A field test of parametric WLAN-fingerprint-positioning methods

A Field Test of Parametric WLAN-Fingerprint-Positioning Methods

Philipp Müller, Matti Raitoharju, and Robert Piché
Department of Automation Science and Engineering, Tampere University of Technology,
P.O. Box 692, 33101 Tampere, Finland
Emails: {philipp.muller, matti.raitoharju, robert.piche}@tut.fi

Abstract—**Fingerprint-based (FP) positioning methods determine a receiver's position using a database of radio signal strength measurements that were collected earlier at known locations. For positioning with WLAN signals, nonparametric methods such as the weighted k-nearest neighbour (WKNN) method are widely used. Due to their large data storage and transmission requirements those methods are infeasible for large-scale mobile device services. In this paper we consider parametric FP methods, which use model-based representations of the survey data. We analyse the positioning performance of those methods using real-world WLAN indoor data and compare the results to those of the WKNN method.**

I. Introduction

In the last decade indoor positioning techniques have received extensive attention, and are nowadays an essential feature of many commercial and public service networks. Since many of these applications have to run on small mobile devices, the positioning algorithms have strict limits on allowed energy, memory, bandwidth, and computational resources.

Positioning in indoor environments can rely on measurements from e.g. an inertial measurement unit (IMU) or from radio signals such as Bluetooth, wireless local area networks (WLAN), or ultra-wideband (UWB). In contrast to Bluetooth and UWB, the infrastructure required for WLAN-based positioning, namely WLAN access points (APs) and receivers in user devices (UEs), is already in place. Instead of using signal propagation time, most WLAN-based positioning algorithms exploit the correlation between the received signal strength (RSS) and the UE's location (see e.g. [1, p. 47]). This choice is supported by the fact that media access control (MAC) addresses of APs and the corresponding RSS values are already contained in transmission data [2, pp. 57 ff.], [3, 4], so no changes to the AP software are required.

Because modelling signal propagation, especially in indoor environments, is rather complex, (nonparametric) location fingerprinting methods are widely applied for positioning [5]. Those methods estimate the UE's position by comparing the list of current AP RSS measurements to a database (called a radio map) of information (called fingerprints) on APs and their corresponding RSS values for known positions.

Fingerprints (also known as location reports, reception reports, or observations) are collected in an offline phase by site survey, war-driving or crowd-sourcing. In addition to the UE's current position, each fingerprint (FP) contains radio characteristic records. When FPs are collected from a WLAN they include, in general, at least identifiers (IDs) for the APs from which the UE received a signal and their corresponding RSS values. For localisation in the online phase, a common approach is to use some variant of the weighted k-nearest neighbour (WKNN) method, a nonparametric estimation method. The idea of the WKNN method is to compute a location estimate as a weighted mean of the k FP locations from the radio map whose vectors of AP signal strength values are in some sense closest to the vector of currently measured AP signal strength values [5]. For overviews on nonparametric location fingerprinting methods we refer the reader to [5, 6] and references therein.

Nonparametric fingerprinting methods have the advantage that modelling the signal propagation is not needed. These algorithms have been shown to be reasonably precise and reliable in indoor environments (see e.g. [5, 7]). Their disadvantage is that they work directly with the FP data, and the size of this database can be a critical issue when FP-based positioning is offered as a large-scale service for mobile devices, especially in cellular telephone networks. For example, the Third Generation (3G) system provides transmission rates of 5 Mbit/s at most; for Second Generation (2G) systems those rates are significantly lower [1, pp. 7 ff.]. For WLANs the size of the database is less critical since these systems provide sufficient data transmission rates. Thus, positioning in real time on a mobile UE may be unfeasible since data transmission from the server might be too time consuming or expensive [8]. In addition, large amounts of data have to be stored on the server [9]. One approach for mitigating this problem is to apply data-compression to the radio map [10, 11]. A more fundamental way to address the issue is to use parametric (model-based) FP methods. Those methods use models with small numbers of parameters to describe the FP data, which besides data transmission is also beneficial for the positioning.

In this paper we present an experimental comparison of some of the parametric FP methods that have been proposed in the literature. We start with a brief summary of the methods. In Section II we look at a parametric FP approach that uses elliptical probability distributions for modelling the area in which an AP's signal can be received. Section III is dedicated to a signal propagation path loss model that is calibrated from FP data. In Section IV we consider an approach that uses mixtures of Gaussian distributions to approximate multimodal distributions. This technique is useful for nonlinear and/or non-Gaussian systems for which traditional approaches such as Kalman filter (KF) and extended KF (EKF) perform poorly. In Section V the performance of these different parametric FP techniques, with and without filtering, is compared in benchmark tests using real-world WLAN measurements in two university buildings. Section VI summarises and concludes

II. Coverage area models

In [6, 8] a computationally light method for parametric fingerprinting is proposed. In order to reduce the size of the FP radio map the authors represent the coverage area (CA; aka reception region) of any AP by an elliptical probability distribution, which can be represented by five parameters [12]. This probability distribution represents only the region in which a signal from the AP can be received; other than an implied reception strength threshold, it gives no information about the RSS. The approach enables fast transmission of the radio map to a UE [6, 8] and fast computation of the UE's position.

