TELECOM PARIS

Wireless sensor network simulation and localization

Author:
Yukun LIU

Email:
yukun.liu@telecom-paris.fr

Supervisor:
Pascal Bianchi
Ons Jelassi

pascal.bianchi@telecom-paris.fr
ons.jelassi@telecom-paris.fr

Contents

1 Introduction

Under the rapid development of Internet-of-things (IoT), location is one important feature in the IoT field and localization is becoming a high value service [1]. The key drivers to the localization have been governmental institutions with an aim to provide emergency services. The support for emergency service has motivated the development of localization strategies [2]. Nonetheless, the exploitation of the location information has also drawn attention from operators due to its commercial potentiality. The location of a device can generate massive functionality to the user, and to the network as well. Civilian applications of localization includes navigation, localization-based service and monitoring. Despite the direct and indirect economical profit may be brought by localization applications, the network operators are cautious in investing due to the installation cost and privacy issue.

Localising a device in an outdoor environment typically utilises the Global Positioning System (GPS). This system is the standard for outdoor localisation and provides metre-level accuracy during normal operations. However, it provides high positioning accuracy at the expense of additional hardware and storage space, while it can quickly deplete the battery on the device, and the location information is not necessarily shared with network operator. For networks with low energy cost sensors, e.g. Low-Power Wide Area Network (LPWAN), GPS will be no longer available. Therefore several other independent systems for outdoor localization are explored.

This project focuses on localizing under LPWAN schema with tranmisson signal at a frequency about 1GHz. It developed a pipeline from WSN simulation to outdoor localization, and proposed a novel method for localizing with transfer scenarios. There is no physical experiment involved in this work and the localization methods are tested based on data generated from simulation.

2 Simulation

The use of Wireless Sensor Networks involves conception, designing and test phase [3]. While for the test phase, building a testbed and running experiments on it are costly, time consuming, and difficult on technical aspect. Moreover, since real experiments involve many uncertainties and factors affecting the result, experimental repeatability is largely compromised [4]. Thus, a simulator is essential for WSNs development and application that it allows to tune configurable parameters and reproduce experimental results. Gaining insights and drawing conclusions from a simulation study is not a trivial task. Nowadays, with the huge variety of available simulators, it is important to identify which simulator suits the most for a particular scenario.

The radio frequency (RF) propagation module in MATLAB describes the behavior of electromagnetic radiation from a point of transmission as it travels through the surrounding environment based on path loss model [5]. It is capable to calculate the coverage, the strength of signal and has great flexibility in setting up the environment and builds the site on open street map. Moreover, it provides functions for channel state information, with operational frequency from 100 MHz to 100 GHz, which fits the need ($\sim$ 1 GHz). Therefore, it is chosen as our ideal simulator tool and all the latter analysis will be drawn from it.

2.1 Simulator settings

To build a complete simulator, three components are required to be set up: transmitters, receivers and a propagation model. In our study, we focus on a simple uplink process, without concerning protocol and encoding/decoding process. Thus transmitters and receivers correspond to sensors and base stations respectively. They can be built up with location (latitude, longitude, elevation) and transmitting frequency. The propagation model describes the transmitting channel, and predicts the propagation and attenuation of radio signals as the signals travel through the environment. After the three components are established, the simulator can load open map information and execute the simulation with the support of the map.

In our implementation, the key characteristics of sites are listed in Table. 1. For simplicity, transmitters and receivers are synchronized. Every transmitter and receiver contains one antenna only (Single-input single-output). Besides, the antenna is assumed to be isotropic that the transmitter radiates uniformly in all directions and the receiver captures signals in all directions. The geolocation and height of sites depend on the spatial scenarios. Some specific cases we used in simulation are presented in Section. 2.2.

Property	Parameter
Antenna type	Isotropic
Antenna size	1
Transmitter power*	10 W
Transmitter frequency*	1 GHz
Receiver sensitivity#	-100 dBm

* - transmitter property only, # - receiver property only

Table 1: Configuration of sites

For propagation, a shooting-bouncing-rays (SBR) method with ray tracing model is chosen. The SBR method launches many rays from transmitters, and the reflection between rays and surroundings are modeled (other interaction types like diffraction, refraction, scattering are ignored). It models both line-of-sight (LOS) and none-line-of-sight (NLOS) conditions, and supports calculation of propagation paths for up to ten path reflections.

