Crowdsensing Influences and Error Sources in Urban Outdoor Wi-Fi Fingerprinting Positioning

Cristian-Liviu Leca *, Ioan Nicolaescu and Petrica Ciotirnae

Military Technical Academy ‘Ferdinand I’, 050141 Bucharest, Romania; ioannic@mta.ro (I.N.);
petrica.ciotirnae@mta.ro (P.C.)
* Correspondence: cristian.leca@mta.ro

Received: 14 November 2019; Accepted: 08 January 2020; Published: 12 January 2020

Abstract: Wi-Fi fingerprinting positioning systems have been deployed for a long time in location-based services for indoor environments. Combining mobile crowdsensing and Wi-Fi fingerprinting systems could reduce the high cost of collecting the necessary data, enabling the deployment of the resulting system for outdoor positioning in areas with dense Wi-Fi coverage. In this paper, we present the results attained in the design and evaluation of an urban fingerprinting positioning system based on crowdsensed Wi-Fi measurements. We first assess the quality of the collected measurements, highlighting the influence of received signal strength on data collection. We then evaluate the proposed system by comparing the influence of the crowdsensed fingerprints on the overall positioning accuracy for different scenarios. This evaluation helps gain valuable insight into the design and deployment of urban Wi-Fi positioning systems while also allowing the proposed system to match GPS-like accuracy in similar conditions.

Keywords: crowdsensing; databases; smartphones; urban positioning; Wi-Fi fingerprinting

1. Introduction

Mobile phone sensing has recently attracted a great deal of attention due to factors such as their ubiquitous presence in everyday life, the addition of sensors not specific to their original design (barometers, accelerometers, light sensors, etc.), allowing them to be used intuitively for mobile sensing. Additionally, their permanent connection to the Internet using mobile data or a Wi-Fi connection has allowed the use of mobile phone sensors in areas such as transportation, tourism, logistics, air quality monitoring, and social networking [1,2]. Mobile phone sensing can have its powers augmented by engaging a large number of participants (a crowd) in contributing for a small cost or toward a common cause (e.g., to monitor road traffic and obtain real-time navigation) by the process of mobile crowdsensing.

We consider that the development of the mobile crowdsensing paradigm is essential to the development of functional Wi-Fi fingerprinting positioning systems for urban and outdoor areas. Previous research on fingerprinting positioning is mainly aimed at designing techniques and systems for indoor areas [3,4].

While the development of indoor location-based services and positioning systems have been motivated by complex, non-line-of-sight (NLOS) environments and the general unavailability of Global Positioning System (GPS) localization, urban crowdsensed fingerprinting positioning will compete in offering accurate positioning with the Global Navigation Satellite System (GNSS) for mobile devices equipped with both types of sensors and will prove to be a problem-solver for energy- and size-constrained IoT devices requiring GPS-like accurate positioning.

Reference [5], cited on the official gps.gov website, states that GPS has an average positioning accuracy of 5 m in rural areas and approximately 16.8 m in urban environments. This sets a clear
accuracy goal for Wi-Fi positioning systems designed for urban environments, consisting of attaining similar or better accuracy with lower costs.

Early research on urban wireless positioning aimed at improving location-sensing accuracy by eliminating typical non-line-of-sight errors while not acknowledging the benefits of applying crowdsensing in enabling a Wi-Fi urban positioning system [6–8].

Other technologies that are readily available in smartphone devices used for crowdsensing, such as Bluetooth [9], cellular [10], accelerometers [11], magnetic sensors [12], or pedometers [13], can also be integrated in fingerprinting positioning systems as has been previously shown.

In this paper, we apply mobile crowdsensing to prove the feasibility of designing a Wi-Fi fingerprinting positioning system for urban areas. The envisioned system benefits from the ubiquitous presence of Wi-Fi access points (APs) in urban areas, as proven in our previous research [14–16]. The Wi-Fi AP received signal strength (RSS) measurements are made with the help of the 802.11n network interface of the mobile phones used for crowdsensing. The measurements used for positioning are coupled with the geographical position where they were captured, thus constructing the fingerprint database. The position of the measurements is known using the GNSS receivers of the smartphones used for crowdsensing.

We propose the usage of the crowdsensing technique as opposed to previous research using automated or grid-based fingerprint collection for indoor positioning systems. Finally, the envisioned positioning system will require continuous feedback from users that choose to contribute to the crowdsensing effort. This will enable:

- Continuous growth and resolution of the coverage area for the positioning systems;
- Coverage of areas restricted from public access (e.g., private homes);
- Constant feedback on the APs situation, allowing for the removal of inactive, mobile, or moved APs;
- A high density of fingerprints in areas most frequented by users, resulting in a higher accuracy of the positioning in areas most likely to be required.

Regarding the aspect of using different smartphones, our research was not aimed at mitigating the effects of distinct hardware and software, as this would prove to be an impossible task for a crowdsensing system. This is caused by the numerous smartphone devices available on the market that would ultimately be able to contribute to the proposed system.

The main challenge specific to the proposed solution lies in proving that fingerprinting positioning can achieve GPS-like accuracy in similar conditions specific to the urban environment. To address this we thoroughly evaluated the accuracy of the fingerprinting positioning system using different sub-datasets extracted from our collected data. For the position estimation, we apply the methods proposed in the papers published by Lohan et al. and Mendoza et al. [17,18].

The remainder of the article is organized as follows: Section 2 covers related work on similar positioning systems and highlights our previous work on the subject. Section 3 introduces the data collection method and the motivation behind choosing a specific test area for our research and discusses the implications of the crowdsensing technique on fingerprint database generation. In Section 4 we present the results of our analysis, highlighting the characteristics of the positioning algorithms used and the influence of the crowdsensed measurements. Section 5 concludes our work and highlights the main contributions of the paper.

