Mobile Location using Database Correlation with Least-Squares and Bayes Filtering

Mohamed Khalaf-Allah[1] and Kyandoghere Kyamakya[2]

[1]University of Hannover, Institute of Communications Engineering, Hannover, Germany
khalaf@ikt.uni-hannover.de

[2]University of Klagenfurt, Department of Informatics Systems, Klagenfurt, Austria
kyamakya@isys.uni-klu.ac.at

Abstract - **Database correlation methods for mobile location could be enhanced by efficient preprocessing of available information. We used least-squares filtering to study the relationship between the predicted signal strength and distance to serving base station (BS) in order to provide more knowledge during the online location process. A Bayes filter has been built to achieve robust online weighting of location candidates. The final location estimation is done using one of three different methods. The positioning accuracy of the whole system using these estimation methods are tested in a suburban area, which is a very common environment in European cities. In such environments, the average cell radius is much larger than of urban cells. This requires the relaxing of FCC location accuracy requirements for suburban cells or the need of more work to meet these accuracies.**

1. Introduction

The obligation for emergency call location set by the FCC [1] and the importance of location-based services (LBS) for mobile computing and internet are the major driving forces of the increasing interest in location techniques for mobile wireless terminals. The main methods of positioning a mobile station (MS) are satellite-based and cellular system based technologies. An overview of these technologies is given in [3] and [4].

In cellular networks, MS positioning is usually achieved by one of the following techniques: (1) Cell-id, (2) Time-of-arrival (TOA), time-difference-of-arrival (TDOA), and enhanced observed time-difference-of-arrival (E-OTD), (3) Angle-of-arrival (AOA), and (4) Network-assisted GPS (A-GPS). These techniques have different properties in terms of accuracy, cost, coverage, system impact, and power consumption.

Cell-id methods are the simplest to implement since no hardware extensions are required at the base or mobile stations. However, positioning accuracy is in the range of hundred meters up to several kilometres depending on the characteristics of the area under concern. TOA-based techniques need mutual synchronization of at least three BSs, which is difficult to achieve. AOA methods require the installation of special antennas at BSs. Moreover, TOA and AOA techniques suffer large positioning errors due to multipath propagation. A-GPS solutions main drawbacks are power consumption, the need of a clear view to four satellites, and the necessity to install additional receivers.

Cell-id techniques still provide an attractive alternative as they utilize only network available information. This is an advantage over the other methods in terms of cost coverage and system impact. There are many ways to improve positioning accuracy of cell-id approaches; one of them is the database correlation method (DCM) [6] – [10]. The key idea of DCM is the construction of a database of reference locations. The reference entries contain values of location sensitive parameters. Construction of the database can be performed by field measurements [6], [7] and [10], or using radio wave propagation prediction tools [8] and [9]. Later a moving MS collects network measurements to be matched with the database entries in order to yield location estimates. Location sensitive parameters employed for database construction include received signal levels (RxLev) [6] – [9] and channel impulse response (CIR) [10]. CIR is used in high bandwidth systems such as UMTS, where accurate timing measurements and the measurement of channel multipath profile are enabled. It is especially useful as a position fingerprint in areas with sufficient amount of distinct topography, whereas in flat areas the CIR characteristics are not more significant than those of RxLev. However, the bandwidth of GSM is too low to allow CIR measurements to contribute efficiently to positioning based on DCM.

We have developed a database correlation method based on RxLev as a location dependent parameter. Two filtering techniques were utilized to further help resolving ambiguities during the positioning process. Least-squares (LS) filtering was used in the off-line phase (preprocessing step) of our technique in order to characterize the relationship between predicted RxLev at some location and the distance of that location to the serving BS. This procedure has been applied to all BSs in our test area. In the online phase, Bayes filtering was employed to determine the importance of every location candidate. We also present three methods for the final location estimation that yield different accuracies. The proposed location algorithm was implemented and applied to measurements of RxLev and TA in a GSM network deployed in a suburban area. Positioning accuracy was investigated offline, and the results are provided in section 4. A scheme of the overall location system is given in Fig. 1.

