Dynamic Nearest Neighbors and Online Error Estimation for SMARTPOS

Philipp Marcus, Moritz Kessel, Martin Werner
Mobile and Distributed Systems Group
Ludwig-Maximilians-University Munich
Munich, Germany
{*philipp.marcus,moritz.kessel,martin.werner*}*@ifi.lmu.de*

***Abstract*—Location-based services are possibly the most popular services with respect to mobility, since they allow for the automated filtering of information relevant to the user. This paper presents a detailed evaluation of SMARTPOS, an indoor positioning system based on deterministic 802.11 fingerprinting and a digital compass. SMARTPOS is accurate enough to supply location estimates for indoor location-based services and can be deployed standalone on a mobile phone. Assuming that the mobile phone is held in front of the body, the system considers the user's orientation to avoid errors caused by the blocking effect of the human body. For location estimation it employs a kNN approach on that part of the fingerprint database that corresponds to the user's current orientation. As an extension to this approach, SMARTkNN is proposed which is based on dynamically selecting the number of nearest neighbors. This improved the mean position error to 1.10 meters and to a maximum position error of 2.65 meters in a 250 square meter environment in comparison to SMARTPOS which achieved a mean position error of 1.16 meters and a maximum position error of 2.74 meters. Furthermore, it is shown that the errors of SMARTPOS are normally distributed. Based on this fact, a novel online error estimator using bivariate Gaussians is proposed which gives the best approximation of the observed errors compared to existing methods. Additionally it was observed, that the density of the underlying radiomap strongly correlates to the maximum error and has a weaker impact on the observed mean error.**

***Keywords*-802.11 Fingerprinting, Orientation Filter, Mobile Phone Positioning, Location-Based Services.**

I. Introduction

In recent years, a trend towards mobility can be recognized. Smartphones, small devices with comparatively high processing power and mobile internet, make it possible to work while traveling, to stay connected to social networks, and to retrieve nearly any information anywhere at any time. One of the most popular mobile services are location-based services (LBS). These are value-added services, which utilize the location of the mobile to present the user with information about its surroundings. Navigation and information services, friend-finder, pet-tracker, and location-based games are only a small part of the number of services and applications filling the app-stores of the world.

The key enabler for LBS is the Global Positioning System (GPS). It enables accurate positioning in outdoor environments, the usage is free of charge, the system is globally available, and most of today's smartphones are equipped with a GPS-receiver. Unfortunately, GPS is not able to track people in indoor environments with acceptable accuracy. Signals might get lost due to attenuation effects of roofs and walls or lead to position fixes of very low accuracy due to multipath propagation.

Even worse, indoor location-based services require much higher precision guarantees than outdoor services. Errors should not exceed a few meters to allow for a differentiation between several floors or rooms. Otherwise, the service could provide information for places, which are quite far away from the actual position of the target. Despite these challenges, many users would appreciate indoor location-based services, especially in large and complex buildings such as museums, shopping malls, airports, hospitals, or university buildings.

Existing indoor positioning techniques can be grouped by their level of precision and the expenses for additional infrastructure. Dedicated indoor positioning systems such as ultra wide band or ultrasonic systems consist of several components with the sole purpose of determining the positions of possibly multiple targets in indoor environments. The precision is often high, but an expensive infrastructure is needed and hence the space where positioning is possible is usually limited to a small area, where higher accuracy compensates the high cost. Another class of systems is built on existing infrastructure such as WLAN, Bluetooth or inertial sensors for positioning. The precision of such systems is limited, but the system can be deployed with few additional expenses.

In this paper, we extend SMARTPOS [1], an indoor positioning system for smartphones based on deterministic WLAN fingerprinting and a digital compass. The system is self-positioning, meaning that the whole positioning process (including all measurements) is carried out on the phone. It achieves a high accuracy within few meters and therefore is able to provide interactive, non-background indoor location-based services with high quality location estimates at no additional expenses. SMARTPOS makes use of the smartphone's orientation (which should correspond to the user's orientation) to avoid errors caused by the blocking effect of the human body. Only those fingerprints are considered for location estimation that were measured while viewing in a similar direction like the user. As an extension to the system, a detailed evaluation of the system's errors has been carried out. Based on the results of that evaluation a novel online error estimation scheme is proposed which enables the system to provide an error estimator as a Gaussian probability density function for each position measurement. Furthermore, a method for dynamically choosing the number of nearest neighbors based on convergence criteria is proposed which obviates the need for empirically determining an optimal number for every environment.

The remainder of this paper is structured as follows: In the next section, an overview of existing indoor positioning systems with focus on WLAN fingerprinting is given. In Section III, the original SMARTPOS is presented and evaluated in detail, stressing the impact of several parameters and decisions on the design of the system. Whether weighted or non-weighted kNN (k-nearest neighbors) in signal space should be carried out, the influence of missing values on the algorithm and the performance gain of including the orientation on SMARTPOS and a Naive Bayesian Estimator are evaluated. In Section IV, the question of the reliability of the positioning method is researched and an online error estimation scheme introduced. Then the influence of the density of fingerprints is analyzed in Section V and a novel algorithm for dynamically choosing the best number of neighbors for every position fix is presented in Section VI. Section VII concludes the paper and gives hints on future work.

II. Related Work

In the past 15 years, a variety of technologies for indoor positioning have been proposed. A good overview of existing indoor positioning systems using radio frequency (RF) technologies such as radio frequency identification (RFID), ultra wide band (UWB), ultra high frequency (UHF), WLAN and Bluetooth is given in [2]. However, the authors do not describe up-to-date systems, which have been developed since 2007. We therefore focus in this section on the recent development and work closely related to our research.

Many pedestrian indoor positioning systems rely on WLAN fingerprinting algorithms [1], [3], [4], [5], which offer position estimates with sufficient accuracy (i.e., 1-3m) while utilizing the existing WLAN infrastructure and therefore avoiding high expenses. These algorithms belong to the area of pattern matching and work in two phases: The first phase is called the calibration phase, where a database is created by the collection of received signal strength indicator (RSSI) at certain reference positions from the surrounding access points (AP). The accumulated information of RSSI, AP and reference position at a specific time/interval is called a fingerprint. In the second phase, positioning is carried out by comparing current RSSI measurements with the previously stored values from the database. Different algorithms calculate the position as the reference position of the nearest fingerprint in signal space [3], the average of the k-nearest neighbors (kNN) with or without the distance in signal space as additional weight [1]. Some algorithms also utilize Bayesian methods [4], [5] based on probability distributions derived by multiple measurements over a length of time. While earlier systems utilize laptops for position determination, the recent trend goes towards smartphones. Martin et al. present one of the first WLAN positioning systems which integrates both offline and online phase on a mobile phone [6].