2. Related Work

Outdoor positioning is usually based on multilateration or triangulation techniques [19] which work better in line-of-sight propagation conditions. These types of methods are not suited for urban environments affected by fading and NLOS propagation [15]. The approach presented in [20] aims at estimating the AP position with the help of information contained in the AP’s name. Indoor positioning systems [21] have also tackled the NLOS problem, one particular promising solution
being the process of signal association with particular locations [22–25]. In this class of fingerprinting positioning methods, a position is solely characterized by its measured signal pattern, consisting of the consecutive collected Wi-Fi measurements. Thus, our proposed crowdsensed fingerprinting scheme is applied similarly without the need of knowing exact AP positions. Our method does not require distance or angle measurements, leading to its suitability for use in urban areas.

Wi-Fi fingerprinting has traditionally exploited Wi-Fi interfaces equipped on mobile devices, not limited to smartphones [26], as well as the ubiquitous presence of Wi-Fi APs [27]. Many of the early proposed systems [28,29] rely on an initial training phase required for constructing the fingerprint database to be used for the positioning phase. Outdoor positioning systems need significant time and effort for the training phase, which crowdsensing systems aim to limit. Crowdsensed systems are suited for outdoor environments where smartphones automatically collect RSS measurements coupled with GPS obtained locations for the training phase. As an alternative, crowdsourced systems have been proposed for indoor systems which require an active contribution from the user. In crowdsourced systems, the users are required to input their exact position in the positioning area using a dedicated app [30–34].

Designing a fingerprinting positioning system for urban environments requires knowledge of the presence and density of Wi-Fi APs in the system’s coverage area. Our previous work on fingerprinting positioning [15] proved the ubiquity of Wi-Fi APs in urban environments and offers insights regarding the data collection methodology, fingerprint database maintenance, and sources of error specific to the environment (e.g., mobile APs).

Achieving seamless urban positioning with no constraints on the indoor/outdoor environment conditions is a significant step towards implementing a global indoor positioning system, as proposed in [35].

As highlighted in previous research [17,18], Wi-Fi fingerprinting positioning systems encounter one major problem, consisting of challenges in implementing and reproducing results published in previous research papers on the subject. To address this issue, we formatted our data to match the proposed format for standard datasets in [17] and we used two of the positioning algorithms proposed in [18]—the weighted-centroid method and the log-Gaussian distance method.

3. Fingerprint Analysis and Test Data Description

The data collected for our research consisted of more than 1 million Wi-Fi measurements collected in Bucharest (Romania) and areas nearby the city by the help of eight different smartphones carried by different users for 6 months. The data collection process was a part of the everyday life of users and they were not required to scan in a specific way, space, or time. This means that the collected measurements were collected while stationary or when the users were moving from one place to another by any means available in the city.

Due to the crowdsensing method, the collected measurements are denser near the city center where the scanning occurred more often. This prompted us to select one of the most frequented areas, shown in Figure 1a during the crowdsensing step for the evaluation of the collected data on the accuracy of Wi-Fi fingerprinting positioning.

The chosen test area has a high density of Wi-Fi measurements and allows for long-term variations in Wi-Fi signal propagation. The area is characteristic to the urban environment of Romania, being placed between two high-traffic boulevards while the buildings in the area are tower apartment blocks. Locations of measurements in the test area are shown in Figure 1b.

The test area, with a surface of 1 km$^2$, scanning resulted in 75,188 fingerprints of 10,072 distinct APs. The average signal level for all fingerprints was $-83.85$ dBm with a standard deviation of 5.46 dB.

The fingerprint coordinates were transformed from the decimal degree format to local coordinates. The decimal degree format is used to express latitude and longitude geographic coordinates as decimal fractions, while the local coordinates are referenced to the upper-left corner of the test area and are expressed in meters on the x and y axes.
The test area dataset and all subsequent sub-datasets were randomly split between training data (85% of the total) and test data (15% of the remaining fingerprints). The training fingerprints were used for estimating the position of the test fingerprints. The resulting estimated position was compared to the test fingerprint position resulting in a positioning error. Figure 2 is an example of training and test data positions with locally referenced coordinates expressed in meters.

Fingerprint collection by the crowdsensing technique has implications for the density and the received signal strength characteristics.

The density of fingerprints influences the accuracy of the crowdsensed positioning system and is closely linked to the number of people that have visited a certain area. Figure 3a displays the density of the gathered fingerprints using an interpolated graph overlaid on the test area map. Each colored dot signifies the presence of a number of APs in the area around it, the color of the dot ranging between blue (1 AP) and red (70 APs).

The average signal level (dBm) of the fingerprints in the test area is shown in Figure 3b. The signal map was generated using the interpolation technique with the average being calculated for a 10-m radius.

The placement of the AP, for example, the top of a building, and the crowdsensing technique influence the signal strength at ground level where the measurements were collected. APs placed on lower floors or at ground level have the chance to be scanned more often and be seen with a higher signal level, as shown in Table 1.
Figure 3. Test area analysis.

Table 1. Frequency of measurements for the APs in the test area.

| Frequency | APs Measured |
|-----------|--------------|
| 1         | 3507         |
| 2         | 1446         |
| 3         | 791          |
| 4         | 575          |
| 5         | 479          |
| 6         | 384          |
| 7         | 297          |
| 8         | 229          |
| 9         | 202          |
| 10–19     | 1137         |
| 20–29     | 492          |
| 30–39     | 232          |
| 40–49     | 108          |
| 50–59     | 60           |
| 60–69     | 50           |
| 70–79     | 34           |
| 80–89     | 21           |
| 90–99     | 10           |
| 100–109   | 11           |
| >110      | 7            |

Figure 3c was obtained by plotting the density of APs scanned less than three times. By comparing Figure 3a,c we observe that APs scanned less than three times were observed at the margins of the test area where the overall density of measurements was lower. This leads to the following observations:
• APs in densely scanned areas are themselves more likely to be scanned more often;
• The crowdsensing method offers consistent results for successive scans of the same area;
• APs’ radio visibility displays a character of stability for repeated scans.

