Intercontinental Geoinformation Days

igd.mersin.edu.tr

The impact of variable neighbor numbers on Wi-Fi fingerprint-based indoor positioning using the KNN and WKNN algorithms

Behlül Numan Özdemir*[1], Ayhan Ceylan [1]

[1]Konya Technical University, Faculty of Engineering and Natural Sciences, Geomatics Engineering Department, Konya, Turkey

Keywords
IPS
Wi-Fi
Fingerprinting
WKNN

ABSTRACT

Indoor positioning is an area where GNSS signals are either not available or very weak to provide sufficient positioning accuracy. We use smart mobile devices, which are technologically advanced today, as a solution to this issue. Despite the fact that they contain GNSS receivers and some also have dual-band chips, they currently do not have solutions for indoor spaces. As a result, we use Wi-Fi infrastructure, which is as widely used as GNSS. Although the purpose of its emergence is wireless communication, it is now one of the most popular indoor positioning applications. This study used the fingerprint approach, which is among the most successful methods of indoor positioning using this technology. We looked at two parameters related to both the positioning and calibration stages. The 2-meter point interval had the lowest mean errors when these parameters, which are the number of neighbors in KNN and WKNN algorithms and the point frequency in the calibration process, were examined. Furthermore, it has been observed that the KNN algorithm produces significant errors as the number of closest neighbors selected increases. Given the method's simplicity, we may conclude that the NN algorithm's results are quite respectable.

1. INTRODUCTION

There has been a rapid increase in indoor positioning applications due to the widespread use of smart phones and their technical advancements (Wi-Fi 6, Dual-Band GPS, inertial sensors etc.). The Wi-Fi fingerprinting method is one of the most well-known of these applications. Consisting of two phases, Offline (Calibration) and Online (Positioning) Phase, fingerprinting method has its strengths and weaknesses. The main drawback of this approach is the Calibration Phase, which takes a long time and necessitates a lot of human effort. Although some studies concentrate on decreasing the necessary manpower (Wu et al., 2015; Yang et al., 2013) or looking for solutions with autonomous robots (Bakri et al., 2020), others propose a variety of time-saving approaches (Bi et al., 2019). To provide an example of how Wi-Fi Fingerprinting differs from other approaches, it makes use of existing WLAN infrastructure, obviating the need for additional hardware. Since almost all smartphones and majority of smart mobile devices have Wi-Fi hardware readily available, this method is highly applicable to anywhere needed.

In the aforementioned calibration phase a point grid is generated homogeneously in indoor space. These are referred to as calibration points, and the signal fingerprint is the vector formed by the measurements made on them. A typical fingerprint consists of 2D point coordinates, received signal strength (RSS) information from nearby WAPs (Wireless Access Point) and floor information. Data from inertial sensors such as magnetic field sensors and accelerometers is also included in the fingerprint in more sophisticated systems. RSS measurements are performed multiple times at each calibration point due to signal fluctuations. The RSS values obtained from an unknown point are compared to the RSS values obtained from all calibration points during the Positioning Phase. While Euclidean signal distance is the most commonly used distance/similarity measure in this comparison, several other distance/similarity measures are also used (Cha, 2007; Torres-Sospedra et al., 2015). Finally, the location of the unknown point is estimated from the point (s) with the smallest signal distance. If a single calibration point is used to determine this smallest signal distance, the nearest neighbor algorithm (NN), and if more than one calibration point is used, K-Nearest Neighbor (KNN) algorithms are used. In this study, for the KNN algorithm

* **Corresponding Author**

*(bnozdemir@ktun.edu.tr) ORCID ID 0000-0001-7351-1870
(aceylan@ktun.edu.tr) ORCID ID 0000-0003-4408-4245

Cite this study

Özdemir B.N. & Ceylan A (2021). The impact of variable neighbor numbers on Wi-Fi fingerprint-based indoor positioning using the KNN and WKNN algorithms. 2nd Intercontinental Geoinformation Days (IGD), 44-47, Mersin, Turkey

and Weighted-KNN algorithm, different K values were tested and the changes in position accuracy were examined. The primary goal of this research is to determine the K number in an indoor space under real-world conditions and to compare the results obtained with weighted measurements.