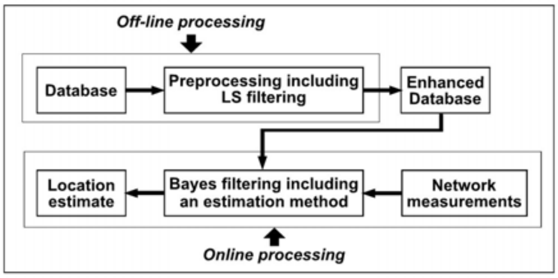

Fig. 1. Scheme of the proposed location system

2. Database Construction and Preprocessing

The database was constructed using a 3D deterministic radio wave propagation prediction model. More details on this model are given in [11]. The database contains predicted RxLev values for a populated suburban area in Hannover, north Germany, at reference locations. Moreover, data about the BSs in the given area were also provided. These include geographical locations, antenna height, azimuth and tilt, effective isotropic radiated power, channel numbers, cell identifiers, etc. Fig. 2 shows an example of the predicted RxLev values for one cell antenna. To increase the usability of the database we carried out two stages of off-line processing (preprocessing). In the first stage (section 2.1), more information from the raw database have been extracted and stored to help enhance the positioning accuracy and performance of the online algorithm. Least-squares processing has been applied to every cell in order to study the relationship between predicted RxLev at some location and the distance of that location to the serving BS (section 2.2).

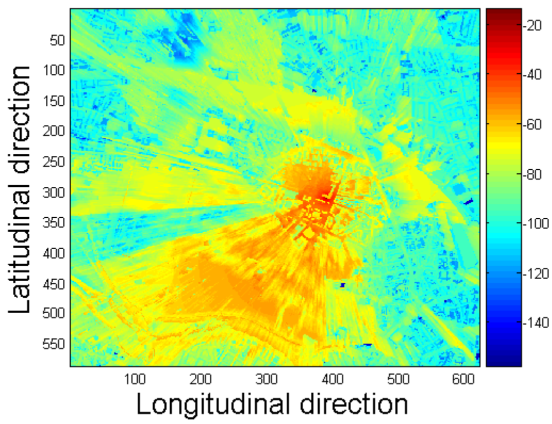

Fig. 2. Map of RxLev (dBm) generated by the radio wave propagation prediction tool for a base station antenna. The simulation is performed over an area of approx. 9 km^2 divided into pixels (621 x 588 pixels in the longitudinal and latitudinal directions respectively) with a resolution of 5 m

In general, the positioning accuracy of DCM depends mainly on three factors: (1) determining where to search in the database, (2) using a reliable technique to evaluate the location candidates, (3) effectively estimating the location from the candidates. The first is achieved by TA measurements, sector information of the serving BS, and the relationship between RxLev and distance to serving BS obtained by the LS regression (see section 2.2). We use Bayes filtering and introduce three location estimation methods (see section 3) to fulfil the second and third factors respectively.

2.1. Database Preprocessing

The pixels served by every cell antenna are determined (Fig. 3) and sorted in an array according to their distances to the cell antenna. This will guarantee that no location outside the coverage area of the cell antenna will be returned by the location algorithm when the deviation between predicted and measured RxLev's are large or when the situation is highly ambiguous due to an increased number of equally probable location candidates. Moreover, the six strongest (in terms of RxLev) neighbouring cell antennas are determined for every pixel. This will help concentrate on the real location candidates taking into account neighbour cell-id's that appear the measurement report. Thus, every array entry contains pixel-id, pixel coordinate (midpoint of the pixel), RxLev's/id's from/of serving and the six strongest cell antennas, and distance of the pixel to serving cell antenna. This constellation simplifies the online location estimation process.