One of the first developed systems for WLAN fingerprinting, RADAR [3], includes already the impact of the user's orientation in the position calculation by obtaining empirical data for multiple orientations. Kaemarungsi et al. further analyze the effects of the user's presence and orientation on RSSI values in [7]. The results show that the attenuation effects of the human body can lower the RSSI by more then 9 dBm. COMPASS [5] is one of the first fingerprinting systems that addresses the problem of attenuation effects caused by the human body by adding a digital compass to the system. In the calibration phase, fingerprints for several selected orientations (typically each $45°$ or $90°$) are collected at reference positions. In the positioning phase, the user's orientation is measured by a digital compass and only those fingerprints with a similar orientation estimate are used for the positioning algorithm. COMPASS presents the most similar approach to the SMARTPOS System. However, we extent the work in serveral directions. By using kNN instead of a Bayesian estimator, the number of measurements carried out for fingerprint creation is massively reduced. While COMPASS reports 20 to 100 measurements for a single fingerprint to correctly estimate the Gaussians, we tested our system with 3-5 measurements. While the COMPASS approach might achieve an even higher accuracy due to the larger training dataset and the inclusion of the RSSI's second moment, it is not well suited for the self creation of databases by the user due to the higher calibration effort. Chan et al. also present a system running on a mobile phone considering the orientation of the user in [8], but apply a technique called Newton Trust Region for further position refinement.

Most up-to-date systems combine WLAN fingerprinting with additional technologies such as inertial sensors to offer more accurate position estimates and continuous tracking functionality [9], [10]. In [9], the authors utilize a particle filter for fusing WLAN fingerprint location estimates with an accelerometer. For the utilization of the SMARTPOS system in Bayesian filtering techniques, a probability distribution needs to be given for each position calculation. Existing approaches [11], [12] often utilize grid based approaches, where the discrete probability distribution is directly obtained by the probability of all grid cells according to a Bayesian model. In [9], Evennou and Marx utilize a Gaussian distribution for particle weighting with the mean located on the WLAN position and a variance based on the deviation of the RSSI.

Lemelson et al. further investigate error estimation of WLAN fingerprinting based position determination in [13]. They propose different schemes which estimate the occuring error as a scalar that can be used to assess the trust of position estimates.

In this paper, we show that the position errors follow a Gaussian with high probability. Based on this result, an online error estimator is proposed that derives a Gaussian probability density function modelling an estimate for the ground truth position relative to the position fix. We then compare our own approach to slightly adapted proposals from Lemelson et al. [13]. In contrast to that paper, Beder et al. propose an offline error estimation method which allows for the calculation of the expected uncertainty of every possible position [14]. Similar to [9] a Gaussian distribution of RSSI values is assumed and the covariance matrix of the fingerprint is used to calculate the expected error.

In addition to the error estimation, this paper proposes a method for dynamically estimating the number of neighbors suitable for position estimation. Roshanaei and Maleki combine in [15] traditional RSSI-based fingerprinting with a method based on Angle of Arrival (AOA) to further reduce the set of the nearest neighbors to those which are located in a certain area. The area is determined by their AOA algorithm using an adaptive antenna array. Altintas and Serif enhance in [16] the neighbor selecton by k-means clustering. The candidate fingerprints are clustered according to their reference points and only the fingerprints of one cluster which has the smallest diameter are returned. Another approach is presented by Shin et al. in [17]. For their weighted nearest neighbor approach, the neighbors are picked from a set of all fingerprints with a distance in signal space below a certain threshold. Furthermore, the mean distance in this set is calculated and only those fingerprints considered for position estimation whose distance is below this mean value. In contrast to related work, the method for dynamically choosing k presented in this paper is based on convergence criteria of the derived position estimates for different values of k. One advantage is, that it is completely independent from deriving thresholds in the signal space.

III. SMARTPOS: A System for Self-Contained Mobile Positioning

In this section, we describe the original SMARTPOS system presented in [1], a system for an accurate and self-contained indoor positioning based on deterministic 802.11 fingerprinting and a digital compass. The system runs stand-alone on a mobile phone and consists of a management module for the creation and maintenance of the fingerprint database and a module for location determination. The latter offers the possibility of modifying several parameters concerning the deterministic location estimation or allows a change of the positioning method to a room-based bayesian approach.

A. Database Creation on a Mobile Phone

During the offline phase, active scans for WLAN signals from surrounding access points (APs) are executed with a mobile phone at several reference positions. The measured signal strength values are enhanced with the viewing direction and the pixel coordinates of the reference position on a bitmap of the floor. The viewing direction is obtained by the digital compass of the smartphone, the position is assigned by tapping on a zoomable and scrollable map displayed on the screen of the mobile. Finally, these values (in the following referred to as fingerprints) are stored in a database. At each reference position, four fingerprints are created, one in the direction of each axis of the specific building. The alignment along the axes of the building instead of the geographic directions is carried out to improve the accuracy of the application in tracking scenarios since most users move along the main axes of a building, e.g., when walking down a corridor. For each fingerprint, five scans are executed and the average of the received signal strengths is stored in the database to reduce the impact of short-time fluctuations. Furthermore, the orientation of the phone, which is derived from the mobile phone's compass, is averaged throughout the sampling time and also stored in the database. This is done to remedy the disturbances of the magnetic field inside of buildings, especially near electronic sources or large amounts of metal.

B. Deterministic Location Estimation

During the online phase, SMARTPOS utilizes a deterministic positioning algorithm based on weighted kNN to estimate the approximate position of the user. WLAN signal strength measurements are carried out in a continuous fashion and for each measurement m the current orientation o of the phone is measured by its digital compass.

The orientation is considered to represent the approximate viewing direction of the user and hence implicitly yields the information about the attenuation of his body. The online RSSI values should therefore not be compared to all fingerprints in the database due to possible influence of the human body, but only to those fingerprints that correspond to a similar viewing direction to o during the offline phase. Since the viewing direction is retrieved from the noisy readings of the compass, the orientation is averaged over the duration of each scan. This mechanism could also be replaced by advanced filtering algorithms to reduce the impact of outliers. SMARTPOS considers only a subset S of all fingerprints in the database containing those with a maximal deviation of $50°$ from o and is therefore able to reduce the number of fingerprints matched in the online phase to an extent of 25% of the database size.

On the remaining subset S of filtered fingerprints, the nearest neighbours in signal space with respect to m are computed. SMARTPOS uses a sophisticated distance metric for the comparison of two RSSI measurements (i.e., the online measurement m and a fingerprint $f \in S$): Each measurement contains the information about all RSSI values with the MAC address of the AP, which sent the signal. Since at a given position only signals of a subset of all access points in the building can be received, the question arises how to treat missing signal strength information in one of two compared measurements. One possibility would be to assign a fixed value *MIN* to the RSSI of all access points missing in one measurement. This mechanism favors combinations of measurements, where signals by an AP are of very small strength in one measurement and missing in the other instead of combinations, where a high RSSI value in one measurement is missing a counterpiece in the other. The value of *MIN* should be below the minimal RSSI value measureable by the device. The other possibility is to ignore all signal strength information missing at least in one of the compared measurements. Based on the results of a detailed evaluation (see Section III-C) SMARTPOS utilizes the second approach, which is expected to be more robust in the case a new AP is turned on or an existing AP is turned off. Nevertheless, a minimum overlap of at least three APs is required to avoid choosing wrong neighbors due to propagation symmetries in larger environments.

