

Database Correlation using Bayes Filter for Mobile Terminal Localization in GSM Suburban Environments

Mohamed Khalaf-Allah
Institute of Communications Engineering
University of Hannover
Hannover, Germany
khalaf@ikt.uni-hannover.de

Kyandoghere Kyamakya
Department of Informatics Systems
University of Klagenfurt
Klagenfurt, Austria
kyamakya@isys.uni-klu.ac.at

Abstract—In this paper we present a database correlation method combined with Bayes filtering as a pattern matching technique for mobile terminal localization in GSM suburban environments. A 3D deterministic radio wave propagation prediction model is used for the database construction. Bayes filter provide a robust technique to deal with multi-hypotheses situations usually occur in cellular localization when the location dependent parameter considered is the radio power level. Furthermore, three methods used to yield a location estimate from the resulted hypotheses are presented and compared according to their performance. Results show very acceptable location accuracies relative to the considered suburban area, which is a typical environment in many populated parts of Germany.

Keywords—Bayes filter; database correlation; mobile location

I. INTRODUCTION

Mobile station (MS) positioning is a key problem in wireless networks. It is the problem of determining the position of a MS using location sensitive parameters. The FCC adopted standards for location accuracy and reliability of emergency calls [1] that further motivated the interest in the field of mobile location, which can be traced back to the 1970s [2]. Furthermore, inexpensive but still accurate MS positioning systems are believed to be of great commercial importance and a lot of research is carried out in the area.

A. Localization Techniques

The terms mobile location, positioning and localization are used interchangeably throughout the literature. An overview of localization methods is given in [3]. MS positioning is usually performed using one of the following methods:

- Time-of-arrival (TOA),
- Angle-of-arrival (AOA),
- Network-assisted GPS (A-GPS),
- Time-difference-of-arrival (TDOA) and enhanced observed time-difference-of-arrival (E-OTD), and
- Enhanced cell-id

The evaluation criteria for the different positioning techniques include accuracy, cost, coverage, system impact, and power consumption.

The main drawback of TOA techniques is the need of mutual synchronization of the BSs, which is difficult to achieve leading to poor location accuracy. AOA methods suffer from large positioning errors caused by multipath propagation. Furthermore, special antennas have to be installed at base stations. Power consumption, the need to a clear view to at least four satellites, and the installation of additional hardware are the main disadvantages of A-GPS solutions. Furthermore, multipath propagation degrades TOA estimation. TDOA based techniques need at least three BSs, which could not be fulfilled in rural areas.

Since enhanced cell-id methods utilize only network available information and don't require any additional hardware installations at BSs or in MSs, they seem to be the first alternative to take into consideration. This is advantageous in terms of cost, coverage and system impact in comparison with other methods. However, the accuracy is ranging from about 100 m up to a couple of kilometers depending on type and characteristics of the area covered by the network.

Database correlation is one way to achieve accuracy improvement of enhanced cell-id positioning methods. They also appear in the literature under the names database comparison, location fingerprinting, pattern recognition and pattern matching. In these techniques, a database of location dependent parameters is constructed using field measurements [4] or radio wave propagation prediction tools [5, 6]. Later a moving MS collects measurements to be compared with the values in the database in order to yield position estimates.

The received signal levels (RxLev) from surrounding BSs are used as a location dependent parameter for database correlation in [4] – [6] and [11]. In [7], the location dependent parameter used is the channel impulse response (CIR). However, the bandwidth of GSM is too small for accurate positioning based on database comparison of the CIR only [7].

B. A New Approach

We utilize a 3D deterministic radio propagation prediction model described in [8] to construct the database. Robust

location estimation depends on three main factors. The first is to assure that searching in the database is restricted to an area where the MS is really located. This could be achieved by utilizing TA (timing advance) and the sector information of the serving cell (antenna azimuth). The second factor relies on the pattern matching technique used to evaluate location candidates. The third is to decide how to yield a location estimate from the available hypotheses.

In this paper we use a non-recursive discrete Bayesian filter (DBF) as a matching technique for the database correlation. The accuracy of three simple methods to yield a location estimate is compared and results from measurements in suburban areas are presented.