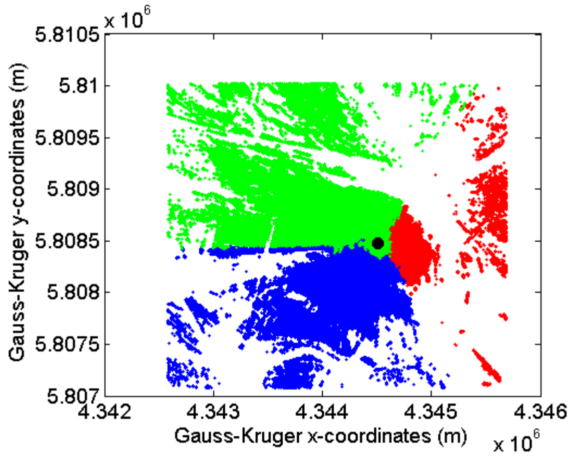

Fig. 3. Preprocessing results are shown for a BS with three sectors. Pixels served by each sector antenna are depicted in different colours. BS location is represented by a black dot

2.2. Least-Squares Batch Processing

Finding the best polynomial curve fit using the least squares (LS) method [12] to a set of data is still considered one of the most significant techniques. It is a procedure developed by Gauss and Legendre [2] and is also known as the least squares filtering or estimation. We applied the LS technique to signals, which can be described by a polynomial. One major concern in utilizing the LS technique is to choose the correct-order polynomial to use to best fit the data set of interest. Therefore, a good knowledge of the problem at hand must be available. This knowledge is based on understanding the dynamics of the problem and on information derived from mathematical techniques previously applied to the problem.

Our goal was to learn about the relationship between $RxLev_i$ at some location i and the distance d_i^* of that location to the serving BS antenna using the prediction data in order to reduce the area (as specified by TA measurements) in which we search for the MS location. For extracting information about the characteristics of the signal under study, we assumed a polynomial model to represent the signal and estimated the coefficients of the selected polynomial by choosing a goodness of fit criterion. The LS criterion is the sum of the squares of the individual discrepancies between the desired polynomial and given data set values. Finally, we minimized the sum of squares of the individual discrepancies in order to obtain the best coefficients for the selected polynomial.

Every BS antenna serves a set of n pixels. For each pixel i we have prediction data $(d_i^*, RxLev_i^*)$ where d_i^* is the distance from the centre of pixel i to its serving BS, and $RxLev_i^*$ is the predicted RxLev at pixel i.

We would like to fit the prediction data with the best polynomial using the LS method, i.e., the prediction data should be manipulated in order to find the polynomial that will best model the relationship between $RxLev_i$ and d_i^*. This relationship could be described as [13]

$$RxLev_i(dBm) = a_0 + a_1 \log_{10} d_i^*(km) \quad (1)$$

And we would like to minimize

$$R = \sum_{i=1}^{n} (RxLev_i - RxLev_i^*)^2$$
$$= \sum_{i=1}^{n} (a_0 + a_1 \log_{10} d_i^* - RxLev_i^*)^2$$

(2)

Where R is the square of the summation of all the residuals or differences between measurements and prediction data.

Expression (1) is a first-order polynomial. Thus coefficients a_0 and a_1 are determined by solving the following system of equations

$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} \log_{10} d_i^* \\ \sum_{i=1}^{n} \log_{10} d_i^* & \sum_{i=1}^{n} (\log_{10} d_i^*)^2 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sum_{i=1}^{n} RxLev_i^* \\ \sum_{i=1}^{n} RxLev_i^* * \log_{10} d_i^* \end{bmatrix} \quad (3)$$

Fig. 4 shows the relationship between the received signal strength and logarithm of distance to serving cell antenna using the database prediction values and the LS regression model, for an example cell. We can see that the complex relationship has been simplified (linearized) by the LS regression procedure.

At time j if the measured signal strength is $RxLev_j$ then an estimated distance to the serving BS antenna d_j is calculated as

$$d_j(m) = 10^{\left(\frac{RxLev_j - a_0}{a_1}\right)} * 1000 \pm e \quad (4)$$

Where e is an estimation error. Combining the information gained from (4) with TA measurements can further specify the area in which we look for location candidates as illustrated in Fig. 5.