Access points scanned less than three times can increase the positioning error significantly, as they are most likely placed at a great distance from the measurement points or are outside the test area.

4. Positioning Accuracy and Significant Error Sources Generated by Crowdsensing

The metric that used for the evaluation of the positioning accuracy of the crowdsensed data was the mean 2D error. The error was calculated as the Euclidean distance between the estimated position and the GPS-determined positions of the test measurements. The distances are expressed in meters using the local coordinates described previously.

As mentioned, the evaluation of the positioning accuracy in an urban test environment for the crowdsensed data was done using the weighted-centroid method [36,37] and the log-Gaussian distance method [18].

4.1. Weighted-Centroid Method

The weighted-centroid method was proposed for fingerprinting positioning in reference [37] as a simple and low-complexity approach. The estimated position is calculated as the weighted average of the AP positions heard in the unknown position.

Using the set of visible APs as \( \mathcal{H} \) and the known AP positions as \( \mathbf{c}_{ap} = (x_{ap}, y_{ap}) \), \( ap = 1, ..., \mathcal{H} \), the weighted-centroid position estimation is calculated as:

\[
\hat{c}_{MS,wc} = \frac{\sum_{ap \in \mathcal{H}} w_{ap} \mathbf{c}_{ap}}{\sum_{ap \in \mathcal{H}} w_{ap}}, \tag{1}
\]

where \( w_{ap} \) are weight functions.

Weighting is done according to the distance to the APs’ positions. Since the RSS can be used as a distance function to the APs, the weights can be replaced by the RSS, using the measurement vector \( \mathbf{m}_n = [m_{n,1}, m_{n,2}, ..., m_{n,N_{ap}}]^T \), \( n = 1, ..., N_{fp} \), where \( N_{fp} \) is the number of measurements collected and \( m_{(x,y),ap} \) is the RSS received from the AP \( ap \) at coordinates \( (x, y) \), resulting in the following RSS-based equation for the weighted-centroid method:

\[
\hat{c}_{MS,wc} = \frac{\sum_{ap \in \mathcal{H}} m_{MS,ap} \mathbf{c}_{ap}}{\sum_{ap \in \mathcal{H}} m_{MS,ap}}. \tag{2}
\]

Since our proposed technique is aimed at using crowdsensed measurements, the APs’ positions will not be known. Thus, the position of all measured APs was estimated as a first step using the weighted-centroid approach, similarly to the above steps.

4.2. Log-Gaussian Distance Method

The log-Gaussian distance method used in our paper was applied similarly as proposed in [38]. The position estimation was attained by the help of a weighted K nearest neighbor calculation on the most similar fingerprints stored in the database. The metric used to extract similar measurements to the one with the unknown position was based on the logarithmic Gaussian distance (LGD). The LGD between two measurement vectors \( \mathbf{p} \) and \( \mathbf{r} \) is:

\[
\text{LGD}(\mathbf{p}, \mathbf{r}) = -\sum_t \log \max(G(p_t, r_t), \epsilon), \tag{3}
\]

where \( G(p, r) \) is the Gaussian similarity between two values \( p \) and \( r \), defined as:
G(p, r) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{(p-r)^2}{2\sigma^2} \right), & \text{if } p \neq 0 \text{ and } q \neq 0, \\ 0, & \text{otherwise}. \end{cases} \quad (4)

4.3. Primary Results

The initial results for both algorithms when applied to the complete test area fingerprint database showed the following mean 2D error:

- 30.017 m for the weighted centroid method;
- 45.414 m for the log-Gaussian distance method.

These mean error values are significantly higher than those reported in previous studies where the test area had a smaller surface and was placed indoors, for example, approximately 10 m in [17] or approximately 6 m in [39]. When compared to previous results reported in papers evaluating positioning accuracy in urban environments, the attained results using crowdsensed data are similar, for example, reference [40] reports an average positioning accuracy between 23.5 and 36 m, while reference [41] reports errors between 35 and 120 m. When comparing indoor fingerprinting positioning research with outdoor fingerprinting positioning research results, we observe that the mean errors of outdoor systems are by an order of magnitude larger than those of indoor systems.

The mean error of outdoor fingerprinting positioning systems can be reduced using database processing, as proposed in [15,42,43], which publish results closing a 15-m average positioning error. The fingerprint database processing in the referenced papers involved eliminating the fingerprints of APs with an estimated coverage radius larger than 300 m and those of APs suspected to be mobile.

4.4. Characterization of the Positioning Algorithms

Each method has its characteristics which influence the resulting positioning error and also the way they benefit from the crowdsensed fingerprints. The implementation of the weighted-centroid method requires an initial step preceding the actual positioning that requires the estimation of the position of all the APs in the fingerprint database. Using crowdsensed data for AP position estimation can easily lead to errors in estimating the AP’s exact location. This is to be expected, as crowdsensed fingerprints are collected alongside roads or near buildings. The initial step mandates the study of the influence number of crowdsensed fingerprints for each AP.

The log-Gaussian method aims at finding the most similar fingerprints with the measurement vector used for positioning by the help of the log-Gaussian distance. The position estimation is then done using the $k$-nearest neighbor ($k$-NN) approach on the $k$ most similar fingerprints. For the evaluation of the method, the number of neighbors was set to $k = 3$. Varying the number of neighbors influences the results of the position estimation algorithm. For example, an increase in the value of $k$ may lead to an estimated position further away from the test position due to the inclusion of less-similar fingerprints that were collected at a greater distance. This can be solved by applying a weight to the $k$-NN algorithm or by evaluating the influence of the crowdsensed data on the position estimation algorithm.