Property	Parameter
Ray tracing method	shooting-bouncing-rays
Max number of reflections	5
Building material	concrete
Surface material	glass

Table 2: Configuration of propagation model

2.2 Spatial scenarios

Spatial scenarios refer to a schema of structure arranging transmitters and receivers in simulation. In general, a square map frame is chosen and the frame is divided into meshes. One example is made in Figure. 2.1. The planar distance is 50m between adjacent transmitters and 200m between adjacent receivers. The height of a transmitter/receiver antenna is set as the geodesic elevation with an offset (1.5m/30m for transmitter/receiver) in our simulation.

The transmitters/receivers sit separately on mesh nodes and transmitters emit signals to all reachable receivers. Subsequently features or measurements will be derived at each of the receiver for further processing.

Figure 2.1: An example of spatial scenarios: a 20×20 transmitter mesh in red and a 5×5 receiver mesh in blue are evenly distributed in a 1km×1km square in London city.

2.3 Feature representation

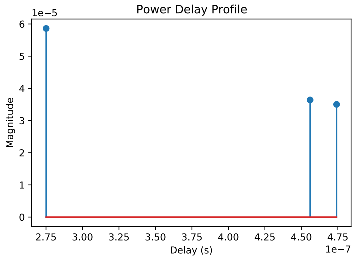
The ultimate objective of the project is to localize a transmitter with the assistance of the network composed by receivers. It is also named network-based localization, determining the device location by using signal measurements performed by the network with respect to the device. The classic measurements such as Time of Arrival (ToA), Received Signal Strength Indicator (RSSI), suffer from significant performance degradation in complex scenarios as being susceptible to multipath effect. Differently, the PHY layer feature, channel state information such as channel impulse response (CIR) can discriminate multipath effect and achieve pervasive localization [6]. Considering our simulation in urban environment, CIR will be an appropriate measurement.

One way to represent the CIR of a multipath channel is by discrete number of impulses shown in Eq. 2.1

$$h(t,\tau) = \sum_{i=1}^{N} c_i(t)\delta(\tau - \tau_i) \tag{2.1}$$

where $h(t)$ is the CIR representation, $c_i(t)$ are the complex attenuation coefficients varying with time, and $\delta(\cdot)$ denotes the dirac function. There are N arrival paths and τ_i is delay corresponding to each path. In implementation, each impulse describes a trace from a transmitter to a receiver, and N is the total number of traces. An example is illustrated in Figure. 2.2.

(a) Ray tracing scenario



(b) Channel profile

Figure 2.2: An example of CIR generation: (a) illustration of signal propagation between sites. Every connected line indicates one ray trace. The site in red color is a transmitter with geolocation (51.5134, -0.0901); The blue site is a receiver located in (51.5132, -0.0909); (b) shows the impulse profile where each impulse corresponds to one trace in (a).

CIR measurements bring rich location information, and also depend largely on environments. It is probable that a receiver captures no signal from a transmitter where the surroundings block the signal. Furthermore, sites in distinct location with diverse surrounding complexity will reflect different number of ray traces and thus different impulses. All these factors may concern data standardization and localization.

3 Localization

3.1 An overview for outdoor localization techniques

Localization is a process to determine the coordinates of a device in a system, which is one of most important subject under the wide diffusion of WSN, and has a variety of applications. There is a rich literature on localization strategies. This study concerns network with low-power sensors in wide area, where radio frequency (RF) is one of the conventional media for positioning [7]. A taxonomy of RF-based localization system is shown in Figure. 3.1. Depending on fundamental principles, the RF-based localization systems can be categorized into range-based and range-free type [2]. Range-based methods base on geometric mapping techniques such as trilateration, triangulation for range estimation and target localization. The range estimation normally relies on measurements such as RSSI, ToA, TDoA from different terminals. It requires at least 3 devices to determine one sensor's position and is vulnerable to multipath effect.

Unlike Range-based algorithms, range-free methods do not require distance estimation among nodes. They are basically implemented from fingerprinting or proximity techniques. The former one generates fingerprint to each location, and compares newly acquired fingerprints against the collected fingerprint database in order to estimate the position. The proximity technique assigns the nodes' location to be the position of the proximal terminal directly while performs less accurate [7, 8].