A. Coverage area estimation

The coverage area is modelled in [6, 8] by a posterior distribution for the ellipse parameters θ given the FP locations $\mathbf{z} = \{\mathbf{z}_1, \mathbf{z}_2, .., \mathbf{z}_n\}$ where the AP was heard. The distribution is given by Bayes' rule

$$p(\theta|\mathbf{z}) \propto p(\mathbf{z}|\theta)p(\theta), \tag{1}$$

where the likelihood and prior are Gaussian. In other words, the CA is modelled by fitting the mean and covariance of a multivariate Gaussian to the data. Alternatively, to obtain a fit that is robust to outliers, the likelihood and prior could be Student-t distributed [12, 13]. A Bayesian formulation of the regression problem has two advantages. Firstly, the Bayesian prior $p(\theta)$ allows one to exploit information about "typical" coverage areas, which is crucial when only a few FPs are available [6, 8]. Such information is available through experimental studies. For example, the typical reception range for WLAN in indoor environments is 20–50 m [2, p. 9]. Secondly, using Bayes' rule allows recursive estimation and updating of estimates [8].

The CA method considered above ignores the specific RSS values corresponding to IDs of observed APs. Hence it is less sensitive to changes in the radio environment than fingerprinting methods that use these values. This gain in robustness, however, comes generally at the cost of lower accuracy compared with nonparametric fingerprinting methods (e.g. WKNN), which besides FP locations and IDs of APs observed in each FP often[1] also store corresponding RSS values, and use them in the positioning phase.

A coverage area method that uses RSS information is proposed in [6, 13]. Instead of storing only one CA per AP in the database, several CAs per AP are stored, which are modeled from FP data that is grouped according to RSS. In [13] the authors examine the use of one, two and three CAs per AP assuming both Gaussian and Student-t distribution for location reports. FPs are grouped based on their RSS values and different CAs are generated using only location reports of their corresponding group. An important feature is that any FP can be part of more than one group. Three different grouping rules are considered: RSS-level, n strongest APs of each FP and $x\%$ strongest APs of each FP; see [13] for details.

[1]Some authors store RSS-based rankings of AP-IDs or RSS ratios or RSS differences rather than the measured RSS values.

B. Positioning using coverage areas

A position estimate for a UE using coverage areas [6, 8, 13] can be obtained by applying Bayes' rule. The position estimate and an uncertainty measure of the estimate can be extracted from a Gaussian posterior probability density function $p(\mathbf{x}|\mathbf{c})$ of the UE position $\mathbf{x}$ given a list $\mathbf{c} = (c_1, c_2, ..., c_N)^T$ of APs observed by the UE in the current position. For the conjugate (i.e. Gaussian) prior pdf of this position, a suitable mean and covariance, which represent prior knowledge on UE's position, could be chosen. In case such information is unavailable, setting the covariance very large is justified [6]. For computing the likelihood $p(\mathbf{c}|\mathbf{x})$ [6, 8, 13] it is assumed that prior probabilities of observing c_n are equal for all $n = 1, ..., N$ and that observations are conditionally independent given $\mathbf{x}$.

III. Path loss models

Path loss (PL) models refer to models of the signal power loss L_P or the received signal strength P_{RSS} along a radio link, averaged over large-scale and small-scale fading [1, p. 127]. In the simplest models the PL depends only on the transmit power and the distance a radio wave travels; more complex models take further factors into account. For an overview of propagation mechanisms and PL models we refer the reader to [1, 2, 14] and references therein.

The relation between the RSS and the radio wave's traveled distance can be used for positioning. From RSS measurements and PL models the distances between a set of reference nodes and the target node are estimated, which then enables estimation of the target node's position. However, the position estimate is sensitive to signal noise and PL model parameter uncertainties because the distance-power gradient is relatively small [15]. Consequently, these estimates are generally less accurate than radio-signal based estimates that are derived using angle of arrival or time delay measurements.

A. Parameter estimation for PL models

It had been shown that it is ill-suited for several real-world applications to assume the parameters of the PL models to be known a-priori [3]. Therefore, the parameters should be estimated, simultaneously with the AP positions (in case they are unknown), from FP data consisting of AP-IDs and corresponding RSS values.

A widely applied PL model (see e.g. Dil and Havinga [4]) that is used for describing the RSS dependency of distance

d between AP and UE in any indoor environment is the log-distance model

$$P_{\text{RSS}}(d) = A - 10n\log_{10}(d) + w, \tag{2}$$

where $A = P_{\text{RSS}}(1)$ (apparent transmission power) and n (PL exponent) are the unknown parameters, and w is a zero-mean Gaussian random variable with variance σ_w^2 used for modelling the shadow fading (aka slow fading). Nurminen et al. [16] estimate Gaussian distributions for AP position as well as A and n of the AP's PL model simultaneously using the Iterative Reweighed Least Squares (IRLS) method. Similarly to the method introduced in Section II, the Bayesian approach used in [16] permits updating AP position estimate and PL model parameters as new fingerprint data becomes available.

The algorithm uses uninformative Gaussian priors. Nurminen at al. [16] argue that one can choose the valid prior mean values for A and n arbitrarily, since for large numbers of FPs the posterior distribution is typically unimodal, which is supported by Li's findings [3]. Nevertheless, for cases with limited data, a well-chosen informative prior is beneficial. Various studies yielded values for the PL parameters (e.g. [17]). For the prior AP position more care should be taken in order to prevent IRLS placing the AP in an area of weak RSS values [16]. However, even with such measures it cannot be guaranteed that the algorithm finds the correct AP position, but covariance matrices yielded by the IRLS give the user a tool to distinguish between reliable and unreliable AP position and PL parameter estimates. To account for correlation in measurement errors the authors add a small constant diagonal matrix for the AP position's covariance matrix. The cross-correlation between AP position and PL model parameters is, however, neglected, mainly to limit the number of parameters.

