Article

Fusion of Channel State Information and Received Signal Strength for Indoor Localization Using a Single Access Point

David Sánchez-Rodríguez [1,2,*,†], **Miguel A. Quintana-Suárez** [2,†],
Itziar Alonso-González [1,2,†], **Carlos Ley-Bosch** [1,2,†] and **Javier J. Sánchez-Medina** [3,†]

[1] Institute for Technological Development and Innovation in Communications,
University of Las Palmas de Gran Canaria, Campus Universitario de Tafira,
35017 Las Palmas de Gran Canaria, Spain; itziar.alonso@ulpgc.es (I.A.-G.); carlos.ley@ulpgc.es (C.L.-B.)

[2] Telematic Engineering Department, University of Las Palmas de Gran Canaria,
Campus Universitario de Tafira, 35017 Las Palmas de Gran Canaria, Spain; mangel.quintana@ulpgc.es

[3] Innovation Center for Information Society, University of Las Palmas de Gran Canaria,
Campus Universitario de Tafira, 35017 Las Palmas de Gran Canaria, Spain; javier.sanchez@ulpgc.es

* Correspondence: david.sanchez@ulpgc.es; Tel.: +34-928-458-047; Fax: +34-928-451-380

† These authors contributed equally to this work.

check for
updates

Abstract: In recent years, indoor localization systems based on fingerprinting have had significant advances yielding high accuracies. Those approaches often use information about channel communication, such as channel state information (CSI) and received signal strength (RSS). Nevertheless, these features have always been employed separately. Although CSI provides more fine-grained physical layer information than RSS, in this manuscript, a methodology for indoor localization fusing both features from a single access point is proposed to provide a better accuracy. In addition, CSI amplitude information is processed to remove high variability information that can negatively influence location estimation. The methodology was implemented and validated in two scenarios using a single access point located in two different positions and configured in 2.4 and 5 GHz frequency bands. The experiments show that the methodology yields an average error distance of about 0.1 m using the 5 GHz band and a single access point.

1. Introduction

Localization systems have provided new services to users in compliance with their positions. In fact, two decades ago, authors in [1] anticipated a rapid increase of applications involving user location awareness. Even so, nowadays, new localization approaches continue to develop due to advancement of new technologies, sensors and communication standards.

The Global Navigation Satellite System (GNSS) is widely used for positioning in outdoor environments. Nevertheless, it does not operate in indoor environments, when there is non line-of-sight (NLOS) with satellites, such as inside dwellings and constructions [2]. Hence, indoor scenarios demand approaches that use other technologies capable of sensing location and achieving high accuracy.

In the literature, different wireless communications standards have been used for indoor positioning. Among them, Wi-Fi [3], ZigBee [4], RFID [5], BLE [6] or ultrasound [7], have been the most used standards. Wi-Fi-based solutions have become very popular due to their wide deployment, low cost, and mature standardization state. Authors in [8] , in addition to providing a

summary of current 802.11 standard with reference to localization, also present the main challenges in time-based positioning systems based on Wi-Fi technologies. Furthermore, they expose a survey on the proposed methods designed to solve the aforementioned challenges, classifying them according to the communication layer in which each method is implemented. Many of these work using a time-based geometric location such as time of arrival (ToA) [9] , round trip time (RTT) [10] or time difference of arrival (TDoA) [11] , even angle of arrival (AoA) [12]. Nevertheless, fingerprinting-based indoor localization systems are in the spotlight, being their benefits the simplicity, low hardware requirements and non requiring additional sensors to infer user location [13,14].

The fingerprint of each scenario is usually built with RSS or CSI measurements received from one or more access points at each reference location. RSS has been widely used for the last two decades, and CSI has been applied for the last ten years. In fact, Radar was one of the first localization systems based on the fingerprinting technique using RSS [15], whereas one of the first localization studies based on CSI was published in [16]. In [17], a paper is presented on how to apply CSI from the fundamentals used in RSS for indoor localization.

CSI is capable of characterizing the channel frequency response between the transmitter and the receiver using the sub-carriers of an OFDM communication. The CSI amplitude and phase of each sub-carrier are influenced by the motions and shifts of the receiver, transmitter, and surrounding objects and humans [17,18]. Hence, CSI provides more fine-grained physical layer information than RSS, and can be also used for localization and tracking services yielding better accuracy. On the other hand, although the RSS is not stable for time providing coarse-grained information, it can offer useful information when it is fused with CSI, improving the accuracy and error distance of the indoor localization systems.

Although CSI provides more fine-grained physical layer information than RSS, in this manuscript a methodology for indoor localization fusing both features from a single access point is proposed to provide better accuracy. In addition, CSI amplitude information is processed to remove high variability information that can negatively influence location estimation. Thus, the main contributions of this work are:

- A methodology based on the fusion of CSI amplitude and RSS information is proposed improving the accuracy of indoor localization approaches that use a single feature.
- CSI amplitudes obtained in MIMO communications systems are processed for removing high variability information and building stable datasets.
- The proposed methodology achieves high accuracy and sub-meter error distance using a single access point.

This manuscript is organized as follows: Section 2 summarizes the related work about indoor localization techniques based on RSS and CSI. Section 3 briefly introduces the CSI and its behaviour in indoor environments. In Section 4 the methodology based on the fusion of CSI amplitude and RSS measurements is proposed. Section 5 describes the scenarios and datasets built in this work. After that, in Section 6, the setup of classifiers, the performance evaluation, the robustness, and the comparison of proposed approach with other research works are described and discussed. Finally, Section 7 contains the conclusion and future works.

2. Related Work

A lot of research on indoor localization using RSS and CSS measurements can be found in the literature. Starting by techniques based on RSS, authors in [19] presented a technical review of several WLAN fingerprinting-based localization solutions and their associated challenges. They start by discussing both fundamentals principles and Wi-Fi based conventional positioning methods. Next, they present current state-of-the-art solutions, and continue by analyzing practical deployment problems to be faced when developing real systems. They classified RSS-based fingerprinting techniques according to propagation models or radio maps.

Also, in [20] the authors presented an extensive review of existing solutions intended based on RSS to improve indoor location estimation. First, the error sources in common localization approaches are considered. Next, their work presents research on several methods for estimating location: starting with probabilistic techniques, continuing with the study of hybrid methods like multimodal fingerprinting, and following with localization methods based on the fusion of spatial contexts. Instead of confining to specific issues, they apply a conceptual framework to propose strategies in order to improve location estimation schemes.

RSS-based techniques are negatively influenced by the multipath effect of indoor environments [21] providing a coarse-grained information. Thus, in order to obtain high accuracy, RSS information from several access points are required [22,23]. Thereby, authors in [24] addressed two main problems in existing algorithms based on weighted K-nearest neighbors (WKNN) and used in Wi-Fi fingerprinting localization: variability in detected AP sets over time and space, and low efficiency in WKNN clustering procedures due to the fact that not all the K neighbours are always near to the user position. They cope with the first problem when calculating fingerprint distance by calculating the similarity of AP sets and combining it with RSS measurements. They solve the second problem by improving the clustering algorithm with the detection and elimination of isolated points with semi-supervised affinity propagation. In recent studies, such as [25], a single access point is used for indoor positioning yielding a good accuracy. Nevertheless, the average error distance can reach up to 3 m, which is a high error value for high demanding applications.

On the other hand, channel state information has been recently employed for indoor positioning systems achieving sub-meter level accuracy, since it provides more fine-grained physical layer (PHY) information of the communication channel. CSI furnishes both the amplitude and phase information of each sub-carrier in OFDM-based communication systems, and these features are used to estimate the location. In recent years, numerous indoor localization systems based on fingerprinting and employing CSI have been proposed. For instance, in [26], the authors used a deep network with four hidden layers using as input the CSI amplitudes. This work yields an error distance of 0.95 m with a single access point. Authors in [27] used a visibility graph to transform the CSI amplitude and phase information of individual sub-carriers into complex networks extracting the network features accordingly. Then, machine learning algorithms are applied to infer the locations, yielding a 96% accuracy.

Authors in [16] proposed a solution which, based on effective CSI values, estimates distance through the use of a developed propagation model, though their system uses also fingerprinting for localization. In the off-line phase, they build the radiomap by evaluating the received signal at each reference position. In the on-line phase, they employ a probability model based on the received strength of each packet by summing the powers of all sub-carriers. They treat the received signal strength at each location as a Gaussian variable, employing Pearson correlation to quantify similarity between measured CSI and stored fingerprints. In terms of accuracy, their system achieves a mean distance error of 0.65 m. In [28], the gathered CSI amplitudes are passed through the k-nearest neighbor (KNN) classifier to compute the location, yielding an error distance about 0.9 m. However, two access points are used in this work.

In [26], the authors presented a CSI fingerprinting scheme based on deep machine learning. In the off-line phase, the system model is obtained by using deep learning to train weights of a deep network. In the on-line phase, a probabilistic methodology is employed to infer the position. CSI fingerprints are obtained by normalizing the sub-carrier amplitudes. Such values are analyzed using four hidden layers of a deep neural network. Depending on the testing environments and conditions, they obtain mean error values varying from 0.95 to 1.8 m.

In [27], authors developed a CSI-based indoor localization technique, combining both intra-subcarrier statistics and inter-subcarrier network features. In the off-line phase, their system considers relationships between subcarriers to extract position information. A Visibility Graph is constructed to model inter sub-carrier correlations. Then, they apply machine learning techniques to

form the fingerprint library. In the on-line phase, testing data are processed similarly, and machine learning algorithms are used to estimate locations. With the best Support Vector Machine (SVM) algorithm, their combined method achieves 95.7% accuracy.

In [29], a device-free CSI indoor fingerprint localization algorithm based on phase difference processing and Principal Component Analysis (PCA) was proposed. In the off-line phase, the algorithm calculates phase differences to correct random time and phase shifts in the communication links. Then, they apply a PCA method to reduce the dimensionality of the denoised data to produce a robust fingerprint database. Lastly, the correspondence between sampling signal and physical locations is established using a back-propagation neural network. In the on-line phase, real-time measured CSI values are classified by means of a back-propagation neural network. The location is estimated by computing a network model based on differential phase processing. They obtain a mean error distance of 1.42 m. In addition, the authors (in another study [30]) used the fusion of CSI amplitude and phase information to enhance the performance of their approach. Besides, to avoid interference with other indoor signals, these features are filtered. Experimental results were tested on two different scenarios yielding an average positioning error between 0.5 and 1.2 m.

