Data Fusion Methods for Indoor Positioning Systems Based on Channel State Information Fingerprinting

Hailu Tesfay Gidey 1, Xiansheng Guo 1,2,*, Ke Zhong 1, Lin Li 1 and Yukun Zhang 1

1 Department of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
2 Yangtze Delta Region Institute (Quzhou), University of Electronic Science and Technology of China, Quzhou 324000, China
* Correspondence: xsguo@uestc.edu.cn; Tel.: +86-132-5810-3244

Abstract: Indoor signals are susceptible to NLOS propagation effects, multipath effects, and a dynamic environment, posing more challenges than outdoor signals despite decades of advancements in location services. In modern Wi-Fi networks that support both MIMO and OFDM techniques, Channel State Information (CSI) is now used as an enhanced wireless channel metric replacing the Wi-Fi received signal strength (RSS) fingerprinting method. The indoor multipath effects, however, make it less robust and stable. This study proposes a positive knowledge transfer-based heterogeneous data fusion method for representing the different scenarios of temporal variations in CSI-based fingerprint measurements generated in a complex indoor environment targeting indoor parking lots, while reducing the training calibration overhead. Extensive experiments were performed with real-world scenarios of the indoor parking phenomenon. Results revealed that the proposed algorithm proved to be an efficient algorithm with consistent positioning accuracy across all potential variations. In addition to improving indoor parking location accuracy, the proposed algorithm provides computationally robust and efficient location estimates in dynamic environments. A Cramer-Rao lower bound (CRLB) analysis was also used to estimate the lower bound of the parking lot location error variance under various temporal variation scenarios. Based on analytical derivations, we prove that the lower bound of the variance of the location estimator depends on the (i) angle of the base stations, (ii) number of base stations, (iii) distance between the target and the base station, \( d_{jr} \), (iv) correlation of the measurements, \( \rho_{ai} \), and (v) signal propagation parameters \( \sigma_C \) and \( \gamma \).

Keywords: indoor positioning; CSI-fingerprinting; parking lots; data fusion; knowledge transfer; principal component analysis; CRLB analysis

1. Introduction

The rapid development of indoor positioning systems (IPS) has been fueled by the emergence of both fifth- and sixth-generation communication systems (5G and 6G) and the internet of things (IoT), as well as the growing commercial interest in location-based services (LBSs). Due to the incredible development of mobile applications, LBSs have gained significant importance in both industrial and commercial applications such as vehicle indoor parking lots [1], indoor navigation [2], self-driving cars [3], security monitoring, and large venue management [4], military use [5], emergency services, tracking, and tourism, and many others [6–8]. Modern vehicles rely heavily on GPS to determine their location; however, GPS-based vehicle positioning frequently fails in typical indoor environments especially underground parking lots [9,10]. This could be justified because the indoor environment setting is described as a more complicated scenario than the outdoor setting, owing to (i) the non-line of sight (NLOS) path as a cause of incoherent propagation caused by various barriers along the transceivers; (ii) inherent heterogeneity of signal distributions caused by the dynamic environment in both temporal and spatial variations; and (iii) severe
signal attenuation and/or shielding of satellite signals [9,10]. Despite this, a study found that 95% of the time, cars are in a parking lot or an indoor environment [11]. As a result of the challenging nature of the indoor positioning problem (IPP), various radio frequency (RF)-based wireless technologies have been developed to address the demand for higher positioning accuracy while remaining computationally efficient.

To achieve the desired goal of an IPS, two major predictors must be considered: (i) the type of signal features used to establish the fingerprint, such as received signal strength (RSS) [12], CSI [13], time of arrival (TOA) [14], time difference of arrival (TDOA) [15], angle of arrival (AOA) [16], and so on; and (ii) the underlying network technology that generates the signal features. Despite the fact that each technology and signal feature has its own trade-off that limits the scalability of its implementation, over the last two decades, various wireless positioning technologies, including but not limited to cellular networks (GSM) [17], WLAN-Wi-Fi [18], Bluetooth [19], RFID [20], ultra-wideband (UWB) [21], ZigBee [22], inertial navigation systems [23], geomagnetic [24], visible light communication (VLC) [25], etc. have been proposed and investigated. Regardless of the complexity of the indoor environment, the rapidly growing commercial interest in indoor location-based services (ILBSs) still dictates an effective approach for positioning systems that considers both the cost and practicality of their implementation. Vehicle positioning is critical for applications such as navigation, driver assistance and autonomous driving. Several alternative systems for indoor vehicle parking have been developed, including, for example, intelligent parking systems [26] combining RFID and WSN technologies into intelligent parking management systems that use RFID as a unique identification number on the passive RFID card of the driver [27], hybrid systems that use external sensors as well as in-vehicle sensors [28], multiple fisheye surveillance cameras [29] and 3D light detection and ranging (LIDAR) scanners [30]. Additionally, camera and artificial intelligence technologies such as Fuzzy K-means (FCM) and particle swarm optimization (PSO) classification have been used as parameter improvements for parking space detection [31]. Moreover, since the BLE (Bluetooth Low Energy) sensor consumes less power, an intelligent parking system through BLE is possible [32].

However, due to the additional infrastructure required for their implementation, most indoor positioning applications (based on Bluetooth, ultra-wideband (UWB), radio frequency identification (RFID), and others) have extra construction costs. Unlike other wireless IPS, which have been criticized for their limited scalability due to the demand for additional hardware devices, RSS-based Wi-Fi fingerprinting is gaining popularity due to its low computational complexity, cost-effectiveness, and ease of implementation and compatibility with existing network infrastructure [33–35]. A study [36] proposed an improved Wi-Fi fingerprinting method to replace the behavior of the indoor GPS environment to provide a reliable method of locating vehicles indoors, while also preserving the architecture of the vehicle locating system and facilitating a smooth transition from the outdoor to the indoor, and vice versa. However, due to the complexity of a typical indoor environment, RSS-based fingerprint indoor positioning is susceptible to signal fluctuations, resulting in inefficiency and poor overall positioning performance. The signal fluctuations of the RSS measurements of both instances of training and testing datasets for Wi-Fi APs of AP 1, AP 2, AP 3, AP 4, and AP 5 were reported to have standard deviations (in dB) of 16.8, 15.9, 14.5, 17.9, and 17.1 and 15.63, 15.14, 14.40, 17.92, and 0.00, respectively, as illustrated in Tables 1 and 2 of [37]. This experimental result [37] confirms that the temporal variations in signal distributions have a significant impact on indoor positioning performance-based RSS fingerprints, owing to the effects of multipath, NLOS, and channel conditions such as fading, shadowing, and scattering. Indoor positioning-based RSS fingerprints (FBIP-RSS), on the other hand, have been characterized as having poor spatial resolution and low dimensional feature spaces, which directly degrades indoor positioning accuracy [38]. As a result, the system fails to achieve the desired accurate and robust positioning estimates due to four critical predictors associated with the RSS-based fingerprint that determine the quality of the IPS: (i) high temporal signal fluctuations, (ii) RSS measurements highly
susceptible to the effect of a typical indoor environment, (iii) low dimensional feature spaces, and (iv) requirements for large size labeled samples [37–41], which is both costly and labor-intensive.

However, as Wi-Fi network technology has advanced, the Channel State Information (CSI) signal feature can now be extracted from commercial Wi-Fi devices using network interface cards (NICs) that can provide multi-channel subcarrier phase and amplitude information, allowing to better characterize the signal propagation model with the help of multiple-input multiple-output (MIMO) and orthogonal frequency division multiplexing (OFDM) technologies. In comparison to RSS signal features, the CSI-signal feature is distinguished by its ability to depict: (a) fine-grained channel features, (b) higher dimension features, (c) the diversified physical layer of both phase and amplitude information, and (d) more robust and temporal stable features. In line with this, a comparative analysis of CSI and RSS was performed [42], and the study revealed that the physical characteristics of CSI can significantly reduce the problem of the RSS, such that (a) multipath effect propagation can be better handled, (b) have strong stability, particularly in a static environment, and are relatively stable to the dynamics of the environment, and (c) reduce radio interference of carrier frequency signals [42]. Thus, given the ability of modern Wi-Fi devices to extract CSI measurements using NICs, as well as the previously mentioned strengths of the CSI signal feature, indoor positioning-based CSI fingerprinting (FBIP-CSI) is attracting significant attention for improving indoor positioning performance.

Nonetheless, due to the complex indoor environment of multipath effects, the CSI-based IPS performance still faces challenges in severe dynamic range and fluctuations among high-dimensional channels [43]. Various CSI fingerprinting-based methods have been proposed and achieved high positioning accuracy in addressing both the impact of the multipath effect as well as signal fluctuations in a dynamic indoor environment [44–46]. A novel indoor localization system, FIFS (fine-grained indoor fingerprinting system), has been proposed to address signal fluctuations caused by multipath effects in typical indoor environments [47]. During the matching stage, a probabilistic model was applied to accurately map the observed CSI values [47]. However, the assumption of signal distribution requires that the signals be normally distributed to estimate the target’s position by the weighted average of the CSI amplitude values, which is a real challenge due to the indoor multipath effect and dynamic environment. Furthermore, several machine learning (ML) algorithms have been used to mitigate the indoor multipath effect, including support vector machine (SVM) [48], random forest (RF) [49], k-nearest neighbors (KNN) [50,51], and visibility graph (VG) based methods [52]. Moreover, a study [53] proposed a low-overhead indoor positioning transfer learning system based on improved TrAdaBoost to mitigate environmental changes and new scenarios. The proposed method [53] is robust in time and space, with lower site survey overhead while maintaining the same positioning accuracy. Even though the proposed ML algorithms have improved location accuracy, they still face computational complexity challenges due to the CSI’s higher dimension features. Additionally, the proposed ML algorithms are challenged to provide robust location fingerprint estimation, because (i) the fingerprint database must be kept up to date for robust location fingerprint estimation, (ii) the algorithms rely heavily on individual fingerprint parameters, and (iii) a large number of labeled CSI samples are required during model training, which is also very expensive.

The primary goal of this article is to improve CSI-based fingerprint indoor positioning by using data fusion methods to represent temporal signal variations of a location estimator in indoor parking lots. Furthermore, the heterogeneous transfer learning method (HetTL) was used in conjunction with principal component analysis (PCA) to reduce time complexity and ensure cost effectiveness by avoiding unnecessarily high costs associated with extra Wi-Fi access points (Wi-Fi APs) deployment (sources for irrelevant signal features) from the model. This study, on the other hand, uses Cramer-Rao lower bound analysis to estimate the lower bound variance for the estimator data fusion method of CSI-based fingerprint indoor positioning. In this study, the contributions are fourfold:
(1) We proposed a data fusion method to represent temporal signal variations by constructing new feature vector spaces based on the most significant predictors and enabling heterogeneous knowledge transfer with the goal of reducing calibration overhead in an indoor parking system.

(2) To efficiently detect parking lots, we used the principal component analysis technique to reduce data noise caused by multipath effects, as the channel state information amplitude or fingerprints received from multiple base stations could be mis-matched with the actual target’s fingerprint patterns. In other words, multiple signals arriving at the receiver end from different paths may cause fingerprint duplication and degrades the overall performance of the system. This refers to the possibility of a high-dimensional curse for CSI-based fingerprinting in indoor parking scenarios.

(3) The Cramer Rao lower bound (CRLB) analysis was used to estimate the lower bound variance for the estimator of data fusion methods of a vehicle’s indoor parking lot or to measure the unbiasedness of the location estimator for an indoor parking system-based channel state information fingerprinting.

(4) We conducted a comparative analysis of the proposed algorithms in terms of the performance of indoor positioning estimation-based channel state information fingerprinting in comparison with the most popular algorithms in the field of machine learning using predictive modeling as a baseline.

The rest of this study is organized as follows: related works are presented in Section 2. Section 3 describes the framework and problem formulation of fingerprint-based indoor positioning with emphasis on data fusion methods, the process of database construction-based CSI-fingerprinting and the system architecture. Evaluation metrics and the CRLB analysis of data fusion methods applied for location estimation are also presented in Section 3. Experimental results and discussions are presented in Section 4. Finally, conclusions are provided in Section 5.

2. Related Works

This section provides a brief overview of fingerprint-based methods and data fusion techniques used to address the indoor positioning problem (IPP), system modeling of IPP-based CSI fingerprinting, performance evaluation metrics for positioning estimation, and challenges that limit the applications of various signal features used in IPP. Since vehicle positioning is essential in applications such as indoor parking lots, indoor navigation, driver assistance and autonomous driving, accurate information about mobility patterns and vehicle trajectories are essential to improve positioning performance. Generally, three positioning algorithms [54] are applied in IPS: (a) triangulation, (b) the proximity, and (c) scene analysis or fingerprinting. The triangulation method uses the geometric properties of a triangle to estimate the target’s location and comprises two types: angulation and lateration, referring to measuring the angle and distance from multiple grid points (GPs), respectively. One can ask how do these algorithms and signal features estimate the position of a target in relation to the transmitting node or source anchor. Or, what signal features exactly do the algorithms used to estimate the target’s location?

Various signal features were proposed and investigated to address IPP, mainly: time of arrival (TOA/TDOA) [55], angle of arrival (AOA) [56], received signal strengths (RSS) [57–59], and channel state information (CSI) [60,61]. In the lateration algorithm, the distance could be acquired indirectly by measuring the received signal strengths indicator (RSSI), time of arrival (TOA), or time difference of arrival (TDOA). Both the signal features of TOA and TDOA are the most accurate techniques, which can filter out multipath effects in an indoor environment situation despite both requiring a line-of-sight (LOS) path along with the transceiver [62,63], which is infeasible in a complex indoor environment. Moreover, they do require high construction costs to be implemented due to their requirements for extra infrastructure investment pertinent to signal directions and need to store precise timing information [64], or these two signal features need to be precisely synchronized [65,66]. Whereas the AOA signal feature does not require time synchronization between measuring
units but the AOA or the angulation method demands extra hardware devices pertinent for signal directions [67].

In contrast, the wireless fidelity signal (Wi-Fi: 802.11) has received significant acceptance for IPS both in academia and industry communities [68] mainly for the following reasons: (a) pervasive penetration of Wireless LAN and deployment of Wi-Fi-enabled mobile devices across the globe (cost-effective as it is adopting the existing wireless network infrastructure); (b) the radio wave covers a wide range with a radius about 300 feet or the widespread of its signal over long distances; and (c) it does not require line-of-sight measurement of base stations [55,69] and achieves high applicability in a complex indoor environment. On the other hand, most indoor localization technologies based on Wi-Fi rely on received signal strengths and can be directly implemented using the existing wireless communications infrastructure without any calibrations. The Wi-Fi received signal strength (RSS) measured in decibel milliwatts (dB) is used to find a relationship between transceivers, or measures the accuracy of localization based on the distances between the mobile user and available Wi-Fi access points [57–59] through the third method. The so-called scene analysis or fingerprinting comprises two phases: training and testing phases. The RSS fingerprints of the so-called radio map are first collected from each Wi-Fi access point at multiple locations within the defined grid points (GPs) and a predictive model is trained to learn the ‘signal-to-location’ relationship (training phase). The learned model is then applied to infer the location of the target based on the new measurement obtained (online phase) [57–59]. Nevertheless, the positioning characteristics-based RSS fingerprinting still has a fundamental problem in accuracy and robustness in IPS, and both temporal and spatial signal fluctuations lead to spontaneous or not robust localization errors. Furthermore, indoor positioning-based Wi-Fi RSS fingerprints have been characterized by low-dimensional feature spaces and a poor spatial resolution, which directly degenerates the indoor positioning performance [70]. In summary, the system fails to achieve the desired accurate and robust positioning estimates due to four critical predictors associated with the RSS-based fingerprint that determines the quality of the IPS: (i) high temporal signal fluctuations, (ii) RSS measurements highly susceptible to the effect of a typical indoor environment, (iii) low-dimensional feature spaces, and (iv) requirements for a large size of labeled samples [37], which is both costly and labor-time-intensive [37–41]. Moreover, RSS is also highly dependent on the used Wi-Fi chipset and how it estimates and reports the RSS value.