B. Positioning using PL models

Once the parameters of the PL model and the positions for all APs are estimated, trilateration or some other nonlinear estimation technique can be used to estimate the position of the UE. In [16] Nurminen et al. test three different methods that use the PL model (2) with real WLAN data in an indoor office environment: a grid method that uses standard Monte Carlo integration, the Metropolis-Hastings algorithm, and the IRLS. All three methods are analysed using both point estimates and Gaussian distributions for A and n. The tests show that assuming Gaussian distributions for the parameters rather than point estimates is, in general, beneficial. These results are not surprising, since the PL model contains approximation errors [18]. If less FP data is available for estimating the PL model those errors, in general, are larger. Therefore, in such situations it should be beneficial, from a theoretical point of view, to assume more uncertainty in the parameter estimates. Furthermore, the PL exponent n can be assumed constant only for a limited time in an environment [3]. Since those changes are minor as long as the environment stays the same, they can be captured to some extent by assuming some uncertainty in the PL exponent estimate.

In terms of computational demand the grid method and the Metropolis-Hastings algorithm have no edge compared with the WKNN that is used as reference, whereas the IRLS is significantly faster and achieves running times close to those of the CA method presented in Section II.

IV. Gaussian mixtures models

A known disadvantage of the CA approach discussed in Section II is that most of the probability mass is located near the centre of the ellipse that is used for describing an AP's coverage area. However, for weak signals the UE is more likely to be close to the edge of the CA. Therefore, CAs yield in such cases rather poor estimates in the positioning phase [6]. In the previous section we looked at approaches that address such issues by taking into account the RSS in addition to the AP-ID by using PL models. Alternatively, we could apply Gaussian mixture (GM) models (aka Gaussian sum models).

A Gaussian mixture is a convex combination of Gaussian density functions $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, namely

$$p(\mathbf{x}) = \sum_{n=1}^{N} \omega_n \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n), \tag{3}$$

where weights ω_n are nonnegative and sum to one. The main theoretical motivation behind GM and filters based on it is that any density function can be approximated, except at discontinuities, by a convex combination of Gaussian densities arbitrarily closely [19].

A. Representing FP data using GMs

In [20] Kaji and Kawaguchi introduce an approach that represent an AP's RSS distribution as a GM model. Although this approach generally will require more data to be stored in the radio map than the CA approach of Section II, it should still require considerably less storage compared with traditional FP databases.

In their algorithm the collected FP data is first transformed into a point distribution, where the point density depends on the signal strength received in a FP (the higher the RSS the higher the point density). Then the parameters of the GM model, namely mean values $\{\boldsymbol{\mu}_n\}_{n=1}^N$, covariance matrices $\{\boldsymbol{\Sigma}_n\}_{n=1}^N$ and component weights $\{\boldsymbol{\omega}_n\}_{n=1}^N$, are optimised by expectation maximisation [21]. Kaji and Kawaguchi point out that their approach allows updating the GM models as new FP data becomes available. They do not provide an equation or rule for determining the number of components N. In our tests in Section V we use $N = \max([K/40], 8)$, where K is the number of FPs in which the specific AP is observed.

B. Positioning using GM approximation of the PL model

In positioning tasks the statistical model is often non-Gaussian and/or significantly nonlinear (see e.g. [22] for a criterion for significant nonlinearity). Therefore the Bayesian recursion is generally unsolvable in closed form [19]. Applying a GM to solve such generally multimodal systems has the advantage that it can follow the multiple peaks of the probability distribution function, unlike the EKF. However, to ensure fast computing times, which would allow real-time positioning on mobile devices, the number of components should be kept small.

Therefore, Müller et al. [23] introduced a generalised version of GM that relaxes the non-negativity restriction on component weights (GGM). For more details see [24]. Assuming isotropic ranging models, i.e. omni-directional AP antennas, the GGM yields a satisfying approximation of the normalised measurement likelihood with only two Gaussian components, namely

$$p(d|\mathbf{x}) \approx \mathcal{N}(\mathbf{x}_u;\ \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \cdot (1 - \bar{c} \cdot \mathcal{N}(\mathbf{x}_u;\ \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)), \quad (4)$$

with component 1 having positive weight and component 2 having negative weight, and $\mathbf{x}_u$ being the position vector contained in state $\mathbf{x}$. Furthermore, the formula $\bar{c} = c \cdot (2\pi)^{\frac{n_u}{2}} \sqrt{\det(\boldsymbol{\Sigma}_2)}$, where $c \leq 1$ and n_u is the dimension of $\mathbf{x}_u$, ensures nonnegativity of (4). To achieve similar approximation quality the traditional GM would require a significantly larger number of components due to the infinite number of peaks of the likelihood, which would prohibit its application for real-time positioning.

The main principle of the GGM is to use the AP location as mean values $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, rather than some of the normalised likelihood's peaks. It then uses the range measurement d, which in case of WLAN RSS measurements is derived from the PL model (2) using the RSS, to determine the components' covariance matrices. Due to assuming isotropic ranging models those covariance matrices are multiples of the identity matrix, i.e. $\boldsymbol{\Sigma}_1 = \sigma_1^2\mathbf{I}$ and $\boldsymbol{\Sigma}_2 = \sigma_2^2\mathbf{I}$ with $\sigma_2 < \sigma_1$. The values σ_1 and σ_2 are determined using a heuristic model, whose parameter(s) are optimised in the offline phase by either minimising the Lissack-Fu distance [23] or the Kullback-Leibler divergence [24] between the exact normalised likelihood and its GGM approximation.

