IndoorSim: Simulation of Wireless-LAN-Based Indoor Positioning Systems Incorporating CAD-Floorplans

Martin Werner
Ludwig-Maximilians-Universität München
Oettingenstraße 67, 80538 München
martin.werner@ifi.lmu.de

Abstract: With this paper, we explain a simulation environment for the popular area of Wireless-LAN-based indoor positioning based on free software. Using GNU Octave and typical radio propagation models, this open source indoor localization environment allows one to develop, analyze, understand and teach indoor positioning algorithms based on CAD-plans. The toolkit allows for calculating complex Wireless-LAN propagation maps, generate sequences of measurements following paths drawn in CAD, generation of room-based or coordinate-based fingerprinting databases as well as statistical analysis including the effect of noise on positioning accurarcy and precision. Furthermore some extraordinary simple data formats are proposed to facilitate exchange between indoor localization researchers and to support the collection of measurements for the indoor positioning community.

1 Introduction

In the last decade many new services for mobile phones emerged which integrate the location of a mobile terminal into the service. With the broad availability of GPS-based smartphones and the growth of fingerprinting localization databases for the outside area, this type of service has become very simple to implement. Due to the strong limitations induced by the small display and difficult on-display input, the integration of location information leads to better service experience and allows for guidance and navigation services.

However, the position of a mobile phone can only be determined for the outside area, as GPS is not available in indoor environments. Nevertheless it could be very useful to have some insight in the geolocation of mobile devices inside buildings or even to have a continuous positioning integrating the outside area and indoor environments.

Due to the complexities in setup and maintenance of indoor positioning systems, wide support for indoor geolocation is not available. However there are many examples of specialized indoor geolocation systems which reach acceptabe accuracy and precision.

In general, indoor geolocation can be determined by measurements of physical parameters which vary with location. There are many systems based on special infrastructure brought into the building. The classical systems were often based on infrared light due to the availability of infrared-technology in these days. Others were using ultrasonics which have very good diffuse propagation charachteristics. Two such systems are the Active Badge System [WHFG92] based on diffuse infrared light emitted from tags and read from a distributed network of sensors and the Cricket Location Support System presented in [PCB00].

These can provide high accuracy and dependability, but induce infrastructure costs. Other systems are based on existing infratsructure such as Wireless-LAN access points. These systems reach only limited accuracy, but do not induce infrastructure costs and hence have a better chance in providing ubiquitous services not limited by an administrative domain.

In this paper, we want to focus on indoor geolocation using Wireless-LAN infrastructure. We present a public domain simulation environment which can be used to design, evaluate and test Wireless-LAN-based positioning systems for indoor environments.

2 Related Work

Now, how can an existing Wireless-LAN infrastructure be used to provide indoor geolocation? Wireless-LAN typically consists of a distributed system of access points, where mobile terminals (clients) can associate with. As the mobile terminal needs to know, that a Wireless-LAN infrastructure is available, access points periodically broadcast beacon messages signaling their presence. To facilitate a good choice of access point, a client typically tries to associate with a Wireless-LAN access point with the best signal. As the channel quality measures are unknown at this point in time, Wireless-LAN clients typically rely on signal strength information which is therefore reported by most Wireless-LAN drivers. As Wireless-LAN uses a narrowband signal for communication, the signal strength can be modelled as a constant number and not depending on the frequency as for example with wideband systems.

From empirical studies, we know, that the signal strength of a Wireless-LAN system depends strongly on the distance between sender and receiver, the propagation path and several effects such as noise, self-interference, distortion, scattering etc.

Though the physical properties of all these signal parameters are well-studied, we can not use them in practice, as we typically do not have a model of the surroundings which is correct enough. Therefore, we typically rely on statistical methods and simplified propagation models to induce distances from signal strength information in some sort of regression analysis and machine learning.

2.1 Wireless-LAN signal propagation models

The most simple propagation model, which can be applied to Wireless-LAN signal, is the propagation in free-space (vacuum). The following equation gives the corresponding equation (λ is the wave length, d denotes distance, the results are given in dB) modelling the physical relationships:

$$E(d) = 10 \log_{10} \left(\frac{4\pi d}{\lambda} \right)^2$$

This equation consists of an application of the "inverse square law" which states, that radiated energy of a point source fades with the square of the distance, as the energy is distributed across the surface of a sphere ($4\pi r^2$), which grows quadratically with the radius. The other component is modelling the aperture of the receiving antenna and is frequency-dependent.