Furthermore, in [31,32], the authors combined the CSI amplitude with the time-of-flight (ToF) and the angle-of-arrival (AoA), respectively. These works yield an average error distance of about 0.6 m, nevertheless, multiple frequency bands and multiple APs are used, respectively. Further, authors in [33] presented a CSI-based localization system with a deep learning based algorithm employing both average amplitudes and estimated angles of arrival for fingerprinting. In the off-line phase, the proposed system uses estimated AoAs and average amplitudes employing deep learning to train and store the weights and building a bi-modal fingerprint database. In the on-line phase, they take in a probabilistic model for estimating location based on the fingerprint database and bi-modal test data. According to their results, mean error varies from 1.57 to 2.15 m employing 5 GHz WiFi band.

In [34], authors employed AoA and RSS measurements for indoor localization. The AoA is estimated based on the CSI obtained from the existing Wi-Fi APs. Then, they evaluate a Multiple Signal Classification (MUSIC) algorithm using either AOA or RSS only, and both together. The approach is validated in a 10 m × 12 m room with four APs, and can achieve sub-meter localization accuracy. However, they need all four APs and the computational cost is proportional to the number of APs.

Authors in [35] presented a Wi-Fi-based positioning system combining both RSS and CSI data with the aim of producing a robust fingerprint. Both in the offline and online phase apply a data pre-processing. First, to reduce noise and multipath effects in CSI they apply a time domain filtering technique. Then, for reducing redundancies, authors divide the communication channel into four sub-channels, averaging channel responses of sub-carriers in each of the sub-channels. Finally, location is estimated by means of an improved WKNN algorithm yielding an average error distance of 1.79 m. Although this work provides a solution similar to the our approach its level accuracy is less, and the data pre-processing and time domain filtering can have a high computational cost.

In [18], a survey of WiFi sensing with CSI is presented and the works are grouped by signal transform and signal extraction techniques, being the Fourier and Discrete Wavelet widely used. In all of them, the analysis of CSI amplitude and phase of the communication links play an important role in processing of data. Moreover, MIMO communications permit to establish multiple transmitter-receiver links at the same time. Therefore, the amount of data to be processed increases compared to systems based on RSS. Thus, in our work, with the aim of removing high variability information, the best transmitter–receiver pair of MIMO communication is analysed and selected in the frequency domain. This processing is carried out using the CSI amplitudes only and avoiding the joint use of the phase, in the early stages of pre-processing.

There is no doubt that substantial progress has been carried out in the advancement of approaches for indoor localization, using RSS measurements, and more recently, by means of CSI dealing with amplitude and phase information of sub-carriers in OFDM systems. Nevertheless, to the best of our knowledge, the research presented in this study is the first to propose a methodology that combines

both CSI amplitude and RSS information from a single access point for indoor localization yielding sub-meter level accuracy.

3. Channel State Information

OFDM is widely used in wireless networks standards, such as IEEE 802.11 g/n/ac [36], IEEE 802.16 [37], LTE [38], 5G New Radio [39] and so on. In OFDM systems, the available spectrum bandwidth is partitioned into multiple orthogonal sub-carriers, and the data are transmitted over those sub-carriers at different frequencies. Thus, CSI gives information on the characteristics of the wireless channel providing information how the signal is propagated from the transmitter to the receiver to a certain frequency carrier. The amplitude and phase of the sub-carriers are affected by multipath, fading, and scatter effects. Furthermore, the presence of objects, humans or movement directly affects both the amplitude and phase of the signal [17,18]. Therefore, CSI picks up the behavior of the wireless signal when transmitted from a transmitter to a receiver and makes it a valid measurement for use in different detection applications in WiFi. It can be used to obtain the channel quality of the communication link, and therefore it provides more fine-grained physical layer (PHY) information. Thus, phenomenons due to the multipath effect, such as scattering and fading, can be easily modelled. The channel model in a multipath transmission can be expressed as:

$$\vec{Y} = H \times \vec{X} + \vec{N} \tag{1}$$

where $\vec{Y}$ is the received signal vector, $\vec{X}$ is the transmitted signal vector, H denotes the channel's frequency response, and $\vec{N}$ represents the white Gaussian noise. H can be computed from $\vec{Y}$ and $\vec{X}$, and it can be expressed in terms of S sub-carriers as:

$$H = \begin{bmatrix} H_1 & H_2 & \cdots & H_S \end{bmatrix} \tag{2}$$

The CSI for each sub-carrier is expressed as:

$$H_i = |H_i| \times e^{j \sin(\angle H_i)} \tag{3}$$

where the amplitude and phase values of the i-th sub-carrier are denoted by $|H_i|$ and $\angle H_i$, respectively. For MIMO systems, where multiple transmitter and receiver antennas are used, the CSI is a $t \times r \times S$ complex matrix, where t represents the numbers of transmitter antennas, and r denotes the numbers of antennas at the receiver.

Figure 1 shows a comparison between 30 sub-carrier CSI amplitudes and the RSS values from a single receiver antenna for 100 consecutive packets at a given location. Specifically, both the CSI and the RSS measurements were collected at the same time and in the same location, about 12 m away from the access point in the line of sight. As can be observed, the CSI measurements of each subcarrier vary slightly while the RSS values show a great variability. Therefore, CSI has better stability over time, which is more appropriate for indoor localization.

On the other hand, in our preliminary experiments, it was noticed that CSI amplitude values show signs of great steadiness for continuously received packets at the same location. Nevertheless, in MIMO communications, the channel frequency response at each antenna was different. This effect can be observed in Figure 2, in which 30 sub-carrier amplitudes for 100 consecutive packets in each receiver antenna is shown. As can be seen, the channel frequency response has different properties for each receiver antenna, that is, the CSI signals are time variance from an receiver antenna to another. Moreover, as it is observed in Figure 2a, the A antenna (red color) has a great variability for consecutive packets, and therefore can introduce noise into the classifier. In addition, it has been observed that measurements acquired at a given location while alternating different laptop orientations (North, South, East and West) present different MIMO communication patterns. Hence, even if the receiver has a wireless card with three antennas, the system does not always achieve 1×3 MIMO transmissions.

It can be observed in Figure 2c where two receiver antennas are only used when the device is facing East, a 1×2 MIMO communication is done. Thus, to reduce the received noise and avoid a high variability at the input of the classifier, in this research, the processing of the CSI matrix has to be carried out in order to select the most stable transmitter-receiver antenna pair.
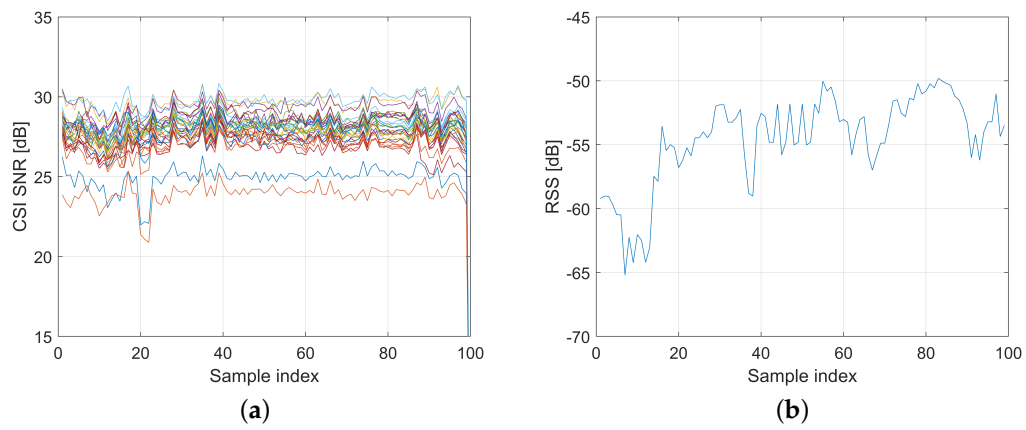

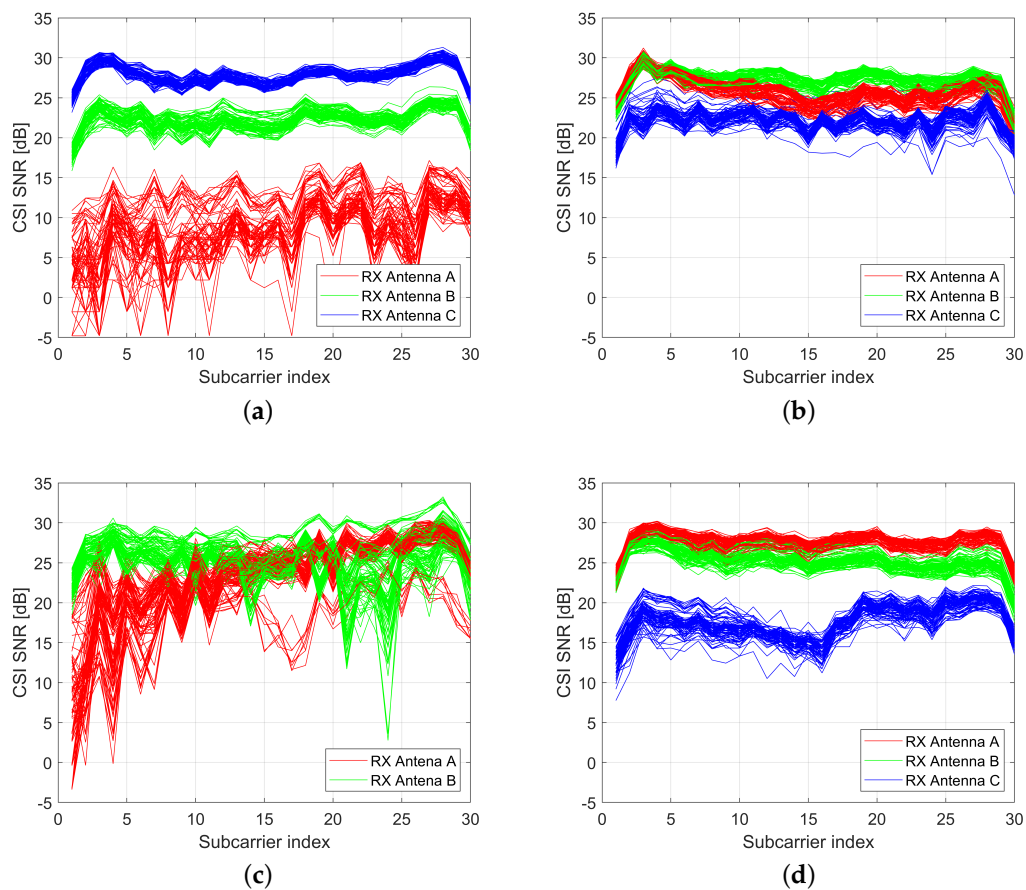**Figure 1.** CSI vs RSS. (**a**) CSI amplitudes. (**b**) RSS.

