Comparison of Different Vector Distance Measure Calculation Variants for Indoor Location Fingerprinting

Guenther Retscher, Julian Joksch

Department of Geodesy and Geoinformation, Research Group Engineering Geodesy, TU Wien, Vienna, Austria

Abstract.

The study-at-hand discusses Wi-Fi location fingerprinting in an indoor environment. Wi-Fi is a predestinated signal-of-opportunity which can be used for positioning of a mobile user as most devices nowadays incorporate a Wi-Fi card and it is available in many buildings and public spaces. For the determination of the user location in the fingerprinting method signal strength observations are carried out in two phases. In the first training phase signal strength measurements from the visible Wi-Fi Access Points are collected to build-up a fingerprint database. In the following positioning phase, a user can be located and tracked if he carries out similar measurements and compares them with the values in the fingerprinting database. For the matching a distance criterion is applied to obtain the best estimation of the users' location. In analytical form the use of nine different vector distances for such an approach is investigated. The selected distances included the Manhattan, Euclidean, Chebyshev, Canberra, Cosine, Sorensen, Hellinger, Chi-square and Jeffrey vector distance. In the test bed in an office environment four multiple-SSID (Service Set Identification) Wi-Fi networks existed at a physical single Access Point location. From the results in this investigation it could be seen that not the use of all signal strength measurements yields to a better positioning solution but the measurements to one network out of the four provides a better performance. The achievable positioning accuracies depend mainly on the selection of the vector distance and matching algorithm. Furthermore, the Access Point architecture and configuration are determinant factors. In most tests in the selected office environment the Cosine vector distance provided the overall best performance followed by the Euclidean and Hellinger distance. Only with the Chebyshev distance significantly larger positioning errors occurred. In average a minimum mean distance error of

LBS 2016
13th Conference on Location-Based Services
Vienna, 14–16 November 2016

Published in "Proceedings of the 13th International Conference on Location-Based Services", edited by Georg Gartner and Haosheng Huang, LBS 2016, 14-16 November 2016, Vienna, Austria.

around 1.4 m could be achieved when using a single network in a multiple-SSID configuration.

Keywords. Location fingerprinting, matching algorithms, vector distance measures

1. Introduction

The number of mobile applications for the processing of location data increases continuously more and more. Since smartphones can receive and process GNSS and Wi-Fi signals the demands on availability, reliability and accuracy for positioning has also increased. Outdoors the accuracy can reach a view meters, but indoors it is usually much lower. Because GNSS signals can only be attenuated received in buildings, other systems had to be developed, which use the same hardware, the cellular phone. Using an existing Wi-Fi technology, an indoor absolute positioning system can be built with little effort and low costs.

In this analytical study the use of the location fingerprinting method is evaluated. A test area in an office environment with a regular grid of reference points was built therefore. At each reference point (RP) the so-called fingerprint, the unique signature of the received signal strength (RSS) of surrounding Wi-Fi Access Points (APs), was measured and stored in a training database. In the next step, fingerprints of off-grid test points TPs were taken in the positioning phase. By the similarity of the fingerprints of test and reference points the position of the test points can be obtained, by using several vector distances and matching methods. In this study deterministic fingerprinting algorithms based on the nearest neighbour (NN), K-nearest neighbour (KNN) and K-weighted nearest neighbour (KWNN) matching algorithms are investigated. The estimated position is then the position of the fingerprint with the minimum vector distance *VD*. Most commonly the Euclidean distance is employed for this task which is calculated for each AP in the positioning phase from the fingerprinting database values obtained in the training phase. The Euclidean distance is a special case of the more universal Minkowski distance. Apart from this vector distances the use of the Manhattan, Chebyshev, Canberra, Cosine, Sorensen, Hellinger, Chi-Square and Jeffrey distance are assessed. The resulting positioning performance of these nine different calculation variants is analyzed and compared in detail. In the field test site RSS measurements to six APs of four different multiple Wi-Fi networks (i.e., multiple-SSID (Service Set Identification) which offer different MAC addresses at a single physical AP) and a combined database of all networks are available for comparison. Thereby four different user orientations were measured in the training phase and two in the positioning phase describing the possible movement directions of the user.

The paper is organized as follows: Firstly, the basics of location fingerprinting are examined and summarized in section 2. It is discussed that fingerprinting is a so-called feature-based localization technology. In section 3 the fingerprinting matching approaches are evaluated and then in section 4 the suitable vector distance *VD* calculation variants. The field campaign set-up and test bed are presented in section 5 followed by a detailed assessment of the achieved positioning results in section 6. Finally, concluding remarks and an outlook on future work are given in section 7.

2. Basics of Wi-Fi Location Fingerprinting

Location fingerprinting is a feature-based positioning method. This term was introduced by Niedermayr & Wieser in 2012 to describe that any type of spatially varying features can be used for positioning. In contrast to common localization methods where usually distances, distance differences or angles are measured so that the coordinates can easily be computed using analytical geometry, the position is obtained by comparison of measured location-dependent features with given reference values associated with specific positions. According to Niedermayr & Wieser (2012) the features need to fulfil the following requirements: (1) the signal field varies significantly with varying location (the spatial gradient should be high), (2) the field is constant in time or its temporal variation is predictable, and (3) the feature corresponding to the field is observable and can be uniquely quantified. The major advantage of this type of positioning technology is that they do not require an unobstructed line-of-sight (LOS) between the mobile user and known reference points (or satellites, in the case of GNSS). Figure 1 illustrates the concept of feature-based positioning.

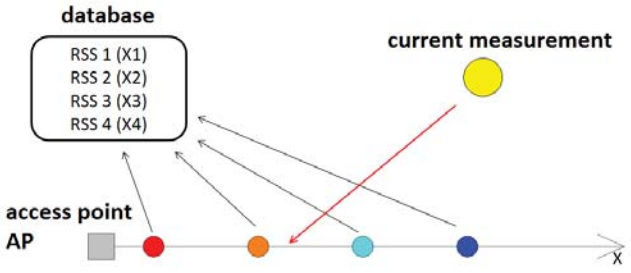

Figure 1. Operational principle of feature-based positioning

Within this positioning procedure a feature is selected, such as the received signal strength (RSS) of transmitters. Using a single RSS measurement, how-

ever, will usually not yield a unique solution. Thus, when using Wi-Fi for instance, the signal power of several Wi-Fi transmitters, i.e., the hotspots or Access Points (APs), has to be measured at the mobile client. From the comparison of the measured RSS and the reference RSS stored in a database the position of the user can be estimated. The reference values can either be derived in advance from georeferenced measurements taken during a mapping phase and stored in a database, or from numerical models. Commonly reference measurements are performed on a number of reference points (RPs) distributed throughout the area of interest in a so-called training or off-line phase instead of the use of simulated models (which yield usually in much lower positioning accuracies). A database of RSS values measured on all RPs is built-up during this phase. The requirement is that the location of the RPs has to be determined in a local (e.g. related to the building in the case of indoor positioning) or global coordinate system (usually required for outdoor positioning). In the positioning or on-line phase, simply speaking, then the current location of the user is obtained by matching of the on-line measured RSS with the values in the RSS database. In other words, with the feature then the user's location is estimated by matching measurements with the closest predetermined location fingerprints included in so-called fingerprinting or radio maps (Niedermayr & Wieser, 2012; Retscher, 2016). The matching methods are described in the following section 3.