The signal feature of the channel state information, however, has emerged as an enhanced wireless channel metric with significant data throughput [42,43] to replace the received signal strength (RSS) for IPS, in which the Wi-Fi networks used MIMO-OFDM techniques whereby data are modulated on multiple channels in different frequencies and simultaneously transmitted among multiple antenna pairs. Apparently, the high dimensional features are possibly produced in WLAN systems due to the MIMO technology integrated into CSI; therefore, an opportunity exists to improve the positioning performance despite facing computational complexity as the main trade-off that needs to be addressed. Ostensibly, the high dimensions of features on their own may not be a positive predictor for the localization process; thus, identifying the most significant predictors is a must; otherwise, some redundant features may inflate or degrade (i.e., cause model overfitting) the system modeling and cause unjustifiable cost for the extra deployment of Wi-Fi access points (Wi-Fi APs). Moreover, some studies based on CSI have demonstrated improved accuracy over RSS for indoor location estimation [42,43], and this can be justified as the CSI reflects the multipath propagation of the signal, to some extent, better than the RSSI.

Towards this end, several CSI-based object detection schemes in WLAN systems have been studied [43,71,72] and shown that the CSI at each subcarrier in OFDM can be used to characterize the target’s behaviors through the two important features of amplitude and phase fluctuations of the targets in a frequency-selective fading channel. Moreover, with the introduction of MIMO technology, high dimensional CSI features are possibly produced in WLAN systems, and it can be considered as an opportunity to improve the
positioning performance, although computational complexity is the main trade-off that needs to be addressed. Additionally, a comparative analysis between CSI and RSSI was conducted [42,73]. The study [42,73] revealed that the physical characteristics of CSI can significantly reduce the problem of RSSI such that (a) multipath effect propagation can be better handled (b) owns strong stability especially in a static environment and relatively stable to the dynamics of the environment (c) reduce radio interference of carrier frequency signals [42,73]. Thus, CSI can present different subcarrier amplitude and phase characteristics for different propagation environments. Notably, the overall structural characteristics of CSI remain relatively stable compared with the RSS signal feature; nevertheless, appropriate signal processing technology (SPT) is required. Hitherto, the CSI-based indoor positioning system still has a challenging problem in severe dynamic range and with fluctuation among high-dimensional channels due to indoor multipath effects [43]. Table 1 below presents the notations used in this study.

Table 1. List of Notations.

| Notation | Description |
|----------|-------------|
| $Y$, $X$, $H$, and $\phi$ | Received signal vector, transmitted signal vector, channel matrix and the AGWN (additive Gaussian white noise) |
| $n_t$, $n_s$, $n_T$, $n_r$, & $n_m = M$ | Number of target instances, number of source instances, number of transmitter antennas, number of receiver antennas, and number of subcarriers |
| $\hat{H}$, $H_k(f_m) = H_m$ | Estimated channel frequency response (CFR) in frequency domain of all subcarriers, the channel state information for each sample of the $m$th subcarrier of $k$th transmitter-receiver pair. |
| $|H_m| < H_m$ | Amplitude of the $m$th subcarriers, phase of the $m$th subcarrier |
| $d$, $\text{CSI}_\text{eff}$ |Propagation distance between the transceiver, effective channel state information measurements |
| $c$, $f_c$, $n$, and $\varphi$ |Radio wave phase velocity, central frequency of CSI, path loss attenuation factor, and environmental factor. |
| $h(t)$, $a_i$, $\theta_i$, $\tau_i$, $N$ and $\delta(t)$ | Channel impulse response (CIR), amplitude values, phase, time delay of the $i$th path, the total number of multipaths and Dirac delta function |
| $\{c_{r_j}^a(t_d), a \in [1,n_a], i \in [1,n_c]\}$ | The $i$th CSI amplitude value at the $r$th RP of the $j$th BS from $a$th antenna. $t_d$ refers to the measurement days. |
| $n_c$, $n_a$ | Number of CSI measurements at each RP, number of antennas of a BS of a receiver |
| $L_{ij}$, $\hat{L}_{ij}$, $r_{ij}^2$ and $\omega_{ij}$ | The position of the $j$th base station, the unknown location of the target, the distance between the $i$th base station and the target, and weight transfer. |
| $\mu \sum_{\rho_{ij}^t}$ |The overall mean values of the CSI measurements, multidimensional covariance of different sources of CSI measurements, the correlation coefficients between the two feature vectors |
| $\lambda$, $C_{\text{TX}}$, $G_{\text{RX}}$ |The wavelength of the transmitted signal, the antenna gains at the transmitter, the antenna gains at the receiver |
| $C_{\text{Rx}}^q(t_1)$, $C_k^e(t_2)$ |The fused CSI measurements collected on different measurement days of two months of September and October 2020 |
| $C_{\text{r_j}}^{\text{final}}(s_d)|_{s_{r_j}}$, $C_{\text{r_j}}^{\text{final}}(s_d)|_{s_{r_j}}$, $C_{\text{r_j}}^{\text{final}}(s_d)|_{s_{r_j}}$, $C_{\text{r_j}}^{\text{final}}(s_d)|_{s_{r_j}}$ |Refined Sources Fingerprint of training dataset, Refined CSI testing dataset, Fused refined CSI training fingerprints, Fused refined CSI testing fingerprints |

System Model

Channel state information (CSI) has emerged as an enhanced wireless channel metric (significantly enhanced data throughput) [74,75] in place of the received signal strength (RSS) for IPS, in which the Wi-Fi networks use the MIMO-OFDM techniques whereby data are modulated on multiple channels (subcarriers) in different frequencies and simultaneously transmitted among multiple antenna pairs (the 802.11 a/g/n standard). The channel response can be extracted from the receivers in the format of CSI, which reveals a set of channel measurements representing the amplitudes and phases of every channel [43,74–76]. Additionally, the receiver signal strength can reflect the channel quality of the transmitter and receiver, which can be analyzed from the CSI obtained from the physical layer. CSI also describes the signal propagation process and shows whether the transmitted signal is
affected by scattering, attenuation and other factors in the propagation process. In general, the CSI can provide more detailed channel information for a sample than the RSSI. Thus, the received signal power after the multipath channel in OFDM systems can be represented as:

\[ Y = HX + \varphi \]  

(1)

where \( Y \) and \( X \) represent the received and transmitted signal vector, respectively, and \( H \) and \( \varphi \) represent the channel matrix and AGWN (additive Gaussian white noise), respectively, such that \( \varphi \sim N(0, \sigma^2 I) \). Where \( I \) is an identity matrix. Thus, the CSI of all subcarriers can be estimated as:

\[ \hat{H} = \frac{Y}{X} \]  

(2)

where \( \hat{H} \) denotes the channel frequency response (CFR) in the frequency domain. In the narrow band flat fading OFDM channel, the channel matrix \( H \) estimated at the receiver represents the physical layer CSI over multiple sub-carriers with the dimension (the format of the received CSI measurements) \( n_T \times n_r \times n_m \); and \( n_T, n_r \) and \( n_m \) represent the number of transmitter antennas, receiver antennas and subcarriers for each antenna pair, respectively. We group the subcarriers of the channel state information for each sample along the transceiver antenna pairs as:

\[ H = \begin{bmatrix} H^1(f_1), H^1(f_2), \ldots, H^1(f_m), & H^2(f_1), H^2(f_2), \ldots, H^k(f_m) \end{bmatrix} \]  

(3)

where \( H^k(f_m) = H_m \) denotes the \( m^{th} \) subcarrier of the \( k^{th} \) transmitter-receiver pair. Thus, for each location, there are a total \( L \) streams for each sample, where \( L = n_t \times n_r \times n_m \). And each group of CSI represents the amplitude and phase of an OFDM subcarrier:

\[ H_m = |H_m|e^{j\angle H_m} \]  

(4)

where, \( |H_m| \) and \( \angle H_m \) represents the amplitude and phase of \( m^{th} \) subcarriers, respectively.

In MIMO [77] systems with \( p \) transmit antennas and \( q \) receive antennas, CSI is a matrix of \( p \times q \) dimension, which can be expressed as follows:

\[ H(f_m) = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1q} \\ h_{21} & h_{22} & \cdots & h_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ h_{p1} & h_{p2} & \cdots & h_{pq} \end{bmatrix} \]  

(5)

where \( H_{pq} \) exists in the form of a complex number, which represents the amplitude and phase of the subcarrier of the antenna stream. In [78] proposed a fine-grained indoor localization based on CSI data and FILA (Fine Grain Indoor Localization) weights the filtered CSI and normalizes the power to the center frequency in the band as:

\[ \text{CSI}_{\text{eff}} = \frac{1}{M} \sum_{m=1}^{M} \frac{f_m}{f_c} |H_m| \]  

(6)

where \( \text{CSI}_{\text{eff}} \) is the effective CSI for distance estimation, \( M \) and \( f_c \) are the number of subcarriers and the calculated center frequency, and \( |H_m| \) is the amplitude of the filtered CSI on the \( m^{th} \) subcarrier. The propagation distance between the transceiver can be represented by effective channel state information as:

\[ d = \frac{1}{4\pi} \left( \frac{c}{f_c |\text{CSI}_{\text{eff}}|} \right)^2 \times \varphi^{1/n} \]  

(7)
where \(d\) is the estimated distance (in meters (m)) between the transmitter and receiver in an indoor environment setting, \(c\) is the radio wave phase velocity (in m/s), \(f_c\) is the central frequency of CSI (in Hertz or cycles/seconds), \(n\) is the path loss attenuation factor (in dB), and \(\varphi\) is the environmental factor (in dB). The environment setting was being conducted in a specified area in controlled fashion, and assumed a constant environmental factor \((\varphi)\) to estimate the distance between the transceiver. Additionally, the environment factor mainly describes the fixed values (Friis transmission formula) to be assumed depending on the environmental setting as urban, free space, indoor (NLOS/LOS), suburban, etc. The idea behind the environmental factor describes that the targets are exposed or shared same experiences within the experimental setting of the defined indoor parking region as a baseline though practically difficult to ensure it. One can derive that the functional relationship between CSI values and the distance is not direct proportional such that:

\[
\text{CSI}_{\text{eff}} = \frac{c}{f_c} \sqrt{\frac{\varphi}{(4\pi d)^n}}
\]

Similarly, the individual path characteristics in a wireless propagation channel are modeled as a temporal linear filter, known as the channel impulse response (CIR) [79]. Given the time invariant channel, the CIR is defined as:

\[
h(\tau) = \sum_{i=0}^{N+1} a_i e^{-j \theta_i} \delta(\tau - \tau_i)
\]

where \(a_i\); \(\theta_i\) and \(\tau_i\) are the amplitude, phase, and time delay of the \(i\)th path of signal propagation, similarly \(N + 1\) and \(\delta(\tau)\) are the total number of multipaths and Dirac delta function, respectively. The CIR is characterized as the channel frequency response (CFR) in the frequency domain, and it has been reported [79] in the commercial off-the-shelf WiFi devices that sampled versions of CFRs are revealed to upper layers in the format of CSI. Thus,

\[
H(f_m) = \sum_{i=0}^{N} a_i e^{-j(f_m \tau_i + \theta_i)}
\]

where \(f_m\) is the frequency of the \(m\)th subcarrier and \(H(f_m)\) is CSI at the \(m\)th subcarrier, and each CSI depicts the amplitude and phase of a subcarrier as:

\[
|H(f_m)| = |H(f_m)| e^{j \theta_m}
\]

where \(|H(f_m)|\) is the amplitude and \(\angle \theta_m\) is the phase of each subcarrier, and the received signal gain (in dB) at each subcarrier is proportional to the amplitude of CSI and can be expressed as:

\[
\hat{H}(f_m) = 20 \log|H(f_m)|
\]

Thus, the CSI amplitude is a measure of the power of the Wi-Fi link between the transceiver. Using Equation (10), from Euler’s property, \(H(f_m)\) can be written as:

\[
H(f_m) = \sum_{i=1}^{N} a_i [\cos(2\pi f_m \tau_i) - j \sin(2\pi f_m \tau_i)]
\]

And the channel state information amplitude for each subcarrier is

\[
|H(f_m)| = \sqrt{\left[ \sum_{i=1}^{N} a_i \cos(2\pi f_m \tau_i) \right]^2 + \left[ \sum_{i=1}^{N} a_i \sin(2\pi f_m \tau_i) \right]^2}
\]
amplitude of the LOS path. Since the attenuation of signal strength along the LOS path is mainly caused by path loss and shadowing \cite{80}, \( a_0 \) can be expressed as:

\[
    a_0 = \sqrt{G_{Tx}G_{Rx}} \frac{\lambda}{(4\pi d_0)^{n/2}} H_0, \tag{15}
\]

where \( \lambda, G_{Rx} \) and \( G_{Tx} \) represent the wavelength of the transmitted signal (in meters), the antenna gains at the receiver (in dB) and transmitter (in dB), respectively. \( d_0 \) denotes the distance of the LOS path, \( n \) is the environmental attenuation factor, and \( H_0 \) represents the attenuation of signal amplitude (in dB) caused by shadowing. The NLOS paths originate from radio reflection and refraction. During each reflection or refraction, only partial energy of the signal is transmitted \cite{80}, which can be measured by a reflection or refraction coefficient \( \xi \). Therefore, based on Equation (15), with the refraction coefficient, the amplitude of the \( m \textsuperscript{th} \) path \( a_m \) can be expressed as:

\[
    a_m = \sqrt{G_{Tx}G_{Rx}} \frac{\lambda}{(4\pi d_m)^{n/2}} H_m \xi^l_m, \tag{16}
\]

where \( \xi \in (0, 1) \) is the reflection coefficient and \( l_m \) is the number of reflections (refractions) along the \( m \textsuperscript{th} \) path and each refraction is assumed to have the same coefficient. The \( d_m \) represents the distance of the \( m \textsuperscript{th} \) non-line of sight (NLOS) path and \( H_m \) denotes the attenuation of signal amplitude caused by shadowing along the \( m \textsuperscript{th} \) path. A simplified wireless propagation model is built by integrating the effects of path loss, shadowing, and multi-path based on Equations (9), (15) and (16).

3. Problem Formulation and Framework

This section presents an overview of the framework and problem formulation of CSI-based fingerprint IPS with emphasis on: (a) data fusion techniques, (b) process of database construction-based CSI-fingerprinting for both instances of training and testing datasets, (c) the system architecture of CSI-based fingerprinting for IPS, (d) approaches of knowledge transfer learning and the CRLB analysis applied to IPS.

