improve accuracy on semi-outdoor space localization

In a semi-outdoor environment, it is challenging to make precise predictions due to limited BLE, and WiFi signals received. To tackle this problem, the weighting scheme algorithm of siFu can be further investigated, allowing siFu to give lower weighting to some relatively unreliable signals. Assuming the signal noise follows Gaussian distribution, a mixture density network(MDN) can be used to estimate the distribution of location and orientation. The corresponding distribution's negative log-likelihood of the objective function can be minimized and find the optimized weighting using gradient descent.

5.3 More signal or information can be fused in this framework

From the previous experimental result, we can see that the siFu framework can achieve satisfactory localization results. Therefore, more novel localization methods can be combined, such as visual light position(VLP) or humidity.

Reference

[1] Q. Niu, N. Liu, J. Huang, Y. Luo, S. He, T. He, S. Chan and X. Luo. 2019. "DeepNavi: A Deep Signal-Fusion Framework for Accurate and Applicable Indoor Navigation". *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*. Vol. 3, No. 3, pp. 99:1-99:24, September 2019.

[2] W. Du, P. Tong and M. Li. 2018. UniLoc: A Unified Mobile Localization Framework Exploiting Scheme Diversity. *IEEE 38th International Conference on Distributed Computing Systems* 07 (2018), 818-829.

[3] F. Hong, Z. Guo, Z. Zhang, M. Wei, Y. Feng and Z. Guo. 2014. WaP: Indoor localization and tracking using WiFi-Assisted Particle filter. *IEEE 39th Conference on Local Computer Network* 10 (2018).

[4] Y. Shu, B. Cheng, G. Shen, C. Zhao, L. Li, and F. Zhao. 2015. Magicol: Indoor localization using pervasive magnetic field and opportunistic WiFi sensing. *IEEE Journal on Selected Areas in Communications* 33, 7 (2015), 1443–1457.

[5] X. Piao et al. 2016. RSS Fingerprint Based Indoor Localization Using Sparse Representation with Spatio-Temporal Constraint. *Sensors (Basel, Switzerland)* vol. 16,11 1845. 3 Nov. 2016.

[6] H. Zou, B. Huang, X. Lu, H. Jiang and L. Xie. 2016. “A robust indoor positioning system based on the procrustes analysis and weighted extreme learning machine". Wireless Communications *IEEE Transactions*, vol. 15, no. 2, pp. 1252-1266, 2016.

[7] Y. Zheng, G. Shen, L. Li, C. Zhao, M Li, and F. Zhao. 2014. Travi-Navi: Self-deployable indoor navigation system. *ACM MobiCom*, pages 471–482, 2014.

[8] H. Wang, S. Sen, A. Elgohary, M. Farid, M. Youssef, and R. Choudhury. 2012. No need to war-drive: unsupervised indoor localization. *ACM MobiSys*, pages 197–210, 2012.

[9] S. He, S-H Chan, L. Yu, and N. Liu. 2015. Calibration-free fusion of step counter and wireless fingerprints for indoor localization. *ACM UbiComp*, pages 897–908, 2015.

[10] H. Zou, C, et al. 2017. Accurate Indoor Localization and Tracking Using Mobile Phone Inertial Sensors, WiFi and IBeacon. *IEEE International Symposium on Inertial Sensors and Systems (INERTIAL)*.

[11] C. Zhang and X. Zhang. 2017. Pulsar: Towards ubiquitous visible light localization. *In Proc. ACM MobiCom*. ACM, 208–221.

[12] Wi-Fi scanning overview. Android Developers. https://developer.android.com/guide/topics/connectivity/wifi-scan