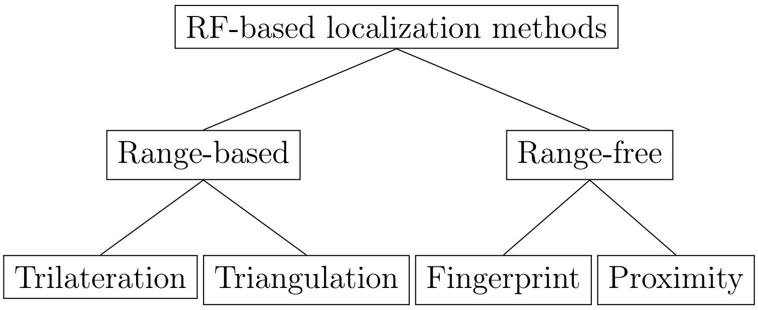


Figure 3.1: A simplified taxonomy of RF-based localization methods

3.2 Fingerprinting-based outdoor localization

Since the concerning scenario is outdoor wide area with low power sensors, a fingerprinting-based localization method is employed, which exhibits multipath effect resistance and can achieve high positioning resolution. The fingerprints database is generated from CIR measurements as stated in Section. 2.3.

3.2.1 Preprocessing

A CIR measurement sample is a complex 2D array. It consists of multiple impulses, and each marks a transmission trace. The fingerprinting database will be represented as a cell, where each entry in the cell corresponds to a CIR between a Tx and Rx, and its label will be the geo-location of devices to locate:

$$CIR\ cell = X = \begin{bmatrix} H_{1,1}, & H_{1,2}, & \cdots & H_{1,R} \\ H_{2,1}, & H_{2,2}, & \cdots & H_{2,R} \\ \vdots & \vdots & \cdots & \vdots \\ H_{T,1}, & H_{T,2}, & \cdots & H_{T,R} \end{bmatrix} \tag{3.1}$$

where T, R denote the number of Tx and Rx, $H_{i,j}$ indicates the CIR measurements from transmitter i to receiver j.

Due to a variety of environment complexity, the shapes of $H_{i,j}$ may vary. To standardize the form of data, we choose to take the magnitude of each measurement $\|H_{i,j}\|$ and pad 0 in each sample to the same size M, as the max number of traces. For training purpose, the measurements are flattened as vectors. Consequently, the fingerprinting dataset will be of shape $T \times (R * M * 2)$. Not surprisingly, the labels of fingerprints are the geo-locations of devices to locate:

$$Y = \begin{bmatrix} lat_1, & lon_1 \\ lat_2, & lon_2 \\ \vdots, & \vdots \\ lat_T, & lon_T \end{bmatrix} \tag{3.2}$$

where (lat_i, lon_i) notes the latitude and longitude of Tx_i.

One dataset created by simulation in London city is shown in Figure. 3.2. It shows the spectrum of CIR values, and it is evidently sparse since 0s are padded for standardization. There are still room left for feature fusion.

Besides, it is verified the magnitude is highly correlated with time delay, which is logic. The lower delay results in lower attenuation.

3.2.2 Experiments

Firstly four ensemble learning methods are chosen for non-linear mapping between the measurements and locations: Random Forest (RF), ExtremeGradientBoost (XGB), LightGradientBoostMachine (LGBM), AdaptiveBoost (ADB). The testbed for data simulation is chosen in a 1km×1km square area, latitude from 51.5108 to 51.5198, longitude from -0.0988 to -0.0844, located in London. The Tx

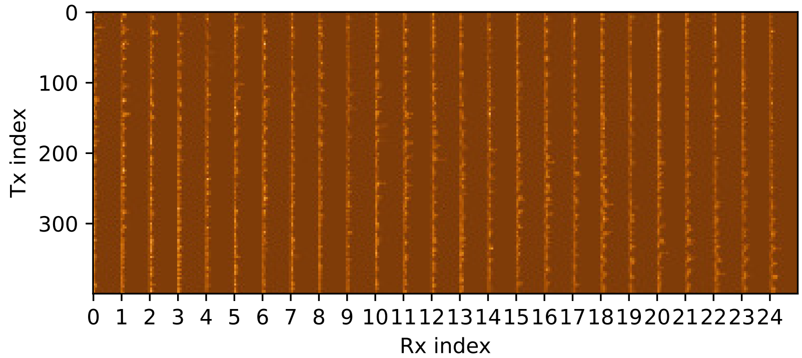


Figure 3.2: An overview of fingerprinting database for London city: 400 Tx and 25 Rx during simulation with 5 max reflections

and Rx are uniformly distributed in a 20×20 and 5×5 mesh in fixed positions as shown in Figure. 2.1, with maximum reflection of 5. The distance between tx mesh, namely grid size, is 50m, which is a criteria to evaluate the algorithms. In addition, the evaluation metric is defined as the distance between prediction and label.