3.1. Data Fusion Techniques

It has been noted in various types of research that RSS fingerprinting based on positioning functionality has a fundamental problem in terms of accuracy and robustness in IPS and that both temporal and spatial signal fluctuations cause spontaneous or inestimable localization errors, regardless of the affordability and ease of implementation. A survey conducted in \cite{81} stated that the fusion of multiple measurements from different sensors has been becoming crucial to improve the positioning performance. It is also clearly noted that the CSI can measure different information to yield a better location estimate than the RSSI \cite{43,82}. In \cite{81}, we have also discussed that positioning or tracking based on a single measurement could aggravate the tracking and/or positioning performance. Moreover, hybrid methods were also proposed to enhance indoor positioning performance including the hybrid-based positioning system of different localization applications such as the combinations of Bluetooth, Wi-Fi, UWB and ZigBee \cite{8}, Wi-Fi with visual light positioning (VLP) \cite{83}, Wi-Fi and Bluetooth low energy (BLE) beacons \cite{84}, and others. By analyzing the limitations of signal strength values (RSSI) fingerprint locations, geometric locations, and inertial navigation locations, an indoor data fusion method based on an adaptive unscented Kalman filter (UKF) was proposed \cite{85}. The algorithm uses a six-position error calibration method and Kalman filter to compensate for the MEMS-SINS data and establishes the correlation between location data and RSSI/geomagnetic data based on the feature sorting vector fingerprint matching method, which leads to improved data stability and indoor location accuracy \cite{85}. In this study, the scope of our work is limited to a positioning system based on the various temporal variations collected over different periods of time based on CSI fingerprints. The real measurements of CSI were considered for our analysis. We claim
that the data fusion technique as illustrated in Figure 1 below is an effective way to further improve the accuracy and robustness of indoor-based positioning given that the different temporal signal variations are aggregated to account and measure their signal differences, which could produce the net effect of their positioning performance. Figure 1 presents the proposed framework of the CSI-based fingerprint data fusion technique for IPS.

Figure 1. The proposed framework of the CSI-based Fingerprint Data Fusion technique for an indoor positioning system.

3.2. System Architecture and Database Fingerprint Construction

We have noted that positioning characteristics-based RSS fingerprinting still has a fundamental problem in accuracy and robustness in IPS, and both temporal and spatial signal fluctuations cause spontaneous or inestimable localization errors. It has been shown that the CSI at each subcarrier in OFDM can be used to characterize the target's behaviors through the two important features of amplitude and phase fluctuations of the targets in a frequency-selective fading channel. Additionally, due to the MIMO technology integrated into CSI, high-dimensional features are possibly produced in WLAN systems and can be considered as opportunity to improve the positioning performance, although computational complexity is the main trade-off needed to be addressed. Of course, the high dimensions of features by its own may not be a positive predictor for the localization process; thus, identifying the most significant predictors is a must otherwise some redundant features may degrade or inflate (i.e., cause for model overfitting) the system modeling. In line with this, a comparative analysis between CSI and RSSI was conducted and the study revealed that the physical characteristics of CSI can significantly reduce the problem of RSSI, such that (a) multipath effect propagation can be better handled (b) strong stability is provided, especially in a static environment, and relatively stable to the dynamics of the indoor environment (c) reduce radio interference of carrier frequency signals.
In this study, as depicted in Figure 2 below we adopted the second method, the so-called scene analysis to construct a database of fingerprints that comprises two main phases: training and testing phases. The CSI fingerprints are first collected from each base station (BSs) located at four different locations (there were four BSs in total) within the defined reference points (RPs) and partitioned into two parts as a training set and testing set for different purposes. Suppose \( \{c_{ij}^{ai}(td), a \in [1,n_a], i \in [1,n_c]\} \) denote the \( i^{th} \) CSI amplitude value at the \( r^{th} \) RP of the \( j^{th} \) BS from \( a^{th} \) receiver antenna (Rx). The \( td \) refers to the time when the data were collected, specifically measured in number of days. \( n_c \) and \( n_a \) are the number of CSI measurements at each RP and number of antennas of a BS, respectively. Let \( C_r^f(t) \) represent the aggregated values of \( c_{ij}^{ai}(td) \), which are collected from all RPs of the \( j^{th} \) BS. Let’s consider that the CSI measurements were collected on different days of two months of September and October, 2020 represented as \( C_r^f(t_1) \) and \( C_r^f(t_2) \), and \( \rho_{ij}^{ab} \) denote the correlation coefficient between the two vectors as given in Equation (31). Now, we can formulate the problem mathematically such that the general fingerprint-CSI based positioning of the indoor environment scenario is divided into \( R \) reference points (RPs). Each RP represents a target’s location and is indexed with a label \( r, (r=0,1,\ldots,R-1) \). There were \( b \) detectable base stations \( (j=1,2,\ldots,b) \) in total. Thus, the \( j^{th} \) CSI amplitude value at the \( r^{th} \) RP of the \( j^{th} \) BS from \( a^{th} \) antenna in a particular day of \( td \) forms a vector and the fingerprint database can be represented as a multi-dimensional matrix such that \( C_r^f(t_d) \):

\[
C_{ij}^{ai}(td) = \{C_{ij}^{ai}(td)\} = \begin{bmatrix}
    c_{01}^{11} & \cdots & c_{01}^{1n} \\
    \vdots & \ddots & \vdots \\
    c_{01}^{(R-1)1} & \cdots & c_{01}^{(R-1)n} \\
    c_{11}^{11} & \cdots & c_{11}^{1n} \\
    \vdots & \ddots & \vdots \\
    c_{11}^{(R-1)1} & \cdots & c_{11}^{(R-1)n} \\
    \vdots & \ddots & \vdots \\
    c_{n1}^{11} & \cdots & c_{n1}^{1n} \\
    \vdots & \ddots & \vdots \\
    c_{n1}^{(R-1)1} & \cdots & c_{n1}^{(R-1)n} \\
\end{bmatrix}
\] (17)

The fingerprint database is described explicitly as: \( C_r^{ai}(td) = \{C_{ij}^{ai}(td); (x_r,y_r)\} ; \) \( r=0,1,\ldots,R-1 ; i=1,2,\ldots,n_i \& j=1,2,\ldots,b \), and \( (x_r,y_r) \) is the corresponding coordinate to the associated location of the CSI signature fingerprint. The training set constituted about 80 percent of the total instances, and the remaining are allotted to the testing dataset. To this end, the predictive models were trained to characterize the ‘signal-to-location’ relationship given the training instances and their corresponding labels of the location were stored in the database fingerprints (offline phase). This process was repeated for all RPs to characterize or store the signal signature of a reference point with their corresponding location. The learned models were then applied to infer the location of the target’s location based on the testing data points or measurements by mapping into the highest likelihood similarity of signal-signature stored during the training phase time. The target instances would be the testing dataset and denoted as \( \{C_r^{ai}\} ; i=1,2,\ldots,n_i \& j=1,2,\ldots,b \) and the training dataset representing the source domain can be represented as: \( C_r^{ai} = \{C_{ij}^{ai}; (x_r,y_r)\} ; \) \( r=0,1,\ldots,R-1 ; i=1,2,\ldots,n_s \& j=1,2,\ldots,b_s \) called the labeled source data. \( n_t \) and \( n_s \) represent the numbers of measurements for the target and source data instances, respectively. Figure 2 below demonstrates the proposed system architecture of the CSI-based fingerprint for indoor positioning system.
3.3. Principal Component Analysis (PCA)

With the advancement of Wi-Fi network technology, the CSI signal feature can now be extracted from commercial Wi-Fi devices using network interface cards (NICs) that can provide multi-channel subcarrier phase and amplitude information, allowing to better characterize the signal propagation model with the help of MIMO-OFDM technologies. Thus, CSI can present different subcarrier amplitude and phase characteristics for different propagation environments and the overall structure characteristics of CSI remain relatively stable compared with the RSS signal feature, though appropriate signal processing technology (SPT) is required. Nevertheless, the CSI-based IPS still has a challenging problem in severe dynamic range and fluctuations among high-dimensional channels mainly due to indoor multipath effects and temporal variations. This could be justified because the spread of the CSI in both the amplitude and phase values measured at a particular reference point is so dynamic in nature. Additionally, the inherent heterogeneity of the environment resulting from multipath effects and time differences of when measurements were taken have a significant effect on the distribution of the CSI values received at each reference point. Thus, there is no guarantee that the signal variations be represented by a single value for a target’s position even with the same device. To address these issues, we proposed a data fusion technique for reducing channel fluctuations and improving target positioning-based heterogeneous knowledge transfer by creating a new feature vector based on the most significant predictors. Furthermore, we used the principal component analysis (PCA) technique to reduce noise and irrelevant or insignificant features, because the CSI amplitude or fingerprints received from multiple base station antennas could be mismatched with the actual target’s fingerprint patterns due to fingerprint duplication. Moreover, with the use of the PCA, we addressed the high-dimensional curse for the CSI-based fingerprint method.
The data preprocessing of PCA was used to reduce computational complexity and ensure cost effectiveness by avoiding unnecessarily high costs associated with extra Wi-Fi access points (Wi-Fi APs) deployment (sources for irrelevant signal features) from the model.

Recall that the \( i \)th CSI amplitude values received at the \( r \)th RP of the \( j \)th BS from \( a \)th antenna in a particular day of \( t_d \) forms a vector, and the fingerprint database collected on different days can be represented as a multi-dimensional matrix. Each vector of the CSI amplitude measurements are assumed to be independent and identically distributed such that the joint probability density function with \( p \) dimension of random vectors of the CSI amplitude values observed at different temporal variations can be defined as:

\[
f\left(C_{ai}^{ij}(s), \mu\right) = \prod_{i=1}^{n} \left\{ \frac{1}{(2\pi)^{p/2}} |\Sigma|^{1/2} e^{-\frac{1}{2}(C-\mu)'\Sigma^{-1}(C-\mu)} \right\}
\]

(18)

where \( C_{ai}^{ij}(s) = C \) and \( \mu \) represents all possible measurements that the CSI amplitude values could take and the mean, respectively, and this joint normal density is denoted as \( C_{ai}^{ij}(s) \sim N(\mu, \Sigma) \). Similarly, the covariance of CSI amplitude measurements from several antennas of a base station can be determined as:

\[
cov(S_z, S_q) = \frac{\sum (S_z - \overline{S}_z)(S_q - \overline{S}_q)}{n-1}
\]

(19)

The variances of the measurements \( S_z \) and \( S_q \) can be computed as:

\[
sd(S_q) = \sqrt{\frac{\sum_{i=1}^{n} (C_{qi} - \overline{C}_q)^2}{n_q - 1}} \quad \text{&} \quad sd(S_z) = \sqrt{\frac{\sum_{i=1}^{n} (C_{zi} - \overline{C}_z)^2}{n_z - 1}}
\]

(20)

where the sample means of the CSI amplitude for the random vectors of \( Z \) and \( Q \) can be given as:

\[
\overline{C}_z = \frac{1}{n_z} \sum_{i=1}^{n_z} C_{ai}^{ij}(s_z) \quad \text{&} \quad \overline{C}_q = \frac{1}{n_q} \sum_{i=1}^{n_q} C_{ai}^{ij}(s_q)
\]

(21)

Equivalently, we can write the above equation in (19) as:

\[
\sum_{i=1}^{n} (C_i - \overline{C}_i)(C_i - \overline{C}_i)^T = \Sigma \Sigma^T
\]

(22)

To this end, the PCA algorithm comprises four major steps:

(a) Standardize each CSI measurement as:
\[ C_{ai}^{ij}(s_z) = C_{ai}^{ij}(s_z) - \frac{1}{n_z} \sum_{i=1}^{n_z} C_{ai}^{ij}(s_z) \]

(b) Calculate the covariance matrix of the CSI sample measurements:
\[ \sum_{i=1}^{n} (C_i - \overline{C}_i)(C_i - \overline{C}_i)^T = \Sigma \Sigma^T \]

(c) Obtain Eigen value decomposition of covariance matrix

(d) Obtain projection matrix

Algorithm 1 is a pseudo code used to derive the new fingerprints of features-based CSI using both data fusion techniques and principal component analysis.
(ii) Although channel state information could help us provide more feature spaces, the patterns to interfere with the actual target’s fingerprint due to random noise or indoor phase received at a reference point from multiple antennas of a base station assumed to be leveraged to the target domain. The Minkowski is a generalized distance metric between most significant features, independent features, and related source knowledge could be and the dependence of CSI measurements on different BS antennas were managed through heterogeneity and multipath effects. Additionally, potentially duplicated CSI fingerprints over several days that could have different patterns due to the inherent environmental minimize the temporal CSI amplitude fluctuations by considering different measurements for higher dimensional curse must be handled. Thus, we proposed the data fusion technique to multipath effect scenarios and ultimately have a negative effect on positioning performance. The grid point can be duplicated, possibly by multiple BSs, and this causes the matching independently and do not interfere with each other. In practice, however, the CSI amplitude be independent such that the RF signals transmitted from different BSs are transmitted for different purposes. Along with this, we have noted that there are two major challenges defined reference points (RPs) and partitioned into two parts as t training set and testing set (BS) located at four different locations (there were four BSs in total) within the 3.4. Heterogeneous Knowledge Transfer

In this study, the CSI fingerprints were first collected from each antenna of a base station (BS) located at four different locations (there were four BSs in total) within the defined reference points (RPs) and partitioned into two parts as t training set and testing set for different purposes. Along with this, we have noted that there are two major challenges that need to be considered in heterogenous transfer learning: (i) the CSI amplitude and phase received at a reference point from multiple antennas of a base station assumed to be independent such that the RF signals transmitted from different BSs are transmitted independently and do not interfere with each other. In practice, however, the CSI amplitude of a grid point can be duplicated, possibly by multiple BSs, and this causes the matching patterns to interfere with the actual target’s fingerprint due to random noise or indoor multipath effect scenarios and ultimately have a negative effect on positioning performance. (ii) Although channel state information could help us provide more feature spaces, the higher dimensional curse must be handled. Thus, we proposed the data fusion technique to minimize the temporal CSI amplitude fluctuations by considering different measurements over several days that could have different patterns due to the inherent environmental heterogeneity and multipath effects. Additionally, potentially duplicated CSI fingerprints and the dependence of CSI measurements on different BS antennas were managed through principal component analysis.

Thus, the objective function was minimized over the new feature spaces such that the most significant features, independent features, and related source knowledge could be leveraged to the target domain. The Minkowski is a generalized distance metric between two vectors and given as:

$$\| r \| = \sum_{t'=1}^{n} \| C_{t'j}^{n} (s_{t'}) - \hat{C}_{t'j}^{n} (s_{t'}^{f}) \|$$

(23)
\[ \| r \|_1 \text{ and } \| r \|_2 \text{ denotes Manhattan and Euclidean distances, respectively. Similarly, } \mathcal{C}_{\alpha r}^f(s_{tr}) \text{ and } \hat{\mathcal{C}}_{\alpha r}^f(s_{ts}) \text{ represent the fused CSI amplitude values of instances of training and testing datasets respectively. The transfer coefficients } (\omega_{r f}) \text{ constraint is to minimize the amplitude measurements of fluctuations between the instances of both the training and testing dataset. The objective function’s equality constraint would assign higher weights to the most related source instances and lower weights to the least related source instances. The new feature vectors were used to minimize the variations of the amplitude values of CSI based fingerprints and the transfer coefficients can be estimated as:} \]

\[
\min_{i'=1}^{n'} \sum_{j'=1}^{n'} \omega_{r f} \| \mathcal{C}_{r f}^i(s_{tr}) - \hat{\mathcal{C}}_{r f}^{i'}(s_{ts}) \|_2^2 \\
\quad \text{s.t. } \sum_{i'=1}^{n'} e^{-\omega_{r f}} = 1, i'= 1, 2, \ldots, n' \tag{24}
\]

We have used the Lagrangian multiplier method to solve the constrained optimization problem of Equation (24), and we assumed the location estimate at the \((t-1)\text{th}\) iteration \(\omega_{r f}^{(t-1)}\) is obtained (mapped into 2D), and we need to estimate the location of the actual target at the \(t\)\text{th} iteration denoted by \(\omega_{r f}^{(t)}\). CSI fingerprints are collected from each Wi-Fi access point at multiple locations within the reference points (RPs) defined during the training phase. A predictive model is trained to characterize the signal-to-location learning relationship. During the testing phase, the learned model is then used to predict the target’s location based on the new received CSI measurement. One can rewrite Equation (24) using the Lagrangian multiplier method as:

\[
L(\omega_{r f}, \lambda) = \sum_{i'=1}^{n'} \sum_{j'=1}^{n'} \omega_{r f} \| \mathcal{C}_{r f}^i(s_{tr}) - \hat{\mathcal{C}}_{r f}^{i'}(s_{ts}) \|_2^2 + \lambda \left( \sum_{j'=1}^{n'} e^{-\omega_{r f}} - 1 \right) = 0 \tag{25}
\]

where \(\lambda\) is the Lagrangian multiplier. By letting the partial derivative of the Lagrangian with respect to \(\omega_{r f}\) and \(\lambda\) be zeros, we obtain:

\[
\frac{\partial}{\partial \omega_{r f}} L(\alpha_{r f}, \lambda) = \begin{cases} \\
0 \\
0 \\
0 \\
0 \\
0 \\
\end{cases}
\]