The particularities of each positioning method coupled with the less-known influence of the crowdsensed fingerprints prompted us to research the following aspects which influence the accuracy of crowdsensed fingerprinting positioning:

- The influence of mobile APs;
- The average number of fingerprints for each AP;
- The signal level of each AP measurement included in a fingerprint;
- The average signal level of the APs;
- The estimated coverage radius of the APs;
- The number of neighbors used for the $k$-NN estimation.
4.5. Influence of Mobile APs

Mobile APs can lead to significant positioning errors. For example, an AP installed in a public transport vehicle could lead to positioning estimation results placed in the part of the town where the AP was scanned. This prompted us to apply a method for removing the fingerprints of mobile APs, while also investigating the effects of their presence on the overall positioning performance.

The identification of mobile APs requires the analysis of the whole fingerprint database, not only of the test area. Mobile APs are classified using a method previously published by the authors in [15]. The method is based on the assumption that APs maximum coverage radius is rarely larger than 300 m. All measurement points for a distinct AP are selected and used to create a cluster. The minimum bounding box and the centroid point of the cluster are calculated. The maximum coverage of the AP is determined to be as the maximum distance between the centroid point and the minimum bounding box. Thus, APs with a range greater than 300 m are considered mobile.

By applying the described method we attained the following results:

• 3506 APs representing 34.8% of the total were scanned once and could not be classified as fixed or mobile. Thus they were eliminated from the comparative dataset;
• 395 APs representing 3.9% of the total were classified as mobile and were eliminated from the comparative dataset.

Table 2 displays the resulting average positioning error for both positioning methods when applied to the dataset with mobile APs removed. The weighted-centroid method displayed a minor increase in the average positioning error, caused by the elimination of APs seen only once while the log-Gaussian method benefited from the removal of the mobile APs.

Table 2. 2D positioning error for the dataset without mobile APs.

| Dataset                  | Weighted-Centroid Mean Error (m) | log-Gaussian Mean Error (m) |
|--------------------------|----------------------------------|-----------------------------|
| Full dataset             | 30.017                           | 45.414                      |
| Mobile AP removed set dataset | 34.245 (−14%)                    | 28.180 (+37%)               |

4.6. Influence of the Average Number of Fingerprints Collected for Each AP

The average number of fingerprints per AP could influence both positioning methods when considering crowdsensing as the collection method. A low number of fingerprints can limit the information available in the AP position estimation step characteristic to the weighted-centroid method while a large number of fingerprints correlated to a large coverage area might reduce the unique character of each fingerprint, leading to an increase in the positioning error for both methods.

To study the previous hypothesis, the dataset was split into subsets using the average fingerprint number for each AP as the criteria as follows:

• Reduced fingerprint dataset containing APs averaging 1 to 5 fingerprints resulting in 13,647 fingerprints of 6798 APs;
• Average fingerprint dataset containing APs averaging 6 to 20 fingerprints resulting in 24,890 fingerprints of 2318 APs;
• Dense fingerprint dataset containing APs averaging 21 to 60 fingerprints resulting in 26,957 fingerprints of 828 APs;
• Very dense fingerprint dataset containing APs averaging more than 61 fingerprints resulting in 10,234 fingerprints of 128 APs.

The resulting mean positioning error for each sub-dataset is shown in Table 3.
Table 3. 2D positioning error influenced by the average fingerprint-to-AP ratio.

| Dataset                        | Weighted-Centroid Mean Error (m) | Log-Gaussian Mean Error (m) |
|--------------------------------|----------------------------------|----------------------------|
| Full dataset                   | 30.017                           | 45.414                     |
| Reduced fingerprint dataset    | 26.916 (+13%)                    | 92.198 (−103%)             |
| Average fingerprint dataset    | 37.896 (−20%)                    | 31.147 (+30%)              |
| Dense fingerprint dataset      | 49.439 (−64%)                    | 38.274 (+15%)              |
| Very dense fingerprint dataset | 76.055 (−153%)                   | 65.929 (−45%)              |

The weighted centroid displayed increased precision for the reduced fingerprint dataset. This is caused by a higher positioning accuracy in scenarios where a reduced number of fingerprints were collected for the same APs in close proximity to each other. This is the case for APs that are visible in a narrow area at street level due to their placement in a distant position. For the datasets with a higher fingerprint-to-AP ratio, the precision was lower, mainly due to a decrease in the AP position estimation caused by the higher number and wider geographical distribution of the APs’ fingerprints.

The log-Gaussian method had a worse result for the reduced and very-dense datasets and significantly better accuracy for the intermediate values. A low number of fingerprints for APs means that less data is available for the statistic characterization of the data, while a large number of fingerprints often leads to the phenomena of geographic diffusion of precision where similar signal levels are found in fingerprints placed far away from each other.

Very dense fingerprint-to-AP ratios led to a decrease in positioning accuracy, leading to the conclusion that collection and storage of large volumes of fingerprints is not always beneficial to the fingerprint positioning system.

4.7. Influence of Fingerprint Signal Level

The signal level’s influence on positioning accuracy was studied by splitting the dataset into subsets using the signal level value expressed using dBm units as a threshold. The subset split aimed at keeping the number of fingerprints in each subset similar. The results of the analysis are displayed in Table 4.

Table 4. 2D positioning error influenced by fingerprint signal level.

| Dataset          | Weighted-Centroid Mean Error (m) | Log-Gaussian Mean Error (m) |
|------------------|----------------------------------|----------------------------|
| Full dataset     | 30.017                           | 45.414                     |
| Subset under −90 dBm | 32.394 (−8%)                   | 118.566 (−161%)           |
| Subset between −86 and −90 dBm | 37.574 (−25%)                  | 78.873 (−74%)             |
| Subset between −81 and −85 dBm | 31.318 (−4%)                   | 71.872 (−58%)             |
| Subset between −76 and −80 dBm | 24.784 (+17%)                  | 82.650 (−81%)             |
| Subset between −71 and −75 dBm | 17.922 (+40%)                   | 122.574 (−169%)           |
| Subset over −70 dBm | 15.489 (±48%)                   | 136.788 (±201%)           |

As shown in Table 4, the log-Gaussian method’s accuracy was significantly lower than that of the full dataset. This was due to the usage of the log-Gaussian distance for determining the similarity between fingerprints and positioning measurements. When splitting the measurements using 5 dBm intervals it becomes more difficult for the method to properly identify similarities and differences between the fingerprints and unknown position measurements, leading to an increase in positioning error. The weighted-centroid method displayed increasing accuracy for signal levels above −80 dBm, as is to be expected for distance-based positioning methods.