Based on the Euclidean distance $d_i = dist(m, f_i)$ in signal space the subset $N \subset S$ of the k nearest neighbours is computed. In addition, SMARTPOS assigns a weight w_i to each fingerprint $f_i \in N, i \in \{1, \ldots, k\}$ which is indirectly proportional to the distance in signal space. It computes after the following formula:

$$w_i = \left(d_i \sum_{j=1}^{k} \frac{1}{d_j} \right)^{-1} \tag{1}$$

It is easy to see that the w_i are normalized since $\sum_{i=1}^{k} w_i = 1$. For the computation of the user's position l, SMARTPOS calculates the weighted average of $l_i, i \in \{1, \ldots, k\}$, l_i being the reference position of the fingerprint f_i:

$$l = \sum_{i=1}^{k} l_i w_i \tag{2}$$

C. SMARTPOS Evaluation

For the evaluation of the SMARTPOS system, two sets of fingerprints were manually collected under laboratory conditions, i.e., without anybody around, in a part of our university building. All RSSI information was gathered with a HTC Desire. The first set is arranged in an approximate grid of 79 reference positions with fingerprints measured in the direction of all four main axes of the building, which results in 316 fingerprints in total (the grey dots in Figure 1). The second set is a much smaller set of 64 fingerprints at 16 pseudo-randomly distributed reference positions (again measured in the direction of all four axes) within the coverage of the database and is used as substitution for online measurements (the black dots in Figure 1). This ensures that our results originate from an identical setting for all the different location estimators. The estimators are evaluated in respect to four criteria according to [2]: the accuracy as the mean position error, the precision as the maximal and the standard deviation, and the complexity as the number of compared fingerprints. The question of scalability, cost and robustness is not considered, since the scalability and the cost are the same in all systems and the robustness is hard to measure. In the following, the results from a detailed evaluation of SMARTPOS in the described setting are presented and discussed. SMARTPOS is evaluated as follows: First, the deterministic kNN approach is analyzed and the settings of several parameters compared to each other. The questions of assigning a weight to the nearest neighbors and whether missing signal strength information should be considered or ignored are discussed and the impact of the user's orientation on accuracy and precision presented. In a consecutive step, an optimal value for k is determined for SMARTPOS. Finally, the usage of orientation information in a Naive Bayesian Estimator is analyzed.

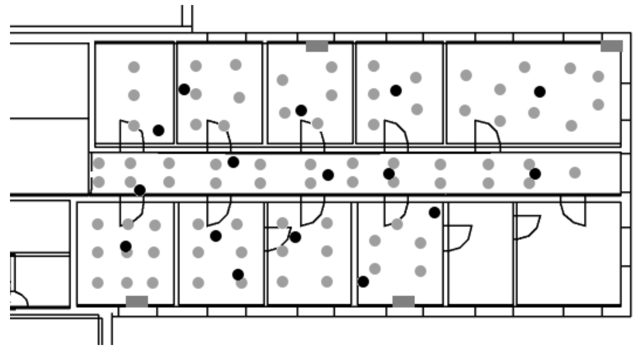
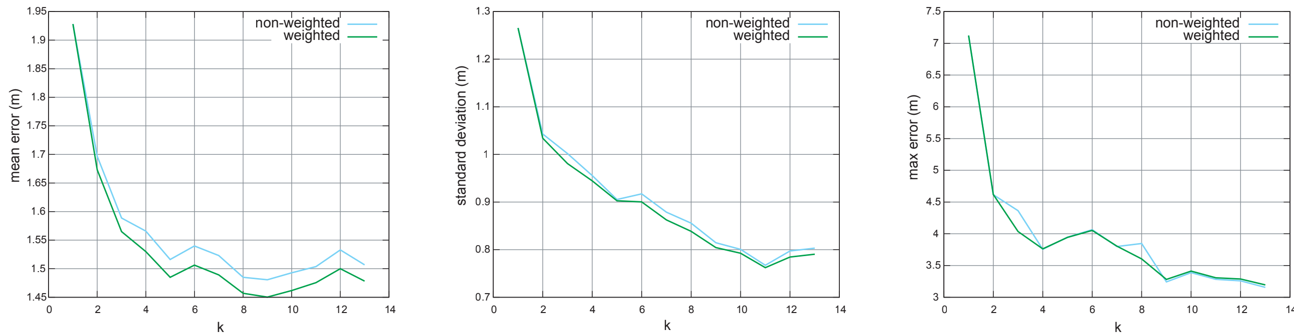


Figure 1: Reference database (gray dots) and online testset (black dots). APs are displayed as grey rectangles.

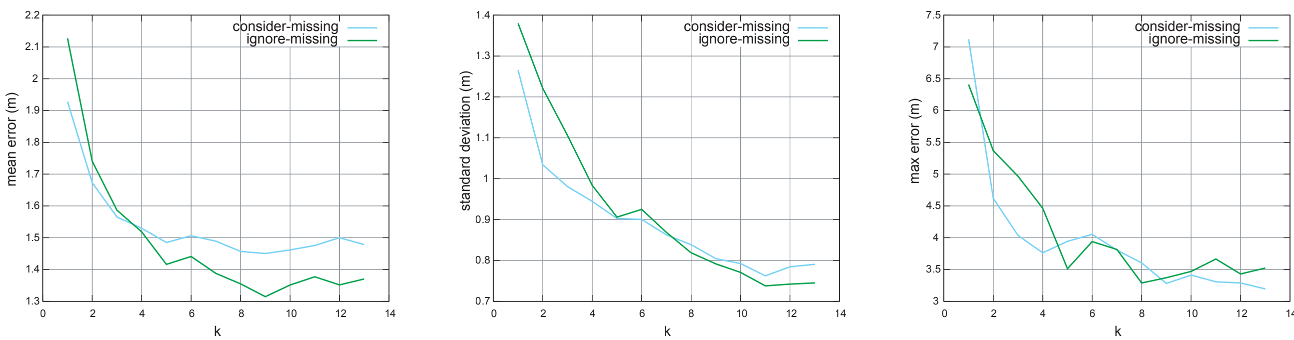
1) Weighted or Non-Weighted kNN: When using a kNN approach together with WLAN fingerprinting one has to decide whether just to compute the center of the nearest neighbors or to add a weight to each of the k-nearest neighbors according to the distance in signal space and then calculate the center of mass. With SMARTPOS, we evaluated both approaches for variable k. Figure 2 shows the results. The weighted approach behaves similarly, but performs better for each $k > 1$. The same applies for the deviation while the maximum error shows no significant difference except for two outliers ($k = 3$ and $k = 8$), for which the weighted approach also performs better. SMARTPOS therefore utilizes a weighted kNN as described in Section III-B.