By adding up the first \(n\) terms in Equation (26), we can obtain:

\[
\lambda = \sum_{i'=1}^{n} \| \mathcal{C}_{r f}^i(s_{tr}) - \hat{\mathcal{C}}_{r f}^{i'}(s_{ts}) \|_2^2 \tag{27}
\]

And substituting Equation (27) into Equation (26) gives the estimated transfer coefficients as:

\[
\omega_{r f} = -\ln \left( \frac{\| \mathcal{C}_{r f}^i(s_{tr}) - \hat{\mathcal{C}}_{r f}^{i'}(s_{ts}) \|_2^2}{\sum_{i'=1}^{n} \| \mathcal{C}_{r f}^i(s_{tr}) - \hat{\mathcal{C}}_{r f}^{i'}(s_{ts}) \|_2^2} \right) \tag{28}
\]

The pseudo code for positioning using the heterogenous knowledge transfer-based data fusion technique is provided in Algorithm 2.
Algorithm 2 The proposed heterogenous knowledge transfer-based CSI-fingerprint indoor positioning system

1. Input: Refined Sources Fingerprint
   
   \[ C_{\text{ref}i}^{a} (s_{d})_{tr} = \{ C_{\text{ref}i}^{a} (s_{1}^{a}); C_{\text{ref}i}^{a} (s_{2}^{a}); \ldots; C_{\text{ref}i}^{a} (s_{t}^{a}); (x'_{r}, y'_{r}) \}; a = 1, n_{a} \; j = 1, 2, \ldots, b \]

2. Input: Refined CSI testing dataset
   
   \[ C_{\text{ref}i}^{al} (s_{d})_{ts} = \{ C_{\text{ref}i}^{al} (s_{1}^{al}); C_{\text{ref}i}^{al} (s_{2}^{al}); \ldots; C_{\text{ref}i}^{al} (s_{t}^{al}); (?, ?) \}; a = 1, n_{a} ; j = 1, 2, \ldots, b \]

3. Output: Fused refined CSI fingerprints \( C_{\text{ref}i}^{al} (s_{f}^{al}) \), position estimates, transfer coefficients

   4. for \( r = 0 : R - 1 \) do
   5.     for \( j = 0 : b - 1 \) do
   6.         for \( a = 0 : n_{a} - 1 \) do
   7.             for \( i = 0 : n - 1 \) do
   8.                 repeat
   9.                     Step 1:
   10.                    Fuse the CSI amplitude fingerprints from all the sources with temporal variations as in Algorithm 1
   11.                     Step 2:
   12.                        Obtain projection matrix: \( G' = (g_{1}, g_{2}, \ldots, g_{e}) \)
   13.                     Step 3:
   14.                        Compute \( \omega_{e, f}^{a} \) by using Equations (24)–(28)
   15.                     end for
   16.                     end for
   17.                     end for
   18.                     end for
   19.                     Train a classifier from \( C_{\text{ref}i}^{al} (s_{f}^{al}) = \{ C_{\text{ref}i}^{al} (s_{e}^{al}); (x'_{r}, y'_{r}) \} \) while considering weights of source domains \( \omega_{e, f}^{a} \)
   20.                     Estimate target’s location \( (x'_{r}, y'_{r}) \) on \( C_{\text{ref}i}^{al} (s_{e}^{al}) \) by applying the trained classifier
   21.                     \[ f \left( \left\{ C_{\text{ref}i}^{al} (s_{e}^{al}); (x'_{r}, y'_{r}) \right\}, \omega_{e, f}^{a} \right) \]
   22.                     return \( \{ C_{\text{ref}i}^{al} (s_{f}^{al}); (x'_{r}, y'_{r}) \}_{tr}; \{ C_{\text{ref}i}^{al} (s_{e}^{al}); (x'_{r}, y'_{r}) \}_{ts}; \omega_{e, f}^{a} \)

3.5. Evaluation Metrics for Positioning

This section presents the metrics applied to evaluate the positioning performance, and we compared the performance of the proposed algorithms against different machine learning algorithms taken as baselines and validated through extensive real-life experiments. The dataset were collected over several days with potential temporal signal fluctuations based on CSI-fingerprinting of the real measurements. In this study, the root mean square error was used to evaluate the positioning performance of the proposed algorithm, and it is defined as:

\[
\text{RMSE} = \left[ \frac{1}{n_{f}} \sum_{f=1}^{n_{f}} \left( (y_{f} - \hat{y}_{f})^{2} + (y_{f} - \hat{y}_{f})^{2} \right) \right]^{1/2}
\]  \hspace{1cm} (29)

where \( [\hat{x}_{r}, \hat{y}_{r}]^{T} \) and \( [x_{r}, y_{r}]^{T} \) are the predicted location estimate and actual location of the target, respectively. And \( n_{f} \) is the total number of samples to be located in the target domain.

3.6. Cramer-Rao Lower Bound (CRLB) Analysis for IP Performance-Based CSI-Fingerprinting

This section presents the analysis of CRLB for location estimation-based channel state information-measurements used to evaluate the performance of indoor positioning system and estimates a lower limit for the variance of any unbiased estimator of an unknown parameter. Additionally, the CRLB is suitable for stationary gaussian parameter estimation [86–89]. The CSI-based fingerprint database is denoted as \( C_{\text{ref}i}^{al} (t_{d}) = \{ C_{\text{ref}i}^{al} (t_{d}); (x_{r}, y_{r}) \}; r = 0, 1, \ldots, R - 1; i = 1, 2, \ldots, n \& j = 1, 2, \ldots, b \) and \( (x_{r}, y_{r}) \) is the corre-
sponding coordinate to the associated location of the CSI signature fingerprint. Thus, we proposed the data fusion technique to represent the temporal CSI amplitude fluctuations by considering different measurements over several days that could have different patterns due to the inherent environmental heterogeneity and multipath effect. Suppose the \( i^{th} \) CSI amplitude values received at the \( r^{th} \) RP of the \( j^{th} \) BS from \( a^{th} \) antenna in a particular day of \( t_d \) forms a vector. Thus, the fingerprint database collected on different days can be represented as a multi-dimensional matrix and each vector of the CSI amplitude measurements are assumed to be independent and identically distributed such that the probability density function of a random vector of the CSI amplitude values observed in particular days or time measurement \( t_d \) can be defined as:

\[
f(C^{ai}_{rj}(s), \mu) = \frac{1}{\sqrt{2 \pi \sigma^2}} e^{-\frac{(C^{ai}_{rj}(s) - \mu)^2}{2 \sigma^2}}
\]

where \( C^{ai}_{rj}(s) = C \) and \( \mu \) represents all possible measurements that the CSI amplitude values could take and the mean, respectively, and this joint normal density is denoted as \( C^{ai}_{rj}(s) \sim N(\mu, \sigma^2) \). Similarly, the correlation of CSI amplitude measurements from several antennas of a base station can be determined as:

\[
corr(C_x, C_y) = \frac{\text{cov}(C_x, C_y)}{sd(C_x) \times sd(C_y)}
\]

In line with this, the distributions of the CSI values measured at a particular reference point are so dynamic in nature due to the inherent heterogeneity of the environment, and time differences in measurements being taken have a significant effect on the distribution of the CSI values received at each RP. As a result, there is no guarantee that the signal fluctuations will be represented by a single value for a specific position even with the same device. The CRLB used the CSI real measurements to analyze the lower bound of the location estimation error, and it is significantly important to characterize the properties of this lower bound to evaluate the impact of different parameters on the accuracy of a target's localization. Moreover, the CRLB analysis can also provide important system design suggestions by revealing error trends with the indoor localization system deployment. To estimate the lower bound of the position error variance, first assume that the position of the \( j^{th} \) base station and the unknown location of the target are denoted as \( L_j = (x_j, y_j)^T \) and \( \hat{L}_j = (\hat{x}_j, \hat{y}_j)^T \), respectively, then the distance between the \( j^{th} \) base station and the target can be defined as \( r_{i,j}^2 = \{ (L_j - L) (L_j - L)^T \} \). It has been stated that the CSI fingerprint measurements follow a normal distribution with mean zero and variance \( \sigma^2 \). Thus, we adopted the assumption of normality and the covariance of the estimator \( L_i \), which is \( n \times 2 \), vector can be defined as:

\[
E\left\{ (L_j - \hat{L}_j) (L_j - \hat{L}_j)^T \right\} = \begin{bmatrix}
E(x_j - \hat{x}_j)^2 & E(x_j - \hat{x}_j) E(y_j - \hat{y}_j) \\
E(y_j - \hat{y}_j) E(x_j - \hat{x}_j) & E(y_j - \hat{y}_j)^2
\end{bmatrix}
\]

The diagonal elements of (32) denote the mean squared errors, and the off-diagonal elements are the covariances between different parameters. Additionally, consider that \( L_j \) is the unknown deterministic parameter, which is to be estimated from \( n \) independent observations (CSI fingerprint-based measurements collected over different periods of time) of \( C^{ai}_{rj}(s) \), each from a distribution according to some probability density function \( f(C^{ai}_{rj}(s); L_j) \). Thus, the CRLB is defined as the variance of any unbiased estimator \( \hat{L}_j \) of \( L_j \) bounded by the reciprocal of the Fisher information matrix \( I(L_j) \) such that:

\[
\sigma^2(\hat{L}_j) \geq [I(L_j)]^{-1}
\]
And, the $I(L_j)$ is given as:

$$I(L_j) = \begin{bmatrix} I_{xx}(L_j) & I_{xy}(L_j) \\ I_{yx}(L_j) & I_{yy}(L_j) \end{bmatrix} = nE_L \left[ \left( \frac{\partial l(C; L_j)}{\partial L_j} \right)^2 \right]$$

(34)

where $I(C_i(s); L_j) = \log f \left( C_i(s); L_j \right)$ is the natural logarithm of the likelihood function for a single sample $C_i(s)$ and $E_L$ denotes the expected value with respect to the density function of $C, f \left( C_i(s); L_j \right)$. But, if $I(C_i(s); L_j)$ is twice differentiable and holds certain regularity conditions, then the Fisher information can also be defined as:

$$I(L_j) = -nE_L \left( \frac{\partial^2 l(C; L_j)}{\partial L_j^2} \right)$$

(35)

By definition, the CRLB of the unbiased estimator $L_j$ can be calculated as:

$$\text{CRLB} = \frac{I_{xx}(L_j) + I_{yy}(L_j)}{|I(L_j)|}$$

(36)

One can describe the CRLB as just the inverse of FIM and rewrite Equation (19) as:

$$\text{Cov}(L_j) \geq |I(L_j)|^{-1} = \begin{bmatrix} I_{yy}(L_j) & -I_{yx}(L_j) \\ -I_{xy}(L_j) & I_{xx}(L_j) \end{bmatrix}$$

(37)

Thus, the variance of the unbiased estimator for $L_j$ is given by:

$$\text{var}(\hat{L}_j) \geq E(x_j - \hat{x}_j)^2 + E(y_j - \hat{y}_j)^2$$

(38)

Now, from the above Equations (33) and (37), one can observe the relationship between the variance of the unbiased estimator $\hat{L}_j$ and FIM, which basically satisfies the CRLB condition such that:

$$\nu(L_j) = E(x_j - \hat{x}_j)^2 + E(y_j - \hat{y}_j)^2 \geq \text{tr} \left\{ |I(L_j)|^{-1} \right\}$$

(39)

$$\text{tr}[I(L_j)]^{-1} = I_{xx}(L_j) + I_{yy}(L_j)$$

(40)

One can observe from Equation (38), the MSE of $\hat{x}_j$ and $\hat{y}_j$ can be given as:

$$\text{MSE}(\hat{x}_j) = E(x_j - \hat{x}_j)^2 \geq \frac{I_{yx}(L_j)}{|I(L_j)|}$$

(41)

$$\text{MSE}(\hat{y}_j) = E(y_j - \hat{y}_j)^2 \geq \frac{I_{xy}(L_j)}{|I(L_j)|}$$

(42)

where $\text{tr} \left\{ |I(L_j)|^{-1} \right\}$ denotes the trace of the inverse FIM and $|I(L_j)|$ represents the determinant of FIM.

In this study, the multiple measurements of CSI-based fingerprints collected on different days were used as different sources to estimate the lower bound of the error variance of the target’s location using the application of CRLB. We considered that the CSI amplitudes collected on various days follow a multivariate normal distribution. This is consistent with the assumption that different signal features including the well-known features such as RSS, CSI, TOA, and AOA follow a normal distribution [89,90]. Recall that the $i^{th}$ CSI amplitude values received at the $r^{th}$ RP of the $j^{th}$ BS from $a^{th}$ antenna in a particular day of
The CRLB of multiple CSI fingerprint-based measurements for localization can also be described as

\[
\text{CRLB}(\mathbf{C}_m^\alpha) = \frac{1}{G} \left[ \frac{u_{jr}^2 + v_{jr}^2}{u_{jr}^2 \times v_{jr}^2 - (u_{jr} v_{jr})^2} \right]
\]

where:

\[
G = [\alpha_1, \alpha_2, \ldots, \alpha_n] \sum_{i=1}^{n} -1 \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}
\]
Scenario 1: Consider eight temporal variations of measurements of month 1 as different sources due to their distribution follow different patterns: Thus, we consider eight different sources for evaluating the positioning performance

\[
G = \begin{bmatrix}
\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8
\end{bmatrix}
\begin{pmatrix}
\sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{18}\sigma_1\sigma_8 \\
\rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \cdots & \rho_{28}\sigma_2\sigma_8 \\
\vdots & \vdots & \ddots & \vdots \\
\rho_{81}\sigma_8\sigma_1 & \rho_{82}\sigma_8\sigma_2 & \cdots & \sigma_8^2
\end{pmatrix}^{-1}
\begin{bmatrix}
\alpha_1 \\
\alpha_2 \\
\alpha_3 \\
\alpha_4 \\
\alpha_5 \\
\alpha_6 \\
\alpha_7 \\
\alpha_8
\end{bmatrix}
\]

(47)

And,

\[
\text{CRLB}\left(C_{ij}^n(s_{i:1→8})\right) = \frac{1}{G} \begin{pmatrix}
u_f^2 + v_f^2 \\
u_f^2 \times v_f^2 - (u_f v_f)^2
\end{pmatrix}
\]

(48)

Scenario 2: Consider five temporal variations of measurements of month 2 as different sources due to their distribution follow different patterns: Thus, we consider five different sources for evaluating the positioning performance

\[
G = \begin{bmatrix}
\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5
\end{bmatrix}
\begin{pmatrix}
\sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{15}\sigma_1\sigma_5 \\
\rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \cdots & \rho_{25}\sigma_2\sigma_5 \\
\vdots & \vdots & \ddots & \vdots \\
\rho_{51}\sigma_5\sigma_1 & \rho_{52}\sigma_5\sigma_2 & \cdots & \sigma_5^2
\end{pmatrix}^{-1}
\begin{bmatrix}
\alpha_1 \\
\alpha_2 \\
\alpha_3 \\
\alpha_4 \\
\alpha_5
\end{bmatrix}
\]

(49)

