

# Jing, Hao and Pinchin, James and Hill, Chris and Moore, Terry (2016) Wi-Fi fingerprinting based on collaborative confidence level training. Pervasive and Mobile Computing, 30 . pp. 32-44. ISSN 1574-1192

## Access from the University of Nottingham repository:

http://eprints.nottingham.ac.uk/33124/1/Pervasive%20and%20Mobile%20Computing %202015.pdf

## Copyright and reuse:

The Nottingham ePrints service makes this work by researchers of the University of Nottingham available open access under the following conditions.

This article is made available under the Creative Commons Attribution Non-commercial No Derivatives licence and may be reused according to the conditions of the licence. For more details see: http://creativecommons.org/licenses/by-nc-nd/2.5/

#### A note on versions:

The version presented here may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the repository url above for details on accessing the published version and note that access may require a subscription.

For more information, please contact eprints@nottingham.ac.uk

1	Wi-Fi fingerprinting based on collaborative confidence level training
2	Hao Jing <sup>1</sup> *, James Pinchin <sup>2</sup> , Chris Hill <sup>1</sup> , Terry Moore <sup>1</sup>
3	1 Nottingham Geospatial Institute, University of Nottingham, NG7 2TU, UK
4	2 Horizon Digital Economy Research, University of Nottingham, NG7 2TU, UK
5	*Email: lgxhj2@nottingham.ac.uk
6	Abstract:
7	Wi-Fi fingerprinting has been a popular indoor positioning technique with the advantage that
8	infrastructures are readily available in most urban areas. However wireless signals are prone to
9	fluctuation and noise, introducing errors in the final positioning result. This paper proposes a new
10	fingerprint training method where a number of users train collaboratively and a confidence factor
11	is generated for each fingerprint. Fingerprinting is carried out where potential fingerprints are
12	extracted based on the confidence factor. Positioning accuracy improves by 40% when the new
13	fingerprinting method is implemented and maximum error is reduced by 35%.
14	Keywords:
15	Indoor positioning, Wi-Fi fingerprinting, collaborative positioning
16	1 Introduction
17	With the advancement in positioning as well as mobile technologies, location based
18	services (LBS) are no longer just trendy fantasies. LBS applications are expanding from military
19	and government sectors rapidly into commercial and civil applications. Therefore, the
20	fundamental requirement for positioning and navigation is becoming more demanding as
21	solutions are required in more complicated environments, where traditional positioning methods
22	such as Global Navigation Satellite Systems (GNSS) fails. This is known as the indoor positioning
23	problem and numerous methods have been explored over the years to improve positioning
24	performance in such environments [1,2]. Wireless network signal based positioning, such as Wi-
25	Fi fingerprinting, have become widely applied in indoor positioning due to high availability of
26	Wi-Fi signals in urban environments [3,4].

27 Yet Wi-Fi positioning is far from the perfect solution. Wi-Fi networks are not positioning 28 dedicated systems thus signals can be unstable, and sometimes unsuitable for positioning. Hence 29 accuracy and robustness cannot be guaranteed [5]. The complete process of Wi-Fi fingerprinting 30 is achieved in two phases, the training phase which must be carried out first to collect received 31 signal strength (RSS) measurements and the positioning phase to obtain positions based on the 32 fingerprints [6,7]. The positioning performance of fingerprinting relies on the applied positioning 33 algorithm as well as the accuracy and details of the fingerprint database. Therefore, in order to 34 achieve accurate positioning, a detailed database is required. This relies on carefully chosen 35 training points across the building as well as sufficient access points (AP) that covers the area of 36 interest, as more AP will give more information on the variation of signals when in different 37 locations. However, training can be very time consuming. Yet it must be retrained and updated 38 whenever the internal building structure or AP locations change [8,9]. Moreover, although many 39 buildings have been setup with dedicated dense wireless network enabling high accuracy 40 positioning, but most indoor environment still lack such coverage.

To reduce the time and human labour required for database training, Wi-Fi simultaneous localisation and mapping (SLAM) had been applied to enable a quicker way of learning the signal pattern around a new environment and allows the system to navigate in a new environment. However inertial measurements and building information is required, and building information may not always be accessible [10-12].

46 A basic requirement in fingerprinting is that the positions of the fingerprints must be 47 accurate and their RSS measurement should be up-to-date. Studies have looked into the possibility 48 of reducing training effort by reducing training points and training time [13]. Authors in [14,15] 49 looks into autonomous crowdsourcing method for training and updating the Wi-Fi database. For 50 crowd-sourced database, the accuracy of estimated positions of the training data is essential. 51 Collaborative positioning improves positioning accuracy and reliability by applying network 52 constraints on user's positioning measurements. Nearby users may form local networks where relative constraints can be applied to adjust and share each other's measurements [16,17]. Authors 53

in [18,19] improve fingerprinting performance by allowing the user to interact with the system to
label locations and changes. However this requires active collaboration with the user who may
not be willing or could potentially make mistakes.

57 This paper looks into reducing the training effort by introducing a collaborative Wi-Fi 58 fingerprint database training (cFPDB) approach, which achieves quicker and more reliable 59 training. Gaussian Process (GP) regression is applied to generate fingerprints for the entire 60 database from a small amount of training data and a confidence factor is produced for each 61 fingerprint indicating how reliable it is. On the other hand, this solution especially addresses Wi-62 Fi positioning problems in locations where dedicated network is unavailable and are covered only 63 by very sparse APs. With very few APs, users may not be able to observe enough signal variation 64 patterns for accurate positioning. An adaptive collaborative fingerprinting algorithm (WARCP) 65 based on the concept of collaborative positioning is also introduced which provides the location 66 reference for fingerprints as well as knowledge on the expected relationship between Wi-Fi 67 measurements collected by nearby users. Positioning flexibility is also improved as users have 68 the option of performing inertial navigation alone, with collaborative ranging aiding or Wi-Fi 69 fingerprint aiding based on available sensors and number of users.