The rest of the paper is organized as follows. Section II describes our Bayes filter based technique. Experimental setup and simulation results are provided in section III. Conclusion and directions for future work are presented in section IV.

II. THE LOCALIZATION ALGORITHM

A. Bayes Filter

Bayes filter (BF) [9] and [10] is a concept that only provides a probabilistic framework for state estimation. There are two different implementations of Bayes filter that differ mainly in the way they represent belief distributions over the state space. The first implementation represents continuous belief distributions, where the second represents discrete distributions.

In the context of mobile terminal localization, we will be concerned with the discrete version of the Bayes filter as the state space is divided into grids (pixels) with a specified resolution. Bayes filter estimates the posterior belief distribution of a MS position given a map (database) of predicted signal strengths and a series of signal strength measurements. It estimates the state of a dynamical system, i.e. partially observable Markov chain, using measurement data. In this context, the dynamical system is the mobile terminal and its environment, the state is the MS position relative to that environment, and measurements, which include TA and RxLev from serving and neighboring base stations.

B. Mathematics of Bayes Filtering

Bayes filter assumes that the environment is Markovian, i.e. past and future data are conditionally independent if the current state is known. The key idea is to estimate a posterior probability density over the state space conditioned on the measurement data. This posterior is called the *belief* and is denoted

$$Bel(s_t) = p(s_t | o_t, a_{t-1}, o_{t-1}, a_{t-2}, \dots, o_0, m) \quad (1)$$

Where

- $Bel(s_t)$ is the MS belief state at time t ,
- s_t is the state at time t ,

- $o_{t \dots 0}$ denote the measurement data delivered from time 0 up to time t ,
- $a_{t-1 \dots 0}$ denote the actions (movements) performed by the terminal user from time 0 up to time $t-1$, and
- m is the model of the environment, i.e. a map (or database) of predicted RxLev.

Measurements and terminal user actions are assumed to occur in an alternative sequence.

The desired posterior is estimated by applying *Bayes rule*¹ and the *theorem of total probability*² to expression (1), and exploiting the *Markov assumption* twice. To begin with applying *Bayes rule*, expression (1) can be transformed to

$$Bel(s_t) = \frac{p(o_t | s_t, a_{t-1}, \dots, o_0, m) p(s_t | a_{t-1}, \dots, o_0, m)}{p(o_t | a_{t-1}, \dots, o_0, m)} \quad (2)$$

The denominator represents the probability of getting the measurement o_t when the terminal user reached a location due to the action a_{t-1} . This probability is constant relative to s_t and is denoted as η , and called the *normalization constant*. Therefore, expression (2) is written as

$$Bel(s_t) = \eta p(o_t | s_t, a_{t-1}, \dots, o_0, m) p(s_t | a_{t-1}, \dots, o_0, m) \quad (3)$$

Applying *Markov assumption* we find that

$$Bel(s_t) = \eta p(o_t | s_t, m) p(s_t | a_{t-1}, \dots, o_0, m) \quad (4)$$

Employing the *theorem of total probability*, the right most term in (4) is expanded by integrating over the state at time $t-1$

$$Bel(s_t) = \eta p(o_t | s_t, m) \int p(s_t | s_{t-1}, a_{t-1}, \dots, o_0, m) p(s_{t-1} | a_{t-1}, \dots, o_0, m) ds_{t-1} \quad (5)$$

Exploiting *Markov assumption* for the second time and noting that the second term in the integration is simply $Bel(s_{t-1})$ we obtain the recursive equation

$$\begin{aligned} & \overline{p(a | b, c) = \frac{p(c | a, b) p(a | b)}{p(c | b)}} \\ & \overline{p(A) = \int p(A | B_i) p(B_i) dB_i} \end{aligned}$$

$$Bel(s_t) = \eta p(o_t | s_t, m) \int p(s_t | s_{t-1}, a_{t-1}, m) Bel(s_{t-1}) ds_{t-1} \quad (6)$$

Where

- $p(o_t | s_t, m)$ is the measurement model, and
- $p(s_t | s_{t-1}, a_{t-1}, m)$ is the MS motion model.