Fingerprinting was firstly employed for Wi-Fi (or WLAN) positioning with the system RADAR in 2000 (Bahl and Padmanabhan, 2000; Kjærgaard, 2008). It is more robust to environmental effects on the RSS than using the RSS-based lateration algorithm. This is because the location fingerprinting algorithm constructs a search space according to the previously-measured RSS distributions in the radio maps. The advantage of constructing a fingerprinting database is that it can be used to consider a great number of detrimental effects from the surrounding environment, such as reflections and obstructions, into the radio maps and thus increases the accuracy for finding the best matching position based on RSS in the positioning phase (Retscher et al., 2012). Figure 2 shows an example for a radio map of a Wi-Fi AP derived from the training measurements using a smartphone in the test site (compare Figure 5).

The main problems in fingerprinting, however, are that the construction of a fine radio map leads to high workload and heterogeneous mobile devices measure RSS differently. Spatial interpolation techniques are usually employed for densification of the radio map and different devices are used for RSS measurements in the training phase to form a joint database (Retscher, 2016).

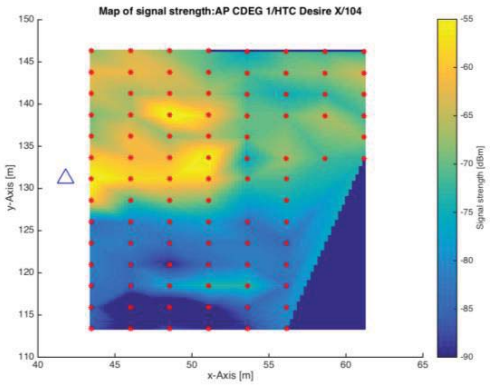

Figure 2. Example of a radio map of a single Wi-Fi Access Point (AP) in the class room of the test site (after Retscher & Roth, 2016)

3. Matching Approaches

The matching methods are classifications which define either which reference RSS values are incorporated into the position of the test points TPs ($\hat{l}_t$) – with weighting where appropriate – or which are rejected. Figure 3 illustrates the principle idea for the definition of the distance relationship between the DB of the training phase and the positioning phase (Retscher & Hofer, 2016). In the simplified case shown here RSS scans are measured to three APs (AP 1, 2 and 3) from two test points TP 1 and 2. The allocation of the positioning scans with the fingerprints in the DB is specified regarding to their corresponding minimum distance. In other words, the one that has the minimum vector distance *VD* is determined as the estimated location according to the selected distance of each training location. In the shown case then the scan is allocated to TP 1.

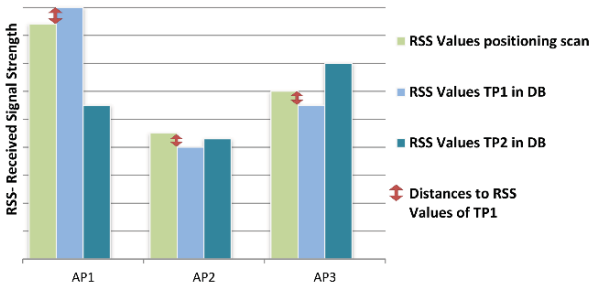

Figure 3. Allocation of RSS scans in the positioning phase to the training fingerprinting database DB (after Retscher & Hofer, 2016)

In this study only deterministic fingerprinting algorithms based on the nearest neighbour (NN), k-nearest neighbour (KNN) and k-weighted nearest neighbour (KWNN) matching algorithms are investigated. These are the most commonly employed algorithms and have been introduced by Bahl and Padmanabhan (2000). Their principle is briefly described in the following.

3.1. Nearest Neighbour (NN)

In this algorithm the distance vector d_t which contains the respective vector distances to all RPs $(d_1, d_2, d_3, \dots, d_R)$ is defined. Afterwards the minimum is determined and the coordinates are assigned to the TP as given in the following mathematical relationship:

$$d_t = \begin{bmatrix} d_{r=1} \\ d_{r=2} \\ d_{r=3} \\ \vdots \\ d_{r=R} \end{bmatrix} \rightarrow \min(d_t) \rightarrow r \tag{1}$$

with $\hat{l}_t = l_r$ where l_r is the coordinate vector of RP_r.

3.2. K-Nearest Neighbour (KNN)

In case of the KNN method a weighting of the nearest neighbours is performed. Thereby the weights are evenly split around the K RPs to be used and afterwards the coordinates l_r are multiplied by the respective weighting factor w_r and are divided by the sum of all w. The weighting factor w_r is the weighting from RP_r described in the form:

$$w_r = \frac{1}{K}, \; K \geq 2 \tag{2}$$

with

$$\hat{l}_t = \sum_{r=1}^{K} \frac{w_r}{\sum_{r=1}^{K} w_r} \cdot l_r. \tag{3}$$

Huang (2014) discusses that the accuracy increases if a K value of up to 10 in maximum is used and then it decreases again. That's why the calculation was carried out in this work with K values up to 10. The empirical determination revealed that $k = 3$ led to the best result (see Figure 7 in section 6.2).

3.3. K-Weighted Nearest Neighbour (KWNN)

For this method a weighting w_r is calculated in dependence of the respective vector distance $VD(s_t, s_r)$:

$$w_r = \frac{1}{|VD(s_t, s_r)|} \tag{4}$$

where s_r is the fingerprint from RP_r and s_t is the fingerprint from TP_t.

Similar as in the KNN method $\hat{l}_t$ is calculated from equation (3) to obtain the coordinate vector.

3.4. Positioning Accuracy

The most important aspect in the analysis and assessment of a localization method is the achievable positioning accuracy. Apart from the requirement to achieve acceptable positioning accuracies, of course also the performance and the costs have to be considered. In the following, it is briefly summarized how the positioning accuracies in the analyses in this study are defined and characterised.

Usually, the Mean Square Error MSE is employed to describe the achievable positioning accuracy. Universal is valid:

$$MSE(\hat{l}_t) = \left|(\hat{l}_t - l_t)^2\right| \tag{5}$$

with $\hat{l}_t = \begin{bmatrix}\hat{x}_t \\ \hat{y}_t\end{bmatrix}$.