This paper firstly introduces the collaborative Wi-Fi fingerprint training method and an analysis on training data is presented to understand how much data is required for generating a reliable database. WARCP is then discussed to achieve positioning based on the collaboratively trained database and ranging constraint between users. Simulations are carried out based on the proposed algorithms and discussed in Section 4. Both training and positioning results are analysed for efficient and reliable Wi-Fi fingerprint training and positioning.

76 2 Collaborative Wi-Fi database training

77 2.1 Wi-Fi fingerprinting

79

80

Wireless network based positioning relies on measuring the signal strength of the received signals. Wireless signal strength will attenuate as it travels from the transmitter (i.e. Wi-Fi APs) to the receiver based on the signal path loss model [3],

$$P_{RX}(d) = P_{d0} - 10n \log_{10} d + a \cdot WAF + \varepsilon$$
<sup>(1)</sup>

81 where  $P_{d0}$  is the RSS in dB at a reference distance, usually 1m, away from the AP, n is the space 82 loss factor which varies in different environments, WAF is the wall attenuation factor and a is 83 the number of obstructions in between the receiver and AP,  $\varepsilon$  is a zero mean Gaussian distributed 84 noise. Positions can be obtained based on computing the change of signal strength from each AP 85 to the receiver. However, wireless signals are quite noisy due to interference and obstructions inside buildings. Therefore the actual observation  $\tilde{P}$  and the expected  $P_{RX}(d)$  from Eq.1 can 86 87 differ up to 20dB. Wi-Fi fingerprinting overcomes this problem by taking advantage of signal 88 disruptions in complicated environments. Although signals are easily disturbed and measurement 89 error  $\varepsilon$  is hard to predict, but as long as the building structure remains unchanged, the disturbance 90 reflected in the signal strength will remain alike in the same location. Therefore, the RSS 91 measurements from each AP form a pattern that reflects a specific location, known as fingerprints. 92 The first step of fingerprinting is the training phase, where a number of locations, known 93 as training points (TPs), are selected within the area of interest and the RSS from all APs are

94 measured at each TP. These are stored into a database as one fingerprint. If the RSS are carefully 95 measured, APs are well spread out and the structure of the building is complicated enough, each 96 fingerprint should be unique referring to one specific location in the building. During the 97 positioning phase, the user measures the current RSS and compares it to the fingerprints in the 98 database. Usually, the mean location of k fingerprints with the smallest difference to the current 99 RSS, known as the k-nearest neighbours (kNN), is returned as the estimated position [20].

100 The biggest problem with fingerprinting is that the training process requires a huge 101 amount of human labour, especially in large complex buildings. This increases the possibility of

- 102 human error and time cost. Moreover, the database needs to be retrained and updated each time
- 103 the infrastructure changes to maintain an up-to-date database for accurate positioning.

#### 104 2.2 Database training

Assuming that the RSS of a location is correlated to the RSS of a nearby TP based on Eq.1, GP is applied to enable faster and more efficient training, which makes the database easier to maintain and update.

For accurate database training, the selected TPs should cover the entire area of interest and RSS should be collected over a period of time on each TP to fully reflect the variance and stability of the signal from each AP. Each fingerprint is typically structured as  $\{(x_n, y_n)|RSS_{n1}, \sigma_{n1}, AP_1 \cdots, RSS_{nm}, \sigma_{nm}, AP_m\}$ , where  $(x_n, y_n)$  is the position of the *n*th fingerprint,  $RSS_{nm}$  is the mean RSS and  $\sigma_{nm}$  is the standard deviation of the *m*th AP at the *n*th TP, AP<sub>m</sub> is the unique identification of the AP, usually the MAC address. The uniqueness of the fingerprint is enhanced by the number of APs found and the amount of RSS collected.

115 As positioning is achieved by comparing RSS to the fingerprints, more fingerprints would 116 mean more detailed database, which potentially results in better positioning. The most 117 straightforward way of increasing fingerprints would be increasing TPs. However, it is almost 118 impossible to cover the entire floor plan with TPs due to the required amount of work. Therefore, 119 the entire area is usually divided into evenly distributed grids and a TP is placed at the centre 120 point of each grid, assuming that the RSS is the same within each grid. Typical grid sizes are 121 1m×1m, 2m×2m [21]. Smaller grids ensure a more detailed database, although it will be more 122 time consuming and laborious.

Based on the path loss model, we can see that the signal strength at each TP is correlated to its distance to the AP. In locations with fewer obstructions, the signals behave according to the model with a small noise. Hence less TPs are required as the RSS can be predicated from RSS at nearby TPs based on Eq.1. GP is applied to predict the RSS of fingerprints that are near to but not on TPs. In areas where training data has already been collected, GP increases the density of the 128 fingerprints without increasing the number of TP. If (x, z) are samples drawn from a noisy process 129 [22],

$$z_i = f(x_i) + \varepsilon \tag{2}$$

130 where each  $x_i$  is an input sample and  $z_i$  is the target or observation value,  $\varepsilon$  is assumed to be a 131 zero mean normally distributed noise. Gaussian process estimates the posterior distributions over 132 the functions f from the training data which is specified by a mean function m(x) and a 133 covariance function, or kernel k(x, x'), which describes the correlation between two input values 134  $x_p$  and  $x_q$ . The squared exponential kernel is applied here,

$$k(x_p, x_q) = \sigma_f^2 \exp\left(-\frac{1}{2\ell} |x_p - x_q|^2\right) + \sigma_n^2$$
(3)

135 where  $\sigma_f^2$  is the signal variance,  $\ell$  is the length scale that describes the strength of correlation over 136 a distance,  $\sigma_n^2$  is the Gaussian observation noise. The RSS measurements and the locations of TPs 137 are input into the system to train for the hyperparameters  $\theta = \langle \sigma_n^2, \ell, \sigma_f^2 \rangle$ , which define the 138 predication functions. The predication process is then carried out based on the predicative 139 distribution

$$p(z_*|x_*, X, Z) = \int p(z_*|f(x_*))p(f(x_*)|x_*, X, Z) \,\mathrm{d}f(x_*) \tag{4}$$

140 The locations of the TPs are input as X while the RSS measured at the TPs are the target values Z. 141 The desired fingerprints cover the building by 1m grids at locations  $x_*$ , and the RSS of the desired 142 fingerprints  $z_*$  is predicted based on the trained predication functions.