Assuming a narrowband signal with constant frequency $f = \frac{c}{\lambda}$ and incorporating this inaccurarcy and all other effects (antenna gain, propagation path, reflection, distortion etc.) into the exponent (^{2} in the equation above) and a reference starting value at a fixed distance d_0, we find the so-called "One-Slope Model", whose model parameters α and $E(d_0)$ are derived from measurements.

$$E(d, \alpha) = E(d_0) - 10\alpha \log_{10} \left(\frac{d}{d_0} \right)$$

The reference distance is typically taken as 1m. Typical values for α are given in the literature [Rap02] ranging from 1.8 to 3. The influence of specific material on Wireless-LAN propagation is studied in [ZBB05].

The next obvious refinement of this model is to add support by a floorplan. In a floorplan, we could calculate the number of walls which have to be penetrated and calculate a common wall attenuation factor. Taking this refinement, we end up with the Motley-Keenan model [MK88], which is used for example in RADAR [BP00]. The refined model equation is given below, where the constant wall attenuation factor is given as W and n denotes the number of walls between reciever and transmitter.

$$E(d, \alpha, n) = E(d_0) - 10\alpha \log_{10} \left(\frac{d}{d_0} \right) - nW$$

Based on the previous propagation models, which are purely based on the direct connecting line between sender and reciever, several models have been proposed and discussed, which take into account other propagation paths. The problem with simulating the physical properties are twofold: First of all the available floorplans are not sufficiently correct and on the other hand the solvation of the involved differential equations with finite difference methods or finite element methods is computationally very expensive. A good way in the middle is the adoption of raytracing techniques, as they give good results with medium calculational overhead [TMDH08]. Another technique, which further reduces from the complexities of raytracing is the calculation of so-called dominant paths [WWW$^+$05] and using the length of these dominant paths as the distance parameter form some distance-based propagation model.

2.2 Noise in RSSI readings

The most critical difficulty in determining the indoor geolocation is given by noise. The recieved signal strength at a fixed location varies over time. This noise consists of inaccuracy of the measurement equipment as well as changes in the environment. The recieved signal strength was measured in a long time experiment with a HTC Desire smart phone and a modern Wireless-LAN infrastructure. The results are given in Figure 1

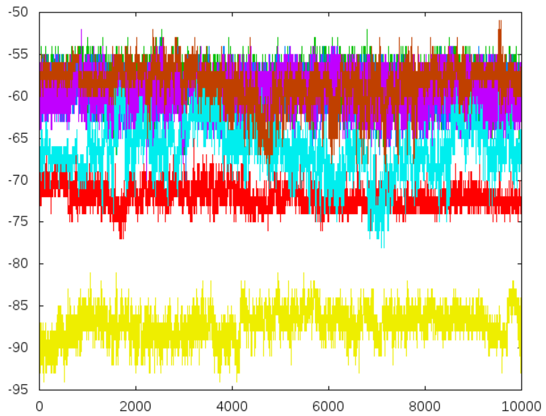

Figure 1: Long term RSSI readings taken from a fixed position in an office environment

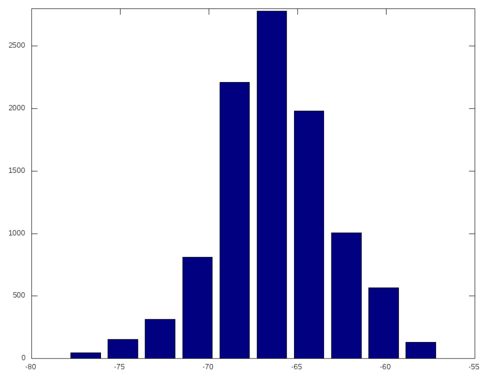

Figure 2: A the histogram for one of the access points

As you can see, the signal strength is fluctuating quickly. Fortunately, the distribution of the measurements is mainly following a gaussian process, as you can see from the histogram given in Figure 2. Furthermore the spectral power density is nearly constant (i.e. the noise is nearly white), as you can see from integrating the fourier transformation of the measurements, which results in a (nearly) straight line.

