WLAN Indoor Positioning Based on Euclidean Distances and Fuzzy Logic

Andreas TEUBER, Bernd EISSFELLER

Institute of Geodesy and Navigation, University FAF, Munich, Germany,
e-mail: (andreas.teuber, bernd.eissfeller)@unibw.de

Abstract - This paper presents an indoor positioning approach based on pattern recognition of IEEE 802.11 signal strength measurements. The approach is divided into the calibration mode and an operational mode when a user demands to be positioned. During the calibration mode signal strength measurements on sites with known coordinates are carried out and stored in a database. The current signal strength pattern of the measurements in the operational mode is then compared to the database by computing Euclidean distances. An effective processing of these Euclidean distances is undertaken by means of fuzzy logic methods in two stages. While in the first stage a number of candidates of proximate calibration sites is investigated the second stage is responsible for an appropriate weighting of their coordinates and yields the positioning result. First test results of two different environments are presented.

1 Introduction

In indoor environments WLAN according to the IEEE 802.11 standard can provide location information by using signal strength measurements to access points (AP) and hence offers a smart synthesis of communication and navigation functionality. Especially the approach of pattern recognition of signal strength fingerprints can obtain location accuracies of 1 to 5 m. Currently, the most acquainted indoor navigation system based on the IEEE 802.11 which is commercially available is the Ekahau Positioning Engine (EPE). According to the manufacturer's instructions the EPE combines signal strength pattern recognition with an attempt to recover the user's history (including boundary conditions like allowed paths and speed) in order to determine the current position [1].

Extensive tests have demonstrated the accuracy potential of the system. Under optimal conditions, which especially means a sufficient density and an appropriate distribution of AP and an indoor environment which is highly subdivided in different rooms, the achieved accuracies have been even slightly better than the prospects mentioned by the manufacturer [2]. However, poorer conditions lead to a worse accuracy. This could be observed in regular structures like galleries, large rooms, and halls. In order to correct the location results for occurring systematic errors a post-processing approach was proposed [3]. Despite an increase of accuracy this approach is inadequate for real-time positioning. Since the proprietary EPE allows no unauthorized intervention with respect to the algorithms, the development of an independent system was required.

Instead of the absolute signal strength in this work the signal-to-noise ratio is used as input parameter. A software capable of recording signal-to-noise ratio measurements of all available AP generates location fingerprints. Rather similar to the EPE during the calibration mode signal strengths of a set of calibration points of known coordinates to be stored in a database is established initially. During the operational mode the current signal-to-noise ratio pattern of the user is compared to the database. Processing algorithms to be developed have the task to determine a user position from the comparison of the signal-to-noise ratio patterns. The next chapters starting with the idea of Euclidean distances describe how the processing algorithms work and how they are implemented.

2 Processing Euclidean Distances

With respect to any detected AP a signal-to-noise ratio difference between the current record and each calibration record of the database can be computed. In the following the current signal-to-noise ratio records are referred to as observation data and the collected signal-to-noise ratio records in the database are referred to as calibration data. The computed signal-to-noise ratio differences can be expressed in a signal-to-noise ratio difference vector in which the number of elements represents the number of AP. The norm of this vector is well-known as the Euclidean distance:

$$d_{j,k} = \sqrt{\left(c_j^{AP_1} - s_k^{AP_1}\right)^2 + \left(c_j^{AP_2} - s_k^{AP_2}\right)^2 + \ldots + \left(c_j^{AP_i} - s_k^{AP_i}\right)^2} \tag{1}$$

with:	c	… calibration data		i	…	total number of access points
	s	… observation data		j	…	index of calibration points
	d	… Euclidean distance		k	…	index of observation points

Considerably, the computed minimum Euclidean distance $d_{j,k} = d_{min}$ obtains a significant indication of the calibration point closest to the current observation point. Depending on the density of calibration points in the environment to be investigated, the coordinates of the calibration point connected to d_{min} could be adopted as the correct solution. However, in general, the calibration procedure is a time-consuming work and therefore the density of calibration points will usually be as low as maintainable. Thus, it is recommendable to compute the user position from a weighted mean of a number of calibration points with lowest Euclidean distances:

$$x_k = \left(\sum_{j=1}^{J} \frac{1}{d_{j,k}}\right)^{-1} \cdot \left[\frac{x_1}{d_{1,k}} + \frac{x_2}{d_{2,k}} + \ldots + \frac{x_J}{d_{J,k}}\right] \tag{2a}$$

$$y_k = \left(\sum_{j=1}^{J} \frac{1}{d_{j,k}}\right)^{-1} \cdot \left[\frac{y_1}{d_{1,k}} + \frac{y_2}{d_{2,k}} + \ldots + \frac{y_J}{d_{J,k}}\right] \tag{2b}$$

with:	x	… x-coordinate		J	…	total number of calibration points to be weighted
	y	… y-coordinate		j	…	index of calibration point
	d	… Euclidean distance		k	…	index of observation point

The single calibration point coordinates serve as inputs which is weighted by the reciprocal of the Euclidean distance. The factor in rounded brackets is necessary in order to normalize the overall sum of the reciprocals of the Euclidean distances.