To understand the required density and location setup of training data for generating accurate fingerprint database, different training methods are compared. A Toshiba laptop is used throughout the trials in this paper for consistency, whose wireless adapter is Intel® Centrino® Advanced-N 6200. Four APs are located on Floor A of Nottingham Geospatial Building (NGB), each transmitting signals on both 2.4GHz and 5GHz frequencies. As signal characteristics are different, thus the signals from different frequencies will be treated separately. A full database consists eight MAC address groups, each denoted as AP1a (2.4GHz), AP1b (5GHz), AP2a, AP2b,

- 150 AP3a, AP3b, AP4a and AP4b respectively. Locations of the APs are indicated in Figure 1. Figure
- 151 2 shows the spread of the RSS for 2.4GHz and 5GHz signals during 30 minutes.



- 152
- 153

Figure 1 TPs selected for full fingerprint database

154 56 TPs are selected to cover the entire accessible area in NGB Floor A to establish the 155 ground truth for the fingerprint database, i.e. the best possible database from conventional training. 156 On average, two TPs are located inside a small office and four to six TPs are located in large 157 rooms. The laptop is placed at each location to collect the Wi-Fi RSS data for around thirty 158 minutes until at least 100 vectors from each of the four APs have been collected. The mean and 159 standard deviation of all the collected RSS from all APs at each TP is obtained and sorted into 160 the training input vector. GP is then applied based on the training data to increase the fingerprint 161 density to  $1m \times 1m$ . The resulting fingerprints are stored into a database, denoted as sDB. The 162 training data were collected while the receiver was static and placed over the TP to obtain more 163 stable information of the signal. Training for the 56 TPs takes around 37 hours in total.



165

Figure 2 Wi-Fi RSS fluctuation in 30 minutes

166 Two rooms are further selected to compare the database quality of different TP densities. 167 Another 24 and 32 TPs are selected respectively in two rooms, R1 and R2, so that the TP density 168 in the rooms are  $1m \times 1m$ . A local database is generated for each room based on this set of TPs. 169 R1 is a small meeting room with no obstructions and simple furnishing. R2 is a heavily obstructed 170 store room with metal shelves and electronic equipment.  $\Delta RSS$  is the difference between the RSS 171 of fingerprints from the two different databases at the same location. The difference for each AP 172 is listed in Table 1.

173 A larger  $\Delta RSS$  is seen in R2 which is the heavily obstructed room. Therefore, signals are 174 noisier and less predictable in such places. Hence more TPs are required to generate better 175 database. However, the difference for 5GHz signal is smaller. This is due to that it is less able to 176 penetrate obstructions and the signal pattern for different locations are more unique.

177

Table 1 Mean  $\Delta RSS$  between fingerprints generated from different TP density (dB)

	AP1		A	AP2		AP3		AP4	
	a	b	a	b	a	b	а	b	
R1	2.65	2.12	3.19	2.78	1.77	3.34	8.92	2.97	
R2	10.94	3.77	8.00	7.65	17.68	12.62	8.16	5.89	

#### 179 2.3 Collaborative training

As the indoor wireless environment can alter caused by changes in the wireless hardware, the building structure or even furnishings. Therefore, each fingerprint database must be maintained and updated. Re-training can still be laborious work even when GP is applied. Collaborative database training (cFPDB) is proposed here to save time and also enhance database quality.

cFPDB fundamentally relies on collaborative positioning between a number of mobile users to estimate the reference positions of the TPs and the relationship between the training data. First of all, the basic collaborative positioning algorithm in cFPDB is introduced. Collaborative positioning constrains the measurement error of users by applying a relative ranging constraint. The basic navigation is achieved from inertial measurements and propagated forward based on the dead reckoning model at each step,

$$\begin{bmatrix} \hat{x}_k \\ \hat{y}_k \end{bmatrix} = \begin{bmatrix} \hat{x}_{k-1} + \hat{s}_{(k|k-1)} \cos \hat{\theta}_{(k|k-1)} \\ \hat{y}_{k-1} + \hat{s}_{(k|k-1)} \sin \hat{\theta}_{(k|k-1)} \end{bmatrix}$$
(5)

where  $[\hat{x}_k, \hat{y}_k]$  is the user position at time k,  $\hat{s}_{(k|k-1)}$  is the estimated step length between time k - 1 and k,  $\hat{\theta}_{(k|k-1)}$  is the heading estimation during the step. Low-cost inertial system measurements tend to drift badly after initialisation [23,24]. Therefore, the inertial measurement errors need to be constrained by external measurements, e.g. relative ranging measurements between users that can be obtained from high precision wireless units such as Ultra-wideband sensors [17]. This estimated position serves as the reference location of the measured RSS, i.e. the location of the TPs during dynamic data collection.

Collaborative training is carried out during the collaborative positioning process and this means that only one RSS is collected at a specific location and no knowledge of the signal fluctuation pattern could be obtained initially. While the signal strength could vary up to 20dB or even more at any single location when the equipment is static, this may increase further when the receiver is moving. In conventional training, the fluctuation pattern is captured by extending the

204

training time over hours or even days. In cFPDB, extracting signal features obtained by different users passing previous TPs during different periods helps to capture this pattern.