Figure 2. Channel frequency response for 100 consecutive packets at a given location but different laptop orientations. (**a**) North. (**b**) South. (**c**) East. (**d**) West.

4. Methodology

The methodology followed in this research work is presented in Figure 3. In each scenario, several reference points are defined and, therefore, the position of each them is known. Then, the methodology is developed in five phases: data collection, CSI amplitude processing, features fusion, fingerprint dataset, and classification. First, data is collected at every defined location, such as CSI and RSS values from an IEEE 802.11 access point. As it was aforementioned, CSI information specifies the amplitude and phase of the signal path between a single transmitter-receiver antenna pair. In addition, 1×3 MIMO transmissions are not always achieved and depending on environment conditions CSI may be more variable. Therefore, in the next phase, titled CSI Amplitude Processing, CSI information is processed to extract the sub-carriers amplitude for the transmission with the smallest variability, that is, the best transmitter–receiver communication pair. After that, a vector is built by fusing these attributes. Next, the fingerprint dataset is built with these vectors, and finally, they are employed at the input of machine learning algorithm to create the model. Next, each phase is described.
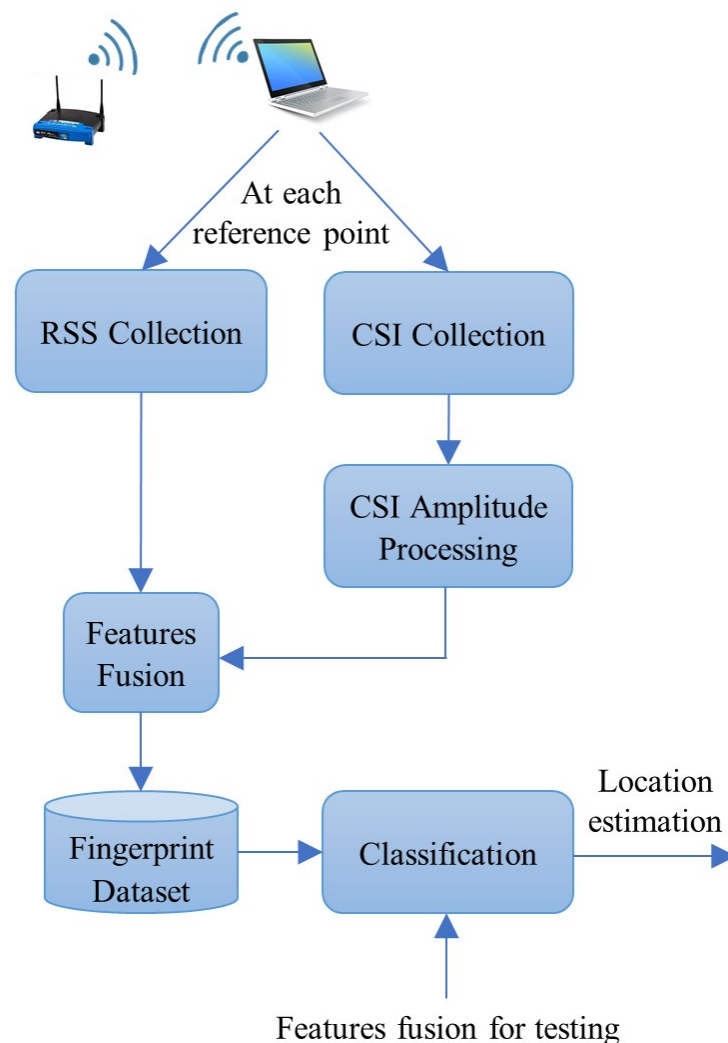

Figure 3. The proposed methodology with CSI and RSS fusion.

4.1. Data Collection

In order to build an indoor localization system based on fingerprinting technique, a task to collect the valuable information that characterizes the indoor scenario has to be addressed. Thus, each defined

reference location is represented by its own fingerprint which in this research is composed by CSI and RSS information collected from a single access point. Equation (4) shows the data collection, where $|H^r_{i,s}|[\tau]$ and $\angle H^r_{i,s}[\tau]$ represent the amplitude and phase of the *s*-th subcarrier of the *r*-th receiver antenna collected at the *i*-th reference location, respectively. In addition, $\varphi_i[\tau]$ is the RSS measurement, in dB, collected at the *i*-th reference location. The index of samples gathered at every location is indicated by $\tau = 1, ..., N$, where *N* represents the total number of samples in each position, *R* indicates the number of reference locations, and values of additional sensors is denoted by θ. Thus, θ indicates the orientation angle of the device if a direction finder is accessible. The number of sub-carriers, *S*, depends on IEEE 802.11 standard employed for communications, and the number of receiver antennas, *r*, depends on wireless network card implementation.

$$\Psi = \begin{pmatrix} |H^1_{1,1}|[\tau] & \cdots & |H^r_{1,S}|[\tau] & \angle H^1_{1,1}[\tau] & \cdots & \angle H^r_{1,S}[\tau] & \varphi_i[\tau] & \theta \\ |H^1_{2,1}|[\tau] & \cdots & |H^r_{2,S}|[\tau] & \angle H^1_{2,1}[\tau] & \cdots & \angle H^r_{2,S}[\tau] & \varphi_2[\tau] & \theta \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ |H^1_{R,1}|[\tau] & \cdots & |H^r_{R,S}|[\tau] & \angle H^1_{R,1}[\tau] & \cdots & \angle H^r_{R,S}[\tau] & \varphi_R[\tau] & \theta \end{pmatrix} \tag{4}$$

On the other hand, as it was aforementioned, CSI is more stable than RSS. The signal strength oscillates around an average value at a every position [21]. Therefore, a huge data collection has to be carried to diminish the RSS instability. However, as it will be demonstrated in the Section 6, the proposed fusion in this research work presents high accuracy with a reduced training dataset.

4.2. CSI Amplitude Processing

As it was said in the previous section, depending on the scenario and the device location, not all receiver antennas are always used in MIMO communications. In addition, the transmitter-receiver pair with higher sub-carrier amplitudes present a greater stability. Hence, with the aim of providing a feature set with small variability at the input of the classifier, the processing of CSI values is carried out in this research, in order to select the best transmitter-receiver pair. For that purpose, the average of each pair is computed as:

$$|CSI| = \frac{1}{S} \sum_{i=1}^{S} \sqrt{|H_i|^2 + \angle H_i^2} \tag{5}$$

where *S* is the number of sub-carrier, and $\angle H_i$ and $|H_i|$ represent the phase and amplitude information of the i-th sub-carrier, respectively. This information is provided by the wireless network card driver as a complex number, and therefore, $|CSI|$ represents the averaged magnitudes of S complex numbers. Once the $|CSI|$ value of each pair is computed, the pair with the greater value is selected. Therefore, the amplitudes of each sub-carrier of the selected pair are fused with RSS value in the next phase.

4.3. Features Fusion

Once the CSI amplitude processing phase has been carried out, both the selected CSI amplitudes, $|H_i|$, and RSS values gathered at each reference location are fused. These features are concatenated into a vector as in Equation (6), where *r* represents the receiver antenna selected in the previous phase, and *i* is the reference location.

$$\begin{bmatrix} |H^r_{i,1}| & \cdots & |H^r_{i,S}| & \varphi_i & \theta \end{bmatrix} \tag{6}$$

4.4. Fingerprint Dataset

Each vector of the features fusion phase represents a row of fingerprint datasets. Thus, the fingerprint dataset contains as many rows as the number of gathered samples, NxR, as it is

showed in Equation (7). In addition, the coordinates or the number of each reference location, RL_j are added to each row. Hence, an observation is represented in each row, RL_j is the nominal class.

$$
\begin{bmatrix}
|H^r_{1,1}|[1] & \cdots & |H^r_{1,S}|[1] & \varphi_1[1] & \theta & RL_1 \\
\vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
|H^r_{1,1}|[N] & \cdots & |H^r_{1,S}|[N] & \varphi_1[N] & \theta & RL_1 \\
|H^r_{2,1}|[1] & \cdots & |H^r_{2,S}|[1] & \varphi_2[1] & \theta & RL_2 \\
\vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
|H^r_{2,1}|[N] & \cdots & |H^r_{2,S}|[N] & \varphi_2[N] & \theta & RL_2 \\
\vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
|H^r_{R,1}|[N] & \cdots & |H^r_{R,S}|[N] & \varphi_R[N] & \theta & RL_R
\end{bmatrix}
\tag{7}
$$

4.5. Classification

Our methodology approaches localization as a classification problem, where the categories to predict the number of reference locations are used to build the fingerprint. Thus, the aim is to predict which one is the closest to a particular unknown localization in each case. In order to validate the proposed methodology, two simple machine learning algorithms were used. The classification algorithms assessed for inferring the location are based on the boosting and the nearest neighbor techniques. The boosting method was implemented with the AdaBoost algorithm [40] using the C4.5 algorithm [41] as a weak classifier. On the other hand, the IBk implementation [42] of the K-Nearest Neighbor (KNN) algorithm [43] was also used as a classifier. More sophisticated algorithms can be used for indoor localization, such as neural networks or convolutional neural networks. However, they spend more computation time, and are not necessarily required to validate the proposal.

5. Datasets

In order to validate the methodology presented in this paper, eight datasets have been employed. The datasets contain both CSI and RSS measurements collected from a single IEEE 802.11 access point in two scenarios. The datasets were built in two places at the University of Las Palmas de Gran Canaria, Spain, concretely in the second floor of the Telematic Engineering Department building, named as Corridor, and also in the Laboratory of Line Transmission, named as Laboratory. In each scenario, four datasets were built varying the position of the access point, A or B, and the frequency band, 2.4 GHz with 20 MHz channel bandwidth, or 5 GHz with 40 MHz channel bandwidth.