From it follows:

$$MSE(\hat{x}_t, \hat{y}_t) = (\hat{x}_t - x_t)^2 + (\hat{y}_t - y_t)^2 \tag{6}$$

and further of the Root Mean Square Error (RMSE):

$$RMSE(\hat{x}_t, \hat{y}_t) = \sqrt{(\hat{x}_t - x_t)^2 + (\hat{y}_t - y_t)^2}. \tag{7}$$

Because it concerns, in this context, the distance error of the ascertained coordinates of the $\hat{l}_t$ of the corresponding test points TPs respective their true coordinates l_t, the error definition $RMSE(\hat{x}_t, \hat{y}_t)$ becomes – as in preceding literature (see e.g. Moghtadaiee & Dempster, 2015) – the distance error *DE* at further down as:

$$DE(TP_t) = \sqrt{(\hat{x}_t - x_t)^2 + (\hat{y}_t - y_t)^2}. \tag{8}$$

To obtain the mean distance error (short: MDE) an average over all TPs is calculated as given in equation (9):

$$MDE(\hat{l}) = \frac{1}{T} \cdot \sum_{t=1}^{T} DE(TP_t). \tag{9}$$

With equations (5) up to (9) all error measures are defined which are used in the analyses in this study. In the following, the application of the matching

approaches in combination with different vector distances *VD* is discussed in more detail.

4. Vector Distance Calculation Variants

The following derivations of the different vector distances *VD* are based on the similarity relation of vectors. For the derivation of the different *VD*s the RSS vector s of a measurement point is described in its universal form:

$$s = \begin{bmatrix} RSS_{\mathrm{p}=1} \\ RSS_{\mathrm{p}=2} \\ RSS_{\mathrm{p}=3} \\ \vdots \\ RSS_{\mathrm{p}=\mathrm{P}} \end{bmatrix} \tag{10}$$

containing the RSS_p values to all visible APs whereby the variable $\mathrm{p} = 1,2, \ldots, \mathrm{P}$ describes the elements of the vector s, namely the averaged RSS values for each AP. p is therefore the number of APs.

Several different approaches are possible to derive *VD*s (see e.g. Machaj & Brida, 2011 or Moghtadaiee & Dempster, 2015). In the following, the derivations are started with the universal Minkowski distance which then leads to three special cases of this *VD*. Furthermore, six other *VD*s are introduced which are used in the comparing performance analysis in this study.

4.1. Minkowski Distance

The following equation describes the universal Minkowski distance L_q:

$$L_q = \sqrt[q]{\sum_{p=1}^{P} \left| s_{t,p} - s_{r,p} \right|^q} \tag{11}$$

where $s_{t,p}$ is the RSS_p of the fingerprint TP_t measured in the positioning phase on the test point to be positioned, $s_{r,p}$ the RSS_p of the fingerprint RP_r measured in the training phase on the reference point, q the norm parameter and L_q is the norm q between two points. The following three vector distances are defined by change of the norm parameter q.

4.2. Manhattan Distance

By choosing $q = 1$ for the norm parameter in the Minkowski distance formula, one receives the Manhattan distance L_1. This VD represents the distance between two points in a right-angled grid (see the two distances in Figure 4 with the same length) and is also called city-block distance, boxcar distance or taxicab distance. This *VD* is used for the calculation of the distance

between two buildings (or crossroad points) in grid-shaped cities, such as in Manhattan, New York – hence the name Manhattan distance (Krause, 1986). The Manhattan distance L_1 results in the following equation:

$$L_1 = \sum_{p=1}^{P} |s_{t,p} - s_{r,p}|. \quad (12)$$

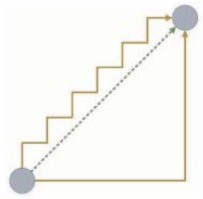

Figure 4. Illustration of two examples for the Manhattan distances with same length (brown lines) and the Euclidean distance (dotted green line) (after Favre-Bulle P, 2015)

4.3. Euclidean Distance

The Euclidean distance L_2 is the most common used vector distance and describes the distance of the shortest direct path between two points in the Euclidean space. This *VD* is also a special case of the Minkowski distance if a value of 2 is used for q. Then L_2 can be derived from L_q in the form:

$$L_2 = \sqrt{\sum_{p=1}^{P} |s_{t,p} - s_{r,p}|^2}. \quad (13)$$

4.4. Chebyshev Distance

By setting $q = \infty$ within formula (11) describing the Minkowski distance, the Chebyshev distance L_∞ is derived. This derivation is also called maximum value distance calculation and determines the maximum absolute difference between the two vector pairs $s_{t,p}$ and $s_{r,p}$. Then L_∞ can be described by the following formula:

$$L_\infty = \max_p |s_{t,p} - s_{r,p}|. \quad (14)$$

4.5. Canberra Distance

The Canberra distance d_{Can} is similar to the Manhattan distance L_1, however, the distance is weighted by the sum of the absolute values. Every summand of d_{Can} has therefore a value between 0 and 1. The mathematical relationship is given by:

$$d_{Can} = \sum_{p=1}^{P} \frac{|s_{t,p} - s_{r,p}|}{|s_{t,p}| + |s_{r,p}|}. \quad (15)$$

4.6. Cosine Distance

The Cosine distance d_{Cos} describes rather the similarity between two vectors than a distance. The right term in equation (16) can have a value between -1 and 1. The higher the value, the more alike are the two vectors. By the deduction of this term by 1, one can interpret the result as a *VD*.

$$d_{Cos} = 1 - \frac{\sum_{p=1}^{P} s_{t,p} \cdot s_{r,p}}{\sqrt{\sum_{p=1}^{P} s_{t,p}^2} \cdot \sqrt{\sum_{p=1}^{P} s_{r,p}^2}}. \tag{16}$$

4.7. Sorensen Distance

The Sorensen distance, also called Bray Curtis distance, is another derivation of the Manhattan distance L_1, where its value is normalized. Then all values are positive and lie between 0 and 1. In case of $d_{Sor} = 0$, $s_t = s_r$ which means that there are two equal vectors. The mathematical relationship is:

$$d_{Sor} = \frac{\sum_{p=1}^{P} |s_{t,p} - s_{r,p}|}{\sum_{p=1}^{P} (s_{t,p} + s_{r,p})}. \tag{17}$$

4.8. Hellinger Distance

For this distance the norm from the square roots of the fingerprint vectors s_t and s_r is used divided by $\sqrt{2}$. This results in:

$$d_{Hell} = \frac{1}{\sqrt{2}} \cdot \left\| \sqrt{s_t} - \sqrt{s_r} \right\|_2. \tag{18}$$

4.9. Chi-square Distance

The Chi-square distance is similar to the Euclidean distance L_2 but is weighted by a factor $\rho_p = \frac{s_{t,p} + s_{r,p}}{2}$ and therefore defined as:

$$d_{Chi} = \sum_{p=1}^{P} \frac{(s_{t,p} - \rho_p)^2}{\rho_p}. \tag{19}$$

4.10. Jeffrey Distance

Finally, for the Jeffrey distance also ρ_p is used and then the *VD* has the form:

$$d_{Jeff} = \sum_{p=1}^{P} \left(s_{t,p} \cdot \log_{10}\left(\frac{s_{t,p}}{\rho_p}\right) + s_{r,p} \cdot \log_{10}\left(\frac{s_{r,p}}{\rho_p}\right) \right). \tag{20}$$

Apart from the Minkowski distance these nine vector distances *VD*s are analysed regarding their performance and achievable positioning accuracies. The following section 5 describes first the indoor test bed and section 6 the major results of the campaign.