3. Bayes Filtering and Location Estimation

Bayes filtering is a concept that provides a probabilistic framework for state estimation [14]. The implementation of Bayes filter depends on the way we represent the *belief distributions* (continuous or discrete) over the state space. In the case of our mobile location problem, the state space is divided into grids (pixels) with a specified resolution. Therefore, a discrete Bayes filter (DBF) [15] was implemented for estimating the location of a mobile terminal. Bayes filter estimates the *posterior* belief distribution of a mobile location at a certain time based on a series of RxLev measurements and a database of RxLev predictions associated with known locations, i.e., it estimates the state of a dynamical system, which is a *partially observable Markov chain* [16], using measurement data. As the mobile environment is assumed to be *Markovian*, past and future measurements are conditionally independent if the current state is known. In this context, the dynamical system is the mobile terminal and its environment, the state is the mobile location in that environment and the network measurements, which include TA and RxLev from serving and neighbouring BSs.

The key idea is to approximate the belief $Bel(s_t)$ at any time t by a set of n weighted location candidates. The belief is the posterior distribution over the state space and is denoted as

$$Bel(s_t) = p(s_t \mid o_t, o_{t-1}, ..., o_0, m) \quad (5)$$

Where $Bel(s_t)$ is the mobile belief state at time t, s_t is the state a time t, $o_{t,...,0}$ denote the measurement data obtained from time *0* up to time t, and m is the model of the environment, i.e., a database of RxLev associated with known locations. A complete mathematical derivation of the proposed Bayes filter is provided in [17].

The belief $Bel(s_t)$ is then approximated as

$$Bel(s_t) \approx \{s^{(i)}, w^{(i)}\}_{i=1,\dots,n} \tag{6}$$

Where $s^{(i)}$ is the location candidate i, and $w^{(i)}$ is a non-negative numeric value called weight that determines the importance of $s^{(i)}$. Location candidates are those which lie within the measured TA range and within the area bounded by $d_j - e$ and $d_j + e$ with respect to the serving cell as explained in section 2.2.

The weight $w^{(i)}$ is calculated for every location candidate according to the *log-normally* distributed measurement model as

$$w^{(i)} = p(o_t \mid s_t, m) = \prod_{j=1}^{M} \frac{1}{\sigma_{RxLev}\sqrt{2\pi}} e^{-\frac{(RxLev_j - RxLev_{DB_j})^2}{2\sigma_{RxLev}^2}} \tag{7}$$

Where $p(o_t \mid s_t, m)$ is the measurement model, M is the number of observed BSs (serving and neighbouring). In GSM networks, $M \le 7$. $RxLev_j$ is the measured RxLev from the *j*-th observed BS, $RxLev_{DB_j}$ is the database RxLev prediction value of the *j*-th observed BS at $s^{(i)}$, and σ_{RxLev} is the standard deviation of RxLev measurements.

The weights are then normalized so that $\sum_{i=1}^{n} w^{(i)} = 1$, thus the belief $Bel(s_t)$ is approximated by a *discrete probability function* defined by the set of location candidates.

The MS movements cannot be directly measured without an extra inertial sensor, which is not the case here. Therefore, the prior belief is initialized as a uniform distribution over the whole candidate locations every time the Bayes filter is run.

The final location estimate $\hat{s}$ is obtained by one of three different methods:

1. Maximum likelihood estimate, where the location estimate is the candidate with the highest weight, i.e.,

$$\hat{s} = \arg\max Bel(s_t) \tag{8}$$

2. Weighted average, where the location estimate is the weighted average of all candidates representing $Bel(s_t)$, i.e.,

$$\hat{s} = \frac{1}{\sum_{i=1}^{n} w^{(i)}} \sum_{i=1}^{n} s^{(i)} \times w^{(i)} \tag{9}$$

Expression (9) is the *mean value* of the posterior distribution $Bel(s_t)$. It will coincide with (8) only in case of *unimodal* and *symmetric* posterior distributions.