4.8. Influence of Average AP Signal Level

The average AP signal level was calculated as the average level of all collected fingerprints for each distinct AP. Again, the dataset was split into subsets using the average signal level as the
criteria, aiming at keeping the number of fingerprints in each set similar. The results of the analysis are displayed in Table 5.

**Table 5.** 2D positioning error influenced by average AP signal level.

| Dataset | Weighted-Centroid Mean Error (m) | Log-Gaussian Mean Error (m) |
|---------|---------------------------------|----------------------------|
| Full dataset | 30.017 | 45.414 |
| Subset under −90 dBm | 32.394 (−8%) | 118.566 (−161%) |
| Subset between −86 and −90 dBm | 36.922 (−23%) | 46.869 (−3%) |
| Subset between −81 and −85 dBm | 42.766 (−42%) | 40.885 (−10%) |
| Subset between −76 and −80 dBm | 46.991 (−56%) | 68.903 (−51%) |
| Subset between −71 and −75 dBm | 32.799 (−9%) | 166.648 (−266%) |

As fewer APs were included in each subset due to the filtering, the weighted-centroid method performed worse in each scenario. The log-Gaussian method had a minor accuracy improvement for APs with an average signal level close to the average signal level of all collected fingerprints.

### 4.9. Influence of Estimated AP Coverage Radius

The AP coverage radius was estimated by the help of the crowdsensing process. The scanning process continued for five months inside the test area, allowing us to assume that street-level coverage was captured with precision. The average coverage radius of APs in the test area was approximately 47 m. The coverage area was determined by the help of the convex hull operation. The convex hull or convex envelope of a set of points \( X \) in a Euclidean plane or space is the smallest convex set that contains \( X \). When applying the convex hull operation to the set of points consisting of the positions of all fingerprints collected for an AP, the result will be an estimation of the area covered by the AP.

The dataset was split into subsets using the estimated coverage radius as the criteria, and the results of the accuracy analysis are displayed in Table 6. The subset split aimed at keeping a similar number of fingerprints for each subset. This was possible for APs with a coverage radius under 80 m. APs with estimated coverage radius larger than 80 m were less numerous in the test area.

The resulting subsets based on the coverage radius had the following dimensions:

- Null radius—6506 fingerprints of 4988 APs;
- Radius between 0 and 30 m—17,618 fingerprints of 2634 APs;
- Radius between 30 and 50 m—19,603 fingerprints of 1328 APs;
- Radius between 50 and 80 m—18,659 fingerprints of 774 APs;
- Radius between 80 and 110 m—8074 fingerprints of 250 APs;
- Radius between 110 and 170 m—3665 fingerprints of 81 APs;
- Radius larger than 110 m—4728 fingerprints of 98 APs;

**Table 6.** 2D positioning error influenced by average AP radius.

| Dataset | Weighted-Centroid Mean Error (m) | Log-Gaussian Mean Error (m) |
|---------|---------------------------------|----------------------------|
| Full dataset | 30.017 | 45.414 |
| Null radius | 11.033 (+63%) | 278.320 (−512%) |
| Radius between 0 and 30 m | 24.385 (+18%) | 21.428 (+47%) |
| Radius between 30 and 50 m | 33.744 (−12%) | 28.127 (+38%) |
| Radius between 50 and 80 m | 47.480 (−58%) | 38.080 (+16%) |
| Radius between 80 and 110 m | 69.585 (−131%) | 57.914 (−27%) |
| Radius between 110 and 170 m | 94.099 (−213%) | 75.373 (−65%) |
| Radius larger than 110 m | 111.839 (−272%) | 97.662 (−115%) |
The positioning performance of the weighted-centroid method benefited from the usage of APs with an estimated coverage radius under 30 m, due to its initial step which requires the estimation of the AP position. A lower-radius AP is less likely to be positioned erroneously.

The log-Gaussian method had significantly greater accuracy for APs with an estimated radius under 50 m. For both methods, an AP coverage radius larger than 80 m resulted in larger positioning errors related to the geographic dilution of precision phenomena.

This analysis allows us to gain insight that can help set an upper limit on the estimated AP coverage area which in turn can limit the volume and effort related to the crowdsensing process.

4.10. Fingerprinting Filtering Summary

As previous filtering shows mixed results for the compared positioning methods, we highlight the results of the weighted-centroid and log-Gaussian methods in Tables 7 and 8 by sorting them according to the 2D positioning error.

| Dataset | Weighted-Centroid Mean Error (m) | Log-Gaussian Mean Error (m) |
|---------|----------------------------------|-----------------------------|
| Null radius | 11.033 (+63%) | 278.320 (−512%) |
| Fingerprint level between −71 and −75 dBm | 17.922 (+40%) | 122.574 (−169%) |
| Radius between 0 and 30 m | 24.385 (+18%) | 21.428 (+47%) |
| Fingerprint level between −76 and −80 dBm | 24.784 (+17%) | 82.650 (−81%) |
| Reduced fingerprint dataset | 26.916 (+13%) | 92.198 (−103%) |
| Full dataset | 30.017 | 45.414 |
| Fingerprint level between −81 and −85 dBm | 31.318 (−4%) | 71.872 (−58%) |
| Radius between 30 and 50 m | 33.744 (−12%) | 28.127 (+38%) |

Table 7 shows that the weighted-centroid method benefited from using fingerprints collected from APs with a radius under 30 m and with signal levels between −71 and −80 dBm, meaning they were collected in the proximity of the AP.