And,

\[
\text{CRLB}\left(C_{ij}^n(s_{i:1→5})\right) = \frac{1}{G} \begin{pmatrix}
u_f^2 + v_f^2 \\
u_f^2 \times v_f^2 - (u_f v_f)^2
\end{pmatrix}
\]

(50)

Scenario 3: Consider the aggregate temporal variations of measurements of various sources of both months into two different sources and calculate the CRLB of those fused different sources for evaluating the positioning performance purpose:

\[
G = \begin{bmatrix}
\alpha_1^f, \alpha_2^f
\end{bmatrix}
\begin{pmatrix}
\sigma^2_1 f & \rho_{12}\sigma_1^f\sigma_2^f \\
\rho_{21}\sigma_2^f\sigma_1^f & \sigma^2_2 f
\end{pmatrix}^{-1}
\begin{bmatrix}
\alpha_1^f \\
\alpha_2^f
\end{bmatrix}
\]

(51)

And,

\[
\text{CRLB}\left(C_{ij}^n(s_{1}^f,s_{2}^f)\right) = \frac{\sigma_1^2 f \sigma_2^2 f - \rho_{12}^2 \sigma_1^f\sigma_2^f}{\alpha_1^2 f \sigma_2^2 f + \alpha_2^2 f \sigma_1^2 f - \rho_{12} \sigma_1^f\sigma_2^f\alpha_1^f\alpha_2^f}
\begin{pmatrix}
u_f^2 + v_f^2 \\
u_f^2 \times v_f^2 - (u_f v_f)^2
\end{pmatrix}
\]

(52)

Now, to estimate the lower bound of the position error variance, first consider that the position of the $j^{th}$ base station and unknown location of the target are denoted as $L_j = (x_j, y_j)^T$ and $\hat{L}_j = (\hat{x}_j, \hat{y}_j)^T$ respectively; then, the distance between the $j^{th}$ base station and $i^{th}$ reference point can be defined as $r_{ij}^2 = \left(\left(L_j - \hat{L}_j\right) \left(L_j - \hat{L}_j\right)^T\right)$, and for simplicity purposes let’s use $\hat{r}_{ij}^2$ instead of $r_{ij}^2$. Similarly, the distance between the $j^{th}$ base station and target point can be represented as $\hat{r}_{ji}^2$. Researchers have found that the CSI fingerprint measurements follow a normal distribution with mean zero and variance $\sigma^2$. Besides, $n_b$ and $\gamma$ represent the number of base stations and path loss attenuation factor, respectively, and $\sigma_C$ is the variance of flat fading, and multipath follows normal distribution.
In [91] have proposed a model for the effective vector of CSI from $j^{th}$ AP measured at RP and given as

$$\ln \text{CSI}_{eff,r}^2 = V_{ij} = \ln \left( \frac{c^2 \delta}{f_r^2 (4\pi r_{ij})^\gamma} \right) + \epsilon_r$$

(53)

where $\delta$ is an environment factor, $\gamma$ is the path loss attenuation factor, $c$ is the radio velocity, and $\epsilon_r$ is a measurement noise and follows normal distribution $\epsilon_r \sim N(0, \sigma_r^2)$. The same device in an unknown location $L_j = (x_j, y_j)$ measures a CSI value of $V_{ij}$ from the same AP given as:

$$\ln \text{CSI}_{eff,d}^2 = V_{ij} = \ln \left( \frac{c^2 \delta}{f_r^2 (4\pi r_{ij})^\gamma} \right) + \epsilon_i$$

(54)

Similarly, $\epsilon_i \sim N(0, \sigma_i^2)$. When the fingerprint is used for localization, the AP locate at an unknown location and we utilize the fingerprint to estimate the coordinate of $L_j$. $H_k$ is defined as the difference between $H_{ej}$ and $H_{ij}$ given by:

$$H_{ij} - H_{ej} = \ln \left( \frac{r_{ij}}{r_{ej}} \right) + \epsilon_d$$

(55)

Thus, the probability density function of the $H_k$ given the location $f(H_k | L_j)$ or the pdf of the estimated location-based CSI-fingerprint is given by:

$$f(H_k(s_k) | L_j) = \prod_{i=1}^n \left\{ \frac{1}{\sqrt{2\pi\sigma_C}} e^{-\frac{1}{2\sigma_C^2}} (H_k(s_k) - \ln \left( \frac{r_{ij}}{r_{ej}} \right))^2 \right\}$$

(56)

where $C_k^2(s_m) = H_k(s_k) - H_{ij}(s_k)$ denotes the difference in the effective vector of the CSI measurement representing the $i^{th}$ target at the $j^{th}$ RP of the $j^{th}$ base station. The CRLB-based CSI-fingerprint based on Equation (45) can also be given as:

$$\text{CRLB}(C) = \frac{1}{\alpha_C} \left[ \frac{u_{ij}^2 + v_{ij}^2}{u_{ij}^2 \times v_{ij}^2 - (u_{ij} v_{ij})^2} \right]$$

(57)

where $\alpha_C = \left( \frac{\gamma}{\sqrt{2\pi\sigma_C}} \right)^2$. Consider the special case of the above equation in (1) and the joint probability density function with $p = 2$ dimensions of random vectors of the CSI amplitude values observed on different temporal variations can be defined as:

$$f(C_k^2(s_1^f, s_2^f)) = \frac{1}{2\pi\sigma_1\sigma_2 \sqrt{1 - \rho^2}} \exp \left\{ -\frac{1}{2(1 - \rho^2)} \left[ \left( \frac{c_1 - \mu_1}{\sigma_1} \right)^2 - 2\rho \left( \frac{c_1 - \mu_1}{\sigma_1} \right) \left( \frac{c_2 - \mu_2}{\sigma_2} \right) + \left( \frac{c_2 - \mu_2}{\sigma_2} \right)^2 \right] \right\}$$

(58)

where $\rho$ denotes the correlation coefficient between the two fused CSI fingerprints collected in month 1 and month 2. Let $\mu = [\mu_{s_1}, \mu_{s_2}]^T$ and $\Sigma = \begin{bmatrix} \sigma_{s_1}^2 & \rho_{s_1}s_2 \sigma_{s_2}^2 \\ \rho_{s_2} \sigma_{s_1}^2 & \sigma_{s_2}^2 \end{bmatrix}$. Similarly, based on the above Equation (57), the probability density function of the estimated location-based CSI-fingerprint for $p = 2$ dimensions can be given as:

$$f(C_k^2(s_1^f, s_2^f)) = \frac{1}{2\pi\sigma_{s_1}\sigma_{s_2} \sqrt{1 - \rho_{s_1s_2}^2}} \exp \left\{ -\frac{1}{2(1 - \rho_{s_1s_2}^2)} [A] \right\}$$

(59)

$$A = \left( \frac{H_1(s_1) - \ln \left( \frac{r_{ij}}{r_{ej}} \right)}{\sigma_{s_1}^2} \right)^2 - 2\rho_{s_1s_2} \left( \frac{H_1(s_1) - \ln \left( \frac{r_{ij}}{r_{ej}} \right)}{\sqrt{\sigma_{s_1}^2}} \right) \left( \frac{H_2(s_2) - \ln \left( \frac{r_{ij}}{r_{ej}} \right)}{\sqrt{\sigma_{s_2}^2}} \right) + \left( \frac{H_2(s_2) - \ln \left( \frac{r_{ij}}{r_{ej}} \right)}{\sigma_{s_2}^2} \right)^2$$

(60)
Thus, one can rewrite Equation (57) for the CRLB based CSI-fingerprint measurements in relation to the angle as follows:

$$\text{CRLB}(\mathbb{C}) = \frac{1}{\alpha^2} \left[ \sum_{r=1}^{R} \left( \frac{\cos \beta_r}{d_r} \right)^2 + \sum_{r=1}^{R} \left( \frac{\sin \beta_r}{d_r} \right)^2 \right] $$

(61)

The CRLB of the fused data from different sources can be calculated based on the concept given in the above Equation (52), and the CRLB or lower bound error of the location for the two fused sources of CSI-based fingerprint measurements collected in two separate months of several days can be given as:

$$\text{CRLB}(\mathbb{C}) = \left[ 1 - \frac{\rho_s^2}{\alpha_s^2} \right] \left[ \sum_{r=1}^{R} \left( \frac{\cos \beta_r}{d_r} \right)^2 + \sum_{r=1}^{R} \left( \frac{\sin \beta_r}{d_r} \right)^2 \right] $$

(62)

In conclusion, the above derivation revealed that the lower bound of the variance of the location estimator depends on (a) the angle of the base stations, (b) number of base stations, (c) distance between the target and base station, $d_j$, (d) correlation of features, $\rho$, and (e) signal propagation parameters, $\sigma_C$ and $\gamma$. Moreover, experimental results have shown that the number of antennas of a base station could affect the lower bound location estimation error by generating a higher dimension of features, unless the most significant predictors are selected; otherwise, the accuracy of positioning performance could be degraded due to the dimensionality curse. This analytical derivation also revealed that the fused data have shown the hybrid effect of the temporal signal variations resulting due to time differences in the measurements being taken.

4. Experimental Results and Discussion

This section presents a number of real-world experiments that were carried out on various occasions, or on a daily basis to measure the temporal signal fluctuations using the fingerprints of the channel state information (interested in variations among the measurement days), and we assess our proposed algorithms as they were applied to these CSI real measurements of various distributions. First, experimental conditions, datasets, and an analysis of the algorithms’ overall performance were presented.

4.1. Experimental Settings

(1) Datasets

The experiments were carried out at Huawei on an area of 75 m$^2$, which contained 225 and 110 reference points (RPs), respectively, all of which were evenly dispersed ($\geq 0.5$ m) from each other as shown in Figure 4. A first survey was conducted in September 2020, and a second survey was conducted in October 2020. For the first month, eight measurements were taken on eight different days. In contrast, five different measurements were taken on five separate days in the second month. To create the entire CSI fingerprint database, four base stations and one transmitter were available to collect channel state information from a location server. The number of reference points collected per day are different in both months. The total number of reference points in each month also differ. Since the measurements under study are naturally unbalanced, our analysis must account for possible discrepancies caused by the imbalance of data. Details of the system description used in the study are provided in Table 2 below. The layout and environmental settings of the real-life experimental scenarios generated for the various datasets are depicted in Figures 4 and 5 below.
Table 2. System description.

| Spectrum | 5.8 GHz |
|----------|---------|
| Bandwidth | 100 MHz |
| Subcarrier bandwidth | 120 kHz |
| Label position frequency | 1 Hz |

Figure 4. Experimental layout conducted at Huawei company for indoor parking-based CSI fingerprints.

Figure 5. Distribution of the reference points for generating CSI-based fingerprints during September and October 2020.
Table 2. System description.

| Parameter               | Value   |
|-------------------------|---------|
| Spectrum                | 5.8 GHz |
| Bandwidth               | 100 MHz |
| Subcarrier bandwidth    | 120 kHz |
| Label position frequency| 1 Hz    |

4.2. Distribution of the Temporal Variations of CSI Amplitude Measurements

Table 3 depicts the distribution of channel state information per label, specifically the number of measurements collected from each reference point (RP), and demonstrates that the label distribution is unbalanced for all datasets collected for the eight different days of September 2020, noted as month 1. In contrast, the number of measurements collected from each reference point (RP) are all equal, indicating that the label distribution for all datasets collected during the five separate days of month 2, October 2020, is balanced. Feature-scaling techniques were used in this study to avoid the dominance of the larger occurrence of labels within the cluster, which would otherwise cause the larger features to dominate the others within the cluster and negatively affects modeling performance. However, for the month 2 data collection, an equal number of CSI fingerprint measurements were recorded for all RPs, and thus the reference point distribution was considered balanced for all datasets collected in October 2020. In general, principal component analysis is recommended to reduce the effect of the dimensionality curse on large datasets with higher dimensions of features where computational complexity is significant. For the use of principal component analysis, features must be standardized. According to Table 3, the total number of RPs collected during the first month is 225, and the total number of labels collected during the second month is 110. The CSI fingerprint measurements features collected during the survey time provide complete information that assisted in determining target positioning, including CSI real measurements, CSI imaginary measurements, latitude, longitude, coordinate systems, time of arrival, angle of arrival, and other relevant data for all scenarios.

Figure 6 depicts the distribution of the principal components and their corresponding reference points of CSI-based fingerprint measurements collected on 16 September 2020, noted as month 1 and day 1 (d1M1). Despite a minor difference caused by temporal signal fluctuations generated within the environment, the CSI-based measurements of both training and testing datasets received from various antennas of a base station at a reference location appear to follow a specific distribution. The first and second principal components of both instances from the training and testing datasets provide the highest proportion of variance explainability to the positioning model and allow one to visualize the effect of variations in the target’s location using the most significant predictors. The first and second principal components could explain the total variance of the system model as a linear combination of the first and second principal components. However, as illustrated in Figure 6c,d, the two principal components of the last principal, which actually constitute about 4% of the total variations in the predictive model for both instances of the training and testing dataset, appear to follow different distributions or the testing dataset appears to fail to represent the training distribution. While principal component analysis is important for minimizing computational complexity, removing significant features has a critical impact on positioning accuracy and negatively impacts positioning performance. A “base model,” according to [38], consists of only two principal components to represent the variational distributions of the target’s prediction, accounting for approximately 16% of the total variance explainability of the model, though [38] achieved 56%. Even though the base model has a lower variance explainability ratio of 16% versus 56%, this finding is consistent with the findings in [38] that lower feature space dimensions can improve computational cost and model simplicity. This clearly demonstrates that the base model was unable to fit our problem with the desired goal, necessitating a larger number of principal components. As a result, we proposed in this study the number of principal components that could account for 95% of the explained variance ratio of system performance in general, but
these dimensions of principal components must also verify if the desired system modeling with the expected positioning accuracy can be maintained. Similarly, Figure 7 depicts the distribution of the principal components and their corresponding reference points of CSI-based fingerprint measurements collected on 9 October 2020, abbreviated as month 2 and day 1 (d1M2).

Table 3. Distributions of channel state information measurements per Label.

| September 2020 Labels (#RPs = 225) | 0 | 1 | 127 | – | 159 | 187 | 224 |
|------------------------------------|---|---|-----|---|-----|-----|-----|
| #of CSI values per Label           |   |   |     |   |     |     |     |
| Dataset 1 (#Labels = 31)           | 793| 811| 742 | – | 836 | 806 | 1183|
| Dataset 2 (#Labels = 6)            | 899| 673| 820 | – | 831 | 842 | 1017|
| Dataset 3 (#Labels = 25)           | 931| 805| 803 | – | 810 | 1060| 1254|
| Dataset 4 (#Labels = 37)           | 798| 841| 863 | – | 930 | 795 | 806 |
| Dataset 5 (#Labels = 36)           | 822| 785| 797 | – | 791 | 804 | 841 |
| Dataset 6 (#Labels = 27)           | 870| 861| 994 | – | 822 | 799 | 833 |
| Dataset 7 (#Labels = 31)           | 1009| 804| 801 | – | 857 | 792 | 832 |
| Dataset 8 (#Labels = 32)           | 814| 804| 819 | – | 799 | 857 | 832 |
| October 2020 Labels (#RPs = 110)  | 0 | 1 | 127 | – | 99 | 107 | 109 |
| #of CSI values per Label           |   |   |     |   |    |    |    |
| Dataset 1 (#Labels = 22)           | 928| 805| 875 | – | 847 | 1128| 903 |
| Dataset 2 (#Labels = 22)           | 824| 823| 816 | – | 803 | 819 | 865 |
| Dataset 3 (#Labels = 22)           | 815| 798| 808 | – | 848 | 829 | 814 |
| Dataset 4 (#Labels = 22)           | 834| 828| 809 | – | 828 | 821 | 912 |
| Dataset 5 (#Labels = 22)           | 1025| 891| 843 | – | 810 | 941 | 1008|

The distribution of the principal components and their corresponding reference points of both training and testing instances of the fused CSI-based fingerprint of real measurements collected over several days in October 2020 is depicted in Figure 8. Unlike the previous daily-based distributions, which were limited to a few specific reference points, the fused data distribution of the entire dataset considered days of month 2 would give us a complete picture of the positioning system. The fused data, on the other hand, represents the entire distribution of the CSI fingerprint measurements over the total defined grid points where positioning performance can be established given the possibility of signal fluctuations from all reference points where targets can be located. The average amplitude and signal fluctuation values are more representative than individual day measurements. Moreover, the total number of reference points collected during the survey period of October 2020 across all days of consideration was 110. Furthermore, despite a slight variation due to temporal signal fluctuations generated within the environment, the fused CSI-based measurements of both training and testing datasets received from
multiple antennas of a base station at a reference point appear to obey a specific distribution. One can also see that the first and second principal components of both instances from the training and testing datasets account for the greatest proportion of the variance explainability of the positioning model. This could explain the total variance of the system model as a linear combination of those most significant features dedicated to positioning performance. However, the two final principal components shown in Figure 8c,d, which account for about 4% of the total variations of the predictive model for both fused instances of the training and testing dataset, appear to follow different distributions. In other words, the fused testing dataset appears to fail to represent the fused training dataset distribution.