3. Trimmed average , where the location estimate is the average of a certain number (k) of the best weighted candidates, i.e.,

$$\hat{s} = \frac{1}{k} \sum_{i=1}^{k} s^{(i)}, \; k < n \tag{10}$$

Note that the belief distribution $Bel(s_t) \approx \{s^{(i)}, w^{(i)}\}_{i=1,\dots,n}$ is sorted according to weights before the final location estimate is calculated. The optimal value of k is determined experimentally.

The performance accuracy of the above mentioned location estimation methods are given in the next section.

4. Experiments and Results

The field tests were performed in an E-Plus GSM 1800 MHz network operating in suburban area in Hannover, north Germany. The database available with us covers an area of about 9 km^2 with 22 cells and only few of them are completely contained in the test area. The pixel resolution is 5 m. The expression *suburban* describes the topology of the environment and its network characteristics and does not indicate a low populated area. Almost all buildings in the test field are lower than the heights of the deployed BS antennas. Cell radius is about 2 km unlike in urban areas, in which it lies between 50 m and 1000 m. These characteristics are very common in European cities due to regulation controlling maximum building height. Accordingly, it is expected to get less accurate location estimates than for urban areas.

The measurements were collected every four seconds by a pedestrian along a route of about 2.4 km with a total of 250 measurement reports. All reports were stamped by a GPS position as ground truth for evaluating the location algorithm.

We investigated the performance of the proposed algorithm using the three methods of location estimation with and without LS processing by offline evaluation of the field measurements. The cumulative distribution functions (CDF) of location error are shown in Fig. 6 and summarized in Table I. The figures show that LS processing of the database increases the location accuracy, because it helps rejecting location candidates that are far away from the true MS Location. The distance estimate of the LS model refines the area determined by TA measurements, in which we look for location candidates by considering the most probable locations associated with a given RxLev value. Moreover, the most accurate estimations are provided by the *trimmed average* method, because it considers only the best candidates of the posterior distribution. The use of *maximum likelihood estimate* is not recommended as it is very sensitive to noisy measurements and database inaccuracies. It has delivered less accurate results even with LS processing. The explanation is that it takes the candidate location with the predicted RxLev value that coincides with the measured RxLev value as the location estimate. This is very unreliable as many candidates have the same predicted RxLev value and choosing the final location estimate depends on the sorting strategy used to arrange all candidates according to their importance. This sensitivity, however, is reduced by the *weighted average* method. Again, averaging only a certain amount (about 10%) of the best weighted candidates (trimmed average) yielded the best results.

To the best of our knowledge, location techniques using database correlation in suburban areas has not yet satisfied the FCC location accuracy requirements if the database is constructed using propagation models rather than field measurements. Propagation models provide less accurate databases, but enables easy maintenance and update process. On the contrary, databases constructed using field measurements are more realistic and accurate, but with higher overhead costs due to the need to cover the whole considered area by field measurements, which is difficult to achieve, maintain and update. Therefore, more accurate estimation techniques should be developed to bridge this gap. However, literature survey confirms that our results still yield the most accurate estimates in such environments so far. The structural nature of these areas and the strength of GSM signals, result in large cell sizes in European wireless layouts, leading to degradation in MS location estimates.

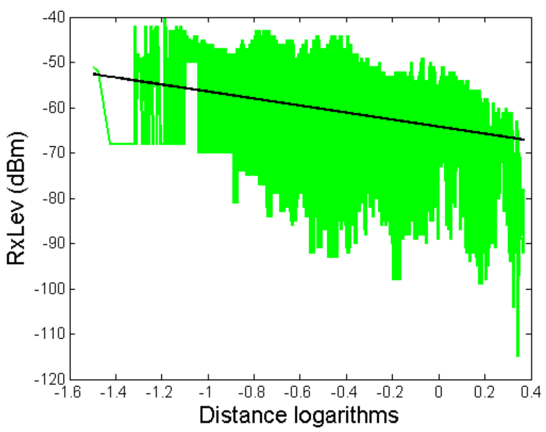

Fig. 4. Relationship between the received signal strength and logarithm of distance to serving antenna for an example cell using the prediction database and the LS linear model