5. Indoor Testbed and Measurements

The indoor test bed is located on the ground floor of a multi-storey office building of the TU Wien – Vienna University of Technology. Figure 5 shows the location of the 93 reference points RPs (illustrated as black dots) distributed in a regular grid with spacing of around 2.5 m between the grid points and the randomly selected six test points TPs (blue dots). The RPs cover mainly three different areas, i.e., parts of a class room in the upper left area, an area with desktop computers in the upper right part and the foyer in the lower part of the test site. The distribution of all six visible Access Points APs CDEG-1 to -6 (indicated as red triangles) is also shown in the Figure. Figure 6 givens some impressions of how the test bed looks like. As can be seen in the left Figure 6, an entresol exists in the foyer whereby the maximum ceiling height is 5 m.

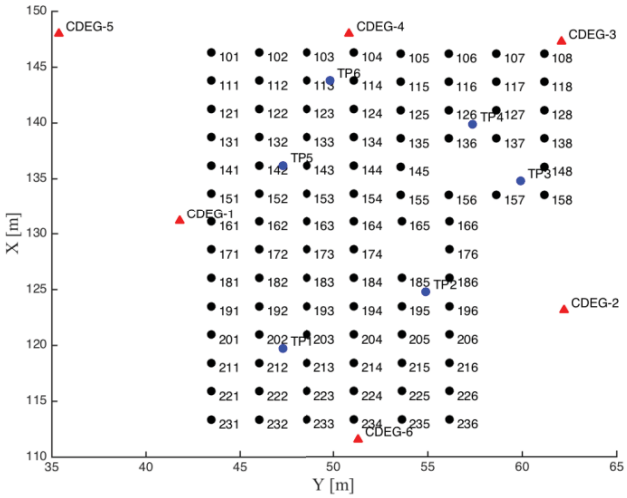

Figure 5. Distribution of reference points RPs (black dots), test points TPs (blue dots) and Access Points APs (red triangles) in the indoor test bed

The three different areas in the test bed have been chosen to provide different conditions for Wi-Fi signal propagation regarding damping and shielding of the radio signals. An other specific feature of the test bed is that in total four different Wi-Fi networks are provided. Besides, these consist of the network eduroam, tunet, tunetguest and wlanipsec. These networks are so-called multiple-SSID (short for Service Set Identifier) Wi-Fi networks. In this case several networks at a physical AP location and a single transmitter as offered. In other words, there are four Wireless Local Area Networks (WLANs) (1, 2, 3 and 4) with native as 1 and mapping to 4 different SSIDs (one, two, three and four) on any of the six APs receivable throughout the whole test bed. A different encryption for the networks is usually applied. On overview about

the characteristics of the four networks is given in Table 1. As can be seen the networks are either encrypted or not. With this set-up the rights are defined how a user may access the Wi-Fi network. Therefore, each network adapter needs his own MAC address. Thus, RSS to 24 different MAC addresses could be measured. Additional RSS measurements to other visible APs in this area were omitted as they could not be received on the majority of the TPs. The measured RSS ranged between -64 to -85 dBm (see section 6.1 for further details).

Figure 6. Impressions from the indoor test bed showing the foyer on the right and the class room on the left

SSID	Characteristics
eduroam	Network for students, staff and participants of this international network
tunet	Network for students and staff as well as visiting scholars and conference participants. The network is encrypted.
tunetguest	Alternative for network tunet, not encrypted.
wlanipsec	Network only for staff of TU Wien while using a VPN (Virtual Private Network) connection.

Table 1. Overview of the characteristics of the four different multiple-SSID networks visible in the test bed

6. Discussion of Evaluation Results

In this section the major results of the investigations are presented. At the beginning of this section general aspects are briefly mentioned and the measurement results and corrections for fingerprinting are discussed in the following. Here the empirical determination of the optimum value for K in the matching approach is elaborated first followed by a detailed discussion of the achieved distance error DE and the mean distance error MDE. The calculations were performed with newly developed MatLab routines (Joksch, 2016).

6.1. General aspects

In the training phase of the fingerprinting approach RSS scans on in total 93 reference points RPs (compare Figure 5) were measured in four different user orientations aligned to the axes of the building. The main reason for measuring in four different orientations is that the Wi-Fi signals are significantly shielded by the human body of the observer if he is located between the AP and the mobile device. For the evaluation, however, only an average over all four orientations was used. The measured RSS values $RSS_{r,p,m}$ are averaged on each RP per SSID using the following relationship:

$$\overline{s}_{r,p} = \frac{\sum_{m=1}^{M_{r,p}} RSS_{r,p,m}}{M_{r,p}} \tag{21}$$

where $RSS_{r,p,m}$ are the measurements m of the RSS to AP_p on RP_r and $M_{r,p}$ is the number of measurements to AP_p on RP_r.

The averaged values have then been summarized in the fingerprint vector s_r on each RP_r in the form:

$$s_r = \begin{bmatrix} \overline{s}_{r,p=1} \\ \overline{s}_{r,p=2} \\ \overline{s}_{r,p=3} \\ \vdots \\ \overline{s}_{r,p=P} \end{bmatrix}. \tag{22}$$

Also on the six test points TPs scanned in the positioning phase the measured RSS values were averaged as given in:

$$\overline{s}_{t,p} = \frac{\sum_{m=1}^{M_{t,p}} RSS_{t,p,m}}{M_{t,p}} \tag{23}$$

where $RSS_{t,p,m}$ are the measurements m of the RSS to AP_p on TP_t and $M_{t,p}$ is the number of measurements to AP_p on TP_t.

Again the averaged values are included in the fingerprint vector s_t on each TP_t:

$$s_t = \begin{bmatrix} \overline{s}_{t,p=1} \\ \overline{s}_{t,p=2} \\ \overline{s}_{t,p=3} \\ \vdots \\ \overline{s}_{t,p=P} \end{bmatrix}. \tag{24}$$

As an example Table 2 presents the fingerprint vectors for the AP CDEG-1 with the four multiple-SSID networks measured on the six test point TP 1 up to TP 6. When looking at the RSS values of the four different networks, i.e., eduroam, tunet, tunetguest and wlanipsec, it can be seen that the RSS values are quite similar and the variations are low compared to short-time fluctuations of the RSS in the test bed as reported by Retscher & Roth (2016) and Retscher & Tatschl (2016a). Also a significant difference in RSS values on the six different TPs can be seen. This is advantageous for positioning using fingerprinting as the fingerprints should be unique on each user locations throughout the test site. For the AP CDEG-6, however, the RSS difference on five of the six TPs is much smaller. It only varies between -65 to -72 dBm. Only the measurements on TP 6 are different with a value of around -80 dBm (see Joksch, 2016 for further details). Then it would be more difficult to estimate the correct user's location if only this single AP would be used. In addition, only on two test points, i.e., TP 5 and 6, no signal could be received from AP CDEG-2. This two TPs are located in the class room and on these locations the Wi-Fi signal of AP CDEG-2, which is outside quite far away (compare Figure 5), was shielded due to the walls and their structure for the whole duration of the field campaign. In this case where no RSS signal can be received a value of -99 dBm is assigned in the fingerprinting vector.