205 The ranging measurement r between the users builds a link between the collected training 206 data. Two thresholds, the separation threshold  $\delta_{sep}$  and integration threshold  $\delta_{int}$ , is defined to 207 identify three different kinds of relationships between the data. If r is above  $\delta_{sep}$ , it would be regarded that the users were not in the same area of interest. Their training data will be stored 208 209 separately and used to generate individual databases. If r is within  $\delta_{sep}$  but above  $\delta_{int}$ , the 210 training data would be considered to be within the same area of interest and used to generate the 211 same database. As it is almost impossible for users to pass the exact same locations during 212 collaborative training, TP is expanded into training areas (TA). Any training data that are within a range of  $\delta_{int}$  would be regarded to reflect the signal pattern for the same TA and adjusted to 213 214 form fingerprints with knowledge of the signal variation for the TA. Figure 3 shows an indication 215 of the relationship between a TA and fingerprint location. The RSS of the fingerprint is assumed to represent the RSS for the entire TA. These thresholds are set according to the expected 216 217 correlation between fingerprints, which can be affected by the environment and the stableness of 218 the RSS in nearby training locations. This will not differ much in the same building, hence the 219 same threshold can be applied throughout.



221Figure 3 A fingerprint representing the training area (Green grid indicating a222training area, red circle is a fingerprint in the TA)

To build the collaborative database dynamically, four different trajectories, denoted as T1, 223 224 T2, T3 and T4, of varying length and locations within NGB Floor A are chosen where training 225 data will be collected during the collaborative positioning phase. Each user follows one of the 226 different routes and collects training data at each step, shown in Figure 4, where blue indicates 227 low RSS and red indicates high RSS values. The data collected along different trajectories are 228 combined to generate collaborative fingerprint databases using GP, denoted as cDB. 229 Collaborative training greatly extends the training data coverage and increases the amount of data 230 for each TA. cDB fingerprints are generated from more sufficient data and longer time span. 231 Hence captures the RSS fluctuation and environment disturbances.



232

233

(a) Collaborative training data of T1 (b) Collaborative training data of T2



234

235

236

(c) Collaborative training data of T3  $\,$  (d) Collaborative training data of T4  $\,$ 

### Figure 4 Collaborative training data from AP1a

237 The RSS of the training data along the dynamic trajectories are compared to the RSS of 238 the TPs of sDB by extracting those that are within a certain distance to the static TPs and 239 measuring the  $\Delta$ RSS. The RSS difference for distances from 1m up to 4m is listed in Table 2, 240 where  $\Delta$ RSS indicates the mean difference, Std dev. indicates the standard deviation of the RSS 241 difference. Signal acquisition is less stable while the receiver is moving thus it can be anticipated that the RSS of the dynamic training data are noisier and differs to that of the static TP RSS. From the data listed in Table 2, we can see that even though the  $\Delta$ RSS between training data is almost 10dB even at 1m, but stays within 15dB up to 3m, which is actually within the RSS fluctuation range. While the difference increases when the distance is 4m, the standard deviation actually drops. This indicates that correlation fails between the two points and their RSS is steadily different as they are too far apart. According to the  $\Delta$ RSS here, the integration threshold can be set to 2m or 3m depending on the environment, i.e. whether it is more like R1 or R2.

249

Table 2  $\Delta RSS$  between dynamic and static TPs (dB)

	1m	2m	3m	4m
ΔRSS	9.85	12.55	13.39	19.36
Std dev.	10.61	10.49	15.91	8.58

250

Three different cDBs are generated and their fingerprints are compared to those of sDB. cDB1 is generated from the training data along T1 and T2; cDB2 is generated from T1 and T3; cDB3 is generated from T1,T2,T3 and T4. Figure 5 plots the mean RSS difference between the fingerprints of each AP in cDB and sDB. As more data is used to generate the database, the fingerprint RSS of the cDB approaches that of the sDB.



#### Figure 5 $\triangle RSS$ between cDB and sDB (dB)

#### 258 **3 Fingerprinting based on confidence level**

#### 259 3.1 Fingerprint confidence factor

260 As dynamic training data contain large signal variances, they should be treated 261 appropriately when applied to generate databases. During the cFPDB process, the system keeps 262 track of all previously and currently collected training data by storing them along the timeline. 263 When new training data is picked up at a TA that has been trained previously, the mean of all 264 RSS from all history data is used as the RSS to generate the fingerprint in GP. The standard 265 deviation of RSS is computed to generate a confidence factor for the fingerprint at the location. 266 The confidence factor vector consists of two values, i.e. the training data difference level  $\Delta_{san}$ 267 and the confidence level  $\eta_{CF}$ .

268 The difference level  $\Delta_{sgn}$  is updated by measuring the sign of the  $\Delta$ RSS between the new 269 RSS at the TA and the RSS of the fingerprint representing the TA. A positive  $\Delta_{sgn}$  indicates that the RSS is increasing and negative  $\Delta_{sgn}$  indicates decreasing RSS. The confidence level  $\eta_{CF}$  for 270 271 each fingerprint in the entire building is generated from the training data standard deviation, 272 which indicates how much signal strength variance to expect at each specific fingerprint location. 273 Smaller  $\eta_{CF}$  means higher confidence with the current RSS of the fingerprint stored in the 274 database. There would not be enough training data to compute the standard deviation of 275 fingerprints at the beginning of a training process. Hence it is given a low  $\eta_{CF}$ , assuming that the 276 collected data is trustworthy. The confidence level only decreases when the amount of training 277 data accumulates and shows obvious fluctuation.