Based on this experiment, we decided to add random white gaussian noise to our model equation to simulate this time fluctuation in measurements. Of course, this approach implies, that algorithm validation has to be done using a multiple of different random noise samples, such that the result is not based on being lucky with respect to noise addition.

2.3 Wireless-LAN positioning algorithms

Radio-based positioning algorithms can be divided into two groups: The first class of positioning algorithms induces a coarse, logical position (e.g. a room label) from the measurements. Therefore a labelled dataset is used for training and a classification algorithm is used for predicting the label of newly arriving measurements. The other class of positioning algorithm uses a database of measurements taken at different positions and a method of numerical prediction such as weighted k-next-neighbours or some regression technique. A direct application of the radio propagation models with a map with the accesspoint locations using the well-known multilateration technique does not work out good enough. The complexities of multipath propagation and errors in the floorplan induce to much difference for this type of localization technique. However it is relatively easy to use, as the distance measurements are combined into a overdetermined linear equation which can be solved by means of least squares regression.

3 Simulation Environment Based on Octave

In this section we describe, how we implemented a simulation environment which takes into account all effects described in the previous chapter and can be used to implement, test and evaluate indoor localization algorithms with simulated data and even with real track data in one homogenuous environment. As the technical basis we decided to use the Octave [Eat02] software suite, which is an open source environment providing a workbench for numerical computations similar to the commercial Matlab software.

3.1 Representing the Simulation Data

For our simulation environment, we need a representation of a CAD-floorplan which is well-suited for use with the calculations, we make on this data basis. We decided to represent a CAD file as an ordered sequence of lines which all have a starting point and an endpoind in two-dimensional coordinates. Layer information is silently dropped. Hence, a CAD-file is represented as a $n \times 4$-matrix where each row of the matrix represents one line. Complex entities such as circles, arcs, splines and bulges are tessellated into line strips shorter than a given constant tessellation threshold.

This representation of CAD is very straightforward, but does not fit to the plotting implementation of Octave. For plotting purposes, the **line**-function expects line lists to be $n \times 2$ matrices of points, which are all interconnected. Plotting multiple lines is done by inserting a point $[\mathrm{NaN}, \mathrm{NaN}]$ into this sequence.

The chosen representation is well-suited for calculating a propagation model based on the access point position, parameters and the Motley-Keenan Model. This is implemented in a module **ins_calculateDiscreteModel**, which takes a floorplan, a position, some global settings for the wall attenuation factor and the access point reference quantities and propagation coefficient. Figure 3 shows how such models might look like based on a trivial floorplan with and without noise.

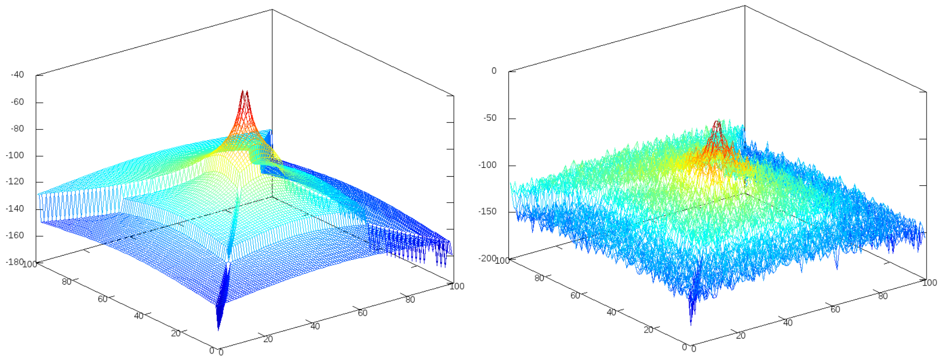

Figure 3: A simple floorplan is used to illustrate propagation modelling using the IndoorSim

For a real floorplan figure 4 shows how complex the propagation can become, if the rooms become a bit more complex. This visualization is without any additional noise. It is simply Motley-Keenan propagation on a CAD floorplan.

For the simulation of a realistic indoor localization scenario, we also need the positions and signal strength (signal power at reference distance) parameters of the access points. These are also modelled as a matrix. In each row of this matrix, the CAD coordinates of the access points and some access point parameters can be specified.