For positioning a GGM for each observed AP is determined, based on the AP's RSS value. Those likelihood approximations are then multiplied with the prior position estimate to get a new Gaussian mixture. Finally, this mixture is collapsed to a single Gaussian to obtain the posterior position estimate and its covariance.

Kaji and Kawaguchi [20] use a particle filter for positioning for their approach, which represents FPs by GM models.

V. COMPARATIVE TESTING

In this section we compare the performance of the parametric fingerprinting and positioning methods described in the previous sections. We evaluated these methods by analysing their WLAN based positioning accuracy for six test tracks located within two buildings of Tampere University of Technology. For two of the tracks measurements were collected several months later than for the other four tracks, which were collected at the same time as the data used for generating the radio maps. Some of the test tracks had floor changes, which were assumed to be known. The radio maps were built separately for each floor. Table I shows for each floor of the two buildings the number of detected APs, the number of FPs, and the number of test points (TP) for the four tracks collected at the same time as the data used for the radio maps. TPs are points on the test tracks that we positioned in our evaluation.

Building	Floor	APs	FPs	TPs
1	1	200	889	19
1	2	289	243	47
1	3	212	160	22
2	1	154	1 530	168
2	2	186	1 582	33
2	3	148	333	19

Table I
DATA SET SIZES. SOME APS COULD BE HEARD ON SEVERAL FLOORS.

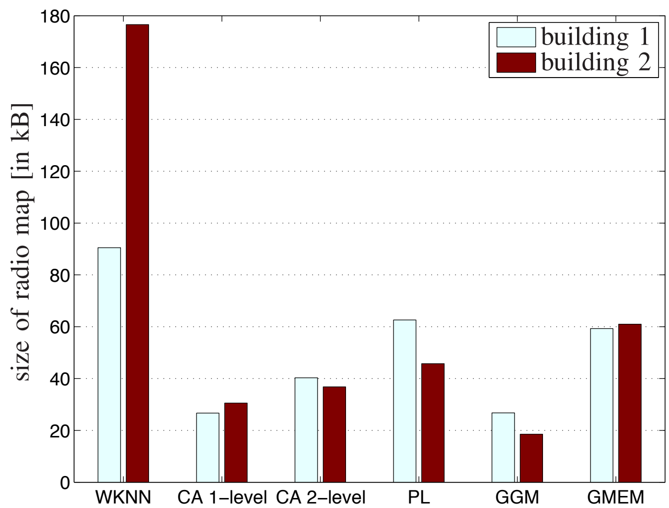
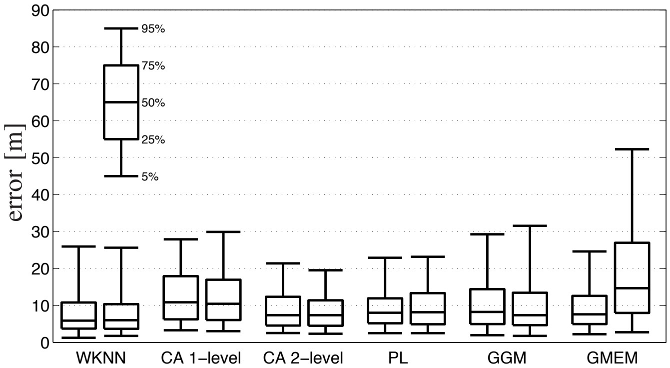


Figure 1. Data storage requirements for radio maps for tested methods in our two test buildings. There were 1 292 FPs in building 1 and 3 445 FPs in building 2

For comparison we implemented CA-based positioning with single CA [8] and 2-level CAs with limit -70 dBm [13], PL model [16], GGM approximation of the PL model [23] and the signal strength estimation model from [20] (denoted GMEM). The standard deviation for RSS based methods was set to 6 dB. In addition to these parametric methods we used a weighted k-nearest neighbours method (WKNN) with $k = 5$ as a reference.

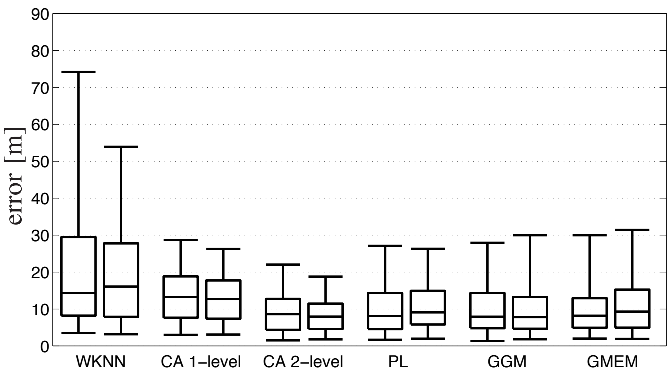
Figure 1 shows the sizes of radio maps for both buildings for each method. The WKNN does not summarise the FPs in any way, and therefore has the highest requirements. All parametric methods reduce the size of the radio map considerably. In our tests the radio map size is reduced between 30% and 90%. However, because the size of the radio map used by WKNN depends on the number of FPs and the size of the radio map used by the other methods depend on the number of APs those numbers cannot be generalised.