SSID	TP 1	TP 2	TP 3	TP 4	TP 5	TP 6
eduroam	-84.0	-81.0	-68.0	-65.0	-56.4	-62.8
tunet	-85.0	-81.5	-68.2	-64.5	-56.0	-63.0
tunetguest	-85.0	-80.4	-66.5	-66.5	-56.0	-64.5
wlanipsec	-82.5	-80.8	-67.5	-65.3	-55.3	-63.0

Table 2. Example for RSS values in [dBm] in the fingerprint vector s_t of one AP CDEG-1 for the four different multiple-SSID networks on the six TPs

6.2. Determination of the K-Value for the Matching Method

For the empirical determination of the optimum value for K for the K-nearest neighbour (KNN) or K-weighted nearest neighbour (KWNN) matching algorithm (see section 3.2 and 3.3 respectively) the Euclidean distance is calculated where all RSS measurements of the four multiple-SSI networks were utilized. As can be seen from Figure 7 the results do not follow the theoretical relationship described in section 3.2. On account of the minimum for a MDE of 2.24 m with $k = 3$, this value is selected for the further evaluation in this study.

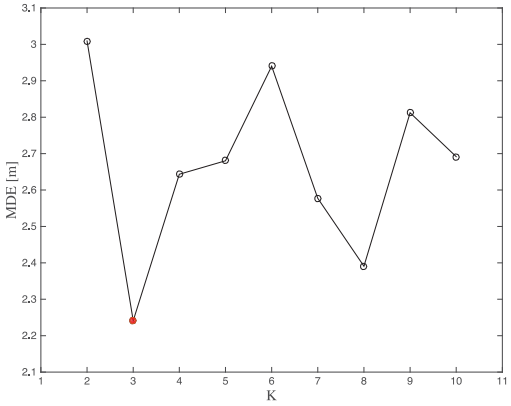

Figure 7. Progression of the MDE in [m] for rising *K*-values from 1 to 10 (minimum obtained *K*-value of 3 is highlighted in red)

6.3. Vector Distance Results Comparison

The different vector distances *VD*s described in section 4 in combination with the matching approaches NN, KNN and KWNN are calculated for every of the six TPs within every SSID network. Then in the following, the *VD*s were sorted in dependence of their dimension and the usable distances were taken. As described in section 3.1 to 3.3, the correct value is the first one for the NN and the first three values for KNN and KWNN method. Afterwards the positions $\hat{l}_t = (\hat{x}_t, \hat{y}_t)$ are calculated using the relationship $\hat{l}_t = l_r$ for the NN where l_r is the coordinate vector of RP_r and respectively equation (3) in the case of the KNN and KWNN matching approach.

In the next step, the resulting distance errors (DE) and mean distance errors (MDEs) calculated with equations (8) and (9) given in section 3.4 of the four different multiple-SSID networks were analyzed. Table 3 presents the DEs in the four different SSID networks and in total whereas Table 4 summarizes the MDEs in the SSID network eduroam of all six TPs for each of the nine *VD*s introduced in section 4. As can be seen from Table 3 if one looks at the column of the total DE, the minimum DE of around 0.30 m is achieved for the Euclidean and Hellinger *VD*s. Furthermore, these two *VD*s do not result in the largest DE. The maximum DE of 8.96 m occurs when using the Chebyshev *VD* in the SSID network tunetguest.

In general, the lowest overall DE occurred in the eduroam network. Thus, in the following only results in this SSID network are presented. Further results concerning the MDE and the four SSID networks are presented in Retscher and Joksch (2016) and have proven that the eduroam network is the one to be chosen. Here it could be seen that the smallest MDE in eight of nine cases occurred. Thereby the smallest MDE is achieved four times for the KWNN

matching approach where *K* was set to 3. For all vector distances the largest MDEs in the eduroam network was obtained if the NN approach is employed. This can also be observed if one looks at Table 4 and Figures 8 to 10. The difference between the smallest MDE and the largest is about 1.10 m (range of 1.40 to 3.50 m). The use of the Cosine distance resulted in a MDE of 1.39 m for the KNN and 1.40 m for the KWNN. The next best results with differences of only 2 to 3 cm was achieved with the Euclidean and Hellinger distance. The Chi-square and the Jeffrey MDE differs only by 8 cm. Also the results with the Manhattan, Sorensen (MDE 1.67 m) and Canberra distance (MDE 1.88 m) lie under the overall MDE of 1.90 m using the KWNN matching approach. With the Chebyshev distance the worst result with a MDE of 2.42 m was achieved.

Table 4 presents the MDEs separately for each of the six test points TP 1 to TP 6 in the eduroam SSID network. The smallest MDE resulted in only 0.43 m on TP 6 using the Cosine, Chi-square and Jeffrey *VD* and the KWNN matching approach. The MDE of the Euclidean distance was slightly higher, i.e., 0.58 m. This cannot be seen as a significant difference as the localization accuracy of Wi-Fi fingerprinting is usually not that high. If one looks at Figures 8 to 10, however, it can be seen that positioning accuracies can be higher, i.e., on the half meter level, than what usually is achieved in many tests reported in the literature. Qualifying it must be mentioned that the performance and achievable positioning accuracies depend very much on the environment and the signal propagation conditions during the measurements. It is always reported that the repeatability might be a problem. A suitable strategy to retrieve this situation might be the use of continuously recorded RSS measurements in the area of interest. Retscher and Tatschl (2016a and b) have developed a differential Wi-Fi positioning approach where RSS scans are performed on selected points in the test bed. This points are equipped with low-cost Raspberry Pi's serving as reference stations as it is done in a differential GNSS network. Then it is possible to derive corrections in real-time which are applied by the mobile client.

6.4. MDE Comparison for Euclidean and Cosine *VD*

Figures 8 to 10 provide a graphical representation of the positioning accuracies for each TP when using the Euclidean and Cosine *VD*. As can be seen the results can be quite different. In general, the KNN and KWNN matching approaches outperform the NN approach. The resulting positioning errors lie in the range of around 0.40 up to 5.40 m (compare Table 4). The best result is obtained on TP 5 in this range using both *VD*s and KNN and KWNN approach. This test point is located in the class room (compare Figure 5). On the other hand, the worst result is achieved under the entresol in the foyer where TP 1 is located using the NN algorithm.