2) Treatment of Missing RSSI: In Section III-B, two approaches for the treatment of missing signal strength



(a) The mean error in meter (b) The standard deviation in meter (c) The maximum error in meter

Figure 2: Comparison of weighted and non-weighted kNN



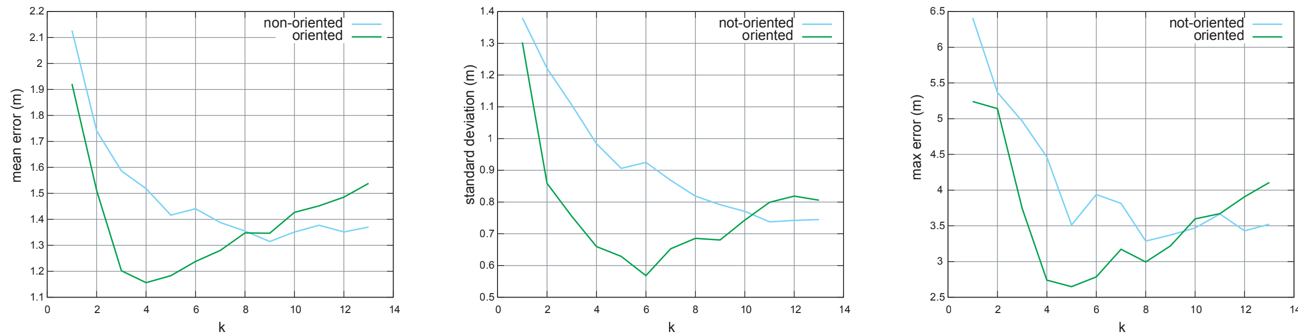
(a) The mean error in meter (b) The standard deviation in meter (c) The maximum error in meter

Figure 3: Comparison of considering and ignoring missing RSSI values

information when comparing two RSSI measurements are described. One considers the information by assigning a minimal value of -100 dBm for the missing RSSI information, the other ignores all RSSI values from APs measured only in one of the two compared measurements. Both approaches were tested for a variable k and the results are presented in Figure 3. The accuracy of a system ignoring missing values is higher than the accuracy of a system considering the information for each $k > 3$ and also offers a minimum mean error for $k = 9$. The deviation only becomes smaller for each $k > 7$ with the minimum for $k = 11$, while the maximum error oscillates and therefore adds little information. Hence, SMARTPOS ignores missing RSSI values as long as signals of at least three common APs have been measured in fingerprint and measurement.

3) Impact of Orientation Information: The most profound innovation of SMARTPOS is the usage of orientation information in a deterministic location estimation system on a smartphone. With the filtering of the fingerprints in the offline database with respect to the orientation information of the user, the complexity of the online matching can be quartered (when using the state of the art four directions for each reference position) and the accuracy and precision increased by a considerable amount. Figure 4 shows the results of the tests. The mean error is much smaller when using the orientation information and also reaches its minimum of 1.16 m for $k = 4$, while the approach without orientation information reaches its minimum of 1.31 m for $k = 9$. The minimal deviation of 0.57 m for $k = 6$ is also much smaller than the minimal deviation of 0.74 m for $k = 11$ without considering the orientation. The same is true for the maximum error, which is minimal for $k = 5$ with a value of 2.65 m when considering the user's orientation, whereas without the orientation information the minimum is 3.29 m for $k = 8$. The much smaller number of k when using the orientation approach can be explained by the fact that the number of fingerprints for comparision is quartered and each online measurement has at most 4 neighbors in the grid, while without the filtering of the user's orientation the number of neighbors can increase to a total of 16 neighbors, because 4 fingerprints are stored for each reference position. In conclusion, SMARTPOS utilizes the orientation information of the user to improve accuracy and precision of the location determination, while reducing the complexity at the same time.

4) Determination of k: Based on our experiments with SMARTPOS, we recommend utilizing an orientation-based weighted kNN approach with k, the number of neighbors,



(a) The mean error in meter

(b) The standard deviation in meter

(c) The maximum error in meter

Figure 4: Comparison of considering and ignoring the user's orientation

set to 4. For the comparison of measurements one should ignore all signal strength information of each AP which is missing in at least one of the measurements when at least three APs are in common. With these parameters, the system offers the lowest mean error of 1.16 m of all possible fixed assignements for k with an acceptable deviation of 0.66 m and a small maximum error of 2.74 m. However, it is shown in Section VI that by dynamically choosing k for each measurement separately the error can be further reduced.

5) Orientation and the Naive Bayesian Estimator: The influence of filtering fingerprints according to their orientation on deterministic kNN positioning has been described. To get a deeper understanding of what influence the reduction of the search space according to the viewing direction has on indoor positioning, we chose to evaluate on the most simple (and often most effective) way of inducing a position from given measurements: Assuming that the variance in measurements is normally distributed, we estimate the mean and variance of a set of measurements taken in the same room and reuse this information for identification.

In order to do so, we assigned a label with each fingerprint specifying the room that it lies in. The long corridor has been cut into three rooms to reduce the variance of measurements in this long area as depicted in Figure 5. Using this labeled data we constructed a Bayesian Estimator, which calculates for each pair of AP and room label the mean RSSI, its standard deviation, weighted sum and precision and reuses them for classification. We tested the classification performance with 10-fold stratified cross-validation training on 90% and evaluation on the remaining 10% of the data.

We used this technique on five different datasets: A dataset for each quadrant and a dataset where a random subset of 25% of all measurements in all directions were taken. In this way we achieve comparable training set sizes.

The results from this experiment are negative: A Bayesian classification of room-labels performs better on the total set of measurements than on the direction-dependent subsets. The results are given in Table I. Hence, for a system based on Bayesian estimation theory, we propose not to use the direction as a filter.

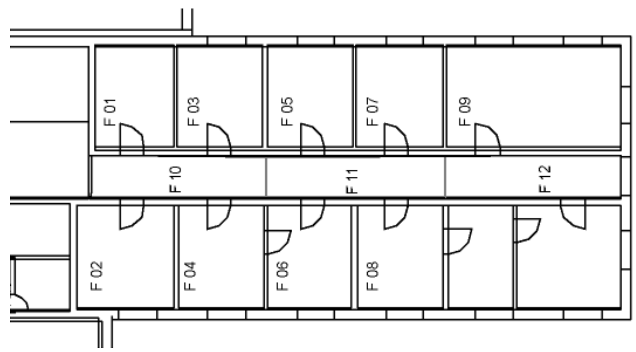


Figure 5: Labeled rooms for the Naive Bayesian Estimator.