Figure 6. Distribution of principal components of datasets collected during month 1 with their corresponding RPs. (a) Training dataset Day1M1, (b) Testing dataset Day1M1, (c) Training dataset Day1M1, (d) Testing dataset Day1M1.
Tables 4 and 5 demonstrate the effect of various feature space possibilities on the amount of variance ratio explainability of the predictive models for indoor parking lots or target localization accuracy, considering the nature of distributions of fused instances of both training and testing data points separately collected over several days in September and October 2020, respectively. The model’s total variations of the fused data could be explained by 16 principal components of both the training and testing datasets. Similarly, the 90% explained variance ratio required 14 principal components to fit the system modeling or indoor parking lot scenario in both fused training and testing phases, and this difference in principal components was found to be insignificant. The first and second principal component distributions account for approximately 23% of the total model variations that could be accounted for by the fused data of both the training and testing datasets, separately. This explains why the two distributions of the base model appear the same, why the CSI fingerprint measurements received at a reference point from different base stations appear to be independent, and why the distribution of each RP seems difficult to characterize with the base model, which consists of only two main components accounting for approximately 23% of the total variance of the predictive model. Although limited and significant predictors could represent the vehicle’s indoor parking system more effectively and cheaply, a more considerable number of dimensions of feature spaces or principal components are required to fit the model. In other words, the base model, which accounts for approximately 23% of the variance ratio’s explainability, does not fully reflect the model’s
variation and thus does not fit the indoor parking system scenario. This is consistent with the finding that the less variance explained ratio by a principal component, the less likely the algorithms are to characterize the ‘CSI amplitude-location’ relationship, and thus fail to estimate the target position or indoor parking lots [92,93].

![Figure 8](image_url)

**Figure 8.** Distribution of principal components of fused data of month 2 with their corresponding RPs. (a) Training Fused data of Month 2, (b) Testing Fused data of Month 2, (c) Training Fused data of Month 2, (d) Testing Fused data of Month 2.

**Table 4.** Effect of feature spaces to the variance account of the predictive model for indoor parking lots using Fused data: M1(#RPs = 225).

| #PCA | Training Data | Testing Data | Variance Explained Ratio |
|------|---------------|--------------|--------------------------|
|      | 2             | 2            | 23.01%                   |
|      | 16            | 16           | 95%                      |
|      | 14            | 14           | 90%                      |
|      | 12            | 12           | 85%                      |
|      | 11            | 11           | 80%                      |

**Table 5.** Effect of feature spaces to the variance account of the Predictive model for indoor parking lots using Fused data M2 (#RPs = 110).

| #PCA | Training Data | Testing Data | Variance Explained Ratio |
|------|---------------|--------------|--------------------------|
|      | 2             | 2            | 19.40%                   |
|      | 16            | 16           | 95%                      |
|      | 14            | 14           | 90%                      |
|      | 12            | 12           | 85%                      |
|      | 11            | 11           | 80%                      |
4.3. Comparative Analysis of Methods

This section presents extensive real-life experiments conducted on various occasions or daily to measure the temporal signal fluctuations observed for indoor parking lots-based CSI fingerprint measurements. We compared our proposed algorithms applied to these CSI-based real measurements of different distributions to popular machine learning algorithms used in prediction tasks. Table 6 depicts the application of several machine learning algorithms to real-world scenarios of indoor parking lots based on CSI fingerprint measurements collected on different days of month 1, specifically September 2020. In this study, we are particularly interested in modeling vehicles’ indoor parking lots based on CSI-based fingerprints of amplitude values from various datasets collected over several days. Data preprocessing was also performed to understand the details of the datasets and gain insights for possible hypothesis or claim generation useful to address the indoor parking problem, and our analysis took those factors into account. Extensive experimentation was performed in this regard on real-life scenarios of the indoor parking phenomenon, and the experimental results revealed that the proposed algorithm was found to be an efficient algorithm with a consistent score of positioning accuracy across the potential dynamics of temporal variations while the average execution testing time was higher than that of the other algorithms, which is the trade-off to be penalized. Furthermore, unlike the support vector machine and neural network algorithms, which consume enormous computational complexity, the proposed algorithm not only improved indoor parking localization accuracy but was also found to be a computationally robust positioning estimator. Results also revealed that the training computational complexity was significantly higher than the testing computational time for all algorithms. However, the support vector machine and neural network algorithms demanded an unusually large amount of computational complexity time during the training and testing phases, following the proposed algorithm.

Table 6. Models vehicle’s indoor parking lots of CSI-based fingerprints of various datasets with different temporal variations.

| Classifiers                  | Day1 (#RPs = 31) | Day2 (#RPs = 6) | Day3 (#RPs = 25) | Day4 (#RPs = 37) | Day5 (#RPs = 36) | Day6 (#RPs = 27) | Day7 (#RPs = 31) | Day8 (#RPs = 32) |
|-----------------------------|------------------|-----------------|------------------|------------------|------------------|------------------|------------------|------------------|
| Decision tree               | 1.52             | 0.67            | 1.32             | 1.80             | 1.66             | 1.59             | 1.83             | 1.93             |
| K-neighbour (KNN)           | 1.64             | 0.80            | 1.35             | 1.87             | 1.87             | 1.78             | 1.99             | 2.04             |
| Support vector machine (SVC)| 1.94             | 1.03            | 2.37             | 2.09             | 1.94             | 2.66             | 2.03             | 2.29             |
| Logistic regression (LR)    | 2.21             | 1.00            | 2.04             | 2.29             | 2.21             | 2.25             | 2.25             | 2.34             |
| Random forest               | 1.43             | 0.60            | 1.18             | 1.60             | 1.51             | 1.49             | 1.75             | 1.81             |
| Neural network (MLP)        | 1.73             | 0.59            | 1.75             | 2.14             | 2.07             | 1.65             | 2.18             | 2.40             |
| The proposed algorithm      | 1.71             | 0.81            | 1.53             | 1.95             | 1.85             | 1.75             | 1.88             | 1.96             |

Dayi (#RPs) . . . > represents the list of days when measurements were being taken with the corresponding number of reference points.

In contrast, results show that both decision tree (DT) and random forest (RF) classifiers performed better in separate datasets of both months (Tables 6–9). However, their positioning accuracy was found to be lower than the proposed algorithm for the fused dataset, which represented the entire temporal signal variations of the indoor environment and was used to make our final decision (Table 10). Furthermore, both the decision tree and random forest models were found to be highly inconsistent with positioning performance across all potential temporal dynamics. This explains why both the DT and RF classifiers were highly susceptible to signal fluctuations caused by the dynamic indoor environment and multipath effects, as these models rely on the randomness of the selected feature. However, the positioning accuracy of these classifiers for all the separate datasets from
both months have improved significantly after transfer learning was applied, but not for the fused dataset. This implies that the estimator of positioning based on this DT and RF classifiers is highly inconsistent and unstable in the dynamic indoor environment. As shown in Table 9, these classifiers’ positioning accuracy was surprisingly reduced after the PCA technique was used for the fused dataset to reduce computational complexity. Thus, overall metrics such as unbiasedness, efficiency, consistency, and average execution testing time have demonstrated that the proposed algorithm is the best algorithm to model indoor parking lots based on CSI fingerprinting of the fused data.

Table 7 illustrates the use of heterogeneous knowledge transfer applied to real-life scenarios of indoor parking lots based on the CSI fingerprint measurements collected on separate days of month 1, September 2020. The experimental results show that the proposed algorithm was found to be an efficient algorithm with a consistent score of positioning accuracy across the potential dynamics of temporal variations that were being observed in the indoor parking problem domain. Additionally, the proposed algorithm has not only managed to improve indoor parking positioning performance through heterogeneous knowledge transfer but also, computationally cost-wise, was relatively efficient, unlike the support vector machine and neural network algorithms, which involve huge computational complexity. Moreover, we observed that the computational complexity was much higher during the application of heterogeneous knowledge transfer for all algorithms in general, and this is the penalty cost to obtain better positioning performance. However, we exceptionally observed that too much computational complexity was demanded by the support vector machine for both the training and testing phases, followed by the neural network algorithm.

Table 7. Models vehicle’s indoor parking lots of CSI-based fingerprints of various datasets with different temporal variations.

| Classifiers                  | Day1 (#RPs = 31) | Day2 (#RPs = 6) | Day3 (#RPs = 25) | Day4 (#RPs = 37) | Day5 (#RPs = 36) | Day6 (#RPs = 27) | Day7 (#RPs = 31) | Day8 (#RPs = 32) |
|------------------------------|------------------|-----------------|------------------|------------------|------------------|------------------|------------------|------------------|
| Decision tree                | 0.68             | 0.31            | 0.39             | 0.80             | 0.74             | 0.71             | 0.81             | 0.86             |
| K-neighbour (KNN)            | 1.50             | 0.73            | 1.22             | 1.71             | 1.68             | 1.61             | 1.82             | 1.88             |
| Support vector machine (SVC)| 1.88             | 0.99            | 2.30             | 2.02             | 1.87             | 2.58             | 1.97             | 2.23             |
| Logistic regression (LR)     | 2.21             | 0.88            | 1.94             | 2.21             | 2.13             | 2.15             | 2.17             | 2.27             |
| Random forest                | 0.67             | 0.29            | 0.55             | 0.76             | 0.73             | 0.70             | 0.82             | 0.86             |
| Neural network (MLP)         | 1.62             | 0.36            | 1.69             | 2.11             | 2.06             | 1.57             | 2.16             | 2.38             |
| The proposed algorithm       | 1.64             | 0.75            | 1.48             | 1.92             | 1.81             | 1.69             | 1.81             | 1.91             |

Dayi (#RPs) ... > represents the list of days when measurements were being taken with the corresponding number of reference points.

Table 8 below demonstrates the use of several machine learning algorithms applied to real life scenarios of indoor parking lots based on the CSI fingerprint measurements collected on separate five days of month 2, October 2020 and the fused data, respectively. Along with this, extensive experimentation applied to real life scenarios of indoor parking phenomenon were conducted, and the experimental results revealed that the proposed algorithm was found to be an efficient algorithm with a consistent score of positioning accuracy across the possible dynamics of temporal variations observed during October 2020 for five separate days. The proposed algorithm has not only managed to improve indoor parking localization accuracy but also computationally cost-wise was also efficient, unlike the support vector machine and neural network algorithms that involve huge computational complexity. Moreover, as noted above, we also observed that the training computational complexity was much higher for all algorithms than testing time in gen-
eral. We have also exceptionally observed too much computational complexity time was demanded by the support vector machine for training and testing phases, followed by the neural network algorithm.

Table 8. Models vehicle’s indoor parking lots of CSI-based fingerprints of various datasets with different temporal variations.

| Classifiers                   | Day1 (#RPs = 22) | Day2 (#RPs = 22) | Day3 (#RPs = 22) | Day4 (#RPs = 22) | Day5 (#RPs = 22) | Fuseddata (#RPs = 110) |
|------------------------------|------------------|------------------|------------------|------------------|-----------------|-----------------------|
| Decision tree                | 1.30             | 1.71             | 1.67             | 1.37             | 1.58            | 2.27                  |
| K -neighbour (KNN)           | 1.39             | 1.91             | 1.74             | 1.47             | 1.76            | 2.29                  |
| Support vector machine (SVC) | 1.65             | 2.32             | 2.20             | 1.84             | 1.95            | 2.07                  |
| Logistic regression (LR)     | 1.89             | 2.39             | 2.19             | 1.88             | 2.25            | 2.30                  |
| Random forest                | 1.08             | 1.51             | 1.42             | 1.21             | 1.42            | 2.28                  |
| Neural network (MLP)         | 1.19             | 1.76             | 1.60             | 1.46             | 1.96            | 2.07                  |
| The proposed algorithm       | 1.52             | 2.02             | 1.76             | 1.53             | 1.80            | 1.97                  |

Dayi (#RPs) ... > represents the list of days when measurements were being taken with the corresponding Number of Reference points.

Table 9 shows the application of heterogeneous knowledge transfer to real-life scenarios of indoor parking lots based on CSI-fingerprint measurements collected on separate five days of month 2, on October 2020, and the experimental results show that the proposed algorithm was found to be efficient with a consistent score of positioning accuracy across the possible dynamics of temporal variations that were observed in indoor parking probing. In contrast to the support vector machine and neural network algorithms, heterogeneous knowledge transfer not only improved indoor parking positioning performance but was also relatively efficient in terms of computational cost. Furthermore, the survey, which was conducted over five days in October 2020, revealed that the computational complexity was much higher during the application of heterogeneous knowledge transfer for all algorithms in general, and this is the trade-off we pay to achieve better positioning performance. Similarly, we observed that the support vector machine demanded an unusually large amount of computational time during the training and testing phases, followed by the neural network algorithm.

Table 9. Models vehicle’s indoor parking lots of CSI-based fingerprints of various datasets with different temporal variations.

| Classifiers                   | Day1 (#RPs = 22) | Day2 (#RPs = 22) | Day3 (#RPs = 22) | Day4 (#RPs = 22) | Day5 (#RPs = 22) |
|------------------------------|------------------|------------------|------------------|------------------|-----------------|
| Decision tree                | 0.57             | 0.75             | 0.75             | 0.61             | 0.70            |
| K -neighbour (KNN)           | 1.26             | 1.72             | 1.60             | 1.33             | 1.62            |
| Support vector machine (SVC) | 1.60             | 2.26             | 2.09             | 1.77             | 1.87            |
| Logistic regression (LR)     | 1.79             | 2.26             | 2.11             | 1.79             | 2.14            |
| Random forest                | 0.52             | 0.75             | 0.67             | 0.56             | 0.68            |
| Neural network (MLP)         | 1.08             | 1.56             | 1.49             | 1.35             | 1.91            |
| The proposed algorithm       | 1.47             | 1.92             | 1.70             | 1.47             | 1.72            |

Dayi (#RPs) ... > represents the list of days when measurements were being taken with the corresponding number of reference points.

On the other side, Table 10 presents models vehicle’s indoor parking lots of CSI-based fingerprints of the fused data-based PCA method. The indoor parking positioning performance significantly improved after transfer learning method was applied for both the
fused data and the separate datasets. The positioning performance for the separate datasets of both months has significantly reduced after principal component was used to minimize computational cost. This revealed that the use of PCA method or considering 95% of the total variations of the model could not represent the entire dynamics of the environment but knowledge transfer could leverage from the training instances to enhance target positioning. However, the positioning performance for the fused data of month 2, October 2020 has maintained equal performance as that of the performances achieved in separate dataset while significant improvement was observed in computational cost. Moreover, heterogeneous knowledge transfer of the fused data based PCA not only improved indoor parking positioning performance but was also significantly efficient in terms of computational cost as can be seen in Figures 8 and 9.