The performance of each method after parameter tuning is illustrated in Figure. 3.4. It shows LGBM method generally performs better that it has lowest mean distance error. While its variance is large that it can severely under-positioning. On the other hand, XGB performs slightly worse in general, but it is relatively robust that its regression error is constrained in a small region. The performance of other two methods wander between LGBM and XGB. AdaBoost least performs in the localization task that its mean distance error is 54.79m, greater than the grid size (50m).

Besides the learning method, the impact of grid size is studied meanwhile. Within the same 1km×1km area and rx setup, the tx mesh is controlled from grid size of 200m to more fine grained: 100m, 50m, 20m. The experiment result is listed in Figure. 3.5. It is logic the positioning error increases as the grid gets coarse. What interests more is the trend that the relative error goes with grid granularity.

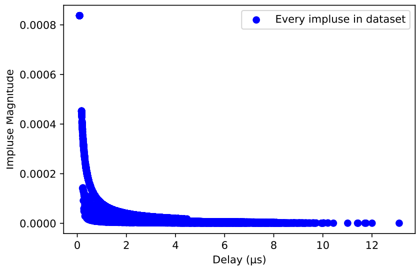
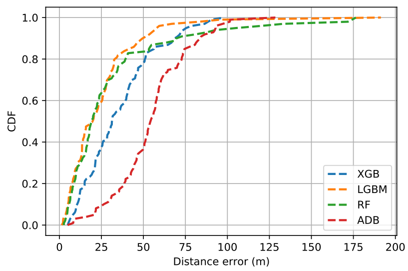
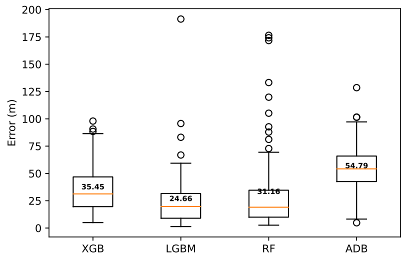


Figure 3.3: Overview of impulse responses: 9239 samples simulated between 400 Tx and 25 Rx in London city



(a) CDF of localization errors



(b) Boxplot of localization errors

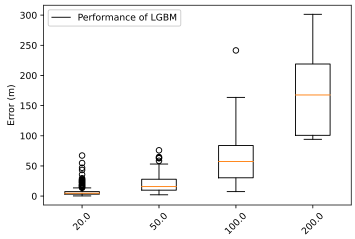
Figure 3.4: Performance evaluation for ensemble methods

3.3 Localization with transferred scenarios

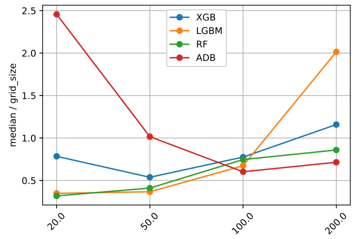
The formal experiments focuses on training model to determine the device position homologous as the database. However, when the trained model is applied to different scenarios, the precision may dramatically drop [9]. Generalizing a model is imperative since in practice only small portion of data can be collected and trained, a model inevitably faces challenges of coming data. Therefore it is needed to enhance the transferability of the model to process the heterologous signal, let's say, from a different map. For example a model trained by fingerprinting database from London, processes the data from Paris. Thankfully some prior information from heterologous signal can be obtained in advance, such as the street map, the position of the station etc, which may help transfer the model.

3.3.1 Transferring framework

Transferring is to train a model with source data and employ it in target data. A novel transfer framework is proposed, combining two phases:



(a) LGBM with varying grid size



(b) Mean/Grid size error w.r.t grid size

Figure 3.5: Performance evaluation w.r.t grid size

Point-to-point distance estimation: Finding a mapping f that maps CIR feature of i-th transmitter $H_i = [H_{i,1}, H_{i,2}, \cdots, H_{i,R}]$ to distances between the transmitter and all the receivers $D_i = [D_{i,1}, D_{i,2}, \cdots, D_{i,R}] \in \mathcal{R}^R$

$$f : H_i \rightarrow D_i$$

In the implementation, the mapping f for distance estimation is replaced by ensemble methods.