		eduroam		tunet		tunetguest		wlanipsec		total	
		min	max	min	max	min	max	min	max	min	max
Manhattan	NN	1.27	5.38	1.27	5.38	1.27	5.38	1.27	6.83	1.27	5.38
	KNN	0.60	2.99	0.95	4.67	0.60	7.20	0.60	4.67	0.60	4.67
	KWNN	0.55	3.01	1.16	4.87	0.71	7.83	0.63	4.79	0.66	4.72
Euclidean	NN	1.27	5.38	1.27	5.38	1.27	5.38	1.79	6.46	1.27	4.57
	KNN	0.42	2.46	0.42	3.90	1.74	7.20	0.42	5.49	0.42	5.49
	KWNN	0.35	2.57	0.32	3.49	1.59	7.73	0.33	5.48	0.31	5.18
Chebyshev	NN	1.27	6.46	1.27	4.57	1.27	8.96	1.79	6.46	1.79	6.34
	KNN	0.60	4.67	0.42	3.90	1.34	4.35	2.15	7.20	0.42	5.56
	KWNN	0.73	4.69	0.34	3.72	1.39	4.72	2.13	7.654	0.35	5.57
Canberra	NN	1.27	5.38	1.27	5.38	1.27	5.38	1.27	6.83	1.27	5.38
	KNN	0.95	2.99	0.95	4.67	0.60	7.20	0.60	4.67	0.60	4.67
	KWNN	0.54	3.03	1.11	4.921	0.72	7.76	0.62	4.82	0.68	4.70
Cosine	NN	1.27	5.38	1.27	5.38	1.27	5.38	1.79	6.46	1.79	4.57
	KNN	0.42	2.46	0.42	3.90	1.74	5.49	0.42	5.49	0.60	5.87
	KWNN	0.43	2.68	0.59	3.02	1.30	4.01	0.29	5.48	0.82	5.62
Sorensen	NN	1.27	5.38	1.27	5.38	1.27	5.38	1.27	6.83	1.27	5.38
	KNN	0.60	2.99	0.95	4.67	0.60	7.20	0.60	4.67	0.60	4.67
	KWNN	0.55	3.01	1.16	4.87	0.71	7.83	0.63	4.80	0.67	4.72
Hellinger	NN	1.27	5.38	1.27	5.38	1.27	5.38	1.79	6.46	1.27	4.57
	KNN	0.42	2.46	0.42	3.90	1.74	7.20	0.42	7.20	0.42	2.99
	KWNN	0.35	2.59	0.30	3.55	1.61	7.67	0.32	7.76	0.30	3.08
Chi-square	NN	1.27	5.38	1.27	5.38	1.27	5.38	1.79	6.46	1.27	4.57
	KNN	0.42	2.46	0.42	3.90	1.74	7.20	0.42	7.20	0.42	2.99
	KWNN	0.43	2.72	0.53	3.21	1.43	8.16	0.27	8.33	0.42	3.17
Jeffrey	NN	1.27	5.38	1.27	5.38	1.27	5.38	1.79	6.46	1.27	4.57
	KNN	0.42	2.46	0.42	3.90	1.74	7.20	0.42	7.20	0.42	2.99
	KWNN	0.43	2.72	0.53	3.21	1.43	8.16	0.27	8.33	0.42	3.17

Table 3. DEs in [m] in the four different SSID networks and in total for the different *VD*s
Green marked values indicate the minimum DE and red the maximum DE of each vector distance *VD*.

		TP 1	TP 2	TP 3	TP 4	TP 5	TP 6
Manhattan	NN	5.38	1.79	4.01	1.79	4.60	1.27
	KNN	2.99	2.15	2.46	0.60	1.52	0.95
	KWNN	3.01	2.21	2.56	0.61	1.07	0.55
Euclidean	NN	5.40	1.79	4.01	1.79	4.57	1.27
	KNN	2.15	2.15	2.46	0.60	0.42	0.95
	KWNN	2.21	2.19	2.57	0.63	0.35	0.58
Chebyshev	NN	1.79	4.01	6.46	1.79	4.57	1.27
	KNN	3.64	2.15	4.67	0.60	2.83	0.95
	KWNN	3.55	2.22	4.69	0.76	2.58	0.73
Canberra	NN	5.38	4.01	4.01	1.79	4.57	1.27
	KNN	2.99	2.15	2.46	1.79	1.52	0.95
	KWNN	3.03	2.22	2.62	1.76	1.11	0.55
Cosine	NN	5.38	4.01	4.01	1.79	4.57	1.27
	KNN	1.79	2.15	2.46	0.60	0.42	0.95
	KWNN	1.91	2.25	2.68	0.65	0.48	0.43
Sorensen	NN	5.40	1.79	4.01	1.79	4.57	1.27
	KNN	2.99	2.15	2.46	0.60	1.52	0.95
	KWNN	3.01	2.21	2.56	0.61	1.07	0.55
Hellinger	NN	5.38	4.01	4.01	1.79	4.57	1.27
	KNN	2.15	2.15	2.46	0.60	0.42	0.95
	KWNN	2.22	2.22	2.59	0.64	0.35	0.57
Chi-square	NN	5.38	4.01	4.01	1.79	4.57	1.27
	KNN	2.15	2.15	2.46	0.60	0.42	0.95
	KWNN	2.30	2.30	2.72	0.68	0.44	0.43
Jeffrey	NN	5.38	4.01	4.01	1.79	4.57	1.27
	KNN	2.15	2.15	2.46	0.60	0.42	0.95
	KWNN	2.30	2.30	2.72	0.68	0.44	0.43

Table 4. MDEs in [m] in the SSID network eduroam of all six TPs for the different *VD*s
Green marked values indicate the minimum MDE and red the maximum MDE of each vector distance *VD*. The underlined values are the smallest MDE.

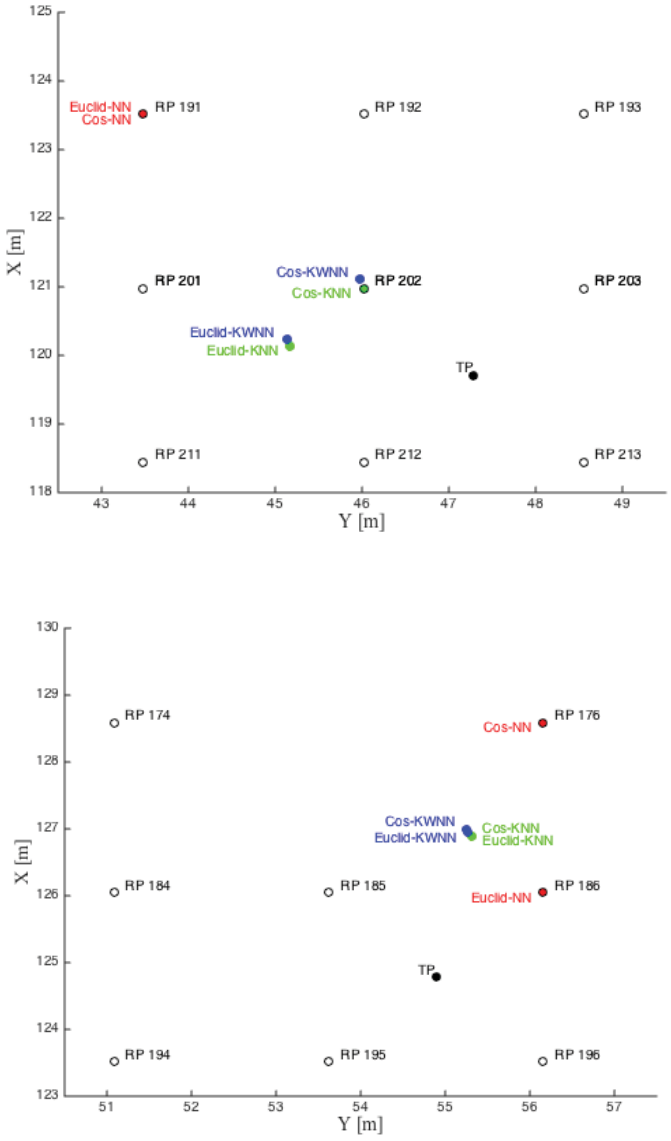

Figure 8. Comparison of the positioning errors in the eduroam network for TP 1 and TP 2 for the Euclidean and Cosine vector distance *VD* and NN, KNN and KWNN matching approach