Detailed analysis of the radio maps revealed that the PL exponent estimate $\hat{n}$ of an AP tends to take values smaller than 2 if the AP has been observed only in a small number of FPs. If the AP has been observed in a larger number of FPs, then a PL exponent estimate smaller 2 is less likely (68% of all APs that were observed in fewer than 100 FPs have $\hat{n} < 2$, but only 27% of all APs that were heard in more than 100



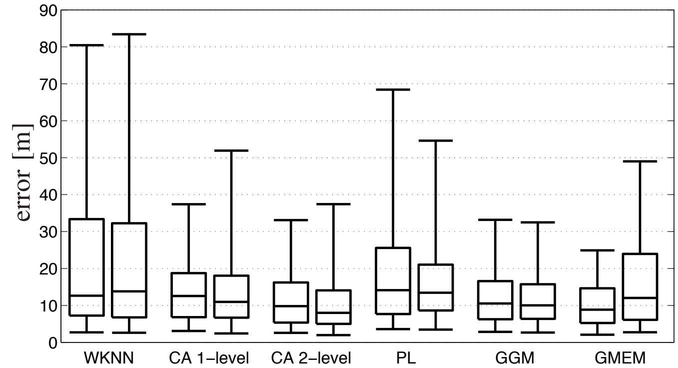
	WKNN static	WKNN filtered	CA 1−level static	CA 1−level filtered	CA 2−level static	CA 2−level filtered	PL static	PL filtered	GGM static	GGM filtered	GMEM static	GMEM filtered
time [ms]	3.2	3.5	1.1	1.3	1.0	1.2	36.8	35.5	52.1	52.3	61.6	31.3
n−cons			4	3	4	4	7	7	17	14	24	1
g−cons			13	8	16	8	19	19	35	28	47	3

Figure 2. Method performances with all data



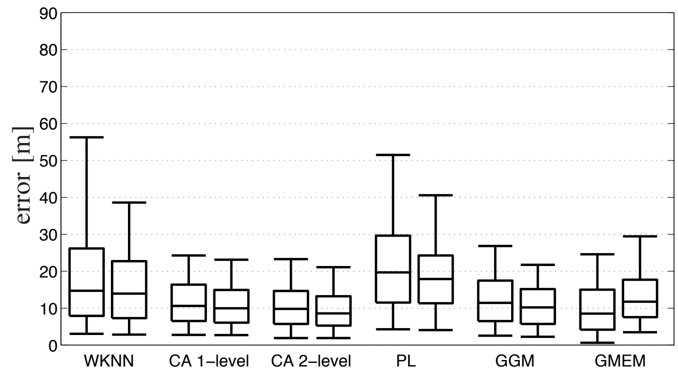
	WKNN static	WKNN filtered	CA 1−level static	CA 1−level filtered	CA 2−level static	CA 2−level filtered	PL static	PL filtered	GGM static	GGM filtered	GMEM static	GMEM filtered
time [ms]	0.6	0.9	0.4	0.6	0.3	0.5	452.4	8.1	11.8	12.0	41.9	7.7
n−cons			26	16	27	18	21	25	31	23	34	18
g−cons			64	46	52	44	40	45	54	44	55	31

Figure 3. Method performances with five strongest measurements



	WKNN static	WKNN filtered	CA 1−level static	CA 1−level filtered	CA 2−level static	CA 2−level filtered	PL static	PL filtered	GGM static	GGM filtered	GMEM static	GMEM filtered
time [ms]	0.4	0.6	0.3	0.4	0.2	0.4	35.4	4.3	3.5	3.7	4.6	3.2
n−cons			37	24	40	31	36	48	50	30	27	35
g−cons			67	53	69	59	60	73	73	61	38	60

Figure 4. Method performances with 90% of APs dropped



	WKNN static	WKNN filtered	CA 1−level static	CA 1−level filtered	CA 2−level static	CA 2−level filtered	PL static	PL filtered	GGM static	GGM filtered	GMEM static	GMEM filtered
time [ms]	0.7	0.9	0.3	0.5	0.3	0.5	417.1	8.4	10.2	10.4	5.5	6.8
n−cons			31	9	32	11	15	10	57	17	13	12
g−cons			62	25	63	26	32	18	88	44	19	23

Figure 5. Method performances when positioning done several months after data for radio map generation was collected

FPs have $\hat{n} < 2$). PL exponents smaller 2 can be explained by the fact that the corridors in which we collected our FPs acted as waveguides [1, p. 66].

The true routes for all six test tracks were measured by clicking a map plot on a touch screen while walking and interpolating between the clicks, and were estimated for both static case and filtered case (i.e. time series). For the filtering we considered the state vector $\mathbf{x}_k$ containing location and velocity of the UE. Both CA-models and GGM were updated using a plain Kalman filter. In addition, we collapsed the GGM to a single component after 5 measurements and after each time step. The GMEM used a grid for static position estimation and a particle filter with 300 particles for the time series estimation; the PL model method used Gauss-Newton for static positioning and a GM filter [25] for time series. In time series the effect of parameter uncertainties varied depending on the location, and therefore was computed in the prior mean of the estimate. The WKNN was given a standard deviation of 10 m for filtering with a Kalman filter.