If the  $\Delta_{sgn}$  for a location is always positive or negative, it would be assumed that the RSS is constantly increasing or decreasing, indicating a possibility of permanent change in the Wi-Fi properties at the location. If the confidence level goes over a given threshold under such conditions, the RSS of the old fingerprint will be replaced with new RSS data.  $\eta_{CF}$  for the fingerprint is then reset to the initial value which represents a high confidence level. If  $\Delta_{sgn}$ changes randomly, the collected RSS is assumed to be within the signal strength random fluctuation range. In such cases, the signal fluctuation range is reflected by  $\eta_{CF}$ . In general, the difference level  $\Delta_{sgn}$  keeps track of the direction of change of the fingerprint along the time axis while the confidence level  $\eta_{CF}$  reflects the expected signal fluctuation at the location, giving the user an updated knowledge of how trustworthy the fingerprints are. The procedure of generating the confidence factor is shown in Figure 6.





#### 290 Figure 6 Flowchart for collaborative fingerprint database training

#### 291 3.2 Collaborative Wi-Fi fingerprinting

In conventional fingerprinting, potential fingerprints are usually found by defining a variance boundary  $\tau_{FP}$  first. If  $RSS_P$  is collected at an unknown location P and  $RSS_{FP}$  are the fingerprints from the database, any fingerprints from the database that fit within

$$RSS_P - \tau_{FP} < RSS_{FP} < RSS_P + \tau_{FP} \tag{6}$$

are extracted as potential fingerprints. However, deciding the value of  $\tau_{FP}$  can be difficult. If the given  $\tau_{FP}$  is underestimated, there is a possibility that no potential fingerprints will be extracted if either  $RSS_P$  or  $RSS_{FP}$  is noisy. Yet if  $\tau_{FP}$  is overestimated, too many potential fingerprints may be found, introducing large location ambiguities.

Fingerprints generated from the cFPDB process take the form of  $\{(x_i, y_i) | AP_1, (RSS_1, \eta_{CF_1}, \Delta_{sgn}) \cdots, AP_n, (RSS_n, \eta_{CF_n}, \Delta_{sgn})\}$ . The confidence factor generated during the training process is used here to help decide the value of  $\tau_{FP}$ , as below

$$\tau_{FP} = a \cdot \eta_{CF} \tag{7}$$

where *a* is a coefficient defining the relationship between the two values. It is adjusted from 1.5 to 3 until potential fingerprints are found. From examining the trial data, it has been found that we might choose a = 1.5 in open areas and a = 3 in heavily obstructed areas.

As the database training is carried out during a collaborative positioning phase, the collaborative measurements can also be applied in fingerprinting when available. Hence the Wi-Fi adaptive collaborative fingerprinting algorithm (WARCP) is proposed. The steps of WARCP are given as below:

## 309 1. Each user propagates based on the DR model as in Eq.5;

At each step, user *i* takes a set of Wi-Fi RSS measurement *RSS<sup>i</sup>*<sub>P</sub> from each AP, if more than one user is found, ranging measurements *r<sub>ij</sub>* are also obtained between user *i* and *j*; *RSS<sup>i</sup>*<sub>P</sub> is stored to update the database; fingerprinting is then performed by considering both the confidence factor and the distance between the potential fingerprints following Eq.6. When *M* and *N* potential fingerprints are found for user *i* and *j*, the distance between pairs of potential fingerprints are measured,

$$dist_{FP} = \sqrt{(x_{FP_m} - x_{FP_n})^2 + (y_{FP_m} - y_{FP_n})^2}, (m = 1, 2, \cdots, M; n = 1, 2, \cdots, N)$$
(8)

316 Fingerprints that obey Eq.9 will remain as potential fingerprints,

$$\left| dist_{FP} - r_{ij} \right| \le \varepsilon_{range} \tag{9}$$

317 Where  $\varepsilon_{range}$  is defined based on the expected noise of the ranging measurement.

3184. Position estimations are obtained from the weighted average of the remaining potential319fingerprint positions.

Fingerprinting reliability is improved here as potential fingerprints are selected according to Eq.6 where  $\tau_{FP}$  changes adaptively. Therefore, a fingerprint with high confidence level, i.e. small  $\eta_{CF}$ , would also be given a small  $\tau_{FP}$ . It would not be chosen as a potential fingerprint unless its  $RSS_{FP}$  is reliable and close to  $RSS_P$ . If a fingerprint's confidence level is low, its possibility of being selected as potential fingerprint is increased as the range of  $RSS_{FP} \pm \tau_{FP}$  is larger. This is to decrease its possibility of being discarded when it differs from  $RSS_P$  due to fluctuation, but its location is actually close to the true location.

### 327 4 Simulations and trials

## 328 4.1 Dynamic training

To examine how data is integrated to update the database, the training data of T4 is collected in two rounds. The first round in the building is part 1(P1) and the second round part 2 (P2). The training data of P1 and P2 are used to generate two individual databases, P1-DB and P2-DB. The combination of P1 and P2 is used to train for another database, T4-DB. The difference in dB for the fingerprints of each database and the fingerprints of sDB is measured and plotted in Figure 7.





337

## Figure 7 $\triangle RSS$ between T4-DB and sDB

338 In most cases, the difference between the dynamic trained database and sDB is reduced 339 when P1 and P2 data is combined together. The difference is further reduced when more data is 340 integrated with T4 to form cDB3. The difference between the fingerprints generated from P1 and 341 P2 is indicated by green circles. However, there are still instances when the difference of RSS is 342 continuously different from each other, resulting in a large difference from sDB, e.g. AP2b. 343 Another instance is AP1b, where the difference between P1 and P2 is not very large, but because 344 both are very different from sDB, their combined data still results in a large bias, as indicated by 345 the yellow  $\nabla$ . During the training process itself, it is hard to decide which data is biased or not. 346 Hence we can only record the variance of all collected data and indicate its likelihood of being at 347 a certain signal strength level.