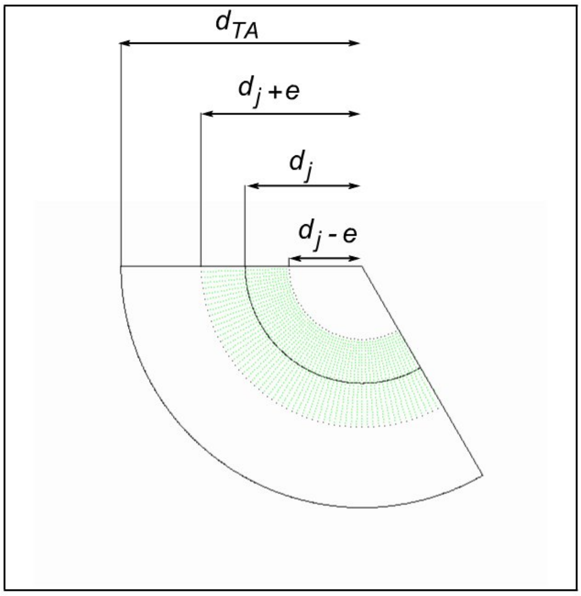

Fig. 5. Exploiting equation (4) to further specify the area in which we search for MS location candidates

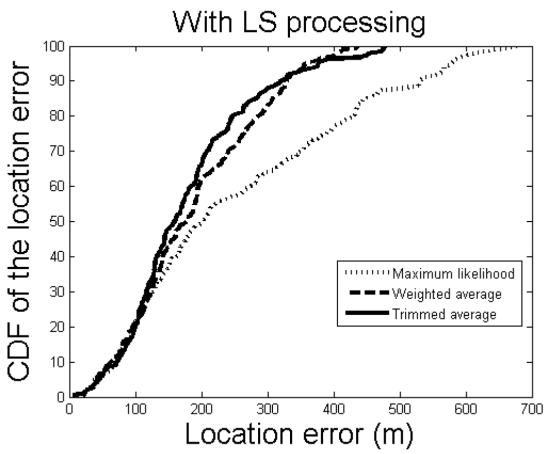

Fig. 6.a

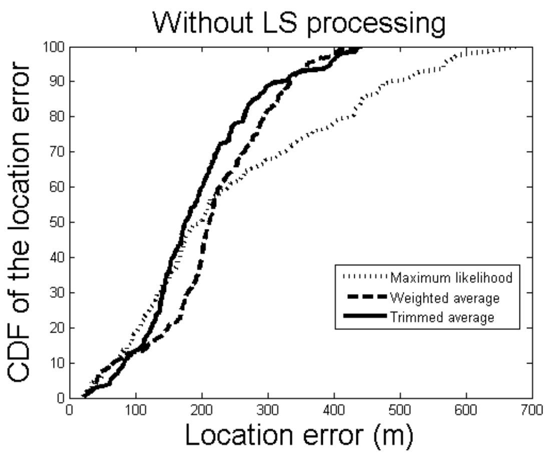

Fig. 6.b

Fig. 6. CDF of location error with and without LS processing using the three location estimation methods

Location Error		Maximum Likelihood	Weighted Average	Trimmed Average
With LS Processing	67%	323 m	231 m	200 m
	95%	579 m	358 m	377 m
	mean	256 m	189 m	179 m
Without LS Processing	67%	295 m	254 m	216 m
	95%	567 m	361 m	395 m
	mean	248 m	217 m	194 m
Improvement due to LS Processing	67%	-9.5%	9%	7%
	95%	-2%	1%	4.5%
	mean	-3%	13%	8%

TABLE I: Location error of the proposed database correlation algorithm with and without the least-squares processing

5. Conclusion

We have presented a database correlation method for mobile location in GSM networks. Off-line preprocessing including least-squares batch filtering of the raw database has been performed to increase the correlation efficiency. The proposed location algorithm utilizes Bayes filtering in the online phase to determine the importance of every location candidate, and the final location estimation is achieved by one of three methods. The accuracy of these methods has been determined experimentally by testing in a suburban environment. The suggested location algorithm provided very acceptable results taking into account the low BS density in the test environment.