2. METHOD

This study was carried out in the Faculty of Engineering and Natural Sciences building of Konya Technical University (Fig. 1). The wireless network infrastructure of the building consists of approximately 80 wireless access points broadcasting 2.4GHz and 5GHz signals.

Figure 1. Aerial view of the building where the measurements were taken

The methodology of the study was carried out in the following order.

I. Obtaining and coordinating the CAD plan of the building
II. Establishment of calibration points as routes along corridors at desired intervals
III. Conducting coordinated signal strength measurements at each calibration point on the routes prepared with the measurement setup and software (Fig. 2).
IV. Conducting coordinated signal strength measurements at random points to test positioning accuracy
V. Analyzing the collected data with the prepared software

Figure 2. Signal acquisition software(left), measurement setup(right)

Using measurements taken with a total station at the corner points, the CAD plan of the building was adjusted to the correct scale and position. NetCAD 8 software was used to create measurement routes and mark calibration points on the plan. The signal strength measurements were completed with the measurement setup by importing the coordinate list of these marked points into the mobile data collection software. To determine the point positioning accuracy, measurements were taken at 237 test points and 1771 calibration points. In addition, the table below shows the number of calibration points for data arranged between 1 to 5 meters to investigate the effect of various neighbor numbers. (Table 1).

Table 1. Total calibration points for each scenario

Gap between two calibration points	Total Calibration Points
1 m	1771
2 m	907
3 m	606
4 m	471
5 m	366

2.1. Position Estimation

The Euclidean distance used to determine similarity is calculated as (1), where P and Q are two signal vectors and j is the vector length.

$$d_{euclidean}(P,Q) = \sqrt{\sum_{i=1}^{j} |P_i - Q_i|^2} \quad (1)$$

When using the nearest neighbor algorithm to determine location, the calibration point with the shortest Euclidean distance is used, and the KNN algorithm selects the closest K number of calibration points, then estimates the point position as the average of the Y and X coordinates of the K number of closest neighbors found (2).

$$P(X,Y) = \frac{1}{K}\sum_{i=1}^{K}(x_i, y_i) \quad (2)$$

In this process, a certain amount of error is made (Δs) since the influence of each neighbor on the point is evaluated equally in this method (Fig. 3).

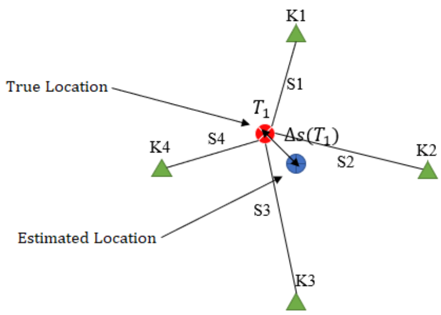

Figure 3. Position estimation with KNN algorithm

The WKNN algorithm, on the other hand, alleviates the problem's impact by weighting neighboring points in proportion to their Euclidean distances (3).

$$w_i = \frac{1/D_i}{\sum_{j=1}^{k} 1/D_j} \quad i = 1,2,\dots,k \tag{3}$$

2.2. Analysis Software

The results of the measurements were analyzed using the software we developed for Wi-Fi fingerprint positioning (Fig. 4). Main features of the program are;

- Thresholding to remove WAPs under certain signal strength
- Plotting signal strength changes over time
- Viewing calibration points on the map
- Interpolation module
- Position estimation with weighted measurements
- Analyzing with Euclidean, Manhattan, Minkowski L3-5 and Sørensen distances
- Different data representation schemes (dBm, Exponential Function and Powed (Torres-Sospedra et al., 2015), Positive, Normalized and Experimental Functions)

Figure 4. Analysis software, plotting screen(top), main screen (bottom)

3. RESULTS AND DISCUSSIONS

To begin with, as the number of neighbors chosen in the KNN algorithm exceeded 2, the average errors increased by up to 6 meters. More importantly, it is seen that the relationship between different neighbor numbers and the accuracies obtained as a result of the change in the calibration point frequency does not change (Fig. 5). This is due to the fact that as the search for the nearest neighbor progresses, less and less relevant points are chosen. This is particularly true in corridor-style indoor spaces. However, this scenario can produce different results in hall-type indoor spaces where the number of relevant neighboring calibration point count is much higher (Shin et al., 2012)