| Dataset | Weighted-Centroid Mean Error (m) | Log-Gaussian Mean Error (m) |
|---------|----------------------------------|-----------------------------|
| Radius between 0 and 30 m | 24.385 (+18%) | 21.428 (+47%) |
| Radius between 30 and 50 m | 33.744 (−12%) | 28.127 (+38%) |
| Average fingerprint dataset | 37.896 (−20%) | 31.147 (+30%) |
| Radius between 50 and 80 m | 47.480 (−58%) | 38.080 (+16%) |
| Dense fingerprint dataset | 49.439 (−64%) | 38.274 (+15%) |
| Average AP signal level between −81 and −85 dBm | 42.766 (−42%) | 40.885 (+10%) |
| Full dataset | 30.017 | 45.414 |
| Average AP signal level between −86 and −90 dBm | 36.922 (−23%) | 46.869 (−3%) |

Similarly to the weighted-centroid method, as Table 8 shows, the log-Gaussian method also benefited from APs with a low coverage radius. Aside from that, the log-Gaussian was characterized by increased accuracy for denser fingerprint subsets than the weighted-centroid method.

As can be seen in both tables, the methods had opposite performances for the same fingerprint subsets, except for the one with APs having a radius between 0 and 30 m. This results in the opportunity of using multiple estimation methods for the envisioned urban fingerprinting positioning systems that can be applied to different areas to be covered by the system. For example, if computed AP statistics show that an area is scanned more densely during the crowdsensing process, the log-Gaussian method should be used, as it outperformed the weighted-centroid method (38.080 m vs. 47.490 m average 2D error). Otherwise, if an area is characterized by a low average number of fingerprints for each AP the weighted-centroid method will outperform the log-Gaussian method (26.916 m vs. 91.198 m average 2D error).
4.11. Influence of the Number of Neighbors Specific to the k-NN Method

The k-NN method is often used for position estimation by fingerprinting systems. The k-NN method was combined with the log-Gaussian distance to select the most similar neighbors in the fingerprint space. Previous papers have proposed different values for $k$, according to the environment that the fingerprinting positioning system was to cover. References [28,44] propose that $k = 3$ or 4 for indoor fingerprinting systems.

The analysis of the influence of the number of neighbors on the accuracy of the positioning was done by varying the number between 1 and 10 and by computing the average 2D positioning error in each scenario. The dataset used contains 15,002 fingerprints of 1929 APs in the test area with an average estimated radius between 0 and 30 m. The results are displayed in Figure 4.

![Figure 4. k-Nearest neighbor (k-NN) average 2D positioning error influenced by the number of neighbors.](image)

According to Figure 4, the average positioning error was smallest for $k = 2$. This result is explained mainly by the data collection method. Collecting Wi-Fi measurements using crowdsensing means that most fingerprints will be captured alongside public roads or in the proximity of buildings, while most APs will be positioned indoors. Using three or more neighbors in the fingerprint space will likely lead to an estimation that is biased towards the two closest fingerprints, and as Figure 4 shows, farther than the real position of the user. When $k = 2$ the positioning error was the smallest due to the probability that the real position was alongside the road on which the two fingerprints were collected.

5. Conclusions

Our analysis shows that the positioning accuracy of outdoor fingerprinting systems can be significantly improved by filtering and optimizing the contents of the fingerprint database. The best results came close to an average 2D positioning error of 21 m for the log-Gaussian distance method. These values do not come close to that achieved by previously published research on indoor systems. Nevertheless, our results show that urban fingerprinting positioning can reach GPS-like accuracy in similar use-cases as reported by reference [5].

The accuracy of crowdsensed fingerprinting positioning systems is shown to be dependent on the density of fingerprints, quantified as the average number of fingerprints collected for each AP. As the average positioning error was computed for the whole test area, including areas that were scanned...
less often and dense, it is safe to assume that for some of the most-scanned roads inside the test area the precision of the fingerprinting system would outperform that of GPS.

Wi-Fi fingerprint positioning systems for urban environments have become an intense topic of research in the context of Internet-of-Things technology. These positioning systems have the ability to offer GPS-like positioning accuracy for energy-constrained devices equipped with a Wi-Fi or 5G radio network interface that is used both for communications and positioning.

To evaluate the performance of urban fingerprint positioning systems, we built our fingerprint database using the crowdsensing technique for data gathering. Since the crowdsensing approach makes it difficult to cover a whole city with a small number of contributors, we chose to conduct our analysis on a dedicated test area specific to the Bucharest urban environment.

We analyzed the crowdsensed fingerprint and highlighted important aspects related to the research problem, including the presence of mobile APs, the influence of the average number of fingerprints collected per AP, the fingerprint signal level, the average street-level AP signal level, the influence of the estimated AP coverage radius, and the performance of the $k$-NN method.

We showed that the chosen positioning methods mostly performed in different ways when particular filters were applied. For example, the weighted-centroid method had a 17.922 m average 2D error for the subset with fingerprints had signal levels between $-71$ and $-75$ dBm, while the log-Gaussian method offered poor results at an average 2D error of 122.574 m. The main differences in the methods’ performances can be assigned to the fact that the weighted-centroid method requires an initial estimation of all of the APs’ positions, while the log-Gaussian method uses a similarity function that can perform poorly in particular situations, as proven by the fingerprint filtering.

We conclude by proving that fingerprint positioning systems can achieve or even out-perform satellite positioning systems in an urban environment, mostly due to the density of Wi-Fi APs and the difficulties that other positioning systems have in these conditions.

**Author Contributions:** C.-L.L. conceived the work, was responsible for data collection and validation, analyzed the data, and wrote the paper. I.N. described the methodology, supervised the research, and revised the paper. P.C. revised the work and was responsible for project administration and funding acquisition. All authors have read and agreed to the published version of the manuscript.