To generate a realistic situation for an indoor positioning algorithm, there are two methods: It is possible to generate a fingerprint database based on a rectangular lattice, to follow a polyline on a special layer in the CAD drawing and to generate a fingerprint at each point in the polyline. Furthermore we have a special module which is able to load room polygon data from CAD, if available. In this situation, the generated fingerprints can be labelled with room labels providing for easy and efficient generation of training databases for classification-based location determination systems.

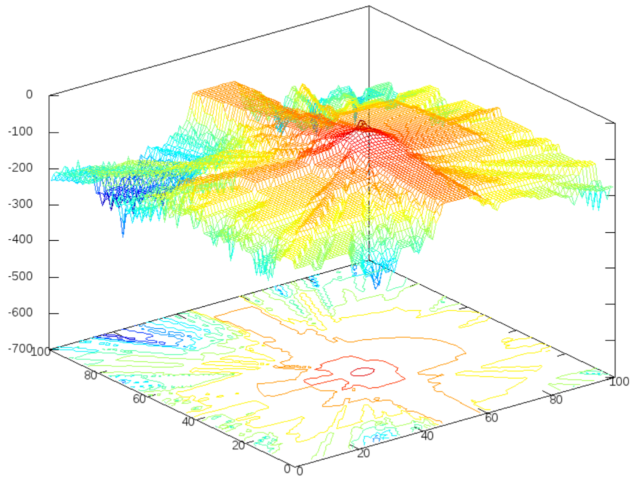

Figure 4: The propagation model for a real floorplan calculated by IndoorSim

3.2 Fingerprints and Tracks

A fingerprint consists of a $(2 + n)$ row vector. The first two entries give the x and y coordinate of the fingerprint position. The rest of the vector consists of the values of RSSI for each of n access points. It is possible to generate these from the simulation framework, to load them from a file and to alter them by random white gaussian noise. A fingerprint track with m points consists of a $m \times (2 + n)$ entries, each line representing one point. Unfortunately it is not possible to have missing values in this matrix. The availability of a RSSI-value reading for a positioning system can be influenced by a global constant (below which values are not used).

3.3 Positoining Algorithms

Within the simulation framework, we implemented several baseline positioning algorithms, which are ready for use. The first one is the classical approach used by RADAR: Weighted k-next-neighbors in signal space. For this positioning algorithm, a grid-based fingerprint database is constructed. To obtain the position of a mobile node from sensor readings, the

k next neighbors (for an integer configuration value k) are used. Then these are weighted by the euclidian distance in signal space to the incoming sensor reading. The positions of these k next neighbours are then combined using a weighted sum $p = \sum w_i p_i$. It is essential to use a well-defined value k for a fingerprint database. If k is too small, the position values get very noisy depending on the resolution of reference values. If the value k is too large, determined positions tend to drift away towards a mean of the positions of the surrounding access points. A typical situation for this positioning algorithm was simulated using a CAD floorplan and the simulation environment.

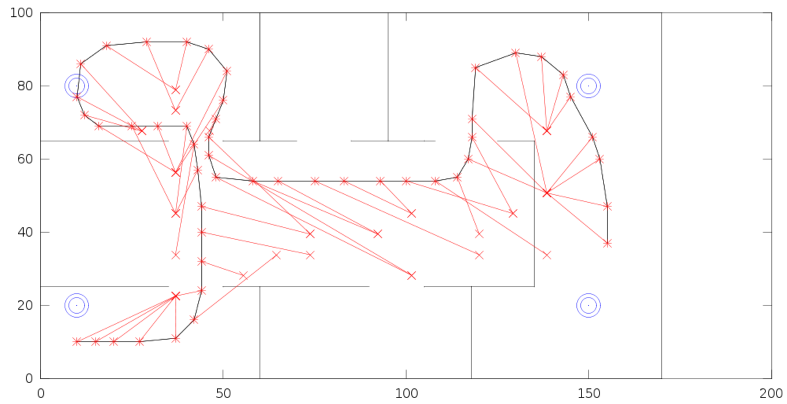

Figure 5: Weighted-kNN on a synthetic floorplan using $k = 3$ and a simulated Wireless-LAN environment.

Figure 5 shows a floorplan, a track inside the floorplan and the deviation of each track-points positioning result from the ground truth. What you can see, is that it sticks near the right path, but with a pretty random deviation into all directions. However, as you can see clearly from the picture, the rooms are assigned correctly in most cases.