To build up the collaborative database cDB3, new training data is stored and compared to old data iteratively. Each time a new data is collected at a TA where data has been collected previously, the variance of the signal strength is measured and applied to generate the confidence factor as described in Section 3.1. Figure 8 plots the confidence level of AP3 for Floor A that is derived by updating the database from T4-P1 with T4-P2, T1 and T2. Blue areas indicate a small  $\eta_{CF}$ , i.e. high confidence in the fingerprint, and red areas vice versa.



## 355

356

## (a) Confidence level map for AP3a (b) Confidence level map for AP3b Figure 8 Fingerprint confidence level map

357 The RSS of the training data is also plotted on the map for reference. The resulting  $\eta_{CF}$  is 358 higher in areas where the training data changes rapidly between each training. Furthermore, the 359 confidence indicator for 5GHz signals is smaller in general than that of 2.4GHz signals. The RSS 360 of the 5GHz signals remain relatively stable for different regions of the building hence the 361 fingerprint pattern is more unique, producing lower $\eta_{CF}$ . 2.4GHz signals, on the other hand, have 362 greater ranging distance and penetrate walls better. However, this causes noisier training data and 363 higher  $\eta_{CF}$ .

364 Three different  $\tau_{FP}$  are chosen to extract potential fingerprints based on the RSS 365 measurements observed at a given location. Figure 9 shows the potential fingerprints extracted 366 when different thresholds are given,  $\tau_{FP} = 5$ ,  $\tau_{FP} = 10$  or  $\tau_{FP} = a \cdot \eta_{CF}$ . As shown in the figure, 367 too many fingerprints are extracted when  $\tau_{FP} = 10$ . Even though the fingerprints close to the true 368 location are extracted, but those that are almost 10m away are also considered as potentials. When  $\tau_{FP} = 5$ , not enough fingerprints that are close to the true location are found. When  $\tau_{FP}$  is set 369 370 adaptively according to  $\eta_{CF}$ , the potential fingerprints are more suitable as all extracted 371 fingerprints are located near the true position. Table 3 lists the average distance from selected 372 potential fingerprints to the true location throughout a whole trajectory when given different  $\tau_{FP}$ 373 and also comparing the results for different frequencies. While 5GHz signal fingerprints are 374 slightly closer to the true location, the best result is still achieved when both frequencies are used.



#### 379 4.2 Collaborative Fingerprint positioning

To evaluate the performance of the fingerprinting method based on the improved 380 381 fingerprint database, an indoor positioning trial is carried out in NGB with two rovers starting 382 from the same point, indicated by the red star in Figure 10. Both rovers wear a foot-mounted IMU 383 to obtain inertial measurements and carry a laptop to collect the Wi-Fi RSS. Relative ranging 384 measurements are simulated based on the indoor performance of UWB units so that the mean is 385 the true distance with a standard deviation of 3m. To enhance the effectiveness of the constraint 386 provided by relative ranging, both rovers start at the same place but travel in different directions 387 so they do not follow each other.





399(a) Positioning error cdf of Rover 1(b) Positioning error cdf of Rover 2400Figure 11 Positioning error cdf for each different algorithm



402 The mean positioning error of each algorithm for both rovers is listed in Table 4, where 403 DR/Wi-Fi indicates results of DR integrated with conventional Wi-Fi fingerprinting, DR/Wi-Fi 404 (cf) indicates results of DR integrated with improved Wi-Fi fingerprinting with confidence factor. 405 Although four APs used for Wi-Fi positioning here seems like a very sparse network for 406 fingerprinting, but the realistic situation is that inside most buildings, the number of APs are only 407 setup to ensure network coverage and not fitted to meet the density requirement for fingerprinting. 408 Results indicate that the accuracy of fingerprinting with confidence factor improves by 36% for 409 Rover 1 and 50% for Rover 2 compared to conventional fingerprinting with very few available 410 APs. The performance of integrating DR with improved Wi-Fi fingerprinting is similar to that of 411 DR integrated with ranging. While Wi-Fi fingerprinting gave slightly better results for Rover 1, 412 ranging results were better for Rover 2. Wi-Fi fingerprinting results are not given alone because 413 with only four APs, the performance is very unstable. Although accuracy can be quite good during 414 some periods, but the robustness is too low for comparison. Furthermore, as Table 4 shows, 415 integrating improved Wi-Fi with DR enhance accuracy compared to conventional Wi-Fi with DR. 416 This clearly shows the improvements on fingerprinting with the improved method.

417

Table 4 Positioning error of different algorithms (m)

	DR/Wi-Fi	DR/Wi-Fi(cf)	DR/Ranging	WARCP
Rover 1	3.67	2.32	2.35	1.76
Rover 2	5.05	2.40	1.76	1.47

418

420 The integration of DR, fingerprinting and ranging, i.e. WARCP, improves positioning 421 accuracy further as the mean error is reduced by around 10% compared to DR/Wi-Fi and 422 DR/ranging integration method. Although the average performance difference between the three 423 integrations is not immense, but the maximum error reduces by 35% when WARCP is 424 implemented. As the accuracy level when integrating DR with Wi-Fi and ranging was around the 425 same level, thus the improvement seen here is once more improvement on fingerprinting. While 426 this study focus on the overall improvement of positioning accuracy when integrating low 427 accuracy inertial measurement with Wi-Fi fingerprinting and collaborative ranging, this system 428 shows more freedom of choosing the appropriate positioning algorithm based on what sensor and 429 measurement is available.