Table I: Evaluation results

Dataset	Number of Fingerprints	Success Rate
All directions	78	79%
North	72	62.5%
West	77	70.13%
East	82	65.85%
South	82	71.95%

IV. Online Error Estimation

Several factors influence the occuring positioning errors of WiFi fingerprinting systems and thus also affect the SMARTPOS system. For many application scenarios, these errors make an online error estimator necessary such as the application of SMARTPOS in Bayesian filters or scenarios of location-based access control [18]. In order to propose a good estimator for SMARTPOS, we first examine the properties of occuring errors in the first part of this section. In the second part, the online error estimator which has been developed for the SMARTPOS system is presented and evaluated.

A. The error distribution of SMARTPOS

In order to determine the real occuring errors, a cross-validation on the recorded reference positions has been performed. In this experiment, each of the 316 fingerprints of the reference database has been used as the current measurement m once and was blacklisted in the process of nearest neighbor selection. This tends to cause slightly increased errors, as the reference positions are reduced by m, but the larger size of samples allows to derive a stronger assumption of the error distribution. The position estimation was based on weighted kNN with $k = 4$, considered the orientation and did not punish missing RSSI, i.e., the optimal SMARTPOS setting. The results are shown in Figure 6 as a 2-dimensional histogram, representing the observed error vectors of position estimates relative to their ground truth position. The results give good incidence to assume that the errors on each axis follow approximately a Gaussian distribution. We ignore any correlations on the axes and even the (obviously existing) differences concerning the deviation, since most services working with inaccurate position information expect a circle as error estimator. Thus we model the occuring errors as twodimensional univariate Gaussians, which is also an important constraint for the proposed estimator as shown later in this section. The individual distribution for each axis is shown in Figure 7. Obviously, the errors on the x-axis tend to have a higher standard deviation which is caused by the fact that the recorded reference positions have a larger extent on the x-axis, as depicted in Figure 1. In order to fortify the assumption of normally distributed errors a Wilk-Shapiro test has been performed for each axis for 50 randomly selected samples of the measured errors. The results showed a test statistic of $W_x = 0.960$ for the x-axis and $W_y = 0.955$ for the y-axis. Given a level of significance of $\alpha = 0.05$, the critical value of W is 0.947 for $n = 50$ which is lower than W_x and W_y. Thus, the assumption of normal distribution can not be rejected for the given level of significane, which allows to define an online error estimator for SMARTPOS based on Gaussians.

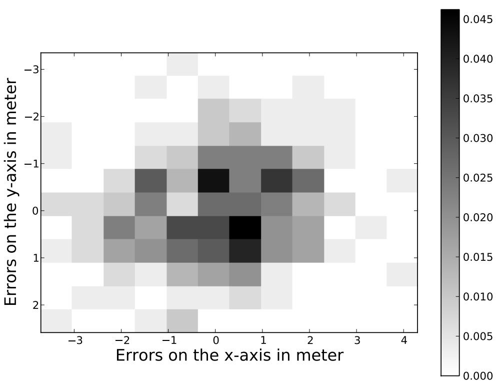


Figure 6: Histogram of the observed error distributions on the x- and y-axis. On each axis, 12 bins have been created for the 316 results.

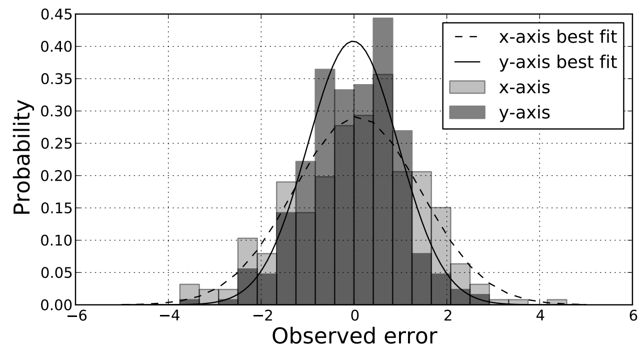


Figure 7: Histogram of the observed error distributions on the x- and y-axis.

B. Estimating positioning errors

As in the last section, the assumption of a normal distribution has been strengthened this section aims at giving an estimation of errors with Gaussians. The mean values correspond to the specific position fix and for each fix, an empirically estimated standard deviation is derived. For $k > 2$, three different error estimation schemes have been defined by Lemelson et. al [13] based on the coordinates $(l_1, l_1, \ldots, l_k)$ in $\mathbb{R}^2$ of k nearest neighbors in signal space. These methods estimate the error, i.e., the geographic distance of the position fix to the ground truth position. For this task, the first method σ_{m1} computes the average geographic distance of the second up to the kth fingerprint to the nearest neighbor:

$$\sigma_{\mathrm{m1}}(l_1, l_2, \cdots, l_k) = \frac{1}{k-1} \sum_{i=2}^{k} \|l_1 - l_i\|_2 \qquad (3)$$

Another modification computes the error estimate as the maximum geographic distance of any selected neighbor to the nearest neighbor:

$$\sigma_{\mathrm{m2}}(l_1, l_2, \cdots, l_k) = \max\left(\bigcup_{i \in \{2,\ldots,k\}} \{\|l_1 - l_i\|_2\} \right) \qquad (4)$$

Finally, a third version estimates the error as the maximum geographic distance of any two fingerprints in the sequence of selected nearest neighbors:

$$\sigma_{\mathrm{m3}}(l_1, l_2, \cdots, l_k) = \max\left(\bigcup_{(i,j) \in \{2,\ldots,k\} \times \{2,\ldots,k\}} \{\|l_i - l_j\|_2\} \right) \qquad (5)$$

As proposed by Lemelson et al. [13], for each of these methods the positioning error is estimated as σ_{mi} with $i \in \{1, \ldots, 3\}$ and the user is assumed to be on a circle with radius σ_{mi} centered at the position fix l. However, in general, the distribution of errors approximately follows a bivariate Gaussian as shown above. In order to allow a more realistic estimation of the occuring error, one could approximate the real error distribution under the assumtion that the errors E_x and E_y on both axes are uncorrelated. Given a position fix $l = (l_x, l_y)$, this allows to define two Gaussians $E_x \sim \mathcal{N}(l_x, \sigma_{mi})$ and $E_y \sim \mathcal{N}(l_y, \sigma_{mi})$ for each $i \in \{1, \ldots, 3\}$ describing a probability distribution for the ground truth position on each axis. In the following, we employ this methodology for defining an univariate Gaussian as error estimator of SMARTPOS.