This effect gives rise to another positioning task of finding the correct position out of a set of measurements. Therefore, the simulation environment is able to load CAD-files in DXF format which contain one closed polyline for each room. Then fingerprints are generated in a regular grid, just as before, but each of them is assigned a room label. The results are then stored in the ARFF format as used by Weka [HFH+09] for data mining. Using the Naive Bayes classificator implemented in weka, room-based positioning using simulated Wireless-LAN received signal strength is possible with this simulation environment. This type of label-based indoor positioning is widely used in practice, as it is much more reliable. A visualization of a room localization is given in Figure 6.

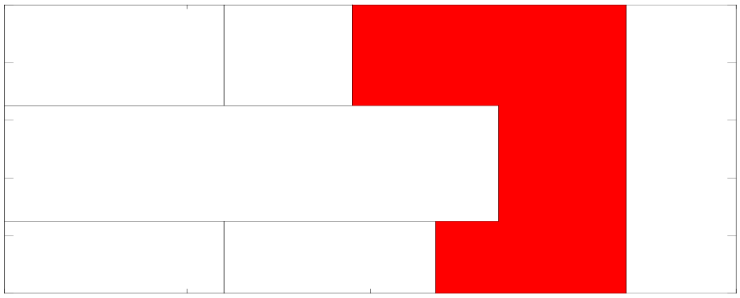

Figure 6: Label-based indoor localization based on polygonal room data

4 Outlook

With our simulation framework it is now possible to simulate Wireless-LAN-based indoor positioning. While it is now mainly a tool for teaching we hope to extend the functionality and make it a widely-used framework for signal-based indoor positioning research. We hope to provide a system which will help beginners to understand and analyze in detail well-known indoor positioning algorithms and signal charachteristics while being powerful enough to facilitate professional use and data exchange. The most important problem for researchers in the area of indoor positioning is the difference between buildings. It is difficult to show, that a special positioning approach is better than a simpler one in general and not only due to overfitting to the test environment. If research would adopt a sufficiently simple data format and share measurements, the whole area of research would be much more reliable and honest. Moreover the implementation of more complex propagation models and the automatic generation of fictional floorplans can help much in comparision of methods. The project is open source and is currently hosted at Google Code [Wer11].

References

[BP00] P. Bahl and V. N. Padmanabhan. RADAR: an in-building RF-based user location and tracking system. In *Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No.00CH37064)*, volume 2, pages 775–784. IEEE, 2000.

[Eat02] John W. Eaton. *GNU Octave Manual*. Network Theory Limited, 2002.

[HFH+09] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The WEKA Data Mining Software: An Update. *SIGKDD Explorations*, 11, 2009.

[MK88] AJ Motley and JMP Keenan. Personal communication radio coverage in buildings at 900 MHz and 1700 MHz. *Electronics Letters*, 24(12):763–764, 1988.

[PCB00] Nissanka B. Priyantha, Anit Chakraborty, and Hari Balakrishnan. The Cricket location-support system. In *MobiCom '00: Proceedings of the 6th annual international conference on Mobile computing and networking*, pages 32–43, New York, NY, USA, 2000. ACM.

[Rap02] T. S. Rappaport. *Wireless communications principles and practices*. Prentice-Hall, 2002.

[TMDH08] Nhut Tran-Minh and Tuan Do-Hong. Application of raytracing technique for predicting average power distribution in indoor environment. In *Communications and Electronics, 2008. ICCE 2008. Second International Conference on*, pages 121 –125, 2008.

[Wer11] Martin Werner. The IndoorSIM Open Source Project, 2011. https://code.google.com/p/indoorsim/.

[WHFG92] Roy Want, Andy Hopper, Veronica Falc ao, and Jonathan Gibbons. The active badge location system. *ACM Trans. Inf. Syst.*, 10(1):91–102, 1992.

[WWW+05] G. Wölfle, R. Wahl, P. Wertz, P. Wildbolz, and F. Landstorfer. Dominant path prediction model for indoor scenarios. In *German Microwave Conference (GeMIC)*, 2005.

[ZBB05] M. L. Zehner, K. Bannicke, and R. Bill. Positionierungsansätze mittels WLAN-Ausbreitungsmodellen, 2005.