An Acer Travelmate 6293 laptop with Intel WiFi Link 5300 802.11n wireless network card was used to gather both CSI and RSS measurements from the access point, which was configured using the 802.11n CSI tool [44]. This laptop has three internal antennas for wireless communications located behind the screen. The CSI tool provides the RSS measured by the receiving NIC at the input of each antenna port, and also provides CSI values in a format that reports the channel matrices for 30 subcarrier groups for every two sub-carriers at 20 MHz or every four sub-carriers at 40 MHz. Each matrix entry is a complex number that specifies the amplitude and phase of the signal path between a single transmitter-receiver antenna pair.

On the other hand, a Linksys 1900ACS access point with four external antennas was used. Therefore, depending on the scenario and device location, up to 4×3 MIMO communications can be achieved, and 90 raw CSI measurements (1×3 MIMO transmission) can be collected for each packet reception. In other words, 30 raw CSI measurements are collected from each reception antenna. If 1×2 or 1×3 MIMO communications are employed, the CSI Amplitude Processing phase of the proposed methodology is in charged of selecting the best transmitter-receiver communication pair.

5.1. Corridor Scenario

In Figure 4, we show the layout area of this scenario, which is approximately 12 m wide and 62 m long. The access point was located at positions A and B, though beacons from 7 unknown access points in the 2.4 GHz frequency band were detected too. The black dots indicate the 38 positions used in the off-line phase, which are allocated with an interval of 3.25 m long and 1.2 metre wide. The CSI and RSS measurements were gathered at each reference location in four orientations (with intervals of 90°). For each access point, orientation angle and reference position, 200 and 450 measurements were consecutively gathered every 250 ms for 2.4 and 5 GHz, respectively. Hence, each off-line dataset has 30,400 and 68,400 instances, respectively. In addition, 18 locations marked by red dots and with an interval about 3.25 m were used for the on-line phase. Each dataset was gathered by taking 220 and 450 measurements for 2.4 and 5 GHz, giving two test datatsets with 15,840 and 32,400 instances, respectively. In all datasets, each instance contains 30 CSI amplitudes of sub-carriers and one RSS value.

○ Training point
● Testing point
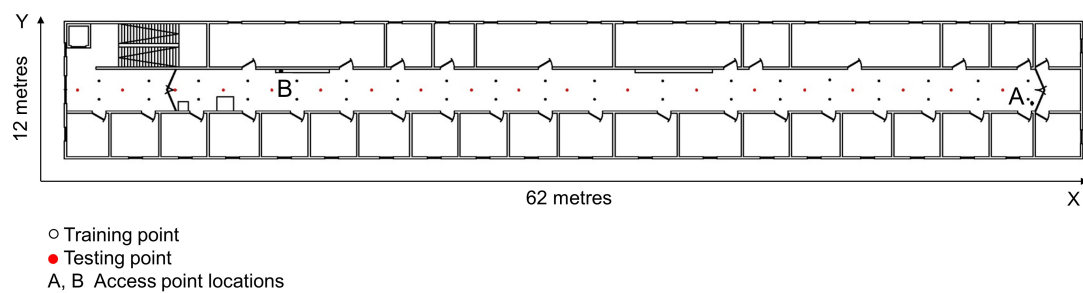A, B Access point locations

Figure 4. Corridor scenario.

5.2. Laboratory Scenario

In Figure 5, we show the layout area of this scenario, which is just about 4.5 m wide and 11.5 m long. The access point was located at positions A and B, though beacons from 4 unknown access points in 2.4 GHz frequency band were detected too. The black dots indicate the 30 positions used in the off-line phase, which are allocated with an interval about 1 metre. The measurements were gathered at each reference location in four orientations (with intervals of 90°). For each access point, orientation angle and reference position, 240 and 250 measurements were consecutively collected every 250 ms for 2.4 and 5 GHz, respectively. Hence, each off-line dataset has 28,800 and 30,000 instances, respectively. In addition, 15 locations marked by red dots and with an interval about 1 m were used for the on-line phase. In each one were also gathered 240 and 250 measurements for 2.4 and 5 GHz, giving two test datatsets with 14,400 and 15,000 instances, respectively. In all datasets, each instance contains 30 CSI amplitudes of sub-carriers and one RSS value.

5.3. Outline of Datasets

The main characteristics of the training datasets employed to verify the fusion of CSI and RSS information are shown in Table 1. In order to demonstrate that the reference locations have a statistically significant relationship with the fused features, a Multivariate Analysis of Variance (MANOVA) was computed. All four of the MANOVA test statistics (Pillai, Wilks, Hotelling and Roy) for all datasets of Table 1 returned a probability less than 0.001, therefore, the features are statistically significant.
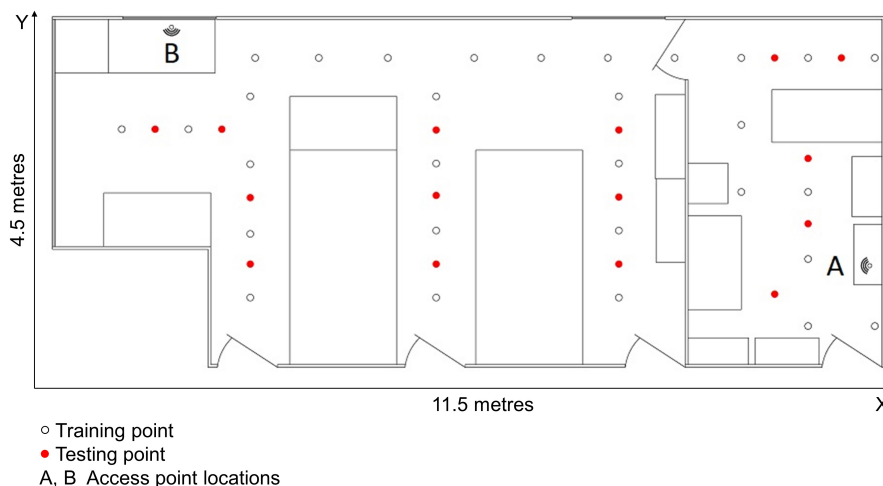
Figure 5. Laboratory scenario.

Table 1. Summary of training datasets.

Scenario	Frequency Band	Access Point	Locations	Instances Per Location and Orientation	Orientations	Instances
Corridor	2.4 GHz	A	38	200	4 (each 90°)	30,400
Corridor	2.4 GHz	B	38	200	4 (each 90°)	30,400
Corridor	5 GHz	A	38	450	4 (each 90°)	68,400
Corridor	5 GHz	B	38	450	4 (each 90°)	68,400
Laboratory	2.4 GHz	A	30	240	4 (each 90°)	28,800
Laboratory	2.4 GHz	B	30	240	4 (each 90°)	28,800
Laboratory	5 GHz	A	30	250	4 (each 90°)	30,000
Laboratory	5 GHz	B	30	250	4 (each 90°)	30,000

6. Results and Discussion

The performance evaluation of the methodology presented in this manuscript is analysed and discussed in this section. The error distance, accuracy and computation time are evaluated using two classical machine learning classifier algorithms, K-Nearest Neighbor (KNN) [42] and AdaBoost [40] using C4.5 algorithm [41] as a weak classifier. Furthermore, in order to evaluate the robustness of our proposal, different experiments were implemented by using distinct sizes of training sets. The proposed methodology has been evaluated using the multi-class accuracy metric. Simply put, accuracy, as in Equation (8), is the ration between the correctly classified observations and the total of observations or instances fed to the model. A correctly classified observation is that one which the output of the model matches with reference location of the observation used for testing. Additionally, in order to evaluate how adrift are badly classified positions, the error comparing the predicted position to the actual closer position is computed, by calculating the Euclidean distance of the two positions, location used for testing and predicted location. Then, the average error distance is computed by Equation (9), being N the number of tests, (x_i, y_i) the i-th actual location and $(\hat{x}_i, \hat{y}_i)$ the predicted location.

$$\text{Accuracy}(\%) = \frac{\text{Instances Correctly Classified}}{\text{Total Number of Test Instances}} \times 100. \tag{8}$$

$$\text{Average Error Distance} = \frac{1}{N} \sum_{i=1}^{N} \sqrt{(x_i, \hat{x}_i)^2 + (y_i, \hat{y}_i)^2} \tag{9}$$

The Weka API [45] was used to developed the proposed methodology. An Intel Core i7 3.4 GHz/32 GB RAM non-dedicated machine, running a MS Windows Operating System, was used to implement the experiments.

6.1. Classifiers Configuration

A wide set of simulations using the proposed methodology were carried out to find the best configuration of each classifier. Therefore, both the CSI amplitude processing and the feature fusion were applied before being used as input to the classifier. The main parameter of each classifier was varied in each simulation. Simulations were carried out using the training datasets described in Table 1, and implementing the 10-fold cross-validation technique. Hence, in each experiment, the dataset is split on 10 partitions, and 10 folds or rounds are run where in each one 9 partitions are used for training the model and the other is used for testing. In each round, a different partition is used for testing. In order to ensure statistical independence, 100 experiments were carried out for each dataset shuffling randomly the input data and averaging the results. Tables 2 and 3 show the average accuracy and the standard deviation for the KNN and AdaBoost classification algorithms, respectively.

Table 2. Accuracy for different values of k parameter (KNN).

Dataset	KNN				
	k = 1	k = 2	k = 3	k = 4	k = 5
Corridor-2.4 GHz-A	**97.14% ± 0.03**	96.32% ± 0.03	96.15% ± 0.03	95.75% ± 0.02	95.32% ± 0.02
Corridor-2.4 GHz-B	**96.19% ± 0.03**	95.16% ± 0.04	94.88% ± 0.05	94.44% ± 0.02	94.21% ± 0.04
Corridor-5 GHz-A	**99.36% ± 0.01**	99.18% ± 0.01	99.11% ± 0.01	98.99% ± 0.01	98.88% ± 0.01
Corridor-5 GHz-B	**98.69% ± 0.01**	98.25% ± 0.01	98.29% ± 0.01	98.08% ± 0.01	97.98% ± 0.01
Laboratory-2.4 GHz-A	**82.87% ± 0.08**	80.09% ± 0.06	81.26% ± 0.07	80.77% ± 0.07	79.95% ± 0.09
Laboratory-2.4 GHz-B	**95.36% ± 0.05**	94.30% ± 0.06	94.58% ± 0.02	93.99% ± 0.05	93.74% ± 0.06
Laboratory-5 GHz-A	**90.47% ± 0.04**	89.06% ± 0.06	89.19% ± 0.06	88.47% ± 0.06	87.42% ± 0.06
Laboratory-5 GHz-B	**96.76% ± 0.03**	96.23% ± 0.05	96.26% ± 0.03	95.10% ± 0.06	95.78% ± 0.04

Table 3. Accuracy for different values of confidenceFactor parameter (AdaBoost + C4.5).