ACKNOWLEDGEMENT

We would like to thank Thomas Kürner, Peter Unger, and E-Plus Mobilfunk GmbH & Co KG for providing the network and prediction data utilized in this research.

REFERENCES

[1] Federal Communications Commission (FCC) Fact Sheet, “FCC Wireless 911 Requirements,” 2001.

[2] S. M. Stigler, “Gauss and the invention of least squares,” *Annals of Statistics*, 9, 1981, pp. 465-474.

[3] H. Laitinen (editor), “Cellular Location Technology,” *Public deliverable of IST/CELLO project*, 2001.

[4] T. S. Rappaport, J. H. Reed, D. Woerner, “Position location using wireless communications on highways of the future,” *IEEE Communication Mag.*, vol. 34, pp. 33-41, Oct 1996.

[5] B. Walke, “*Mobile radio networks: networking and protocols*,” Chichester [u.a.]: Wiley, 2000, Ch. 3.

[6] H. Laitinen, J. Lähteenmäki, T. Nordström, “Database Correlation Method for GSM Location,” presented at *VTC2001-Spring*, Rhodes, Greece, May 2001.

[7] F. Erbas, K. Kyamakya, K. Jobmann, “On the user profiles and the prediction of user movements in wireless networks,” presented at *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC’02)*, Lisbon, Portugal, 2002.

[8] H. Schmitz, M. Kuipers, K. Majewski, P. Stadelmeyer, “A new method for positioning of mobile users by comparing a time series of measured reception power levels with predictions,” presented at *VTC2003-Spring*, Jeju, South Korea, May 2003.

[9] D. Zimmermann, J. Baumann, M. Layh, F.M. Landstorfer, R. Hoppe, G. Wölfle, “Database Correlation for Positioning of Mobile Terminals in Cellular Networks using Wave Propagation Models,” presented at *VTC2004-Fall*, Los Angeles, USA, Sept. 2004.

[10] T. Nypan, “*Mobile terminal positioning based on database comparison and filtering*,” Dissertation at the Norwegian University of Science and Technology, 2004-65, July 2004.

[11] Th. Kürner, A. Meier, “Prediction of outdoor and outdoor-to-indoor coverage in urban areas at 1.8 GHz,” *IEEE Journal on Selected Areas on Communications*, Vol. 20, No.3, pp. 496-506, April 2002.

[12] J. Mendel, “*Lessons in estimation theory for signal processing, communications, and control*,” Prentice Hall PTR, Englewood Cliffs, NJ, 1995.

[13] J. D. Parsons, “*The mobile radio propagation channel*,” Wiley, Chichester, 2nd ed., 2001, ch. 4.

[14] J. Pearl, “*Probabilistic reasoning in intelligent systems: networks of plausible inference*,” Morgan Kaufmann Publishers, San Mateo, CA, 1988.

[15] D. Fox, J. Hightower, L. Liao, D. Schulz, G. Borriello, “Bayesian Filters for Location Estimation,” *IEEE Pervasive Computing*, vol. 2, no. 3, 2003.

[16] K. Chung, “*Markov chains with stationary transition probabilities*,” Springer Publisher, Berlin, Germany, 1960.

[17] M. Khalaf-Allah, and K. Kyamakya, “Database Correlation using Bayes Filter for Mobile Terminal Localization in GSM Suburban Environments,” in *Proc. IEEE 63rd Semi-Annual Vehicular Technology Conference (VTC2006-Spring)*, May 7-10, 2006, Melbourne, Australia.