The methods were tested in four different scenarios:

- Fig. 2: full data
- Fig. 3: only the APs with five strongest signals were used for positioning
- Fig. 4: 90% of APs were dropped pseudorandomly to check how the methods perform in situations with low AP density
- Fig. 5: data for generating the radio maps and data for positioning were collected with a time gap of several months to evaluate the methods' performance degradation over time

In Figs. 2 – 5 we present quantiles with box plots for the positioning errors, absolute time for one position estimate and consistency values that can be used to evaluate the accuracy of the estimated position's covariance matrix that is reported by a method. For the n-cons (normal consistency [26, p. 235 ff.]) values we assumed Gaussian distributed positioning errors, and computed how often the errors were within the 50% ellipse, i.e.

$$(\hat{\mathbf{x}}_u - \mathbf{x}_u)^T \mathbf{P}_k^{-1} (\hat{\mathbf{x}}_u - \mathbf{x}_u) \leq \chi_2^2(0.5) = 1.3863, \quad (5)$$

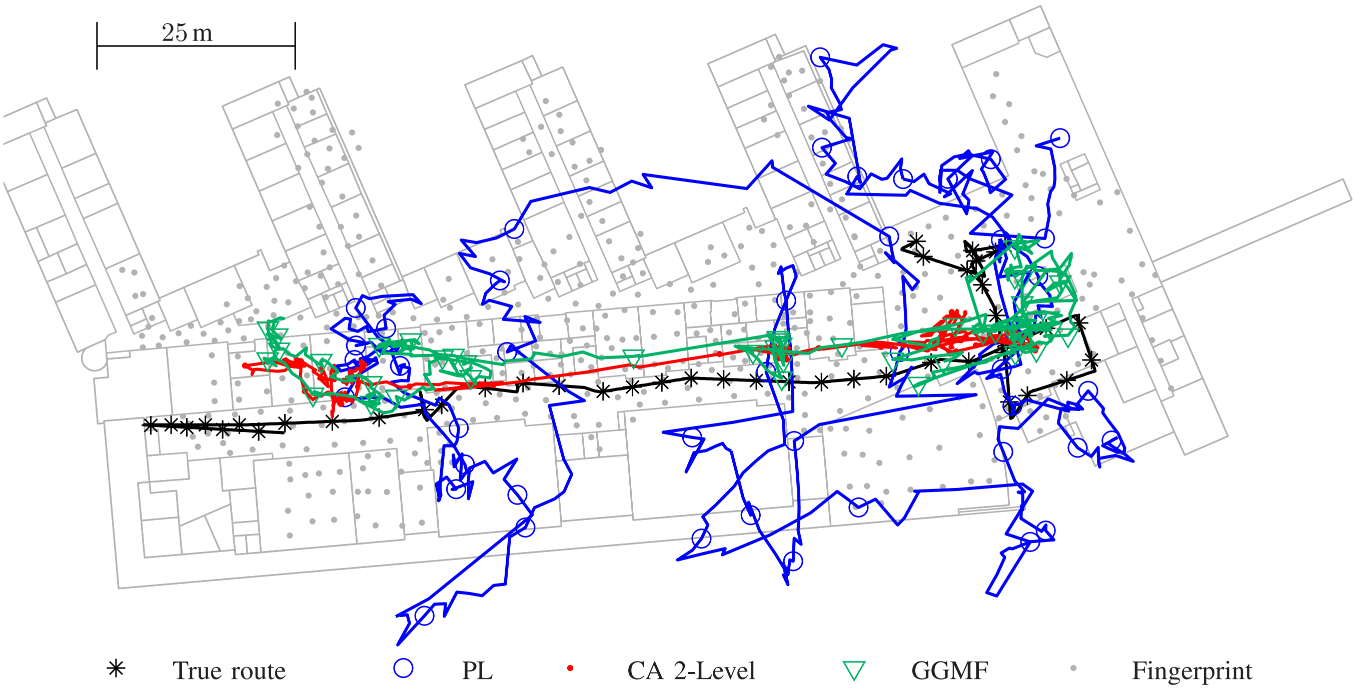


Figure 6. Filtererd routes of selected methods when data for generating the radio map and data for positioning where collected with a time gap of several months.