For the SMARTPOS system, we propose a new error estimator which also derives an univaiate Gaussian centered at l but in contrast to σ_{mi} with $i \in \{1, \ldots, 3\}$ is not only based on the positions of the k nearest neighbors, but also on their corresponding weight and the position estimate. The basic idea is to capture the closeness of the nearest neighbors to the derived position fix, i.e., to derive an estimate for the precision. Given a measurement m at the ground truth position *gtp*, the only information that is online accessible is the estimated position l and the weight w_i of each fingerprint f_i according to the measurement m. For both axes, the standard deviation σ_{m4} is estimated as the weighted average of the distance of the l_i with $i \in \{1, \ldots, k\}$ to the position estimate l:

$$\sigma_{\text{m4}}(l, l_1, l_2, \ldots, l_k) = \sum_{i=1}^{k} w_i \|l - l_i\|_2 \tag{6}$$

Again, a cross-validation was performed as described above with an additional computation of the error estimations σ_{mi} for the described estimators $1-4$. To evaluate each of these, we propose to standardize the set of observed error distances on each axis with the corresponding σ_{mi}. This allows to compare the standardized samples against the standard normal distribution $\mathcal{N}(0, 1)$. The more likely these samples were drawn from $\mathcal{N}(0, 1)$, the better the given estimator. Thus, for each position fix $l = (l_x, l_y)$ for each axis, the standard score has been computed according to $(gtp_x - l_x)/\sigma_{mi}$ and $(gtp_y - l_y)/\sigma_{mi}$. The fits of the standardized samples to $\mathcal{N}(0, 1)$ have been evaluated using qq-plots, which are depicted in Figure 8 for the x-axis and in Figure 9 for the y-axis. Compared to the straight line $y = x$, the results indicate that the derived Gaussians for σ_{m4} show the best approximation of the real error within the quantiles from -2σ to 2σ for both axes. It can also be seen that the reduction to univariate Gaussians does not prevent the estimator from fitting very well to the real error distribution on both axes. The proposed approach thus returns more accurate error estimations than the methods σ_{m1}, σ_{m2} and σ_{m3}. The few outliers suggest that the derived Gaussians tend to underestimate the probability of large real errors. These underestimations also occur with the other evaluated methods and indicate that large errors follow another not even necessarily Gaussian distribution. Nevertheless, the results show that based on σ_{m4}, the derived probability

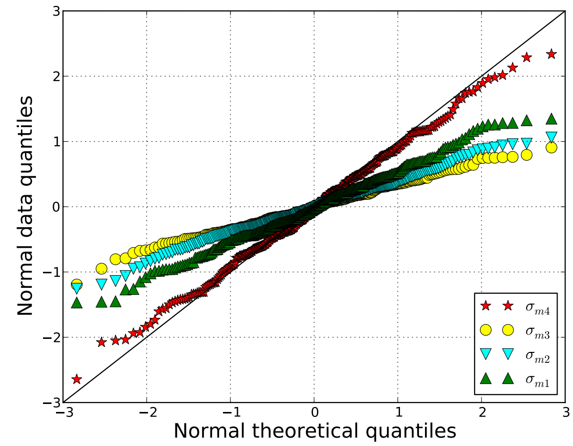


Figure 8: The qq-plot of the generated test data on the x-axis.

density functions for the ground truth position of a position fix correlate with the real error very well. The estimators 1-3

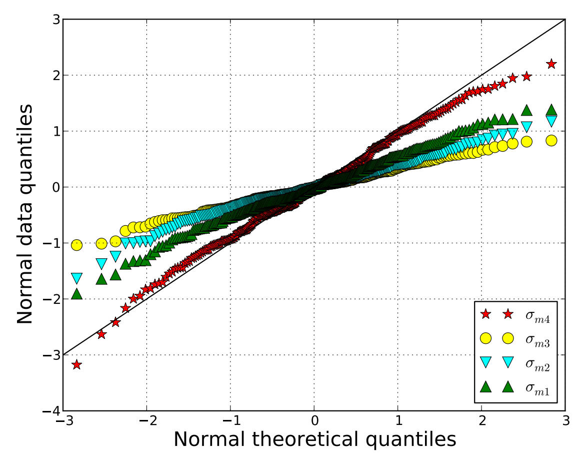


Figure 9: The qq-plot of the generated test data on the y-axis.

tend to underestimate the error with respect to the straight line $x = y$ even within the 2σ quantiles. Clearly, σ_{m3} tends to highly underestimate the errors in nearly all cases. The methods σ_{m2} and σ_{m1} perform slightly better but also tend to underestimate the errors much more than σ_{m4}.

However, we also experimented with the presented error estimators for $k \neq 4$ and observed that still the proposed estimator σ_{m4} has the best fit to $x = y$ but tends to increasingly underestimate the occuring errors with a growing k. Given the obtained results, we suggest that for a given global parameter of k, the best error estimator should be selected using the presented methodology. Compared to existing error estimators, the derived probability density functions in SMARTPOS are expected to yield more robust results in real applications.

V. Density of Radiomaps

As the mean and maximum errors of SMARTPOS are subject to the number of selected nearest neighbors, the density of recorded fingerprints in the underlying radiomap plays an important role. Hence, an interesting aspect is how the mean and maximum errors correlate to this density. To examine this correlation, an experiment has been conducted where these measures have been computed using the online testset against the presented reference database, whose density has been reduced in each iteration by 5%. In each iteration, the 5% of fingerprints to remove have been randomly chosen. If a fingerprint has been picked for removal, the other 3 fingerprints on its location have been removed too. This experiment has been conducted 20 times in sequence. The measured errors for each iteration have been merged and are depicted in Figure 10. The observed

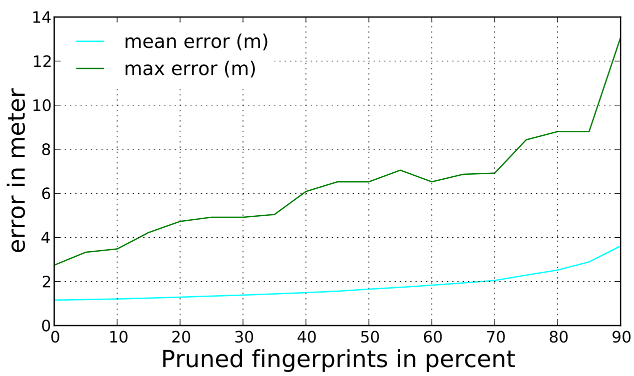


Figure 10: The average positioning error in meters for iteratively reduced fingerprint densities.

results indicate, that the density of the radiomap has much stronger influence on the maximum error as on the mean error. In detail, the maximum error already doubles for a reduction of the reference positions of $35-40\%$ while the same holds for the mean error for a reduction of about 80%. However, both, the mean and the maximum error triple for a reduction of 90%. The results give good reason to assume, that a more dense reference database would only show low impact on the mean error, as the mean error seems to be converging for higher densities. However, as the maximum error is increased by approximately 0.5 m for a reduction of only 5%, we expect that the maximum error of SMARTPOS could be even further reduced by a reference database with a higher density.