### Table 10. Models Vehicle’s indoor parking lots of CSI-based fingerprints of fused data based PCA method.

| Data Collection Period: Month 2, October 2020 | 95% Explained Variance Ratio (EVR) |
|--------------------------------------------|-------------------------------------|
| Day5 (#RPs = 22) | Fused Data (#RPs = 110) |
| Decision tree | 14.69 | 2.29 |
| K-Neighbour (KNN) | 15.09 | 2.30 |
| Support vector machine (SVC) | 12.56 | 2.09 |
| Logistic Regression (LR) | 12.51 | 2.14 |
| Random forest | 15.73 | 2.31 |
| Neural Network (MLP) | 14.30 | 2.23 |
| The proposed algorithm | 1.78 | 1.91 |

TL . . > represents transfer learning.

Moreover, the following Table 11 shows the parameters specification being used by each algorithm in the study.

### Table 11. Parameters specification used for each algorithm.

| SNo. | Parameters | Descriptions |
|------|------------|--------------|
| Decision tree | Default | DecisionTreeClassifier |
| K-Neighbour (KNN) | n_neighbors = 20 | C-regularization parameter |
| Support vector machine (SVC) | kernel = ‘poly’, C = 0.6 | |
| Logistic Regression (LR) | Default | Random Forest Tree |
| Random forest (RF) | n_estimators = 6 | Multi-layer Perceptron classifier |
| Neural Network (MLP) | MLPClassifier, random_state = 10, max_iter = 500 | |

### 4.4. Computational Complexity Analysis

This section compares the computational complexity of our proposed algorithm, which was applied to the fused CSI-based real measurements of different distributions collected in separate days of two months, to popular machine learning algorithms used in prediction tasks to represent the temporal signal variations observed for indoor parking lots. The algorithms in this study were executing on a laptop computer equipped with an AMD Ryzen 3 3200U CPU (2.60 GHz) and 16 GB of RAM. The complexity of an algorithm is primarily determined by two scenarios of complexity: time and space. We used both the functional analysis of Big O and average elapsed execution testing time to compare the time complexity of the algorithms. The Big O functional analysis of time complexity provided us with the overall worst-case computational cost analysis, and all of the algorithms we used have an average of worst-case time complexity of O(n^2). However, the proposed algorithm’s average elapsed execution testing time is much higher than that of the other algorithms that we used. Figure 9 illustrates the comparative analysis of computational testing time of the fused data before and after PCA was applied. We noted that the fused data represents the entire distribution of the CSI fingerprint measurements over the
total defined grid points where positioning performance could be established given the possibility of signal fluctuations that could come from all reference points where targets could be located. Accordingly, the computational testing time for the proposed algorithm is higher followed by the neural network (MLP) and SVC. However, after PCA was applied, the computational testing time was significantly reduced or improved while maintaining the same order.

![Elapsed Time Graph](image_url)

**Figure 9.** Comparative analysis of computational testing time of the fused data before and after PCA applied.

Figure 10 depicts the performance analysis of classifiers applied to indoor parking lots based on fused data of CSI-fingerprints. We have applied three scenarios to analyze the positioning performance as a) the fused data b) the fused data based PCA and c) the fused data in conjunction with PCA and transfer learning. Thus, the proposed algorithm (scenario c) has the best positioning performance with minimum root mean square error applied to indoor parking lots based on the fused data of the CSI-fingerprint measurements.

![RMSE Graph](image_url)

**Figure 10.** Performance analysis of classifiers applied to indoor parking lots based on fused data of CSI-fingerprints.

5. Conclusions

In this study, we considered various scenarios of temporal variations that generated the CSI-based fingerprinting measurements applied to indoor environment settings aimed at vehicles’ indoor parking lots. Along with this, extensive real-life experiments were
conducted at Huawei company in a different time with an area of 75 m² constituting 225 and 110 reference points (RPs) in total. The data were collected over separate dates on September and October 2020, respectively. Each RP was equidistant (>=0.5 m) from the next reference point. The number of measurements considered from each antenna of a base station was unequal in size. Similarly, the number of labels was not totally balanced. Thus, our analysis used the feature scaling technique to avoid possible discrepancies created due to the imbalance of data. To this end, we proposed a heterogeneous data fusion method based on positive knowledge transfer to represent the different temporal variation scenarios of CSI-based fingerprint measurements generated in a complex indoor environment targeting indoor parking lots, while reducing training calibration overheads.

Extensive experiments were carried out with real-world scenarios of the indoor parking phenomenon. Experimental results revealed that the proposed algorithm proved to be an efficient algorithm with consistent positioning accuracy across all potential variations. The proposed algorithm not only improves indoor parking location accuracy, but it is also a computationally robust and efficient estimator for a dynamic indoor environment unlike the decision tree and random forest algorithms, which are significantly affected by the temporal signal fluctuations.

The results also show that the training computational complexity for all algorithms was much higher than the overall testing time, and that the proposed algorithm required higher computational complexity for the training and testing phases, followed by support vector machine and neural network algorithms. However, an interesting finding was observed for the vehicle’s indoor parking lots of CSI-based fingerprints of the fused data-based PCA method. The indoor parking positioning performance significantly improved after the transfer learning method was applied to both the fused data and separate datasets. Nevertheless, the positioning performance for the separate datasets of both months was significantly reduced after the principal component method was used to minimize computational cost. This revealed that the use of the PCA method, or considering 95% of the total variations of the model, could not represent the entire dynamics of the environment, but knowledge transfer could be leveraged from the training instances to enhance target positioning. In contrast, the positioning performance for the fused data of month 2, October 2020 maintained equal performance as that achieved in separate dataset, while significant improvement was also observed in computational cost. Moreover, heterogeneous knowledge transfer of the fused data-based PCA not only improved indoor parking positioning performance, but was also very efficient in terms of computational cost as depicted in Figures 8 and 9. This exactly coincides with our claim that the fused data is significantly important for representing the signal fluctuations-based CSI-fingerprints in a dynamic environment, typically in an underground parking plot.

The CRLB analysis technique was also applied to estimate the lower bound of the positional error variance aimed at indoor parking lots. Similarly, different scenarios of temporal variations were also considered for CRLB analysis of CSI-based fingerprint measurements applied to indoor environment settings, such as vehicles’ indoor parking lots. Thus, the analytical derivation of the CRLB analysis revealed that the lower bound of the variance of the location estimator depends on (a) the angle of the base stations, (b) the number of base stations, (c) the distance between the target and base station, \( d_j \), (d) correlation of features, \( \rho_{ai} \), and (e) the signal propagation parameters \( \sigma_C \) and \( \gamma \). Moreover, experimental results have shown that the number of antennas of a base station could affect the lower bound of the variance of the location estimator by generating a higher dimension of features unless the most significant predictors are selected; otherwise, the accuracy of positioning performance could be degraded due to the dimensionality curse. This analytical derivation also revealed that the fused data have shown the hybrid effect of the temporal signal variations that could come through time differences when measurements were being taken.

The database consists of values of different signal features as clearly mentioned on Section 4.2. The CSI fingerprint measurements collected during the survey provide information to aid in target positioning, including real CSI measurements, imaginary CSI
measurements, latitude, longitude, coordinate systems, time of arrival, angle of arrival and other relevant data for all scenarios. Thus, the original database is a vast one consisting of various information. But, we extract data from a database that suits our research goal. Following are the recommendations we forwarded as future research directions:

1. Even though we limited our scope to CSI-amplitude information based on our objective, the phase information of CSI can also be used as fingerprints to location.
2. Fusion of various signal measurements also could result in a robust and efficient estimator for parking lots, although advanced signal processing technology is required in real life to minimize computational cost.
3. Correlation feature extraction of the various signal metrics can also be considered in addressing signal fluctuations resulted due to the dynamic environment.
4. Effective data preprocessing approaches are highly recommended in improving the positioning performance.
5. Handling the data dimensionality curse could also improve the computational complexity. Thus, various approaches of data reduction techniques are highly important and worth further investigations.

Author Contributions: Conceptualization, H.T.G., X.G., K.Z., L.L. and Y.Z.; Data curation, L.L. and Y.Z.; Formal analysis, H.T.G., X.G. and K.Z.; Funding acquisition, X.G.; Investigation, H.T.G., X.G. and K.Z.; Methodology, H.T.G., X.G. and K.Z.; Project administration, X.G. and K.Z.; Resources, L.L. and Y.Z.; Supervision, X.G.; Visualization, K.Z. and Y.Z.; Writing—original draft, H.T.G.; Writing—review & editing, X.G., K.Z., L.L. and Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant (No. 62171086) and in part by the Municipal Government of Quzhou under Grant (No. 2021D002).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset used for this study are available upon request to the corresponding author.

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

References
1. Elhousni, M.; Huang, X. A survey on 3d lidar localization for autonomous vehicles. In Proceedings of the 2020 IEEE Intelligent Vehicles Symposium (IV), Las Vegas, NV, USA, 19 October–13 November 2020; pp. 1879–1884.
2. Pecoraro, G.; Di Domenico, S.; Cianca, E.; De Sanctis, M. CSI-based fingerprinting for indoor localization using LTE signals. EURASIP J. Adv. Signal Process. 2018, 2018, 49. [CrossRef]
3. Geiger, A.; Lenz, P.; Urtasun, R. Are we ready for autonomous driving? the kitti vision benchmark suite. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 3354–3361.
4. Spiekermann, S. General Aspects of Location-Based Services. In Location-Based Services; Elsevier: San Francisco, CA, USA, 2004; pp. 9–26.
5. Mirama, V.E.; Diez, L.E.; Bahillo, A.; Quintero, V. A Survey of Machine Learning in Pedestrian Localization Systems: Applications, Open Issues and Challenges. IEEE Access 2021, 9, 120138–120157. [CrossRef]
6. Sithole, G.; Zlatanov, S. Position, location, place and area: An indoor perspective. ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. 2016, 3, 89–96. [CrossRef]
7. Buczkowski, A. Location-Based Services—Applications. 2011. Available online: https://geoawesomeness.com/knowledge-base/ location-based-services/location-based-services-applications/ (accessed on 28 December 2021).
8. Guo, X.; Ansari, N.; Li, L.; Duan, L. A hybrid positioning system for location-based services: Design and implementation. IEEE Commun. Mag. 2020, 58, 90–96. [CrossRef]
9. Enge, P.; Misra, P. Special issue on GPS: The global positioning system. Proc. IEEE 1999, 3, 172.
10. Bulusu, N.; Heidemann, J.; Estrin, D. GPS-less low-cost outdoor localization for very small devices. IEEE Pers. Commun. 2000, 7, 28–34. [CrossRef]
11. Barter, P. Cars Are Parked 95% of the Time. Let’s Check. 2013. Available online: http://www.reinventingparking.org/2013/02/ carsare-parked-95-of-time-lets-check.html (accessed on 24 November 2016).
12. 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.

13. Wu, K.; Xiao, J.; Yi, Y.; Chen, D.; Luo, X.; Ni, L.M. CSI-based indoor localization. IEEE Trans. Parallel Distrib. Syst. 2012, 24, 1300–1309. [CrossRef]

14. Reddy, H.; Chandra, M.G.; Balamuralidhar, P.; Harihara, S.G.; Bhattacharya, K.; Joseph, E. An improved time-of-arrival estimation for WLAN-based local positioning. In Proceedings of the 2007 2nd International Conference on Communication Systems Software and Middleware, Bangalore, India, 7–12 January 2007; pp. 1–5.

15. Gentner, C.; Jost, T. Indoor positioning using time difference of arrival between multipath components. In Proceedings of the 2007 7th IEEE International Conference on Indoor Positioning and Indoor Navigation, Montbeliard, France, 28–31 October 2013; pp. 1–10.

16. Tay, B.; Liu, W.; Zhang, D.H. Indoor angle of arrival positioning using biased estimation. In Proceedings of the 2009 7th IEEE International Conference on Industrial Informatics, Cardiff, UK, 23–26 June 2009; pp. 458–463.

17. Ahriz, I.; Oussar, Y.; Denby, D.; Dreyfus, G. Carrier relevance study for indoor localization using GSM. In Proceedings of the 2010 7th Workshop on Positioning, Navigation and Communication, Dresden, Germany, 11–12 March 2010; pp. 168–173.

18. Roos, T.; Myllymäki, P.; Tirri, H.; Misikangas, P.; Sievänen, J. A probabilistic approach to WLAN user location estimation. Int. J. Wirel. Inf. Netw. 2002, 9, 155–164. [CrossRef]

19. Chang, K.H. Bluetooth: A viable solution for IoT? [Industry Perspectives]. IEEE Wirel. Commun. 2014, 21, 6–7. [CrossRef]

20. Wang, C.S.; Huang, C.H.; Chen, Y.S.; Zheng, L.J. An implementation of positioning system in indoor environment based on active RFID. In Proceedings of the 2009 Joint Conferences on Pervasive Computing (IPCC), Tamsui, Taiwan, 3–5 December 2009; pp. 71–76.

21. Alarifi, A.; Al-Salman, A.; Alsaleh, M.; Alnafessah, A.; Al-Hadhrami, S.; Al-Ammar, M.A.; Al-Khalifa, H.S. Ultra wideband indoor positioning technologies: Analysis and recent advances. Sensors 2016, 16, 707. [CrossRef]

22. Hu, X.; Cheng, L.; Zhang, G. A Zigbee-based localization algorithm for indoor environments. In Proceedings of the 2011 International Conference on Computer Science and Network Technology, Harbin, China, 24–26 December 2011; Volume 3, pp. 1776–1781.

23. Yasir, M.; Ho, S.W.; Vellambi, B.N. Indoor positioning system using visible light and accelerometer. J. Lightwave Technol. 2014, 32, 3306–3316. [CrossRef]

24. Shu, Y.; Bo, C.; Shen, G.; Zhao, C.; Li, L.; Zhao, F. Magicol: Indoor localization using pervasive magnetic field and opportunistic WiFi sensing. IEEE J. Sel. Areas Commun. 2015, 33, 1443–1457. [CrossRef]

25. Karunatilaka, D.; Zafar, F.; Kalavally, V.; Parthiban, R. LED based indoor visible light communications: State of the art. IEEE Commun. Surv. Tutor. 2015, 17, 1649–1678. [CrossRef]

26. Mainetti, L.; Palano, L.; Patrono, L.; Stefanizzi, M.L.; Vergallo, R. Integration of RFID and WSN technologies in a Smart Parking System. In Proceedings of the 2014 22nd International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, Croatia, 17–19 September 2014; pp. 104–110.

27. Tsiroupolou, E.E.; Baras, J.S.; Papavassiliou, S.; Sinha, S. RFID-based smart parking management system. Cyber-Phys. Syst. 2017, 3, 22–41. [CrossRef]

28. Gao, Y.; Liu, S.; Atia, M.M.; Noureldin, A. INS/GPS/LiDAR integrated navigation system for urban and indoor environments using hybrid scan matching algorithm. Sensors 2015, 15, 23286–23302. [CrossRef]

29. Kim, S.T.; Fan, M.; Jung, S.W.; Ko, S.J. External Vehicle Positioning System Using Multiple Fish-Eye Surveillance Cameras for Indoor Parking Lots. IEEE Syst. J. 2020, 15, 5107–5118. [CrossRef]

30. Wolcott, R.W.; Eustice, R.M. Visual localization within lidar maps for automated urban driving. In Proceedings of the 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, IL, USA, 14–18 September 2014; pp. 176–183.