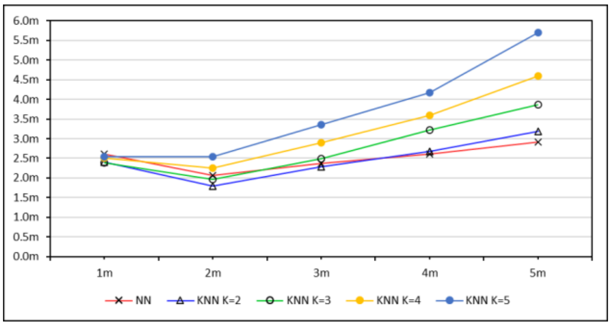

Figure 5. Mean errors of KNN with different neighbor numbers and calibration point gaps

The WKNN algorithm, which has lower average errors, produces similar effects (Fig. 6). WKNN is rather more effective in this situation because it decreases the effect of less important points on position estimation.

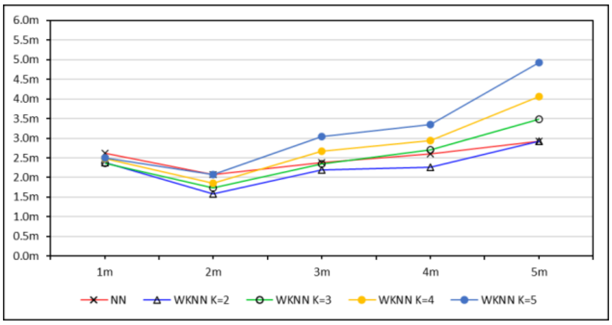

Figure 6. Mean errors of WKNN with different neighbor numbers and calibration point gaps

When each case is looked at individually to decide the best point spacing, it is clear that the 2-meter distribution has the lowest mean errors. In second place, distributions of 1- and 3-meter intervals generated the best results (Fig. 7-8-9-10). It is also clear that the NN algorithm produces respectable result in all ranges, especially above 3 meters.

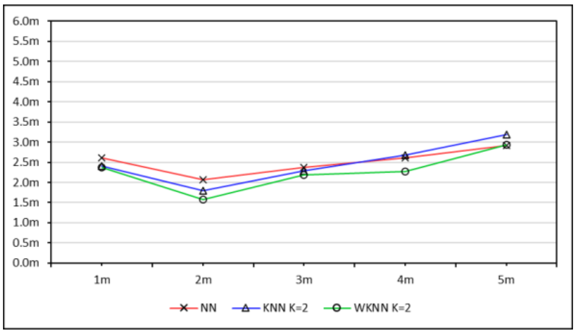

Figure 7. Mean errors of different neighbor numbers for each algorithm, K=2

Total test point count and their locations are clearly linked to the NN algorithms success, which generates comparable average errors at intervals greater than 3 meters. The reason for this is that although the distance between successive calibration points is large in a corridor type closed area, we can say that the closest calibration point to a point in that corridor will be equal or closer to the selected interval.