**Funding:** The work was funded by the Romanian Government UEFISCDI 9SOL project (PN-III-P2-2.1-SOL-2017-08-010).

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

**References**

1. Han, Q.; Liang, S.; Zhang, H. Mobile cloud sensing, big data, and 5G networks make an intelligent and smart world. *IEEE Netw.* 2015, 29, 40–45. [CrossRef]
2. Choudhury, T.; Borriello, G.; Consolvo, S.; Haehnel, D.; Harrison, B.; Hemingway, B.; Hightower, J.; Koscher, K.; LaMarca, A.; Landay, J.A.; et al. The mobile sensing platform: An embedded activity recognition system. *IEEE Pervasive Comput.* 2008, 7, 32–41. [CrossRef]
3. Pei, L.; Zhang, M.; Zou, D.; Chen, R.; Chen, Y. A survey of crowd sensing opportunistic signals for indoor localization. *Mob. Inf. Syst.* 2016, 2016, 401291. [CrossRef]
4. Zhou, X.; Chen, T.; Guo, D.; Teng, X.; Yuan, B. From one to crowd: A survey on crowdsourcing-based wireless indoor localization. *Front. Comput. Sci.* 2018, 12, 423–450. [CrossRef]
5. Van Diggelen, F.; Enge, P. The worlds first gps mooc and worldwide laboratory using smartphones. In Proceedings of the 28th international technical meeting of the satellite division of the institute of navigation (ION GNSS+ 2015), Tampa, FL, USA; 14–18 September 2015; pp. 361–369.
6. Lee, H.; Shim, J.Y.; Kim, H.S.; Li, B.; Rizos, C. Feature extraction and spatial interpolation for improved wireless location sensing. *Sensors* 2008, 8, 2865–2885. [CrossRef]
7. Lee, H.K.; Rizos, C. A framework for calibrating NLOS error to support LBS in urban environments. In Proceedings of the 2003 International Symposium on GPS/GNSS, Washington, DC, USA, 23–24 October 2003; pp. 69–78.

8. Lee, H.; Li, B.; Rizos, C. Implementation procedure of wireless signal map matching for location-based services. In Proceedings of the Fifth IEEE International Symposium on Signal Processing and Information Technology, Athens, Greece, 21 December 2005; pp. 429–434.

9. Faragher, R.; Harle, R. Location fingerprinting with bluetooth low energy beacons. *IEEE J. Sel. Areas Commun.* 2015, 33, 2418–2428. doi:10.1109/JSAC.2014.1356987. [CrossRef]

10. Wigren, T. Adaptive enhanced cell-ID fingerprinting localization by clustering of precise position measurements. *IEEE Trans. Veh. Technol.* 2007, 56, 3199–3209. [CrossRef]

11. Yu, N.; Zhao, S.; Ma, X.; Wu, Y.; Feng, R. Effective Fingerprint Extraction and Positioning Method Based on Crowdsourcing. *IEEE Access* 2019, 7, 162639–162651. [CrossRef]

12. Kwak, M.; Hamm, C.; Park, S.; Kwon, T.T. Magnetic Field based Indoor Localization System: A Crowdsourcing Approach. In Proceedings of the 2019 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Pisa, Italy, 30 September–3 October 2019; pp. 1–8.

13. Chen, Z.; Zou, H.; Jiang, H.; Zhu, Q.; Soh, Y.; Xie, L. Fusion of WiFi, smartphone sensors and landmarks using the Kalman filter for indoor localization. *Sensors* 2015, 15, 715–732. [CrossRef]

14. Leca, C.L.; Ciotirnae, P.; Rincu, C.I.; Nicolaescu, I. Characteristics of crowdsourcing for outdoor radio fingerprinting positioning. In Proceedings of the 2017 9th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Targoviste, Romania, 29 June–1 July 2017; pp. 1–4. [CrossRef]

15. Leca, C.L. Ubiquity of Wi-Fi: Crowdsensing Properties for Urban Fingerprint Positioning. *Adv. Electr. Comput. Eng.* 2017, 17, 131–136. [CrossRef]

16. Leca, C.L. Overview of Romania 802.11 wireless networks security. In Proceedings of the 2017 9th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Targoviste, Romania, 29 June–1 July 2017; pp. 1–4. [CrossRef]

17. Lohan, E.S.; Torres-Sospedra, J.; Leppäkoski, H.; Richter, P.; Peng, Z.; Huerta, J. Wi-Fi Crowdsourced Fingerprinting Dataset for Indoor Positioning. *Data* 2017, 2, 32. [CrossRef]

18. Mendoza-Silva, G.M.; Richter, P.; Torres-Sospedra, J.; Lohan, E.S.; Huerta, J. Long-Term WiFi Fingerprinting Dataset for Research on Robust Indoor Positioning. *Data* 2018, 3, 3. [CrossRef]

19. Gentile, C.; Alsindi, N.; Raulefs, R.; Teolis, C. Geolocation Techniques: Principles and Applications; Springer Science & Business Media: London, UK, 2012.

20. Wang, J.; Luo, J.; Pan, S.J.; Sun, A. Learning-Based Outdoor Localization Exploiting Crowd-Labeled WiFi Hotspots. *IEEE Trans. Mob. Comput.* 2019, 18, 896–909. doi:10.1109/TMC.2018.2849416. [CrossRef]

21. Basiri, A.; Lohan, E.S.; Moore, T.; Winstanley, A.; Peltola, P.; Hill, C.; Amirian, P.; e Silva, P.F. Indoor location based services challenges, requirements and usability of current solutions. *Comput. Sci. Rev.* 2017, 24, 1–12. [CrossRef]

22. Wu, C.; Yang, Z.; Liu, Y.; Xi, W. WILL: Wireless indoor localization without site survey. *IEEE Trans. Parallel Distrib. Syst.* 2012, 24, 839–848.

23. Liu, H.; Gan, Y.; Yang, J.; Sidhom, S.; Wang, Y.; Chen, Y.; Ye, F. Push the limit of WiFi based localization for smartphones. In Proceedings of the 18th annual international conference on Mobile computing and networking, Istanbul, Turkey, 22–26 August 2012; pp. 305–316.