As the actions performed by the terminal user (pedestrian) cannot be directly obtained, we have decided to implement equation (6) non-recursively. The prior or initial belief $Bel(s_{t-1})$ is therefore initialized by a uniform distribution over the state space for every run of the filter.

C. Implementation of the DBF

As already mentioned a discrete version of the Bayes filter has been implemented. The basic idea is to represent the *belief* $Bel(s)$ at any time by a set of n weighted samples (position hypotheses) distributed according to $Bel(s)$ as follows

$$Bel(s) \approx \{s^{(i)}, w^{(i)}\}_{i=1, \dots, n} \quad (7)$$

Where

- $s^{(i)}$ are samples of the hypothesized positions of the MS, and
- $w^{(i)}$ are non-negative numeric values called *weights* and sum up to 1. These *weights* determine the importance of each sample.

The continuous belief $Bel(s)$, thus, is approximated by a discrete probability function defined by the set of samples. The weights $w^{(i)}$ are calculated for every sample according to the measurement model mentioned above as

$$w^{(i)} = p(o_t | s_t, m) = \prod_{j=1}^M \frac{1}{\sigma_{RxLev} \sqrt{2\pi}} e^{-\frac{(RxLev_j - RxLev_{DB_j})^2}{2\sigma_{RxLev}^2}} \quad (8)$$

Where

- M is the number of the observed BSs (main and neighboring),
- σ_{RxLev} is the standard deviation of the measured RxLev,
- $RxLev_j$ is the measured RxLev from the j -th observed BS, and

- $RxLev_{DB_j}$ is the database RxLev prediction value of the j -th observed BS at $s^{(i)}$.

The final location estimate \hat{s} is calculated from $Bel(s)$ using one of three methods:

1. Maximum likelihood estimate (MLE), i.e. taking the sample with the highest weight as the location estimate. This is the estimate at which the posterior is *maximum*.

$$\hat{s} = \arg \max Bel(s) \quad (9)$$

2. Weighted average estimate (WAE), i.e. taking the weighted average of all samples representing the belief distribution as the location estimate. This is the *mean value* of the posterior distribution. It will coincide with the MLE only in case of *unimodal* and *symmetric* distributions.

$$\hat{s} = \frac{1}{\sum_{i=1}^n w^{(i)}} \sum_{i=1}^n s^{(i)} \times w^{(i)} \quad (10)$$

3. Trimmed average estimate (TAE), i.e. taking the average of the k best weighted samples as the location estimate. Where $k < n$.

$$\hat{s} = \frac{1}{k} \sum_{i=1}^k s^{(i)} \quad (11)$$

The calculation flow of the whole localization process is illustrated in Figure 1. The accuracy performances of the final location estimation methods are given in the following section.

III. EXPERIMENTS AND RESULTS

A. Experimental Setup

The measurements were taken in an E-Plus GSM network by a pedestrian walking along a route of about 2.4 km in a suburban area in Hannover, Germany. They were collected every 4 seconds with a total number of 250 measurement reports. The reports included TA and RxLev information from the serving and neighboring cells. Every report was stamped by a GPS position as a location reference to evaluate our algorithm. The considered area is about 9 km² and contains 6 base stations (each with 3 sectors) and 4 indoor antennas. The average cell radius is about 2 km and the pixel resolution of the provided RxLev database is 5 m.

B. Simulation Results

We investigated the accuracy of our localization algorithm by off-line simulations using the three proposed methods for location estimation. The cumulative distributions of the

localization error of these methods are depicted in Figure 2 and summarized in TABLE I.