31. Ichihashi, H.; Notsu, A.; Honda, K.; Katada, T.; Fujiyoshi, M. Vacant parking space detector for outdoor parking lot by using surveillance camera and FCM classifier. In Proceedings of the 2009 IEEE International Conference on Fuzzy Systems, Jeju, Korea, 20–24 August 2009; pp. 127–134.

32. Mackey, A.; Spachos, P.; Plataniotis, K.N. Smart parking system based on bluetooth low energy beacons with particle filtering. IEEE Syst. J. 2020, 14, 3371–3382. [CrossRef]

33. Lin, C.H.; Chen, L.H.; Wu, H.K.; Jin, M.H.; Chen, G.H.; Gomez, J.L.G.; Chou, C.F. An indoor positioning algorithm based on fingerprint and mobility prediction in RSS fluctuation-prone WLANs. IEEE Trans. Syst. Man Cybern. Syst. 2019, 51, 2926–2936. [CrossRef]

34. Costilla-Reyes, O.; Namuduri, K. Dynamic Wi-Fi fingerprinting indoor positioning system. In Proceedings of the 2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Busan, Korea, 27–30 October 2014; pp. 271–280.

35. Qin, S.; Guo, X. Robust Source Positioning Method With Accurate and Simplified Worst-Case Approximation. IEEE Trans. Veh. Technol. 2021, 71, 1891–1900. [CrossRef]

36. Dinh-Van, N.; Nashashibi, F.; Thanh-Huon, N.; Castelli, E. Indoor Intelligent Vehicle localization using WiFi received signal strength indicator. In Proceedings of the 2017 IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM), Nagoya, Japan, 19–21 March 2017; pp. 33–36.

37. Gidey, H.T.; Guo, X.; Zhong, K.; Li, L.; Zhang, Y. An Online Heterogeneous Transfer Learning Methods for Fingerprint-Based Indoor Positioning. Sensors. submitted.
38. Gidey, H.T.; Guo, X.; Li, L.; Zhang, Y. Heterogeneous Transfer Learning for Wi-Fi Indoor Positioning Based Hybrid Feature Selection. Sensors 2022, 22, 5840. [CrossRef] [PubMed]
39. Wang, X.; Gao, L.; Mao, S.; Pandey, S. CSI-based fingerprinting for indoor localization: A deep learning approach. IEEE Trans. Veh. Technol. 2016, 66, 763–776. [CrossRef]
40. Shi, Z.; Wei, L.; Xu, Y. CSI-based Fingerprinting for Indoor Localization with Multi-scale Convolutional Neural Network. In Proceedings of the 2021 IEEE 3rd Eurasia Conference on IOT, Communication and Engineering (ECICE), Yunlin, Taiwan, 29–31 October 2021; pp. 233–237.
41. Li, L.; Guo, X.; Zhang, Y.; Ansari, N.; Li, H. Long Short-Term Indoor Positioning System via Evolving Knowledge Transfer. IEEE Trans. Wirel. Commun. 2022, 7, 5556–5572. [CrossRef]
42. Yang, Z.; Zhou, Z.; Liu, Y. From RSSI to CSI: Indoor localization via channel response. ACM Comput. Surv. 2013, 46, 1–32. [CrossRef]
43. Chapre, Y.; Ignjatovic, A.; Seneviratne, A.; Jha, S. Csi-mimo: Indoor wi-fi fingerprinting system. In Proceedings of the 39th Annual IEEE Conference on Local Computer Networks, Edmonton, AB, Canada, 8–11 September 2014; pp. 202–209.
44. Ferrand, P.; Decurninge, A.; Guillaud, M. DNN-based localization from channel estimates: Feature design and experimental results. In Proceedings of the GLOBECOM 2020—2020 IEEE Global Communications Conference, Taipei, Taiwan, 7–11 December 2020; pp. 1–6.
45. Ness, A.; Adhikari, B.; Hussain, F.; Fernando, X.N. A survey of machine learning for indoor positioning. IEEE Access 2020, 8, 214945–214965. [CrossRef]
46. Regani, S.D.; Xu, Q.; Wang, B.; Wu, M.; Liu, K.R. Driver authentication for smart car using wireless sensing. IEEE Internet Things J. 2019, 7, 2235–2246. [CrossRef]
47. Xiao, J.; Wu, K.; Yi, Y.; Ni, L.M. FIFS: Fine-grained indoor fingerprinting system. In Proceedings of the 2012 21st International Conference on Computer Communications and Networks (ICCCN), Munich, Germany, 30 July–2 August 2012; pp. 1–7.
48. Wu, Z.L.; Li, C.H.; Ng, J.K.Y.; Leung, K.R. Location estimation via support vector regression. IEEE Trans. Mob. Comput. 2007, 6, 311–321. [CrossRef]
49. Wang, Y.; Xi, C.; Zhang, X.; Yang, D. WiFi indoor localization with CSI fingerprinting-based random forest. Sensors 2018, 18, 2869. [CrossRef]
50. Tran, Q.; Tantra, J.W.; Foh, C.H.; Tan, A.H.; Yow, K.C.; Qiu, D. Wireless indoor positioning system with enhanced nearest neighbors in signal space algorithm. In Proceedings of the IEEE Vehicular Technology Conference, Montreal, QC, Canada, 25–28 September 2006; pp. 1–5.
51. Sobehy, A.; Renault, E.; Mühlethaler, P. CSI-MIMO: K-nearest neighbor applied to indoor localization. In Proceedings of the ICC 2020—2020 IEEE International Conference on Communications (ICC), Dublin, Ireland, 7–11 June 2020; pp. 1–6.
52. 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]
53. Yong, Z.; Chengbin, W. An indoor positioning system using Channel State Information based on TrAdaBoost Transfer Learning. In Proceedings of the 2021 4th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE), Changsha, China, 26–28 March 2021; pp. 1286–1293.
54. Han, D.; Jung, S.; Lee, M.; Yoon, G. Building a practical Wi-Fi-based indoor navigation system. IEEE Pervasive Comput. 2014, 13, 72–79.
55. Qin, S.; Guo, X. IoT edge computing-enabled efficient localization via robust optimal estimation. IEEE Internet Things J. 2022, 1. [CrossRef]
56. Dardari, D.; Chong, C.C.; Win, M. Threshold-based time-of-arrival estimators in UWB dense multipath channels. IEEE Trans. Commun. 2008, 56, 1366–1378. [CrossRef]
57. He, S.; Chan, S.H.G. Wi-Fi fingerprint-based indoor positioning: Recent advances and comparisons. IEEE Commun. Surv. Tutor. 2015, 18, 466–490. [CrossRef]
58. Chang, D.C.; Fan, M.W. AoA target tracking with new imm pf algorithm. In Proceedings of the 2014 IEEE 57th International Midwest Symposium on Circuits and Systems (MWSCAS), College Station, TX, USA, 3–6 August 2014; pp. 729–732.
59. Wang, B.; Zhou, S.; Liu, W.; Mo, Y. Indoor localization based on curve fitting and location search using received signal strength. IEEE Trans. Ind. Electron. 2014, 62, 572–582. [CrossRef]
60. Xiao, J.; Zhou, Z.; Yi, Y.; Ni, L.M. A survey on wireless indoor localization from the device perspective. ACM Comput. Surv. 2016, 49, 1–31. [CrossRef]
61. Shahmansoori, A.; Garcia, G.E.; Destino, G.; Seco-Granados, G.; Wymeersch, H. Position and orientation estimation through millimeter-wave MIMO in 5G systems. IEEE Trans. Wirel. Commun. 2017, 17, 1822–1835. [CrossRef]
62. Yin, J.; Wan, Q.; Yang, S.; Ho, K.C. A simple and accurate TDOA-AOA localization method using two stations. IEEE Signal Process. Lett. 2015, 23, 144–148. [CrossRef]
63. Sharp, I.; Yu, K.G. Indoor TOA error measurement, modeling, and analysis. IEEE Trans. Instrum. Meas. 2014, 63, 2129–2144. [CrossRef]
64. Alavi, B.; Pahlavan, K. Modeling of the TOA-based distance measurement error using UWB indoor radio measurements. IEEE Commun. Lett. 2006, 10, 275–277. [CrossRef]
65. Hernandez, A.; Badorrey, R.; Choliz, J.; Alastruey, I.; Valdovinos, A. Accurate indoor wireless location with IR UWB systems: a performance evaluation of joint receiver structures and TOA based mechanism. *IEEE Trans. Consum. Electron.* 2008, 54, 381–389. [CrossRef]
66. Wen, F.; Liang, C. Fine-grained indoor localization using single access point with multiple antennas. *IEEE Sens. J.* 2014, 15, 1538–1544.
67. Taponecco, L.; D’Amico, A.A.; Mengali, U. Joint TOA and AOA estimation for UWB localization applications. *IEEE Trans. Wirel. Commun.* 2011, 10, 2207–2217. [CrossRef]
68. Liu, H.; Darabi, H.; Banerjee, P.; Liu, J. Survey of wireless indoor positioning techniques and systems. *IEEE Trans. Syst. Man Cybern. Syst. Part C—Appl. Rev.* 2007, 37, 1067–1080. [CrossRef]
69. Basri, C.; El Khadimi, A. Survey on indoor localization system and recent advances of WIFI fingerprinting technique. In *Proceedings of the 2015 6th International conference on multimedia computing and systems (ICMCS)*, Marrakech, Morocco, 29 September–1 October 2016; pp. 253–259.
70. Liu, W.; Cheng, Q.; Deng, Z.; Chen, H.; Fu, X.; Zheng, X.; Wang, S. Survey on CSI-based indoor positioning systems and recent advances. In *Proceedings of the 2019 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Pisa, Italy, 30 September–3 October 2019; pp. 1–8.
71. De Bast, S.; Guevara, A.P.; Pollin, S. CSI-based positioning in massive MIMO systems using convolutional neural networks. In *Proceedings of the 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, Antwerp, Belgium, 25–28 May 2020; pp. 1–5.
72. Feng, C.; Arshad, S.; Yu, R.; Liu, Y. Evaluation and improvement of activity detection systems with recurrent neural network. In *Proceedings of the 2018 IEEE International Conference on Communications (ICC)*, Kansas City, MO, USA, 20–24 May 2018; pp. 1–6.
73. Hsieh, C.H.; Chen, J.Y.; Nien, B.H. Deep learning-based indoor localization using received signal strength and channel state information. *IEEE Access* 2019, 7, 33256–33267. [CrossRef]
74. Xia, P.; Zhou, S.; Giannakis, G.B. Adaptive MIMO-OFDM based on partial channel state information. *IEEE Trans. Signal Process.* 2004, 52, 202–213. [CrossRef]
75. Molisch, A.F. *Orthogonal Frequency-Division Multiplexing (OFDM)*; Wiley: New York, NY, USA, 2011.
76. Juntti, M.; Vehkapera, M.; Leinonen, J.; Zexian, V.; Tujkovic, D.; Tsumura, S.; Haru, S. MIMO MC-CDMA communications for future cellular systems. *IEEE Commun. Mag.* 2005, 43, 118–124.
77. Biglieri, E.; Calderbank, R.; Constantinides, A.; Goldsmith, A.; Paulraj, A.; Poor, H.V. *MIMO Wireless Communications*; Cambridge University Press: Cambridge, UK, 2007.
78. Wu, K.; Xiao, J.; Yi, Y.; Gao, M.; Ni, L.M. FILA: Fine-grained indoor localization. In *Proceedings of the 2012 IEEE INFOCOM*, Orlando, FL, USA, 25–30 March 2012; pp. 2210–2218.
79. Wu, Z.; Xu, Q.; Li, J.; Fu, C.; Xuan, Q.; Xiang, Y. Passive indoor localization based on CSI and naive bayes classification. *IEEE Trans. Syst. Man Cybern. Syst.* 2017, 48, 1566–1577. [CrossRef]
80. Molisch, A.F. *Wireless Communications*; Wiley-IEEE Series; John Wiley & Sons: Hoboken, NJ, USA, 2011.
81. Guo, X.; Ansari, N.; Hu, F.; Shao, Y.; Elikplim, N.R.; Li, L. A survey on fusion-based indoor positioning. *IEEE Commun. Surv. Tutor.* 2019, 22, 566–594. [CrossRef]
82. Wang, X.; Gao, L.; Mao, S.; Pandey, S. DeepFi: Deep learning for indoor fingerprinting using channel state information. In *Proceedings of the 2015 IEEE Wireless Communications and Networking Conference (WCNC)*, New Orleans, LA, USA, 9–12 March 2015; pp. 1666–1671.
83. Kanaris, L.; Kokkinis, A.; Liotta, A.; Stavrou, S. Combining smart lighting and radio fingerprinting for improved indoor localization. In *Proceedings of the 2017 IEEE 14th International Conference on Networking, Sensing and Control (ICNSC)*, Calabria, Italy, 16–18 May 2017; pp. 447–452.
84. Antevski, K.; Redondi, A.E.; Pitic, R. A hybrid BLE and Wi-Fi localization system for the creation of study groups in smart libraries. In *Proceedings of the 2016 9th IFIP wireless and mobile networking conference (WMNC)*, Colmar, France, 11–13 July 2016; pp. 41–48.
85. Wang, S.; Hu, D.; Sun, X.; Yan, S.; Huang, J.; Zhen, W.; Li, Y. A data fusion method of indoor location based on adaptive UKF. In *Proceedings of the 2018 7th International Conference on Digital Home (ICDH)*, Guilin, China, 30 November–1 December 2018; pp. 257–263.
86. Zhang, Z.; Xie, L.; Zhou, M.; Wang, Y. CSI-based Indoor Localization Error Bound Considering Pedestrian Motion. In *Proceedings of the 2020 IEEE/CIC International Conference on Communications in China (ICCC)*, Chongqing, China, 9–11 August 2020; pp. 811–816.
87. Zhao, Y.; Li, X.; Xu, C.Z. ER-CRLB Analysis for Indoor Multi-Antenna Localization System. In *Proceedings of the 2018 IEEE/CIC International Conference on Communications in China (ICCC)*, Beijing, China, 16–18 August 2018; pp. 379–383.
88. Hossain, A.M.; Soh, W.S. Cramer-Rao bound analysis of localization using signal strength difference as location fingerprint. In *Proceedings of the 2010 IEEE INFOCOM*, San Diego, CA, USA, 14–19 March 2010; pp. 1–9.
89. Jiang, Q.; Qiu, F.; Zhou, M.; Tian, Z. Benefits and impact of joint metric of AOA/RSS/TOF on indoor localization error. *Appl. Sci.* 2016, 6, 296. [CrossRef]
90. Catovic, A.; Sahinoglu, Z. The Cramer-Rao bounds of hybrid TOA/RSS and TDOA/RSS location estimation schemes. *IEEE Commun. Lett.* **2004**, *8*, 626–628. [CrossRef]

91. Gui, L.; Yang, M.; Yu, H.; Li, J.; Shu, F.; Xiao, F. A Cramer–Rao lower bound of CSI-based indoor localization. *IEEE Trans. Veh. Technol.* **2017**, *67*, 2814–2818. [CrossRef]

92. Wang, J.; Wang, X.; Peng, J.; Hwang, J.G.; Park, J.G. Indoor Fingerprinting Localization Based on Fine-grained CSI using Principal Component Analysis. In Proceedings of the 2021 Twelfth International Conference on Ubiquitous and Future Networks (ICUFN), Jeju, Korea, 17–20 August 2021; pp. 322–327.

93. Kukolj, D.; Vuckovic, M.; Pletl, S. Indoor location fingerprinting based on data reduction. In Proceedings of the 2011 International Conference on Broadband and Wireless Computing, Communication and Applications, Barcelona, Spain, 26–28 October 2011; pp. 327–332.