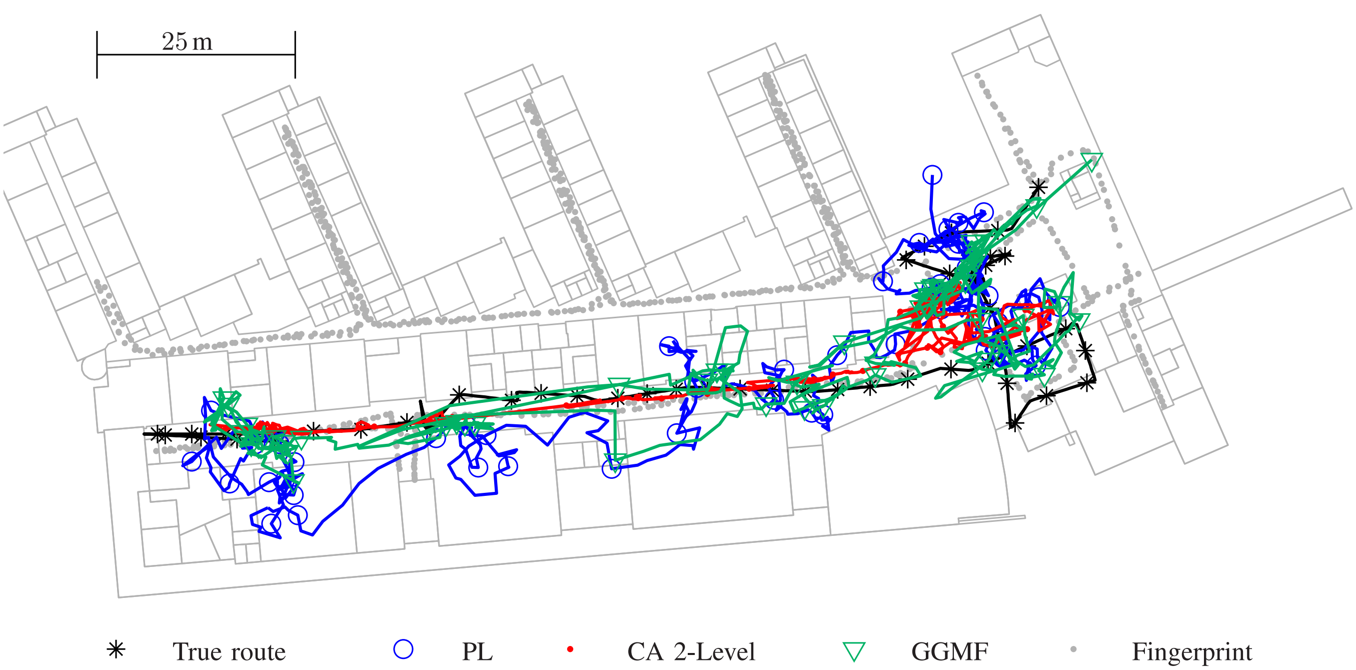


Figure 7. Filtererd routes of selected methods when data for generating the radio map and data for positioning where collected within one month.

where $\hat{\mathbf{x}}_u$ is the estimated UE position, $\mathbf{P}_k$ its covariance matrix and $\mathbf{x}_u$ the true UE position. This measure may be used for checking the error estimate in both ways (if it is too small or large) as long as the distribution is close to the normal distribution. In g-cons (general consistency [27]) we computed how often the errors were within 50% for any distribution using the modified Chebyshev inequality, namely

$$(\hat{\mathbf{x}}_u - \mathbf{x}_u)^T \mathbf{P}_k^{-1} (\hat{\mathbf{x}}_u - \mathbf{x}_u) \leq \frac{2}{0.5} = 4. \qquad (6)$$

When using all of the data all parametric methods were inconsistent, with n-cons values far from the desirable 50% and g-cons values far from the 50% that can be interpreted as minimum requirement (a g-cons of 60% is not necessarily better/worse than 55%), and there are no significant differences between the accuracies of the different methods, except for filtered GMEM. The computation time for static GMEM is higher than for filtered GMEM because it is computed on a grid, whereas the filtered GMEM uses a particle filter. The results suggest that the 300 particles proposed in [20] was too few. Using only the five strongest measurements improved the consistency and reduced the relative computing time for all methods. The large time value for static PL can be explained by the facts that in two (of 308) TPs the convergence was extremely slow and that our implementation did not restrict the number of iterations. Since the GGM's computational demand depends in exponential manner on number of measurements [24] the reduction in computation time for static and filtered GGM could be expected, although in our tests the dependence is not exponential due to collapsing a GGM after five measurements to a single Gaussian. At the same time the positing accuracy degraded significantly only for WKNN and the CA 1-level approach. This is evidence for dependency of the measurements. In the test building there were some Multiple Input Multiple Output (MIMO) APs that produced dependent measurements.

Fig. 4 reveals that the more sophisticated approaches (PL, GGM and WKNN) perform worse or similar than the relatively simple CA methods for scenarios with low AP density. The same holds for the scenario in which the radio map was outdated (compare Fig. 5).

One possible reason for the static and the filtered GGM's poor performance in all four scenarios (compared with their performance in [23] and in [24]) might be that we used a different approach for determining the covariance matrices of the GGM's two Gaussian components, since our tests where carried out in a WLAN rather than in cellular telephone or UWB network. We believe that there exist better approaches than the heuristic we used, but more research on this topic will be necessary. A deeper analysis of the GGM can be found in [24].

Fig. 6 shows an example for positioning with outdated radio maps. None of the filtering methods provide satisfactory positioning accuracy in the lower vertical corridor. One reason for the poor performance in that area is that the radio maps are missing FPs from the southern, central part of the building. Another reason may be that the newer data used for positioning contains weaker RSS values than the older data used for radio map generation, which causes the distance from UE to APs to be overestimated. This causes the PL approach, which relies heavily on the RSS, to position the UE outside the building at several occasions.

Using an up-to-date radio map (Fig. 7) the filtering methods provide satisfying positioning accuracy with the PL approach still struggling. This shows how critical it is to have accurate PL model parameter estimates.

VI. Conclusion

In this paper we presented an overview of parametric fingerprinting and positioning methods, and tested them with real WLAN data for different test tracks and scenarios. Besides their positioning accuracies and consistencies, we also compared the storage requirements for their radio maps with that of the WKNN (as an example of a nonparametric FP method).

All parametric methods enable a significant reduction in the size of the radio map used in the positioning phase. In addition, our tests show that all parametric methods, except the CA 1-level and the filtered GMEM method, provide similar positioning accuracy as the nonparametric WKNN in case of a high CN density and when using all available measurements. When using only the five strongest measurements their computation time drops significantly. Furthermore, all parametric methods still show similar positioning performances, while the WKNN's performance degrades considerably. This means that the parametric methods achieve satisfying positioning accuracy even with few observable APs. When the AP density is low or the mapping data is outdated then the simple CA techniques achieve at least similar positioning accuracy than the more sophisticated parametric techniques and the WKNN. Thus, the CA technique gives the best trade-off between accuracy and computational demand. The other parametric methods are, like the WKNN, more vulnerable to harsh environments.

Acknowledgment

Philipp Müller acknowledges the financial support of the TUT Doctoral Programme. Matti Raitoharju is supported by Tampere Doctoral Programme in Information Science and Engineering, Nokia Inc., Nokia Foundation and Jenny and Antti Wihuri Foundation.