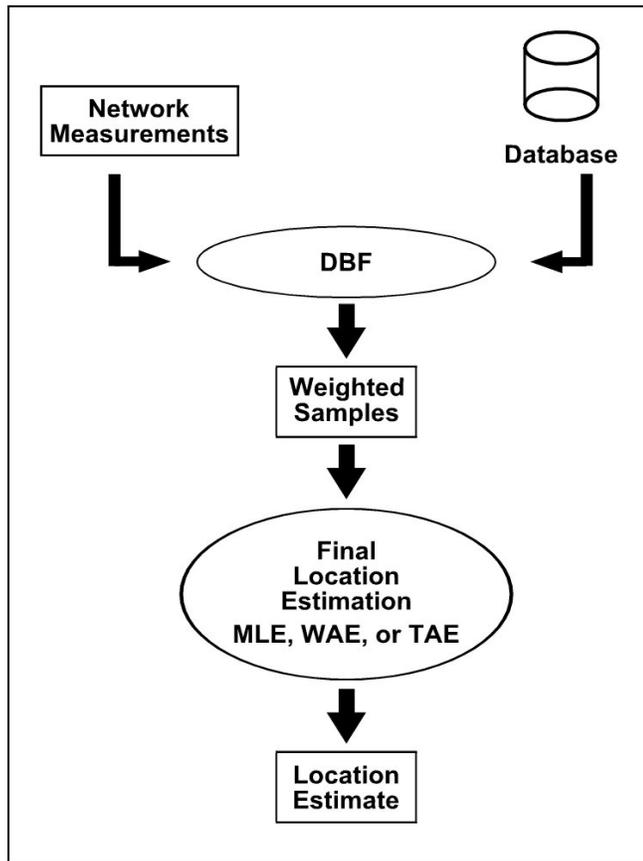


Figure 1. Schematic diagram of the localization algorithm showing the flow of calculations.

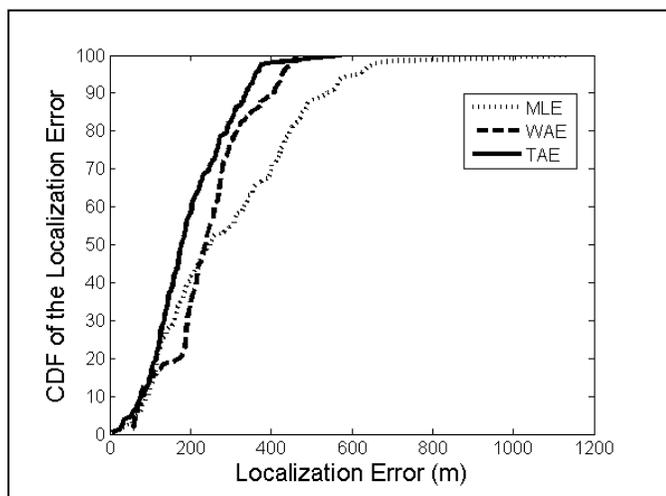


Figure 2. Cumulative distribution functions (CDF) of the localization error using the proposed methods for location estimation.

TABLE I. ACCURACY OF THE PROPOSED METHODS FOR LOCATION ESTIMATION

| Localization Error | MLE | WAE | TAE |
|--------------------|-------|-------|-------|
| 67% | 378 m | 275 m | 229 m |
| 95% | 610 m | 428 m | 361 m |
| mean | 294 m | 240 m | 197 m |

The results show that taking the MLE as the location estimate is not recommended as this is highly affected by noisy RxLev measurements and database inaccuracies. The WAE provides better estimations as it considers all samples with respect to their weights, thus reducing the effects that degraded the previous estimation method. However, taking the average of a specified proportion (in our simulations $k = 10\%$ of the samples) of the best weighted samples still yields better location estimates than the WAE. This has provided less sensitivity to erroneous RxLev measurements and helped neglecting outlier samples. The simulations have also shown that the real MS location is almost always in the region of the 10% best weighted samples. The localization accuracy of our proposed technique is better than those reported in [5], [11] and [6] for the suburban case using their database correlation methods.

In [5] and [11], the authors did not give details about the propagation model utilized in establishing their database. Furthermore, the area of their trial zone, the number of cells, and the average cell radius are not mentioned. Mean localization error of their proposed method is 270 m in non-urban areas. The database in [6] was constructed for the non-urban test area using the Hata-Okumura model, which is less accurate than the deterministic model [8] applied in building up our database. Moreover, the pixel resolution of their database is 20 m. The trial area is 50 km² with 43 cells, and the average cell radius is approximately the same as ours. Their localization error has a standard deviation of 602 m and a mean of 532 m. Note that in [5], [11] and [6], the suburban and rural areas are not handled separately.