VI. Dynamically Calculating the Optimal Number of Nearest Neighbors

As our previous experiment has shown, the number k of fingerprints considered for the position estimation in SMARTPOS has a strong impact on the mean and maximum error of the positioning system. However, it may

function *SMARTkNN*(m, *min_k*)
  l ← *weighted_center_of_mass*(*get_NN*(*1*, m))
  k ← *min_k*
  while $k <$ |*fingerprints*| **do**
    kNN' ← *get_NN*(k, m)
    l' ← *weighted_center_of_mass*(*kNN'*)
    kNN" ← *get_NN*($k+1$, m)
    l" ← *weighted_center_of_mass*(*kNN"*)
    if $\|l" - l'\|_2 < \|l' - l\|_2$ **then**
      l ← *l'*
      $k \leftarrow k+1$
    else
      return *l*
    end if
  end while
  return *l*
end function

Figure 11: The proposed *SMARTkNN* algorithm.

depend on the environment and therefore scenarios exist where no analysis of an optimal k has been performed or is even impossible. Additionally, the covered site might have very diverse properties with respect to the density of recorded fingerprints, the number of receivable access points, or building specific singularities, which makes a fixed global k too inflexible. Even for very uniform scenarios, like the SMARTPOS test environment, a dynamic k might decrease the mean error. To test this hypothesis, we propose SMARTkNN as an extension of the SMARTPOS system. Its pseudo-code is shown in Figure 11.

The algorithm works by iteratively increasing k, and computing a position fix for the current k, $k-1$ and $k+1$ based on Formula 2. It iteratively continues up to that value of k after which the position fixes start to diverge. First, the location l of the nearest neighbor is determined and stored in the variable l. The variable k is initialized with *min_k*. Within the loop, the k and $k+1$ nearest neighbors are determined and corresponding position fixes l' and l'' are computed. If the distance of l' to its predecessor l is smaller than the distance to its successor l'', the loop terminates and returns k. This represents the first optimum for the number of nearest neighbors, as for larger values of k the position fixes begin to diverge. In the other case, the position fixes seem to be converging and the search for a larger k is continued. The search also terminates if the value of $k+1$ corresponds to the number of recorded fingerprints.

The SMARTkNN algorithm was evaluated with the online testset against the reference positions for lower bounds *min_k* $\in \{2,3,4\}$ and was compared to the $k=4$ strategy and the *optimal* k strategy. The latter is suitable for evaluating the proposed algorithm as the theoretical optimal k can be used as a reference value. The results are depicted in Table II and as a histogram comparing the

Table II: Evaluation results

Method	Mean. Error	Max. Error	Std. Deviation
fixed $k = 4$	1.16 m	2.74 m	0.65 m
dynamic $k \geq 4$	1.17 m	2.79 m	0.60 m
dynamic $k \geq 3$	1.10 m	2.65 m	0.65 m
dynamic $k \geq 2$	1.31 m	5.14 m	0.83 m
optimal k	0.66 m	2.14 m	0.51 m

number of chosen k in Figure 12. The optimal number for k is distributed over a large interval while SMARTkNN only picked maximally 8 nearest neighbors and thus had only limited overhead compared to a strategy with a fixed value for k. Furthermore, the complexity of kNN lies much more within the distance calculation to every possible fingerprint and the sorting of the results than in the position calculation for given neighbors.

Compared to the SMARTPOS algorithm for a fixed k, SMARTkNN could slightly reduce the mean error from 1.16 m to 1.10 m and the maximum error from 2.74 m to 2.65 m for a lower bound of *min_k* $= 3$ without increasing the standard deviation. A lower bound of 2 or 4 could not improve the results compared to a fixed $k = 4$. Given Figure

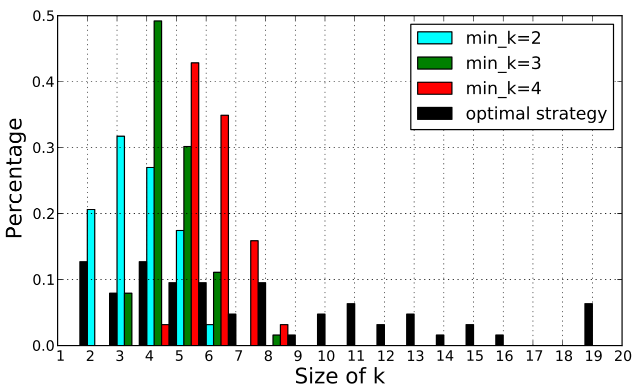


Figure 12: The distribution of the dynamically chosen number of nearest neighbors for SMARTkNN compared to the optimal strategy.

13, the cumulative error distribution of SMARTkNN with $k \geq 3$ has the best fit to the optimal strategy and especially a better fit than SMARTPOS with $k = 4$. An interesting aspect is the mean size of the dynamic k: *min_k* $= 3$ resulted in a mean of 3.5 for k with a standard deviation of 0.90, which subsequently indicates, that values from $k = 3$ to $k = 5$ were preferably selected. This fits quite well to the results observed in Figure 4, where the mean error had its minimum for $k = 4$ with very similar values for $k = 3$ and $k = 5$. For *min_k* $= 2$ and 4, in each case the mean size of k was further away from this minimum with mean values for k of 2.5 and 4.7. The optimal strategy had a mean value of $k = 4.8$ with a deviation of 6.6.

Concluding, the SMARTkNN algorithm showed improved

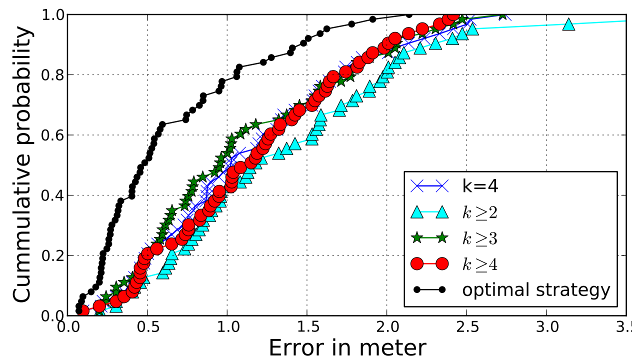


Figure 13: The cumulative error distribution of SMARTkNN compared to SMARTPOS and the optimal strategy.

results compared to SMARTPOS. However, even here we have a strong dependence on the new parameter *min_k*. As the evaluation results indicate, a minimum value of *min_k* $= 3$ should be chosen in the presented scenario. The lower bound of 3 also theoretically accounts for reducing the mean error compared to a lower k as the number of possible candidate points for position fixes is largely increased compared to $k = 2$ or $k = 1$ using the proposed computation over weighted center of mass.