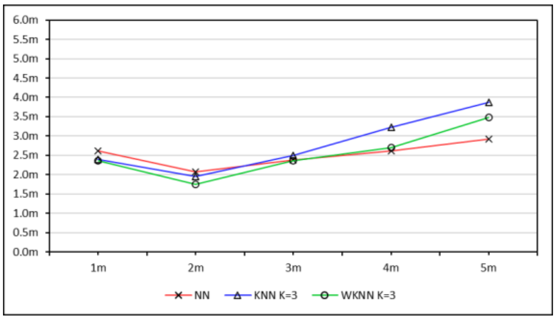

Figure 8. Mean errors of different neighbor numbers for each algorithm, K=3

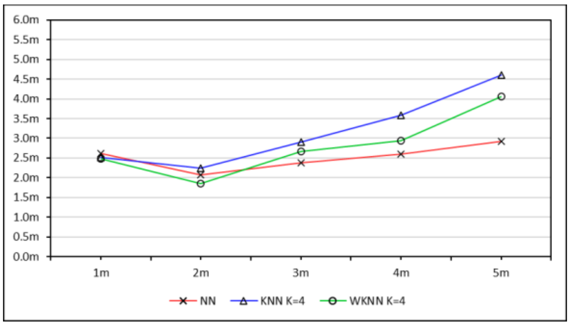

Figure 9. Mean errors of different neighbor numbers for each algorithm, K=4

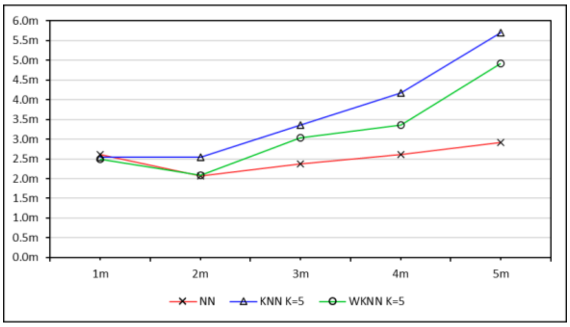

Figure 10. Mean errors of different neighbor numbers for each algorithm, K=5

In addition, the standard deviations obtained from each scenario are shown in Table 1 and Table 2.

Table 1. Standard deviations of NN and KNN algorithms

Gap between two calibration point	NN	KNN K=2	KNN K=3	KNN K=4	KNN K=5
1 m	2.45	2.33	2.22	2.34	2.29
2 m	2.38	1.79	1.70	1.80	1.87
3 m	2.16	2.05	1.70	2.17	2.27
4 m	2.71	2.20	2.40	2.61	2.85
5 m	2.42	2.76	2.65	3.11	3.52

Table 2. Standard deviations of NN and WKNN algorithms

Gap between two calibration point	NN	WKNN K=2	WKNN K=3	WKNN K=4	WKNN K=5
1 m	2.45	2.32	2.21	2.31	2.25
2 m	2.38	1.81	1.65	1.81	1.79
3 m	2.16	2.01	1.64	1.99	2.09
4 m	2.71	2.19	2.23	2.40	2.53
5 m	2.42	2.62	2.39	2.75	3.18

4. CONCLUSION

The application of Wi-Fi fingerprint indoor position determination was investigated in this study. Since the calibration phase is a time-consuming and labor-intensive process, it has been determined that selecting the best point range to determine the number of calibration points is critical. According to this research, the most appropriate calibration point density for the application area is around 2 meters.

It has been observed that as the number of closest neighbors increases, the point positioning accuracy declines. Although this is related to the geometry of the indoor space (Corridor-Type) and fixed K-value, algorithms based on dynamic K-value can provide better results in various scenarios.

We can conclude that the NN algorithm achieves results that are almost as good as KNN and WKNN but it produces higher standard deviation.

REFERENCES

Bakri M Q, Ismail A H, Hashim M S M, Muhamad Azmi M S, Safar M J A & Marhaban M H (2020). Design and Development of a Service Robot for Wi-Fi RSSI Fingerprint Data Collection. *IOP Conference Series: Materials Science and Engineering*, *920*, 012003. https://doi.org/10.1088/1757-899X/920/1/012003

Bi J, Wang Y, Li Z, Xu S, Zhou J, Sun M & Si M (2019). Fast Radio Map Construction by using Adaptive Path Loss Model Interpolation in Large-Scale Building. *Sensors*, *19*(3), 712. https://doi.org/10.3390/s19030712

Cha S H (2007). *Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions*. *1*(4), 8.

Shin B, Lee J H, Lee T & Kim H S (2012). Enhanced Weighted K-Nearest Neighbor Algorithm for Indoor Wi-Fi Positioning Systems. *International Journal of Networked Computing and Advanced Information Management*, *2*(2), 15–21. https://doi.org/10.4156/ijncm.vol2.issue2.2

Torres-Sospedra J, Montoliu R, Trilles S, Belmonte Ó & Huerta J (2015). Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems. *Expert Systems with Applications*, *42*(23), 9263–9278. https://doi.org/10.1016/j.eswa.2015.08.013

Wu C, Yang Z, Xiao C, Yang C, Liu Y & Liu M (2015). Static power of mobile devices: Self-updating radio maps for wireless indoor localization. *2015 IEEE Conference on Computer Communications (INFOCOM)*, 2497–2505. https://doi.org/10.1109/INFOCOM.2015.7218639

Yang S, Dessai P, Verma M & Gerla M (2013). FreeLoc: Calibration-free crowdsourced indoor localization. *2013 Proceedings IEEE INFOCOM*, 2481–2489. https://doi.org/10.1109/INFCOM.2013.6567054