Dataset	AdaBoost + C4.5			
	c = 0.20	c = 0.25	c = 0.30	c = 0.35
Corridor-2.4 GHz-A	**94.99% ± 0.09**	94.93% ± 0.09	94.84% ± 0.07	94.82% ± 0.06
Corridor-2.4 GHz-B	**94.15% ± 0.04**	94.08% ± 0.08	94.12% ± 0.07	93.95% ± 0.11
Corridor-5 GHz-A	98.29% ± 0.03	98.29% ± 0.04	**98.35% ± 0.03**	98.34% ± 0.03
Corridor-5 GHz-B	97.26% ± 0.04	**97.36% ± 0.03**	97.19% ± 0.04	97.26% ± 0.05
Laboratory-2.4 GHz-A	79.21% ± 0.15	79.35% ± 0.11	79.28% ± 0.14	**79.39% ± 0.08**
Laboratory-2.4 GHz-B	94.83% ± 0.07	94.78% ± 0.06	**94.98% ± 0.07**	94.72% ± 0.06
Laboratory-5 GHz-A	86.26% ± 0.16	**86.37% ± 0.13**	86.32% ± 0.13	85.74% ± 0.11
Laboratory-5 GHz-B	93.98% ± 0.11	94.01% ± 0.11	**94.04% ± 0.08**	93.26% ± 0.11

For the KNN algorithm, the k parameter was varied from k = 1 to k = 5. On the other hand, the experiments carried out for the boosting tehcnique were implemented varying the confidenceFactor parameter, c, from 0.20 to 0.35, in steps of 0.05. As can is shown, the best results using the KNN algorithm were obtained with k = 1, and as k increases the accuracy decreases slightly. For the AdaBoost algorithm, all configurations have almost identical accuracy with a maximum variation of 0.2%.

In the next sections, the parameter value that yields the best accuracy was used for the experiments.

In order to ensure that the differences between results obtained for different k and c parameters are statistically significant, a Student's dependent *t*-test analysis was computed. The obtained *p*-values are smaller than 0.05, except between k = 2 and k = 3. Therefore, this indicates that the null hypothesis is rejected.

6.2. Performance Evaluation of the Proposal

In this section, the performance evaluation of the proposed fusion of CSI and RSS information is discussed, comparing the achieved results with the case when features fusion are not used. Results were obtained using the simulation configurations aforementioned, and 100 experiments were carried out for each dataset to ensure statistical independence. Tables 4 and 5 show the performance results obtained using independent features, RSS and CSI, respectively. Table 6 shows performance results using the proposed methodology in this work. The average value of accuracy, error distance and the training time to build the model, as well as their standard deviations are indicated in all tables. As can be observed, and as expected, experiments using only RSS values obtain the worst performance. In this case no more than 76% accuracy is yielded. Nevertheless, if the CSI measurements are only used, the accuracy increases considerably. Indeed, a 99.15% accuracy is yielded in the corridor scenario using the access point at A location, configured at 5 GHz frequency band. Depending on the dataset, this means and increment ranging from 15% to 40% in accuracy. Hence, it is demonstrated that CSI values provide more precise and stable information to be used in indoor localization systems.

On the other hand, when the fusion of CSI and RSS information is used, the accuracy also increases compared when CSI is only employed. As can be observed, the best results are achieved when using the proposed fusion, both in terms of accuracy and error distance. In the corridor scenario, the best average error distance is about 0.39 m and 0.09 m for 2.4 and 5 GHz, respectively. In the laboratory scenario, where abundant multipath signal propagation exists, the best average error distance is about 0.15 m and 0.11 m for 2.4 and 5 GHz, respectively. In most of experiments, the results improve between 2% and 9% accuracy. Although the performance of indoor localization system based on only RSS data is poor, RSS information provides useful information when it is fused with CSI, thus improving the accuracy and error distance of the system.

Table 4. Average performance results using only RSS.

Dataset	KNN			AdaBoost + C4.5		
	Accuracy	Error Distance	Training Time	Accuracy	Error Distance	Training Time
Corridor 2.4 GHz-A	75.78% ± 0.06	4.22 m ± 9.96	28.8 ms ± 1.46	75.33% ± 0.05	4.33 m ± 10.12	2.8 s ± 0.10
Corridor 2.4 GHz-B	75.55% ± 0.06	5.24 m ± 12.71	29.2 ms ± 1.76	75.19% ± 0.05	5.32 m ± 12.08	2.8 s ± 0.12
Corridor 5 GHz-A	60.23% ± 0.03	7.47 m ± 12.64	32.5 ms ± 2.38	60.10% ± 0.04	7.48 m ± 12.64	4.7 s ± 0.30
Corridor 5 GHz-B	70.88% ± 0.04	3.11 m ± 6.79	32.2 ms ± 1.60	70.57% ± 0.04	3.15 m ± 6.83	8.1 s ± 0.46
Laboratory 2.4 GHz-A	42.95% ± 0.10	1.75 m ± 2.00	29.8 ms ± 1.04	42.67% ± 0.09	1.75 m ± 1.99	0.5 s ± 0.01
Laboratory 2.4 GHz-B	74.79% ± 0.07	0.77 m ± 1.64	30.3 ms ± 1.18	74.30% ± 0.07	0.78 m ± 1.66	2.9 s ± 0.11
Laboratory 5 GHz-A	46.79% ± 0.07	1.42 m ± 1.70	29.1 ms ± 1.28	46.31% ± 0.06	1.43 m ± 1.69	0.6 s ± 0.04
Laboratory 5 GHz-B	71.11% ± 0.05	0.94 m ± 1.81	30.7 ms ± 2.47	70.63% ± 0.07	0.95 m ± 1.82	2.7 s ± 0.12

Table 5. Average performance results using only CSI.

Dataset	KNN			AdaBoost + C4.5		
	Accuracy	Error Distance	Training Time	Accuracy	Error Distance	Training Time
Corridor 2.4 GHz-A	94.94% ± 0.05	0.74 m ± 4.87	30.5 ms ± 1.47	92.05% ± 0.13	1.16 m ± 5.84	53.2 s ± 2.70
Corridor 2.4 GHz-B	93.69% ± 0.02	1.10 m ± 6.48	30.3 ms ± 1.53	91.14% ± 0.09	1.64 m ± 7.82	55.7 s ± 2.62
Corridor 5 GHz-A	99.11% ± 0.01	0.17 m ± 2.65	36.6 ms ± 2.06	97.91% ± 0.03	0.40 m ± 3.84	127.1 s ± 5.54
Corridor 5 GHz-B	97.76% ± 0.01	0.36 m ± 3.26	36.4 ms ± 1.53	95.58% ± 0.05	0.73 m ± 4.59	151.8 s ± 7.37
Laboratory 2.4 GHz-A	73.11% ± 0.08	0.94 m ± 1.87	29.3 ms ± 1.54	68.60% ± 0.21	1.09 m ± 1.97	64.7 s ± 3.84
Laboratory 2.4 GHz-B	89.13% ± 0.05	0.40 m ± 1.39	29.2 ms ± 1.08	87.14% ± 0.14	0.48 m ± 1.51	44.1 s ± 2.48
Laboratory 5 GHz-A	86.86% ± 0.05	0.44 m ± 1.32	30.6 ms ± 1.59	82.24% ± 0.12	0.59 m ± 1.48	63.5 s ± 3.01
Laboratory 5 GHz-B	94.29% ± 0.02	0.19 m ± 0.98	30.1 ms ± 1.44	90.66% ± 0.11	0.33 m ± 1.28	47.9 s ± 2.72

Table 6. Average performance results using the fusion of CSI and RSS.

Dataset	KNN			AdaBoost + C4.5		
	Accuracy	Error Distance	Training Time	Accuracy	Error Distance	Training Time
Corridor 2.4 GHz-A	97.14% ± 0.03	0.39 m ± 3.31	36.8 ms ± 1.06	94.99% ± 0.09	0.74 m ± 4.36	52.2 s ± 4.76
Corridor 2.4 GHz-B	96.19% ± 0.03	0.78 m ± 5.69	37.5 ms ± 1.73	94.15% ± 0.04	1.22 m ± 6.88	54.1 s ± 4.68
Corridor 5 GHz-A	99.36% ± 0.01	0.09 m ± 1.70	46.6 ms ± 2.66	98.35% ± 0.03	0.26 m ± 2.82	120.4 s ± 13.2
Corridor 5 GHz-B	98.69% ± 0.01	0.13 m ± 1.61	46.0 ms ± 2.66	97.36% ± 0.03	0.26 m ± 2.11	126.0 s ± 15.8
Laboratory 2.4 GHz-A	82.87% ± 0.08	0.50 m ± 1.30	39.5 ms ± 3.23	79.39% ± 0.08	0.61 m ± 1.42	67.0 s ± 3.52
Laboratory 2.4 GHz-B	95.36% ± 0.05	0.15 m ± 0.86	39.3 ms ± 2.25	94.98% ± 0.07	0.17 m ± 0.89	36.6 s ± 2.04
Laboratory 5 GHz-A	90.47% ± 0.04	0.27 m ± 0.96	39.3 ms ± 2.41	86.37% ± 0.13	0.38 m ± 1.08	64.3 s ± 4.48
Laboratory 5 GHz-B	96.76% ± 0.03	0.11 m ± 0.74	40.2 ms ± 1.62	94.04% ± 0.08	0.19 m ± 0.93	41.7 s ± 1.78

In addition, it can be appreciated that the elapsed time to train the model is lower when the datasets contain only RSS values, due to the fact that each instance is only formed by a single value. If the dataset is composed by CSI data, each instance has 30 values, and therefore, the number of computations to build the model is greater. This consideration is better appreciated when AdaBoost algorithm is implemented.