VII. Conclusion and Future Work

In this paper, we presented SMARTPOS, a positioning system on a smartphone based on deterministic WLAN fingerprinting and a digital compass. SMARTPOS utilizes a weighted kNN approach with $k = 4$ and with a distance metric in signal strength space, which ignores RSSI values from access points visible only at one fingerprint. Furthermore, we proposed SMARTkNN, an extension for SMARTPOS, which uses a dynamic k instead of a fixed number of nearest neighbors. It iteratively increases k, computes a position fix for the current k, $k - 1$ and $k + 1$ and continues with increasing k until the position fixes start to diverge. In this algorithm, the first position fix is initialized with the position of the nearest neighbor, while the minimum number of nearest neighbors involved in the next fixes is a parameter of SMARTkNN.

To give an impression of the system's performance, we analyzed the impact of several parameters on SMARTPOS. We conclude that a weighted approach results in more accurate and precise results than a non-weighted approach. Ignoring missing RSSI values provides better results than assigning a minimal value, at least for higher values of k. In our setting, this was the case for $k > 3$ in the oriented approach and for $k > 7$ in the approach without the user's orientation. With adding the user's orientation, SMARTPOS is able to reduce the mean positioning error to 1.16 m and the variance to 0.66 m. The maximal error in this case is 2.74 m, which is 55 cm smaller and therefore much better than the minimal maximum error of 3.29 m in all experiments without the orientation information. We therefore conclude that the user's orientation should be considered in deterministic 802.11 fingerprinting. The error was reduced even more by introducing SMARTkNN: The mean error was reduced to 1.10 m and the maximum error to 2.65 m without increasing the variance. The cumulative error distributions for several minimum values of k were compared to the optimal strategy and we found that the best results could be obtained for the strategy with $k \geq 3$. However, compared to a fixed $k = 4$, the improvement was quite small but we assume that the main advantage of SMARTkNN unfolds in very diverse scenarios, where a fixed k yields too much inflexibility.

Furtheremore, we examined the reduction of the density of the underlying fingerprint database. We found out that the reduction has a large negative effect on the maximum error but a much smaller influence on the mean error. The reduction of about $35 - 40\%$ doubled the maximum error while the density has a weaker impact on the mean error which doubles for a reduction of 80%.

Last but not least, the error distribution for SMARTPOS was evaluated and found to be normally distributed on each axis. Given this information, a novel online error estimator was defined which employs the weighted average distance of the nearest neighbors to the position fix to derive a Gaussian probability distribution. The derived distributions are univariate Gaussians centered at the position estimate. An evaluation of with several existing estimators showed that our approach gives the best approximation of the real errors. The quality depends nevertheless on the parameter k and degrades for $k \neq 4$ in our testenvironment.

REFERENCES

[1] M. Kessel and M. Werner, "SMARTPOS: Accurate and Precise Indoor Positioning on Mobile Phones," in *Proceedings of the International Conference on Mobile Services, Resources, and Users (MOBILITY'11)*, 2011, pp. 158–163.

[2] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 37, no. 6, pp. 1067–1080, 2007.

[3] P. Bahl and V. N. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system," in *19th Annual Joint Conference of the IEEE Computer and Communications Societies*, ser. INFOCOM 2000, vol. 2, pp. 775–784.

[4] M. Youssef and A. Agrawala, "The horus wlan location determination system," in *3rd International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys 2005, pp. 205–218.

[5] T. King, S. Kopf, T. Haenselmann, C. Lubberger, and W. Effelsberg, "Compass: A probabilistic indoor positioning system based on 802.11 and digital compasses," in *1st International Workshop on Wireless Network Testbeds, Experimental Evaluation and Characterization*, ser. WiNTECH 2006, pp. 34–40.

[6] E. Martin, O. Vinyals, G. Friedland, and R. Bajcsy, "Precise indoor localization using smart phones," in *International Conference on Multimedia*, ser. MM 2010, pp. 787–790.

[7] K. Kaemarungsi and P. Krishnamurthy, "Properties of indoor received signal strength for wlan location fingerprinting," in *1st Annual International Conference on Mobile and Ubiquitous Systems*, ser. MobiQuitous 2004, pp. 14–23.

[8] E. C. L. Chan, G. Baciu, and S. C. Mak, "Orientation-based wi-fi positioning on the google nexus one," in *6th International Conference on Wireless and Mobile Computing, Networking and Communications*, ser. WiMob 2010, pp. 392–397.

[9] F. Evennou and F. Marx, "Advanced integration of wifi and inertial navigation systems for indoor mobile positioning," *EURASIP Journal on Advances in Signal Processing*, vol. 2006, pp. 1–11, 2006.

[10] M. Kessel, M. Werner, and C. Linnhoff-Popien, "Compass and wlan integration for indoor tracking on mobile phones," in *The Sixth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM'12)*, 2012, pp. 1–7.

[11] Widyawan, "Learning data fusion for indoor localisation," Ph.D. dissertation, Cork Institute of Technology, 2009.

[12] O. Woodman and R. Harle, "RF-based Initialisation for Inertial Pedestrian Tracking," in *Proceedings of the 7th International Conference on Pervasive Computing (Pervasive 2009)*, 2009, pp. 238–255.

[13] H. Lemelson, M. Kjaergaard, R. Hansen, and T. King, "Error estimation for indoor 802.11 location fingerprinting," in *4th International Symposium on Location and Context Awareness*, ser. LoCA 2009, pp. 138–155.

[14] C. Beder, A. McGibney, and M. Klepal, "Predicting the expected accuracy for fingerprinting based wifi localisation systems," in *Proceedings of the 2nd International Conference on Indoor Positioning and Indoor Navigation (IPIN 2011)*, 2011.

[15] M. Roshanaei and M. Maleki, "Dynamic-knn: A novel locating method in wlan based on angle of arrival," in *Proceedings of the 2009 IEEE Symposium on Industrial Electronics and Applications (ISIEA 2009)*, 2009, pp. 722–726.

[16] B. Altintas and T. Serif, "Improving rss-based indoor positioning algorithm via k-means clustering," in *Proceedings of the 11th European Wireless Conference*, 2011, pp. 681–685.

[17] B. Shin, J. Lee, T. Lee, and H. Kim, "Enhancing weighted k-nearest neighbor algorithm for indoor wi-fi positioning systems," in *Proceedings of the 8th International Conference on Computing Technology and Information Management (ICCM 2012)*, 2012, pp. 574–577.

[18] P. Marcus, M. Kessel, and C. Linnhoff-Popien, "Securing mobile device-based machine interactions by employing user location histories," in *Security and Privacy in Mobile Information and Communication Systems Social Informatics and Telecommunications Engineering (MOBISEC'12)*, 2012, pp. 81–92.