We thank Henri Nurminen for his advice and for provision of Matlab code for his approach [16].

References

[1] A. F. Molisch, *Wireless Communications*, 2nd ed. Wiley - IEEE, January 2011.

[2] R. Mautz, "Indoor positioning technologies," Habilitation Thesis, http://e-collection.library.ethz.ch/eserv/eth:5659/eth-5659-01.pdf, February 2012.

[3] X. Li, "RSS-based location estimation with unknown pathloss model," *IEEE Transactions on Wireless Communications*, vol. 5, no. 12, pp. 3626–3633, December 2006.

[4] B. Dil and P. Havinga, "RSS-based localization with different antenna orientations," in *2010 Australasian Telecommunication Networks and Applications Conference (ATNAC)*, 2010, pp. 13–18.

[5] V. Honkavirta, T. Perälä, S. Ali-Löytty, and R. Piché, "A comparative survey of WLAN location fingerprinting methods," in *Proceedings of the 6th Workshop on Positioning, Navigation and Communication 2009 (WPNC'09)*, March 2009, pp. 243–251.

[6] L. Koski, R. Piché, V. Kaseva, S. Ali-Löytty, and M. Hännikäinen, "Positioning with coverage area estimates generated from location fingerprints," in *Proceedings of the 7th Workshop on Positioning, Navigation and Communication 2010 (WPNC'10)*, March 2010, pp. 99–106.

[7] J. Machaj, R. Piché, and P. Brida, "Rank based fingerprinting algorithm for indoor positioning," in *2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, September 2011, pp. 6–11.

[8] L. Koski, T. Perälä, and R. Piché, "Indoor positioning using WLAN coverage area estimates," in *2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, September 2010, pp. 1–7.

[9] L. Wirola, T. A. Laine, and J. Syrjärinne, "Mass-market requirements for indoor positioning and indoor navigation," in *2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, September 2010, pp. 1–7.

[10] L. Wirola, L. Wirola, and R. Piché, "Bandwidth and storage reduction of radio maps for offline WLAN positioning," in *2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, October 2013.

[11] S. Eisa, J. Peixoto, F. Meneses, and A. Moreira, "Removing useless APs and fingerprints from WiFi indoor positioning radio maps," in *2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2013, pp. 739–745.

[12] R. Piché, "Robust estimation of a reception region from location fingerprints," in *2011 International Conference on Localization and GNSS (ICL-GNSS)*, June 2011, pp. 31–35.

[13] M. Raitoharju, S. Ali-Löytty, R. Piché, and M. Dashti, "Positioning with multilevel coverage area models," in *2012 International Conference on Indoor Positioning and Indoor Navigation*, http://URN.fi/URN:NBN:fi:tty-201310311408, November 2012.

[14] H. Hashemi, "The indoor radio propagation channel," *Proceedings of the IEEE*, vol. 81, no. 7, pp. 943–968, July 1993.

[15] S. Gezici, Z. Tian, G. B. Biannakis, H. Kobayashi, A. F. Molisch, H. V. Poor, and Z. Sahinoglu, "Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 70–84, July 2005.

[16] H. Nurminen, J. Talvitie, S. Ali-Löytty, P. Müller, E. S. Lohan, R. Piché, and M. Renfors, "Statistical path loss parameter estimation and positioning using RSS measurements in indoor wireless networks," in *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, http://urn.fi/URN:NBN:fi:tty-201310311407, November 2012.

[17] S. Shrestha, E. Laitinen, J. Talvitie, and E. S. Lohan, "RSSI channel effects in cellular and WLAN positioning," in *Proceedings of the 9th Workshop on Positioning, Navigation and Communication 2012 (WPNC'12)*, March 2012, pp. 187–192.

[18] M. Hata, "Empirical formula for propagation loss in land mobile radio services," *IEEE Transactions on Vehicular Technology*, vol. 29, no. 3, pp. 317–325, August 1980.

[19] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, vol. 7, no. 4, pp. 465–479, July 1971.

[20] K. Kaji and N. Kawaguchi, "Design and implementation of WiFi indoor localization based on Gaussian mixture model and particle filter," in *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, November 2012.

[21] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[22] S. Ali-Löytty and N. Sirola, "Gaussian mixture filter in hybrid navigation," in *Proceedings of The European Navigation Conference GNSS 2007*, May 2007, pp. 831–837.

[23] P. Müller, S. Ali-Löytty, M. Dashti, H. Nurminen, and R. Piché, "Gaussian mixture filter allowing negative weights and its application to positioning using signal strength measurements," in *Proceedings of the 9th Workshop on Positioning, Navigation and Communication 2012 (WPNC'12)*, http://urn.fi/URN:NBN:fi:tty-201311061423, March 2012, pp. 71–76.

[24] P. Müller, H. Wymeersch, and R. Piché, "UWB positioning with generalized Gaussian mixture filters," *IEEE Transactions on Mobile Computing*, 2014 (in press).

[25] M. Raitoharju and S. Ali-Löytty, "An adaptive derivative free method for Bayesian posterior approximation," *IEEE Signal Processing Letters*, vol. 19, no. 2, pp. 87–90, February 2012.

[26] Y. Bar-Shalom, R. X. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation, Theory Algorithms and Software*. John Wiley & Sons, 2001.

[27] S. Ali-Löytty, N. Sirola, and R. Piché, "Consistency of three Kalman filter extensions in hybrid navigation," in *Proceedings of The European Navigation Conference GNSS 2005*, Munich, Germany, Jul. 2005.