Regardless to the dataset, classifiers based on KNN algorithm yields the best results. In addition, as it can be noticed, the results obtained using the 5 GHz frequency band yield a better accuracy compared to the 2.4 GHz band. It is due to the fact that the 2.4 GHz band is often completely saturated, and because the 5 GHz signal attenuates very quickly and therefore the interference are less noticeable. Amplitude of signals in this frequency band tend to decay faster over distance, specially when passing through walls. It is also noticed that the accuracy of the system depends on the access point location, even when LOS is feasible. Also, the datasets built when using a NLOS access point give rise to difficulties getting a precise location. This can be appreciated in the laboratory scenario when using the access point located at A position, since in this case most of reference locations are in NLOS referred to that AP.

Lastly, Figures 6 and 7 show the cumulative distribution function (CDF) for different datasets using the KNN algorithm with K = 1, for each access point position, A and B, respectively. As can be seen, regardless of access point location, the fusion of CSI and RSS information has the best performance, yielding an error of around 1 m with a 95% or above of probability in the corridor scenario. Furthermore, the error distance is about 2 m with a 90% or greater probability in the laboratory scenario. Similar values are achieved using only the CSI values, although the difference with the proposed methodology is more noticeable in the laboratory scenario due to the internal distribution of possible locations. In addition, as it was aforementioned, localization estimations using only RSS yield the worst results. Hence, the proposed methodology obtains the best accuracy, yielding an error distance about 1 m at the 90th percentile or above in most of experiments. Furthermore, these results are achieved by using a single access point. Lastly, it is important to note that the best accuracies were achieved using the 5 GHz frequency band due to it is less saturated than 2.4 GHz band.
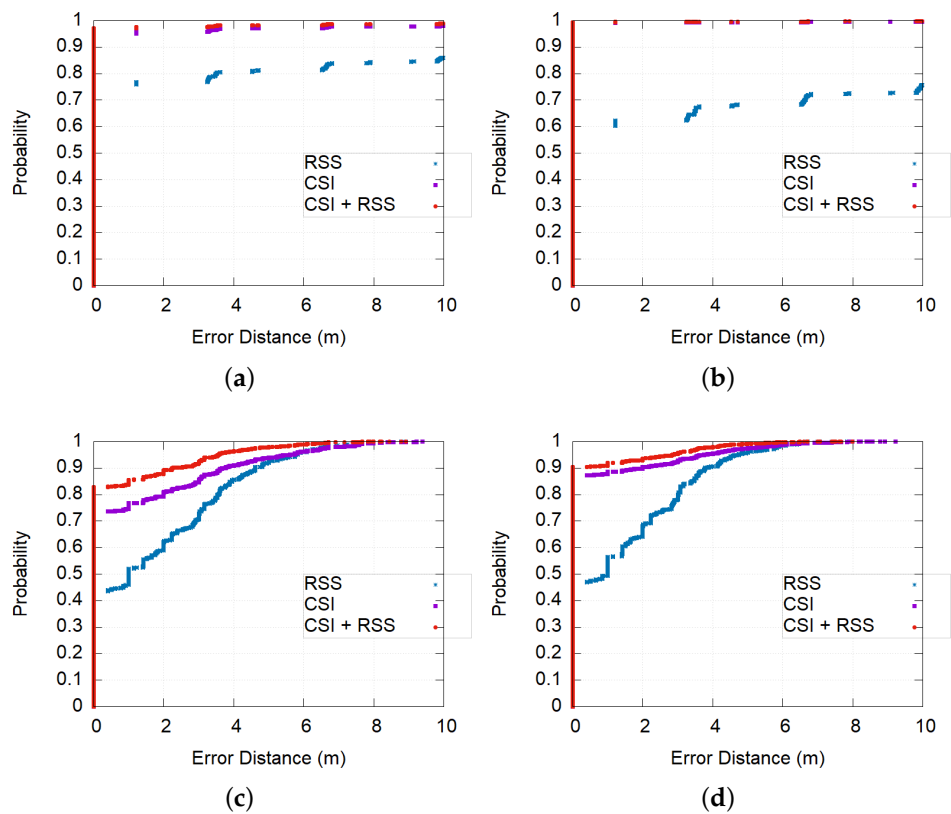
Figure 6. CDF for access point in A location using KNN algorithm with K = 1. (**a**) Corridor 2.4 GHz. (**b**) Corridor 5 GHz. (**c**) Laboratory 2.4 GHz. (**d**) Laboratory 5 GHz.
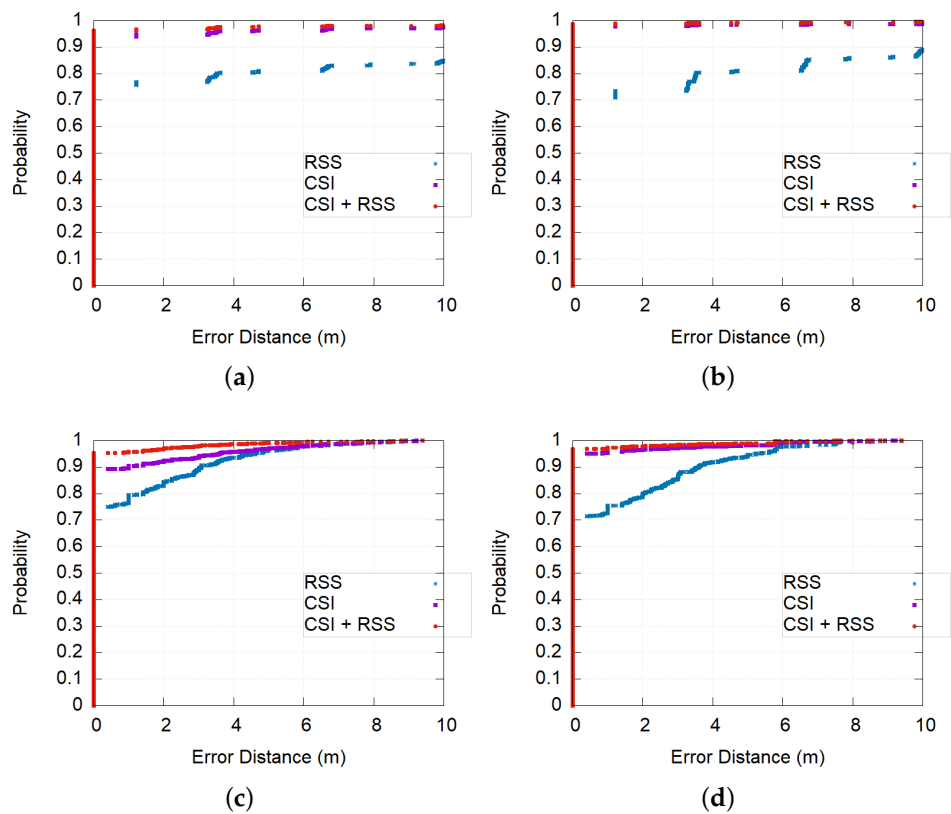
Figure 7. CDF for access point in B location using KNN algorithm with K = 1. (**a**) Corridor 2.4 GHz. (**b**) Corridor 5 GHz. (**c**) Laboratory 2.4 GHz. (**d**) Laboratory 5 GHz.

6.3. Robustness

The robustness of the proposed fusion of CSI and RSS information was evaluated by varying the size of training dataset from 5% to 80%. In these simulations, the datasets with better performance were used, that is, the datasets built with the access point located in A position in the corridor scenario, and with the access point located in B position in the laboratory scenario and using KNN algorithm. Results were obtained using the simulation configurations aforementioned, and 100 runs were carried out for each dataset to ensure statistical independence.

The average experimental results achieved by varying the training size and their standard deviations are shown in Table 7. As can be observed, the performance results improves when the training size does, obtaining an excellent accuracy and error distance with a reduced dataset size. Thus, for datasets using the 2.4 GHz frequency band, an accuracy about 90% is achieved using 10% of the total size. For 5 GHz datasets, the proposed fusion can be also robust because only a 4% accuracy loss is reached using a 10% of training samples. Furthermore, it is noticeable that an accuracy about 90% or above is yielded with a 10% of training size.

Table 7. Robustness of methodology using the KNN classifier with K = 1.

Dataset	Training Size	Accuracy	Error Distance	Training Time	Precision	Recall	F-Measure	ROC Area
Corridor 2.4 GHz-A	5%	43.44% ± 0.14	9.31 m ± 13.61	25.6 ms ± 5.04	0.466	0.434	0.426	0.710
	8%	72.24% ± 0.11	4.04 m ± 9.93	25.2 ms ± 1.12	0.729	0.722	0.719	0.857
	10%	89.66% ± 0.08	1.53 m ± 6.43	25.9 ms ± 1.28	0.897	0.897	0.894	0.947
	20%	93.06% ± 0.08	1.00 m ± 5.22	27.2 ms ± 1.06	0.931	0.931	0.931	0.964
	40%	95.24% ± 0.04	0.67 m ± 4.32	29.2 ms ± 0.96	0.953	0.952	0.951	0.976
	60%	96.25% ± 0.04	0.52 m ± 3.78	30.1 ms ± 1.42	0.963	0.963	0.963	0.981
	80%	96.84% ± 0.04	0.44 m ± 3.52	31.9 ms ± 1.17	0.969	0.969	0.968	0.984
Corridor 5 GHz-A	5%	46.31% ± 0.04	9.30 m ± 14.11	25.8 ms ± 6.43	0.507	0.464	0.461	0.725
	8%	77.79% ± 0.05	3.43 m ± 9.52	26.0 ms ± 1.05	0.788	0.788	0.775	0.885
	10%	97.17% ± 0.03	0.46 m ± 3.74	26.1 ms ± 1.31	0.972	0.972	0.971	0.986
	20%	98.25% ± 0.01	0.27 m ± 2.88	28.8 ms ± 1.20	0.983	0.983	0.982	0.991
	40%	98.88% ± 0.01	0.16 m ± 2.22	31.4 ms ± 1.73	0.989	0.989	0.988	0.994
	60%	99.11% ± 0.01	0.13 m ± 1.97	35.3 ms ± 1.49	0.991	0.991	0.991	0.995
	80%	99.32% ± 0.01	0.10 m ± 1.84	37.9 ms ± 0.90	0.993	0.993	0.993	0.997
Laboratory 2.4 GHz-B	5%	42.97% ± 0.18	1.73 m ± 2.13	25.3 ms ± 5.09	0.434	0.431	0.421	0.706
	8%	69.33% ± 0.13	0.97 m ± 1.86	24.9 ms ± 0.82	0.696	0.692	0.688	0.841
	10%	87.15% ± 0.11	0.41 m ± 1.31	25.7 ms ± 1.29	0.872	0.872	0.871	0.934
	20%	90.96% ± 0.05	0.29 m ± 1.12	27.0 ms ± 2.12	0.910	0.910	0.9.09	0.953
	40%	93.46% ± 0.06	0.21 m ± 0.98	29.4 ms ± 1.15	0.935	0.935	0.934	0.966
	60%	94.43% ± 0.05	0.18 m ± 0.92	29.9 ms ± 1.26	0.944	0.944	0.944	0.971
	80%	95.08% ± 0.10	0.16 m ± 0.87	31.1 ms ± 1.18	0.950	0.950	0.950	0.974
Laboratory 5 GHz-B	5%	44.29% ± 0.16	1.77 m ± 2.19	24.8 ms ± 2.41	0.453	0.443	0.428	0.712
	8%	71.94% ± 0.17	0.88 m ± 1.76	25.7 ms ± 1.56	0.725	0.719	0.716	0.855
	10%	91.04% ± 0.09	0.28 m ± 0.98	25.7 ms ± 1.21	0.911	0.910	0.908	0.954
	20%	93.67% ± 0.04	0.20 m ± 0.96	27.2 ms ± 1.32	0.937	0.937	0..935	0.968
	40%	95.20% ± 0.05	0.17 m ± 0.89	29.2 ms ± 1.32	0.951	0.951	0.950	0.975
	60%	95.72% ± 0.04	0.16 m ± 0.88	30.2 ms ± 1.11	0.957	0.957	0.957	0.978
	80%	95.92% ± 0.07	0.15 m ± 0.88	31.5 ms ± 1.11	0.959	0.959	0.958	0.979