Optimal matching: Finding a function that determines i-th transmitter position whose distance with all Rx matches most as the estimated distances D_i, with knowing Rx positions P.

$$\begin{aligned} &\arg\min_{z \in \mathcal{R}^2} \sum_{i=1}^{R} (d(z, P_i) - D_i)^2 \\ s.t. \quad & z(0) \in (-90, 90), \ z(1) \in (0, 180) \end{aligned}$$

where $P \in \mathcal{R}^{2\times R}$, is a 2D array for Rx positions, $d(,)$ is a function measuring euclidean distance based on two geolocations. It is a deterministic method and requires no prior training.

The general architecture of transferring a model to a known scenario is illustrated in Figure. 3.6. It includes two parts: offline training and online testing. Firstly a source database is used to train a model for distance estimation in advance. Then the model is directly taken to estimate transmitter-wise distance, and then employ an optimizer, with addition of receivers' location information to find the matching position.

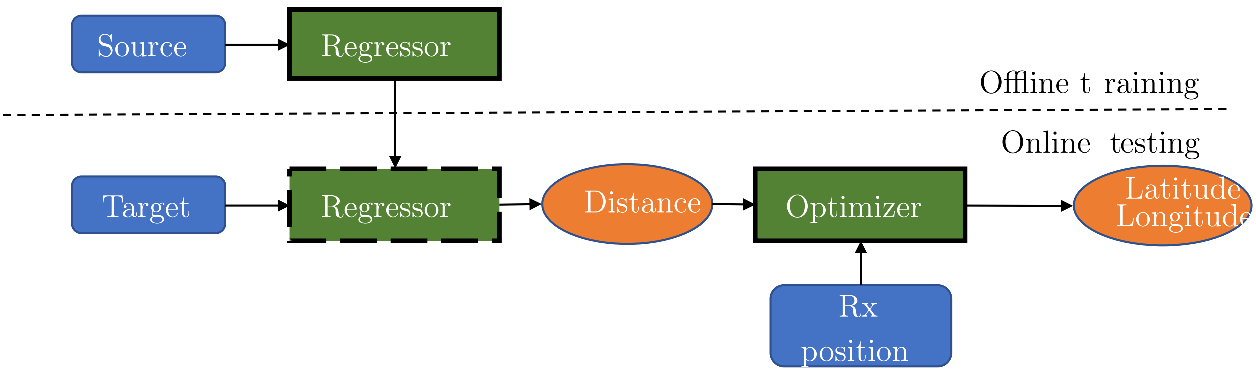
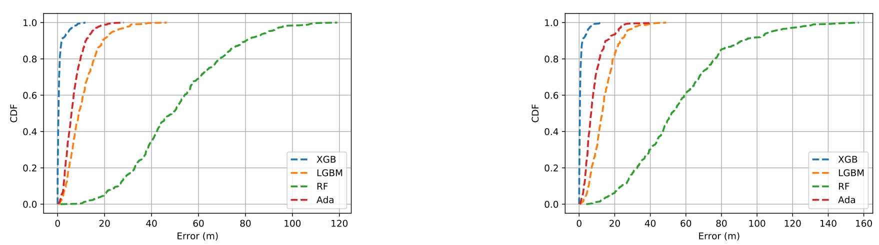


Figure 3.6: Transferring structure

3.3.2 Experiments

For experiments, London is chosen the data source testbed, and Paris and Toulouse are transferred scenarios. The setup of sites are the same: 20×20 Tx mesh and 5×5 Rx mesh uniformly dispersed in a 1km×1km square area with 50m grid size. There are 400 samples for each fingerprinting database.

Firstly, the impact of regressor is tested. Four ensemble methods are selected with positioning tasks in Paris and Toulouse. Following the workflow in Figure. 3.6, the regressor is feed by London fingerprinting database, then tested with target data and followed by a optimizer to determine the position. The result is presented in Figure. 3.7. XGB outperforms on other methods. Considering the grid size of 50m, the results of XGB, LGBM and AdaBoost are acceptable.

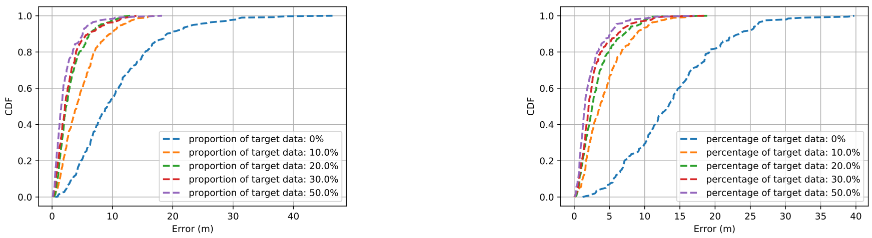


(a) Positioning with Paris data (b) Positioning with Toulouse data

Figure 3.7: Positioning with different scenarios and regressors

Besides, the impact of supplemented target data is explored. Different proportion of target data is to source fingerprinting database in training in order to augment the database. The result is shown below with LGBM as the regressor.