430 As the indoor environment is complicated and prone to change, the WARCP algorithm 431 allows users start its own navigation by using only the inertial measurement from a mobile device 432 (i.e. DR). As more nearby users are found, DR/Ranging can be applied to enhance positioning 433 accuracy by constraining the inertial drift. While collaborative positioning is performed, users 434 can help to train for a Wi-Fi fingerprint database through the collaborative training process which 435 can be stored on to a central server and shared to all local users. When users lose the relative 436 constraint from nearby users, collaborative positioning can no longer be applied, but 437 fingerprinting can be performed when the database is available, which achieves almost the same 438 accuracy level as collaborative positioning. When both Wi-Fi signals and relative ranging 439 measurements are available to the user, the collaborative fingerprinting phase of WARCP can be 440 applied to obtain positioning estimation, which is the most accurate.

WARCP allows the system to search for different choice of integration when different signals and measurements are available. Integrating Wi-Fi and relative ranging with inertial navigation not only enhance the positioning accuracy, but most importantly, the positioning robustness is improved as positioning estimation can be obtained even in changing environments where signals are intermittently available.

#### 446 **5** Conclusions

447 Wi-Fi fingerprinting is a popular method for indoor positioning as Wi-Fi signals are 448 widely available in most urban areas and infrastructures are already well established. However, 449 Wi-Fi signals are not positioning dedicated, hence suffer instability and disturbance from the 450 changing environment and obstructions, which can cause instability in positioning accuracy. 451 Inertial measurements from mobile devices are useful to indoor positioning users as they are 452 available regardless of the environment. But due to the large heading drift of low-cost inertial 453 sensors, errors must be constrained by external measurements to achieve reasonable positioning. 454 This paper presents an improved Wi-Fi fingerprinting method for both phases of fingerprinting 455 which is fundamentally based on indoor pedestrian inertial navigation but also enhance inertial 456 navigation performance, as described in cFPDB and WARCP.

457 During the training phase, cFPDB is applied where the RSS measurements from a number 458 of mobile users are collected during different periods and different locations. Measurements are 459 either sorted into training data for different TAs depending on the distance between the location 460 of the data, where fingerprints are generated and updated based on the data of each TA. When 461 updated with new measurements, the fingerprints in the database are given a confidence factor 462 which indicates both the long-term change direction of the RSS and the expected short-term signal 463 fluctuation at the location of the fingerprint. The positioning procedure is then carried out using 464 WARCP, which allows the system to navigate using available measurement flexibly. Potential 465 fingerprints are extracted based on the confidence factor associated with each fingerprint and then 466 further selected by ranging measurements when available.

With the proposed methods, fingerprint database can be setup during the positioning phase when users enter an environment without prior database. Previous databases can also be updated by gathering information from surrounding users. Both history data and new data are applied to update the database so users not only know the current RSS of the fingerprints but also have an idea of how much signal variance to expect at each location. Therefore, during the positioning phase, fingerprints are selected based on whether the current measurement lies withinthe RSS range of the fingerprint.

Fingerprint based positioning is improved by 40% compared to conventional fingerprinting when the confidence factor is considered. By applying WARCP, which includes ranging measurement constraint, positioning error is further reduced, especially the maximum error which is reduced by 35%. The application of integrating ranging measurements with fingerprinting during training and positioning gives the user more freedom of choosing positioning algorithms based on what information is available. The training effort of fingerprint database is also greatly reduced as training data can be obtained from crowdsourcing.

481 **Reference** 

- 482 [1] D. Niculescu, B. Nath, Position and orientation in ad hoc networks, Ad Hoc Networks 2
  483 (2) (2004) 133–151. doi: 10.1016/S1570-8705(03)00051-9
- 484 [2] R. Harle, A Survey of Indoor Inertial Positioning Systems for Pedestrians, IEEE
  485 Communications Surveys and Tutorials 15 (3) (2013) 1281–1293.
  486 doi:10.1109/SURV.2012.121912.00075.
- 487 [3] F. Evennou, F. Marx, Advanced Integration of WiFi and Inertial Navigation Systems for
  488 Indoor Mobile Positioning, EURASIP Journal on Advances in Signal Processing 2006
  489 (2006) 1–12. doi:10.1155/ASP/2006/86706.
- 490 [4] E. Mok, B. Cheung, An Improved Neural Network Training Algorithm for Wi-Fi
  491 Fingerprinting Positioning, ISPRS International Journal of GeoInformation 2 (3) (2013)
  492 854–868. doi:10.3390/ijgi2030854.
- 493 [5] K. Kaemarungsi, P. Krishnamurthy, Analysis of WLANs received signal strength
  494 indication for indoor location fingerprinting, Pervasive and Mobile Computing 8 (2) (2012)
  495 292–316. doi:10.1016/j.pmcj.2011.09.003.
- 496 [6] E. Mok, G. Retscher, Location determination using WiFi fingerprinting versus WiFi
  497 trilateration, Journal of Location Based Services 1 (2) (2007) 145–159.
  498 doi:10.1080/17489720701781905.
- H. Leppakoski, S. Tikkinen, J. Takala, Optimizing radio map for WLAN fingerprinting, in:
  Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS), 14-15
  October 2010, IEEE, Kirkkonummi, 2010, pp. 1–8. doi:10.1109/UPINLBS.2010.5654332.