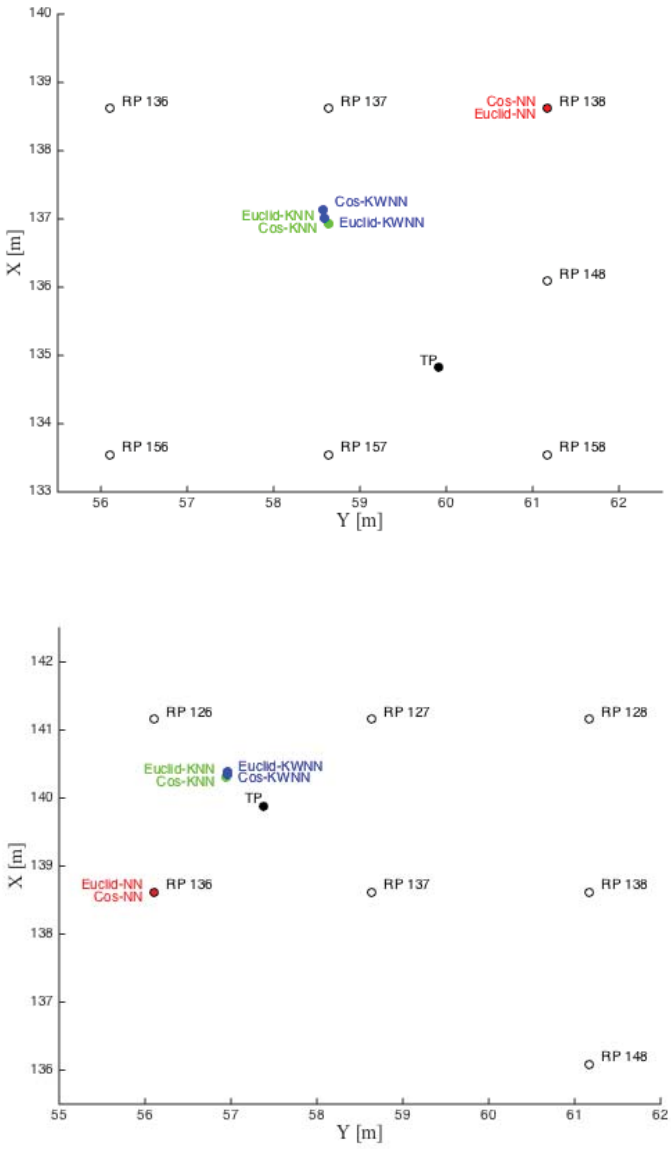

Figure 9. Comparison of the positioning errors in the eduroam network for TP 3 and TP 4 for the Euclidean and Cosine vector distance *VD* and NN, KNN and KWNN matching approach

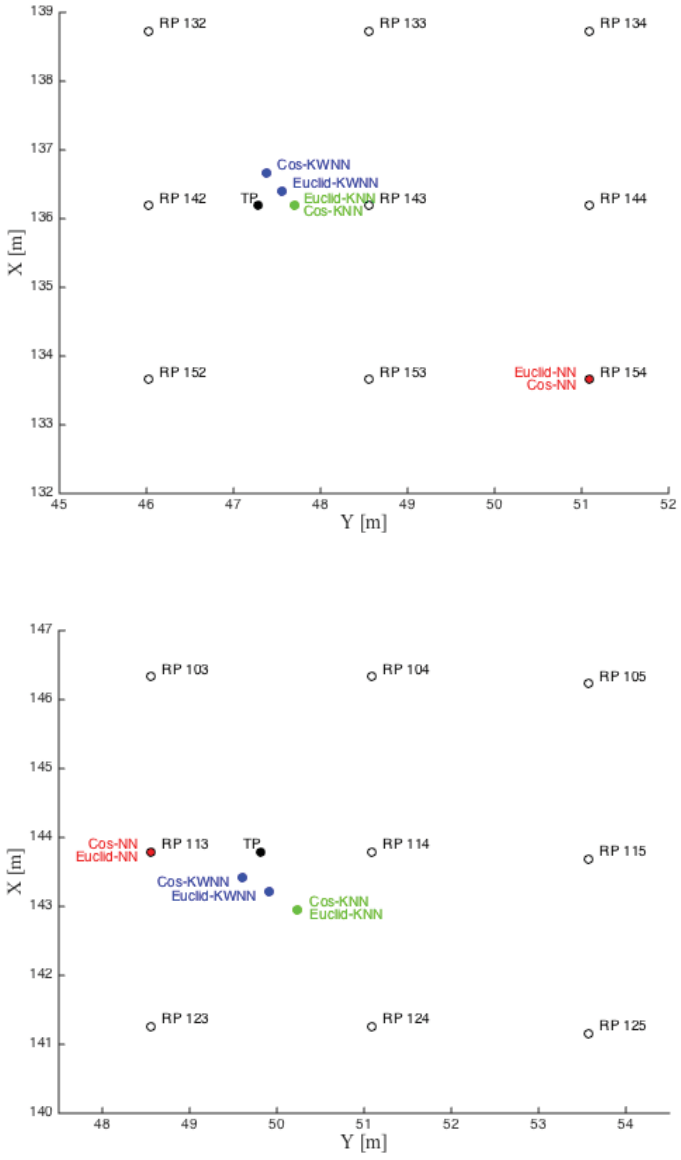

Figure 10. Comparison of the positioning errors in the eduroam network for TP 5 and TP 6 for the Euclidean and Cosine vector distance *VD* and NN, KNN and KWNN matching approach

6.5. Discussion of the Major Outcome

For the assessment of the quality of positioning using Wi-Fi fingerprinting the distance errors DEs and mean errors MDEs were analyzed. It was found that one network should be selected if multiple-SSID networks are existing. In the building of TU Wien where four multiple-SSI networks are provided the network eduroam resulted in the best positioning accuracy and performance. It is not recommended to average over all existing networks. This would lead to a false weighting of the observations.