From Figure. 3.8, it is interesting to witness the evident improvement with



(a) Positioning in Paris with supplemented data

(b) Positioning in Toulouse with supplemented data

Figure 3.8: Positioning with proportional supplemented target data

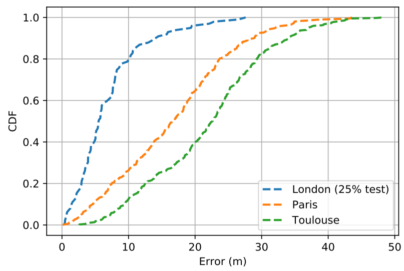


Figure 3.9: Positioning for different cities

supplemented target data. The more target data are supplied, the more accurate the positioning will be.

Lastly, the impact of heterogeneity of data is analyzed. It is vital to verify if heterogenous data will moderate the performance. 75% of data from London (Source) is taken for training the regressor (set as LGBM), and the other 25% is left as target data for testing and comparing with database from Paris and Toulouse. Although it should be noted the test size is unbalanced among the three groups. There are 100 samples for London test, and 400 for Paris and Toulouse. The result is demonstrated in Figure. 3.9. It turns out the model performs better on its homologous data indeed.

A quantitative summation of the localization performance under proposed framework in transferred scenarios is introduced in Table. 3. The localization error is well controlled considering a granularity of 50m.

City	Mean Error (m)	Maximum Error (m)	$p(\|e\| \leq 20m)$
Paris	16.6	43.6	64.0%
Toulouse	21.9	48.2	40.6%

Table 3: Localization performance in transferred scanerios

4 Conclusion

In this project, we deployed a simulator for outdoor WSN, based on Matlab module, and generated multiple fingerprinting database with channel state information measurements. Ensemble learning methods are adpated to train the fingerprint map and the relation between localization precision and grid size is explored. Furthermore, a novel localization scheme is proposed to deal with transferring scenarios. Specifically, a regressor is trained offline for distance estimation in first step. And an optimizer is designed to localize a fitting position by distances. The experimental results demonstrate that the proposed schema can achieve satisfactory localization precision.

Nevertheless, the work is never finished. The method to process CSI measurements could be properly improved. More realistic scenrios can be taken into consideration such as noise, diffraction etc. Last but not least, the experiments yet are built on simulation. Physical experiments should be conducted to validate the methods.

References

[1] Quoc Duy Vo and Pradipta De. A survey of fingerprint-based outdoor localization. *IEEE Communications Surveys Tutorials*, 18(1):491–506, 2016.

[2] José A. del Peral-Rosado, Ronald Raulefs, José A. López-Salcedo, and Gonzalo Seco-Granados. Survey of cellular mobile radio localization methods: From 1g to 5g. *IEEE Communications Surveys Tutorials*, 20(2):1124–1148, 2018.

[3] Michel Bakni, Luis Manuel Moreno Chacón, Yudith Cardinale, Guillaume Terrasson, and Octavian Curea. WSN simulators evaluation: an approach focusing on energy awareness. *CoRR*, abs/2002.06246, 2020.

[4] Fei Yu. A survey of wireless sensor network simulation tools. 2011.

[5] Inc. The MathWorks. *Communication Toolbox*. Natick, Massachusetts, United State, 2022.

[6] Zheng Yang, Zimu Zhou, and Yunhao Liu. From rssi to csi: Indoor localization via channel response. *ACM Comput. Surv.*, 46(2), dec 2013.

[7] Josyl Mariela Rocamora, Ivan Wang-Hei Ho, Wan-Mai Mak, and Alan Pak-Tao Lau. Survey of csi fingerprinting-based indoor positioning and mobility tracking systems. *IET Signal Processing*, 14(7):407–419, 2020.

[8] Adeniran Ademuwagun and Verdicchio Fabio. Reach centroid localization algorithm. *Wireless Sensor Network*, 09:87–101, 01 2017.

[9] Qiang Yang, Yu Zhang, Wenyuan Dai, and Sinno Jialin Pan. *Transfer Learning*. Cambridge University Press, 2020.