On the other hand, Precision, Recall and F-Measure measurements were also computed. Precision is the ration between true positives and all instances classified as positives (true and false). In other words, Precision represents what portion of the predictive positives are actual positives. Recall is the ration between true positives and all of the actual positives (true positives and false negatives). In other words, what proportion of the cases with positive condition are detected as positive. F-Measure represents the harmonic mean of precision and recall, to consider in a balanced fashion both how good a predictor is in detecting all of the actual positives, and not producing false positives. F-Measure is calculated using the following equation:

$$\text{F-Measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{10}$$

As can be observed, these measurements follow a similar behavior related with accuracy. Thus, these measurements increase with the training size, keeping the values above 0.91 when using only a 20% of training size. Moreover, the ROC (Receiver Operating Characteristic) Area is 0.95 or above for a 20% of training size, which means that there is a 95% chance that the model will be able to distinguish among classes. Hence, that value indicates that the system has a good criterion to measure the discriminative ability among classes. These results show that the proposed fusion in this manuscript is also effective when reducing the size of the training set. The standard deviations of Precision, Recall, F-Measure and ROC Area measurements are very low, less than 0.01, and therefore are not shown in the table.

6.4. Performance Evaluation with Test Dataset

In order to evaluate the usefulness of the fusion of CSI and RSS values, new simulations were carried out employing testin locations that were different from those used for model training. That is, the red dots marked in both scenarios, Figures 4 and 5, were used for testing, and the black dots were used for training the system. These simulations were also made with the training datasets with better performance, that is, the datasets built with the A location access point in the corridor scenario, and with the access point located in B position in the laboratory scenario. All experiments were performed using the KNN algorithm with K = 1.

Table 8 shows the average error distance. In a perfect classifier, the locations returned by the model should be the closest training locations to the test points. Thus, for the corridor and laboratory scenario this shortest distance is about 1.75 m and 0.5 m, respectively. However, as can be observed the error distance is relatively higher than foreseen, and close to the training locations corresponding to the second shortest distance to the test points. Although the error distance may seem great for both scenarios, it must to be taken into account that the training dataset of the corridor scenario has a high granularity (3.25 m by 1.25 m), and the laboratory scenario has a non-linear distribution with abundant multipath propagation. Therefore, the error distance may be reduced by defining a smaller grid, that is, adding more training points. Also, the use of sophisticated classifiers such as neural networks could improve the accuracy of the proposed methodology.

Table 8. Performance results using another dataset for testing.

Dataset	Two Shortest Distances between Train and Test Point	Error Distance $\pm$ std
Corridor 2.4 GHz-A	1.75 m/5.03 m	5.11 m $\pm$ 12.78
Corridor 5 GHz-A	1.75 m/5.03 m	4.75 m $\pm$ 9.23
Laboratory 2.4 GHz-B	0.5 m/1.5 m	2.32 m $\pm$ 1.39
Laboratory 5 GHz-B	0.5 m/1.5 m	2.09 m $\pm$ 1.20

6.5. Related Work Comparison

Table 9 shows a comparison of our proposed methodology versus some other works in indoor localization literature. Keeping in mind that the evaluation setup, the collection procedures and the granularity of fingerprinting is different in each scenario of the related papers, a direct comparison should be not carried out. However, in order to highlight the results of the fusion of CSI and RSS, it can be observed that our work yields similar and comparable results when the model is evaluated with training points [28,29,46]. In addition, when independent points are used to evaluate the model, the other works yield best results but employing two or more access points [22,23,32] or using a low granularity for the fingerprinting [25,26,30,35]. Only, authors in [47] achieve a better error distance but using deep learning and regression. However, to our credit a simple and lightweight classifier, as KNN, is used in our approach.

Table 9. State of the art comparison.

Research Work	Algorithm	Data	APs	RL	Distance between RL	Error Distance	Evaluation points
[22]	KPCA, Clustering and Maximum Likelihood	RSS	14	115	2 m	1.76 m	Independent
[23]	KNN based on Spearman distance	RSS	4	400	0.5 m	3.25 m	Independent
[25]	WKNN	RSS	1	406	0.25 m	1.69 m	Independent
[46]	Random Forest	CSI amplitude	4	25	1 m	0.17 m	Training
[28]	KNN	CSI amplitude	2	19	1.5–2 m	0.98 m	Training
[26]	Deep network	CSI amplitude	1	50	0.5 m	0.95 m	Independent
[47]	Regression with deep learning	CSI amplitude	1	15	1.2 m	0.97 m	Independent
[32]	Lateration	CSI amplitude and AoA	3, 4 and 5	55	0.5–2 m	1.9, 0.8 and 0.6 m	Independent
[29]	Neural Network	CSI phase and PCA	1	25	1 m	1.43 m	Independent
[30]	Bayesian classification	CSI amplitude and phase	1	25	0.8 m	0.5 m	Independent
[35]	WKNN	CSI amplitude and RSS	1	138	0.5 m	1.79 m	Independent
This work	KNN	CSI Amplitude and RSS	1	30	1 m	0.11 m / 2.09 m	Training / Independent
				38	1.2 and 3.2 m	0.09 m / 4.75 m	Training / Independent

7. Conclusions

In this manuscript, the PHY layer CSI amplitude and RSS information are used for Wi-Fi fingerprinting. The novelty of our approach lies in the fusion of both of the aforementioned measurements, gathered from a single access point, and the CSI processing to select the transmitter–receiver antenna pair with smaller variability. The performance evaluation of the proposed methodology indicates that high-quality datasets for training the system can be generated, even when using reduced dataset sizes, leading to satisfactory localization accuracy. The proposal was verified and validated with extensive experiments carried out in two different real scenarios, a large corridor and a laboratory with abundant multipath propagation, using both the 2.4 and 5 GHz frequency bands. According to the experimental results, the proposed system yields high accuracy, with achieved results for accuracy above 96% and obtaining the best average error distance of about 0.10 m when using the 5 GHz band in both scenarios. Further, using the 2.4 GHz band, excellent results were reached, yielding error distance values of about 0.40 m in the corridor and within 0.15 m in the laboratory.

On the other hand, interference caused by signals present in indoor environments could be removed by applying some filter to both the CSI amplitude and phase information. Hence, in our ongoing work, we expect to improve the accuracy of the system by using a pre-processed CSI information and employing other alternate machine learning algorithms.

Author Contributions: Conceptualization, D.S.-R. and M.A.Q.-S.; methodology, D.S.-R., M.A.Q.-S. and I.A.-G.; software, D.S.-R. and M.A.Q.-S.; validation, I.A.-G., C.L.-B. and J.J.S.-M.; formal analysis, D.S-R., M.Q.-S., I.A.-G., C.L.-B and J.J.S.-M.; investigation, D.S.-R., M.A.Q.-S., I.A.-G., C.L.-B. and J.J.S.-M.; resources, D.S.-R. and M.A.Q.-S.; writing—original draft preparation, D.S.-R., M.Q.-S., and C.L.-B.; writing—review and editing, D.S.-R., M.A.Q.-S., I.A.-G., C.L.-B. and J.J.S.-M.; supervision, D.S.-R., M.A.Q.-S. and I.A.-G.; project administration, D.S.-R.; funding acquisition, D.S.-R.; All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AoA	Angle of Arrival
AP	Access Point
BLE	Bluetooth Low Energy
CDF	Cumulative Distribution Function
CSI	Channel State Information
GNSS	Global Navigation Satellite System
KNN	K-Nearest Neighbor
LTE	Long Term Evolution
MANOVA	Multivariate analysis of variance
MIMO	Multiple Input Multiple Output
NLOS	Non Line of Sight
OFDM	Orthogonal Frequency Division Multiplexing
PCA	Principal Component Analysis
RFID	Radio Frequency Identification
ROC	Receiver Operating Characteristic
RSS	Received Signal Strength
SVM	Support Vector Machine
ToF	Time of Flight