If one looks at the results using the nine different vector distances *VD*s which were investigated no significant differences between most of them are found. Thus, it cannot be recommended in general, which *VD* should be chosen as the right selection depends very much on the surrounding environment and the present interference conditions affecting the propagation of the Wi-Fi signal. The most commonly employed Euclidean distance in fingerprinting showed no significant MDE difference to the slightly better performing Cosine vector distance (see also Retscher & Joksch, 2016). Figures 8 to 10 have shown that the resulting positions on the TPs obtained with the KNN and KWNN are mostly the closest to the true location of the TP. The DE is then around half a meter. Only the NN positions are further away from the true location.

7. Concluding Remarks and Outlook

Current investigations are focused on the additional integration of continuous long-time measurements to consider temporal and spatial signal variations of the Wi-Fi signal propagation. A concept and first test results of using reference stations which continuously scan and measure the RSS of the surrounding APs are presented in Retscher & Roth (2016). Another possibility to increase the performance of fingerprinting is the additional use of the compass data from the smartphone sensors and an orientation dependent fingerprinting DB.

Apart from the standard deterministic fingerprinting approaches further ongoing investigations are considering probabilistic approaches (see e.g. Honkavirta et al., 2009). The Mahalanobis distance is a suitable *VD* in this respect and has been tested by Ettlinger & Retscher (2016) for a combination of Wi-Fi positioning with fingerprinting using present ambient geomagnetic fields. In this case, the nearest neighbours are determined by using conditional probabilities.

Further future work concerns performance tests in 3D environments using Wi-Fi fingerprinting with the different vector distances. Field tests reported

in the paper from Retscher and Hofer (2016) were performed in the same office building where 3D scenarios, such as the navigation of a mobile user from the building entrances to an office in the third floor, have been tested. In the evaluation only the Euclidean distance and the NN approach have been used so far and a combined solution of all four available multiple-SSID networks. Thus, in the further analysis it should be focused on the eduroam SSID network with an eventual usage of the KNN or KWNN approach.

Integration with the inertial sensors embedded in the smartphone is also a promising strategy. If the measurements of the accelerometer and gyroscope are used continuous positioning and navigation via dead reckoning (DR) is possible. In this case, the drift of the inertial navigation sensors can then be compensated if a Wi-Fi positioning solution is available. Furthermore, in DR, for instance, the step length is adapted if the user climbs stairs or uses an elevator as identified by the Wi-Fi absolute positioning system. Retscher & Hofer (2016) have demonstrated that this strategy is the right direction for localization of mobile devices. Furthermore, the additional use of the barometric pressure sensor which can nowadays be found in many smartphones is also planned. Then an additional determination of the altitude of the user is possible (see Retscher, 2007 for further details and processing strategy).

References

Bahl P & Padmanabhan V N (2000) 'RADAR: An In-building RF-based User Location and Tracking System. INFOCOM 2000, Proceedings of the 19th Annual Joint Conference of the IEEE Computer and Communications Societies, Tel-Aviv, Mar 26-30, Vol. 2, 775-784

Ettlinger A F & Retscher G (2016) Positioning Using Ambient Magnetic Fields in Combination with Wi-Fi and RFID. 7th International Conference Indoor Positioning and Indoor Navigation IPIN 2016, October 4-6, Alcalá de Henares, Madrid, Spain

Favre-Bulle P (2015) Design and Implementation of an Adaptive Indoor Positioning System Based on Received and Interpolated Signal Strengths. Master Thesis, TU Wien

Hofer H & Retscher G (2016) Combined Out- and Indoor Navigation with Smart Phones Using Intelligent Check Points. this conference

Honkavirta V, Perälä T, Ali-Lötty S & Piche R (2009) A Comparative Survey of WLAN Location Fingerprinting Methods. WPNC'09, 243-251

Huang H (2014): Post Hoc Indoor Localization Based on RSS Fingerprinting in WLAN. Master thesis, University of Massachusetts, U.S.A.

Joksch J (2016) Vergleich von unterschiedlichen Distanzberechnungen bei der Indoor-Positionierung mittels Fingerprinting-Verfahren. Bachelor Thesis, TU Wien, Austria (in German)

Kjærgaard M B (2008) Location Fingerprinting. Retrieved August 2016, from http://wiki.daimi.au.dk/mikkelbk/location_fingerprinting.wiki

Krause E F (1986) Taxicab Geometry; an Adventure in Non-Euclidean Geometry, 1st publ. New York, NY: Dover Publ.

Machaj J & Brida P (2011) Performance Comparison of Similarity Measurements for Database Correlation Localization Method. 3rd International Conference Intelligent Information and Database Systems ACIIDS 2011, Part II, 452–461

Moghtadaiee V & Dempster A G (2015) Vector Distance Measure Comparison in Indoor Location Fingerprinting. IGNSS Symposium, July 14-16, Gold Coast, Australia

Mok E & Retscher G (2007) Location Determination Using WiFi Fingerprinting Versus WiFi Trilateration. Journal of Location Based Services, 1:2, 145-159

Niedermayr S. & Wieser A. (2012) Combination of Feature-based and Geometric Methods for Positioning. in: Schwieger V., Böttinger S., Zheng B. (eds.) Proceedings of the 3rd International Conference on Machine Control & Guidance, March 27-29, Stuttgart, 301-310

Retscher G. (2007) Augmentation of Indoor Positioning Systems with a Barometric Pressure Sensor for Direct Altitude Determination in a Multi-storey Building. Journal of Cartography and Geographic Information Science (CaGIS), 34: 4:305-310

Retscher G (2016) Indoor Positioning. Chapter 9, Encyclopedia of Geodesy, Springer Verlag

Retscher G & Hofer H (2016) Wi-Fi Location Fingerprinting Using an Intelligent Checkpoint Sequence. FIG Working Week, May 2-6, Christchurch, New Zealand

Retscher G & Joksch J (2016) Analysis of Nine Vector Distances for Fingerprinting in Multiple-SSID Wi-Fi Networks. 7th International Conference Indoor Positioning and Indoor Navigation IPIN 2016, October 4-6, Alcalá de Henares, Madrid, Spain

Retscher G & Roth F (2016) Wi-Fi Fingerprinting with Reduced Signal Strength Observations from Long-time Measurements. this conference

Retscher G & Tatschl T (2016 a) Differential Wi-Fi – A Novel Approach for Wi-Fi Positioning Using Lateration. FIG Working Week, May 2-6, Christchurch, New Zealand

Retscher G & Tatschl T (2016 b) Indoor Positioning Using Wi-Fi Lateration – Comparison of two Common Range Conversion Models with Two Novel Differential Approaches. IEEE Xplore, 2016 Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS), November 3-4, Shanghai, PR China

Retscher G, Zhu M. & Zhang K (2012) RFID Positioning. in Chen R. (ed.): Ubiquitous Positioning and Mobile Location-Based Services in Smart Phones. Hershey PA, U.S.A.: IGI Global, 69-95

Contacts

Guenther Retscher and Julian Joksch

Department of Geodesy and Geoinformation, Research Group Engineering Geodesy, TU Wien, Gusshausstrasse 27-29 E120/5, 1040 Vienna, Austria
E-Mail: guenther.retscher@tuwien.ac.at
or jjoksch@me.com