IV. CONCLUSION AND OUTLOOK

Mobile terminal localization has become an active area of research recently since many services for mobile networks rely on an accurate positioning of the user terminal. The optimal solution would be to achieve localization with acceptable accuracy utilizing the available information in the existing network without any hardware modifications to base and mobile stations. The database correlation method is a popular technique to provide such a cost-effective solution. The accuracy of these techniques depends mainly on the method utilized for matching location sensitive patterns and yielding final location estimates (MLE, WAE, and TAE), characteristics and type of the area covered by the cellular network (density of cells), the technique used to construct the database (deterministic or empirical), and the pixel resolution.

In this paper we presented simulation results of mobile terminal localization in a GSM network deployed in a suburban

area. The localization algorithm is based on database correlation and Bayes filtering as a pattern matching technique. The database is constructed by a 3D deterministic radio wave propagation tool used for network planning.

The accuracy of three methods to yield location estimates has been presented and compared. Considering only the estimate with the highest calculated likelihood (i.e., maximum likelihood estimate) yielded lower accuracy than taking all hypotheses into account according to their weights (i.e., mean value of the posterior distribution). However, averaging a sufficient number of hypotheses has further improved the localization accuracy. It has been shown that using this method a mean localization error of 197 m and a standard deviation of 229 m were reached in a GSM network with low BS-to-area ratio as is the case in suburban and rural areas. This result is better than those already reported in the literature using similar techniques in similar environments.

Bayes filtering provides an efficient probabilistic framework for parameter estimation in multi-hypotheses situations. Furthermore, since database correlation methods do not need any hardware extensions at BSs or in MSs, the proposed localization algorithm is very effective in terms of cost, accuracy and system impact.

Experiments for car and indoor cases will be conducted. Furthermore, experiments in urban and dense urban areas for pedestrian as well as for car and indoor situations will be carried out to validate the proposed localization technique and check its capability in such scenarios. It is also planned to utilize map-matching techniques in order to further enhance the accuracy of our localization algorithm.

ACKNOWLEDGMENT

We would like to thank Thomas Kürner, Peter Unger, and E-Plus Mobilfunk GmbH 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] G. D. Ott, "Vehicle location in cellular mobile radio system," *IEEE Trans. Vehicle Technology*, vol. VT-26, pp. 43-46, Feb 1977.
- [3] T. S. Rappaport, J. H. Reed, and D. Woerner, "Position location using wireless communications on highways of the future," *IEEE Communication Mag.*, vol. 34, pp. 33-41, Oct 1996.
- [4] H. Laitinen, J. Lähteenmäki, and T. Nordström, "Database Correlation Method for GSM Location," *VTC 2001 Spring*, Rhodes, Greece, May 2001.
- [5] H. Schmitz, M. Kuipers, K. Majewski, and P. Stadelmeyer, "A new method for positioning of mobile users by comparing a time series of measured reception power levels with predictions," *VTC 2003 Spring*, Jeju, South Korea, May 2003.
- [6] D. Zimmermann, J. Baumann, M. Layh, F.M. Landstorfer, R. Hoppe, and G. Wölfle, "Database Correlation for Positioning of Mobile Terminals in Cellular Networks using Wave Propagation Models," *VTC 2004 Fall*, Los Angeles, USA, Sept. 2004.
- [7] T. Nypan, "Mobile terminal positioning based on database comparison and filtering," Dissertation at the Norwegian University of Science and Technology, 2004-65, July 2004.
- [8] Th. Kürner and 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.
- [9] D. Fox, J. Hightower, L. Liao, D. Schulz, and G. Borriello, "Bayesian Filters for Location Estimation," *IEEE Pervasive Computing*, vol. 2, no. 3, 2003.
- [10] S.J. Russell and P. Norvig, "Artificial Intelligence: A Modern Approach," 2nd ed., Prentice Hall, 2002.
- [11] M. Kuipers, K. Majewski, H. Schmitz, and P. Stadelmeyer, "Improved Method for Positioning Mobile Phones Based on Series of Measured Reception Power Levels," *Proc. of SPIE Location Services and Navigation Technologies*, vol. 5084, pp. 19-27, Orlando, FL, USA, April 24, 2003.