References

1. Want, R.; Schilit, B. Expanding the horizons of location-aware computing. *IEEE Comput.* **2001**, *34*, 31–34. [CrossRef]
2. Liu, H.; Darabi, H.; Banerjee, P.; Liu, J. Survey of wireless indoor positioning techniques and systems. *IEEE Trans. Syst. Man, Cybern. Part C (Appl. Rev.)* **2007**, *37*, 1067–1080. [CrossRef]
3. Górak, R.; Luckner, M. Automatic Detection of Missing Access Points in Indoor Positioning System. *Sensors* **2018**, *18*, 3595. [CrossRef] [PubMed]
4. Ou, C.W.; Chao, C.J.; Chang, F.S.; Wang, S.M.; Liu, G.X.; Wu, M.R.; Cho, K.Y.; Hwang, L.T.; Huan, Y.Y. A ZigBee position technique for indoor localization based on proximity learning. In Proceedings of the 2017 IEEE International Conference on Mechatronics and Automation (ICMA), Takamatsu, Japan, 6–9 August 2017; pp. 875–880.
5. Wu, X.; Deng, F.; Chen, Z. RFID 3D-LANDMARC Localization Algorithm Based on Quantum Particle Swarm Optimization. *Electronics* **2018**, *7*, 19.
6. Zhuang, Y.; Yang, J.; Li, Y.; Qi, L.; El-Sheimy, N. Smartphone-based indoor localization with bluetooth low energy beacons. *Sensors* **2016**, *16*, 596. [CrossRef]
7. Nguyen, T.S.; Nguyen, T.N.; Tran, Q.S.; Huynh, T.H. Improvement of Ultrasound-Based Localization System Using Sine Wave Detector and CAN Network. *J. Sens. Actuator Netw.* **2017**, *6*, 12. [CrossRef]
8. Makki, A.; Siddig, A.; Saad, M.; Bleakley, C. Survey of WiFi positioning using time-based techniques. *Comput. Netw.* **2015**, *88*, 218–233. [CrossRef]
9. Hoene, C.; Willmann, J. Four-way TOA and software-based trilateration of IEEE 802.11 devices. In Proceedings of the 2008 IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications, Cannes, France, 15–18 September 2008; pp. 1–6.

10. Günther, A.; Hoene, C. Measuring round trip times to determine the distance between WLAN nodes. In Proceedings of the International Conference on Research in Networking, Waterloo, ON, Canada, 2–6 May 2005; pp. 768–779.

11. Schwalowsky, S.; Trsek, H.; Exel, R.; Kerö, N. System integration of an IEEE 802.11 based TDoA localization system. In Proceedings of the 2010 IEEE International Symposium on Precision Clock Synchronization for Measurement, Control and Communication, Portsmouth, NH, USA, 27 September–1 October 2010; pp. 55–60.

12. Rong, P.; Sichitiu, M.L. Angle of arrival localization for wireless sensor networks. In Proceedings of the 2006 3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks, Reston, VA, USA, 28 September 2006; Volume 1, pp. 374–382.

13. Zhang, S.; Guo, J.; Luo, N.; Wang, L.; Wang, W.; Wen, K. Improving Wi-Fi Fingerprint Positioning with a Pose Recognition-Assisted SVM Algorithm. *Remote Sens.* **2019**, *11*, 652. [CrossRef]

14. Yang, F.; Xiong, J.; Liu, J.; Wang, C.; Li, Z.; Tong, P.; Chen, R. A pairwise SSD fingerprinting method of smartphone indoor localization for enhanced usability. *Remote Sens.* **2019**, *11*, 566. [CrossRef]

15. Bahl, P.; Padmanabhan, V.N. RADAR: An in-building RF-based user location and tracking system. In Proceedings of the IEEE INFOCOM 2000, Conference on Computer Communications, Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No. 00CH37064), Tel Aviv, Israel, 26–30 March 2000; Volume 2, pp. 775–784.

16. Wu, K.; Xiao, J.; Yi, Y.; Chen, D.; Luo, X.; Ni, L.M. CSI-based indoor localization. *IEEE Trans. Parallel Distrib. Syst.* **2013**, *24*, 1300–1309. [CrossRef]

17. Yang, Z.; Zhou, Z.; Liu, Y. From RSSI to CSI: Indoor localization via channel response. *ACM Comput. Surv. (CSUR)* **2013**, *46*, 1–32. [CrossRef]

18. Ma, Y.; Zhou, G.; Wang, S. WiFi sensing with channel state information: A survey. *ACM Comput. Surv. (CSUR)* **2019**, *52*, 1–36. [CrossRef]

19. Khalajmehrabadi, A.; Gatsis, N.; Akopian, D. Modern WLAN fingerprinting indoor positioning methods and deployment challenges. *IEEE Commun. Surv. Tutor.* **2017**, *19*, 1974–2002. [CrossRef]

20. Shang, J.; Hu, X.; Gu, F.; Wang, D.; Yu, S. Improvement schemes for indoor mobile location estimation: A survey. *Math. Probl. Eng.* **2015**, *2015*, 397298. [CrossRef]

21. Kaemarungsi, K.; Krishnamurthy, P. Analysis of WLAN's received signal strength indication for indoor location fingerprinting. *Pervasive Mob. Comput.* **2012**, *8*, 292–316. [CrossRef]

22. Luo, J.; Fu, L. A Smartphone Indoor Localization Algorithm Based on WLAN Location Fingerprinting with Feature Extraction and Clustering. *Sensors* **2017**, *17*, 1339.

23. Xie, Y.; Wang, Y.; Nallanathan, A.; Wang, L. An Improved K-Nearest-Neighbor Indoor Localization Method Based on Spearman Distance. *IEEE Signal Process. Lett.* **2016**, *23*, 351–355. [CrossRef]

24. Hu, X.; Shang, J.; Gu, F.; Han, Q. Improving Wi-Fi indoor positioning via AP sets similarity and semi-supervised affinity propagation clustering. *Int. J. Distrib. Sens. Netw.* **2015**, *11*, 109642. [CrossRef]

25. Kokkinis, A.; Kanaris, L.; Liotta, A.; Stavrou, S. RSS indoor localization based on a single access point. *Sensors* **2019**, *19*, 3711. [CrossRef]

26. Wang, X.; Gao, L.; Mao, S.; Pandey, S. CSI-based fingerprinting for indoor localization: A deep learning approach. *IEEE Trans. Veh. Technol.* **2017**, *66*, 763–776. [CrossRef]

27. Wu, Z.; Jiang, L.; Jiang, Z.; Chen, B.; Liu, K.; Xuan, Q.; Xiang, Y. Accurate Indoor Localization Based on CSI and Visibility Graph. *Sensors* **2018**, *18*, 2549. [CrossRef] [PubMed]

28. Chapre, Y.; Ignjatovic, A.; Seneviratne, A.; Jha, S. CSI-MIMO: An efficient Wi-Fi fingerprinting using channel state information with MIMO. *Pervasive Mob. Comput.* **2015**, *23*, 89–103. [CrossRef]

29. Dang, X.; Ren, J.; Hao, Z.; Hei, Y.; Tang, X.; Yan, Y. A novel indoor localization method using passive phase difference fingerprinting based on channel state information. *Int. J. Distrib. Sens. Netw.* **2019**, *15*, 1550147719844099. [CrossRef]

30. Dang, X.; Si, X.; Hao, Z.; Huang, Y. A Novel Passive Indoor Localization Method by Fusion CSI Amplitude and Phase Information. *Sensors* **2019**, *19*, 875. [CrossRef] [PubMed]

31. Vasisht, D.; Kumar, S.; Katabi, D. Decimeter-level localization with a single WiFi access point. In Proceedings of the 13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16), Santa Clara, CA, USA, 16–18 March 2016; pp. 165–178.

32. Kotaru, M.; Joshi, K.; Bharadia, D.; Katti, S. Spotfi: Decimeter level localization using wifi. In Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication, London, UK, 17–21 August 2015; pp. 269–282.

33. Wang, X.; Mao, S. Deep Learning for Indoor Localization based on Bi-modal CSI Data. *Appl. Mach. Learn. Wirel. Commun.* **2019**, *81*, 343.

34. Tian, Z.; Li, Z.; Zhou, M.; Jin, Y.; Wu, Z. PILA: Sub-meter localization using CSI from commodity Wi-Fi devices. *Sensors* **2016**, *16*, 1664. [CrossRef]

35. Zhao, L.; Wang, H.; Li, P.; Liu, J. An improved WiFi indoor localization method combining channel state information and received signal strength. In Proceedings of the 2017 36th Chinese Control Conference (CCC), Dalian, China, 26–28 July 2017; pp. 8964–8969.

36. Perahia, E.; Stacey, R. *Next Generation Wireless LANs: 802.11 n and 802.11 ac*; Cambridge University Press: Cambridge, UK, 2013.

37. Kamali, B. *AeroMACS: An IEEE 802.16 Standard-Based Technology for the Next Generation of Air Transportation Systems*; Standards Information Network IEEE Press: Hoboken, NJ, USA, 2018; pp. 189–258.

38. Li, D.; Lei, Y.; Zhang, H. A Novel Outdoor Positioning Technique Using LTE Network Fingerprints. *Sensors* **2020**, *20*, 1691. [CrossRef]

39. Yu, H.; Lee, H.; Jeon, H. What is 5G? Emerging 5G mobile services and network requirements. *Sustainability* **2017**, *9*, 1848. [CrossRef]

40. Freund, Y.; Schapire, R.E. Experiments with a new boosting algorithm. *Icml* **1996**, *96*, 148–156.

41. Quinlan, J.R. *C4.5: Programs for Machine Learning*; Morgan Kaufmann Publishers: San Francisco, CA, USA, 1993.

42. Aha, D.W.; Kibler, D.; Albert, M.K. Instance-based learning algorithms. *Mach. Learn.* **1991**, *6*, 37–66. [CrossRef]

43. Cover, T.; Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27. [CrossRef]

44. Halperin, D.; Hu, W.; Sheth, A.; Wetherall, D. Tool release: Gathering 802.11 n traces with channel state information. *ACM SIGCOMM Comput. Commun. Rev.* **2011**, *41*, 53. [CrossRef]

45. Frank, E.; Hall, M.; Witten, I. The WEKA Workbench. Available online: https://www.cs.waikato.ac.nz/ml/weka/Witten_et_al_2016_appendix.pdf (accessed on 22 April 2020).

46. Wang, Y.; Xiu, C.; Zhang, X.; Yang, D. WiFi indoor localization with CSI fingerprinting-based random forest. *Sensors* **2018**, *18*, 2869. [CrossRef]

47. Xiang, C.; Zhang, S.; Xu, S.; Chen, X.; Cao, S.; Alexandropoulos, G.C.; Lau, V.K. Robust Sub-Meter Level Indoor Localization With a Single WiFi Access Point—Regression Versus Classification. *IEEE Access* **2019**, *7*, 146309–146321. [CrossRef]