- 502 [8] S. Bak, S. Jeon, Y.-J. Suh, C. Yu, D. Han, Characteristics of a large-scale WiFi radiomap 503 in indoor localization, IEEE. and their implications 2013, 1 - 5. pp. 504 doi:10.1109/NOF.2013.6724506.
- 505 [9] B. Viel, M. Asplund, Why is Fingerprint-based indoor localization still so hard?, in:
  506 Pervasive Computing and Communications Workshops (PERCOM Workshops), 2014
  507 IEEE International Conference on, 24-28 March 2014, Budapest, 2014.
- 508 [10] B. Ferris, D. Hahnel, D. Fox, Gaussian Processes for Signal Strength-Based Location
  509 Estimation, in: In Proc. of Robotics Science and Systems, 2006, Philadelphia, Pennsylvania,
  510 2006.
- [11] R. Faragher, R. Harle, SmartSLAM An Efficient Smartphone Indoor Positioning System
  Exploiting Machine Learning and Opportunistic Sensing, in: ION GNSS+ 2013,
  Proceedings of the 26th International Technical Meeting of The Satellite Division of the
  Institute of Navigation, 16-20 September 2013, ION, Nashville, TN, 2013, pp. 1006–1019.
- P. Robertson, M. G. Puyol, M. Angermann, Collaborative Pedestrian Mapping of Buildings
  Using Inertial Sensors and FootSLAM, in: Proceedings of the 24th International Technical
  Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2011),
  September 20 23, 2011, Portland, OR, 2011, pp. 1366 1377.
- 519 [13] J. Krumm, J. Platt. Minimizing Calibration Effort for An Indoor 802.11 Device Location
  520 Measurement System, Microsoft Research Technical Report, November, 2003.
- 521 [14] Y. Zhuang, Z. Syed, J. Georgy, N. El-Sheimy, Autonomous smartphone-based WiFi
  522 positioning system by using access points localization and crowdsourcing, Pervasive and
  523 Mobile Computing. doi:10.1016/j.pmcj.2015.02.001.
- 524 [15] Y. Zhuang, Z. Shen, Z. Syed, J. Georgy, H. Syed, N. El-Sheimy. Autonomous WLAN 525 Heading and Position for Smartphones, in: Position, Location and Navigation Symposium 526 PLANS 2014. 5-8 May 2014, IEEE/ION, pp.1113-1121. doi: 527 10.1109/PLANS.2014.6851481.
- L.-W. Chan, J.-R. Chiang, Y.-C. Chen, C.-N. Ke, J. Hsu, H.-H. Chu, Collaborative
  Localization: Enhancing WiFi-Based Position Estimation with Neighborhood Links in
  Clusters, in: Pervasive Computing: Proceedings of 4th International Conference, Pervasive,
  Dublin, Ireland, 7-10 May 2006, Vol. 3968, Springer Berlin Heidelberg, Berlin, Heidelberg,
  2006, pp. 50–66.

- 533 [17] H. Jing, J. Pinchin, C. Hill, T. Moore, A novel weighting approach to collaborative indoor
  534 positioning, in: Proceeding of European Navigation Conference (ENC-GNSS 2014),
  535 Rotterdam, 2014.
- J.-G. Park, B. Charrow, D. Curtis, J. Battat, E. Minkov, J. Hicks, S. Teller, J. Ledlie.
  Growing An Organic Indoor Location System, in: Proceedings of the 8th International
  Conference on Mobile Systems, Applications, and Services (MobiSys '10), June 2010. San
  Francisco, CA, pp. 271-284. doi: 10.1145/1814433.1814461.
- P. Bolliger, K. Partridge, M. Chu, M. Langheinrich. Improving Location Fingerprinting
  Through Motion Detection and Asynchronous Interval Labeling, in: Location and Context
  Awareness, 4th International Symposium, LoCA 2009, Tokyo, Japan, May 7-8, 2009,
  Springer Berlin Heidelberg, pp. 37-51. doi: 10.1007/978-3-642-01721-6\_3.
- 544 [20] B. Shin, J. H. Lee, T. Lee, H. S. Kim, Enhanced weighted K-nearest neighbor algorithm
  545 for indoor Wi-Fi positioning systems, in: Computing Technology and Information
  546 Management (ICCM), 2012 8th International Conference on, Vol. 2, 2012, pp. 574–577.
- 547 [21] W. Liu, Y. Chen, Y. Xiong, L. Sun, H. Zhu, Optimization of Sampling Cell Size for
  548 Fingerprint Positioning, International Journal of Distributed Sensor Networks 2014 (2014)
  549 1–6. doi:10.1155/2014/273801.
- 550 [22] C. Rasmussen, C. Williams, Gaussian processes for machine learning, Adaptive
  551 Computation and Machine Learning, MIT Press, Cambridge, Mass, 2006.
- F. Weimann, G. Abwerzger, B. Hofmann-Wellenhof, A Pedestrian Navigation System for
  Urban and Indoor Environments, in: Proceedings of the 20th International Technical
  Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2007),
  September 25 28, 2007, ION, Fort Worth, TX, 2007, pp. 1380 1389.
- K. Abdulrahim, C. Hide, T. Moore, C. Hill, Aiding Low Cost Inertial Navigation with
  Building Heading for Pedestrian Navigation, Journal of Navigation 64 (02) (2011) 219–
  233. doi:10.1017/S0373463310000573
- 559

## 561 Tables

	AI	P1	A	P2	A	P3	A	P4
	a	b	а	b	а	b	a	t
R1	2.65	2.12	3.19	2.78	1.77	3.34	8.92	2.9
R2	10.94	3.77	8.00	7.65	17.68	12.62	8.16	5.8

		1m	2m	3m	4m
-	ΔRSS	9.85	12.55	13.39	19.36
	Std dev.	10.61	10.49	15.91	8.58

Table 2 ARSS bets dv mic and static TPs  $(d\mathbf{R})$ 

500	Table 5 T ingerprinting error for different $t_F p$ (iii)						
-	τ	5dB	5dB 10dB	$a\cdot\eta_{\mathit{CF}}$			
	€ <sub>F</sub> p	500	Toub	Dual	2.4GHz	5GHz	
-	Error	16.48	15.51	9.07	11.37	9.69	
569							
570							

## Table 3 Fingerprinting error for different $\tau_{FP}$ (m)

	DR/Wi-Fi	DR/Wi-Fi(cf)	DR/Ranging	WARC
Rover 1	3.67	2.32	2.35	1.76
Rover 2	5.05	2.40	1.76	1.47