24. Sun, W.; Liu, J.; Wu, C.; Yang, Z.; Zhang, X.; Liu, Y. MoLoc: On distinguishing fingerprint twins. In Proceedings of the 2013 IEEE 33rd international conference on distributed computing systems, Philadelphia, PA, USA, 8–11 July 2013; pp. 226–235.

25. Xiao, Z.; Wen, H.; Markham, A.; Trigoni, N.; Blunsom, P.; Frolik, J. Non-line-of-sight identification and mitigation using received signal strength. *IEEE Trans. Wirel. Commun.* 2014, 14, 1689–1702. [CrossRef]

26. Lui, G.; Gallagher, T.; Li, B.; Dempster, A.G.; Rizos, C. Differences in RSSI readings made by different Wi-Fi chipsets: A limitation of WLAN localization. In Proceedings of the 2011 International Conference on Localization and GNSS (ICL-GNSS), Tampere, Finland, 29–30 June 2011; pp. 53–57.

27. Sapiezynski, P.; Stopczynski, A.; Gatej, R.; Lehmann, S. Tracking human mobility using wifi signals. *PLoS ONE* 2015, 10, e0130824. [CrossRef]
28. Bahl, P.; Padmanabhan, V.N. RADAR: An in-building RF-based user location and tracking system. In Proceedings of the INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies, Tel Aviv, Israel, 26–30 March 2000; Volume 2, pp. 775–784. [CrossRef]

29. Youssef, M.; Agrawala, A. The Horus WLAN location determination system. In Proceedings of the 3rd international conference on Mobile systems, applications, and services, Seattle, WA, USA, 6–8 June 2005; pp. 205–218. [CrossRef]

30. Bolliger, P. Redpin-adaptive, zero-configuration indoor localization through user collaboration. In Proceedings of the first ACM international workshop on Mobile entity localization and tracking in GPS-less environments, Orlando, FL, USA, September 2008; pp. 55–60.

31. Park, J.g.; Charrow, B.; Curtis, D.; Battat, J.; Minkov, E.; Hicks, J.; Teller, S.; Ledlie, J. Growing an organic indoor location system. In Proceedings of the 8th international conference on Mobile systems, applications, and services, San Francisco, CA, USA, 15–18 June 2010; pp. 271–284.

32. Ferris, B.; Fox, D.; Lawrence, N.D. WiFi-SLAM Using Gaussian Process Latent Variable Models. IJCAI. 2007, Volume 7, pp. 2480–2485. Available online: https://www.semanticscholar.org/paper/WiFi-SLAM-Using-Gaussian-Process-Latent-Variable-Ferris-Fox/600376bb397c5f3d8d33856b3a6db314b6bda67d (accessed on 11 January 2020).

33. Rai, A.; Chintalapudi, K.K.; Padmanabhan, V.N.; Sen, R. Zee: Zero-effort crowdsourcing for indoor localization. In Proceedings of the 18th annual international conference on Mobile computing and networking, Istanbul, Turkey, 22–26 August 2012; pp. 293–304.

34. Wang, H.; Sen, S.; Elgohary, A.; Farid, M.; Youssef, M.; Choudhury, R.R. No need to war-drive: unsupervised indoor localization. In Proceedings of the 10th international conference on Mobile systems, applications, and services, Lake District, UK, 26–29 June 2012; pp. 197–210. [CrossRef]

35. Jung, S.; Lee, G.; Han, D. Methods and Tools to Construct a Global Indoor Positioning System. IEEE Trans. Syst. Man, Cybern. Syst. 2018, 48, 906–919. [CrossRef]

36. Laitinen, E.; Lohan, E.S. On the choice of access point selection criterion and other position estimation characteristics for WLAN-based indoor positioning. Sensors 2016, 16, 737. [CrossRef]

37. Razavi, A.; Valkama, M.; Lohan, E.S. K-means fingerprint clustering for low-complexity floor estimation in indoor mobile localization. In Proceedings of the Globecom Workshops (GC Wkshps), San Diego, CA, USA, 6–10 December 2015; pp. 1–7. [CrossRef]

38. Cramariuc, A.; Huttunen, H.; Lohan, E.S. Clustering benefits in mobile-centric WiFi positioning in multi-floor buildings. In Proceedings of the 2016 International Conference on Localization and GNSS (ICL-GNSS), Barcelona, Spain, 28–30 June 2016; pp. 1–6.

39. Wu, C.; Yang, Z.; Liu, Y. Smartphones based crowdsourcing for indoor localization. IEEE Trans. Mob. Comput. 2015, 14, 444–457. [CrossRef]

40. Li, B.; Quader, I.J.; Dempster, A.G. On outdoor positioning with Wi-Fi. Positioning 2008, 1. [CrossRef]

41. Quader, I.J.; Li, B.; Feng, W.; Dempster, A.G. Use of fingerprinting in Wi-Fi based outdoor positioning. Proc IGNSS Sydney 2007. Available online: https://pdfs.semanticscholar.org/c6fd/f92a834da0f3f580cd454af7ed8953d331ed7.pdf (accessed on 11 January 2020).

42. Wang, J.; Tan, N.; Luo, J.; Pan, S.J. WOLoc: WiFi-only outdoor localization using crowdsensed hotspot labels. In Proceedings of the INFOCOM 2017-IEEE Conference on Computer Communications, Atlanta, GA, USA, 1–4 May 2017; pp. 1–9. [CrossRef]

43. Sapiezynski, P.; Gatej, R.; Mislove, A.; Lehmann, S. Opportunities and challenges in crowdsourced wardriving. In Proceedings of the 2015 ACM Conference on Internet Measurement Conference, Tokyo, Japan, 28–30 October 2015; pp. 267–273. [CrossRef]

44. Torres-Sospedra, J.; Moreira, A. Analysis of Sources of Large Positioning Errors in Deterministic Fingerprinting. Sensors 2017, 17, 2736. [CrossRef] [PubMed]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).