The simple application of position determination based on Euclidean distances leads to feasible results without the need for extended computing capacities. For signal strength records of only one epoch the application of the Euclidean distances is very simple. If records of more than one epoch are available, it will be possible to determine supplementary information like the mean, the median, the maximum, the minimum or the standard deviation of the signal strength measurements. In the following it is assumed that these multi-epoch data may contain additional information exploitable for position determination. This approach which will make use of the Fuzzy Set theory will be explained in detail in chapter 4.

3 Fuzzy Set Theory

In science, it is common to work with accurate data and models. Since scientists are aware of the nature of imprecise information due to the limitations and restrictions of the used sensors and the applied processing methods to obtain the necessary information they attempt to make an assessment of precision by using terms like "variance" and "probability". However, in many situations of everyday life it is sufficient to characterize a condition by rather simple terms like "*the water is **cold**, **warm**, or **hot**.*" In most of the cases this information is satisfying for further actions. Fuzzy logic theory assimilates this idea and every condition like *cold* or *warm* is called a *membership function*. A set of membership functions is connected to a *linguistic term*, in this case the temperature.

The main objective of processing fuzzy sets is to calculate the grade of accordance of the current data set with the membership functions of the linguistic term. Initially, the user of fuzzy set theory has to define the linguistic terms as well as their membership functions. In practice, the user has to describe the condition *warm* quantitatively and qualitatively. Apart from the mean or centre temperature of *warm* the way of transition to the neighboured membership functions *cold* and *hot* has to be defined. It is conceivable that a membership function yields only a sharp output at a singular point like it is the case using triangular or Gaussian functions as membership functions (Fig. 1 left). Then the transition range embraces the whole area between the singular maximum points of the membership functions. The very most part of the temperature values is connected to a fuzzy output value. If the membership function is a trapezoidal function the fuzzy nature is restricted to those temperature values connected to the edges of the trapezoid (Fig. 1 right).

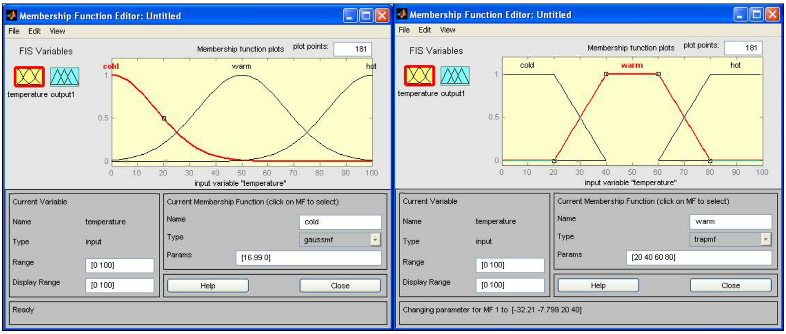

Fig. 1 Fictive linguistic term temperature and its membership functions *cold*, *warm* and *hot*.

Of course, the information on the temperature of the water is only one linguistic variable in order to obtain – for instance – the efficiency of the run of a washing machine. Another linguistic term could be the amount of the deployed washing powder. In order to achieve an assessment of the result (or the output) it is necessary to establish rules. The following trivial sounding rule is conceivable:

*If the temperature of the water is **hot**, the cleaning result will be **good**.*

The *if-clause* of the rule is always connected to the related membership function (e. g. *hot*) of the input, the *main-clause* of the rule is always connected to the related membership function (e.g. *good*) of the output. By means of a set of these rules and the obligatory definition of the output membership functions a confident statement on the cleaning result should be possible. All computations presented in the next chapters were conducted using the Fuzzy Logic Toolbox of MATLAB. Scientific papers on fuzzy set theory can be found among others in [4],[5],[6].

4 Two-Stage Fuzzy Set Design for WLAN Positioning (Airport Hangar Environment)

4.1 First Stage

Since the propagation of the radio signals is massively affected by multipath – a desired property for communication purposes – for WLAN positioning the principle of pattern recognition was preferred. Preparatory measurements had led to the result that pattern recognition yields accuracies twice better than the computation of free space loss. This is even more remarkable considering that these tests had been carried out in a large hall (airport hangar) without obstacles or partition walls. It has to be emphasized that the fuzzy system described in this chapter 4, especially the second stage, is designed for this kind of environment.

In chapter 2 the significance of the Euclidean distances has been pointed out. It has been foreshadowed that in case of multi-epoch data additional information of the signal-to-noise measurements can be used. For the approach of this paper these are:

1. the maximum signal-to-noise ratio (`snrmax`)
2. the standard deviation of the signal-to-noise ratio (`snrstd`)
3. the mean signal-to-noise ratio (`snrmean`)
4. the overall sum of all signal-to-noise ratios to all AP (`snrsum`)

On the fundament of these linguistic terms and their membership function a set of rules is established in order to express human assessments like, e. g.:

> *If the Euclidean distance of the mean of the signal-to-noise ratio is* ***good****, the matching between the calibration and the observation point will be* ***good****.*

The four linguistic terms mentioned above were connected to three membership functions each. Trapezoidal functions were used as membership functions. For each of the four linguistic terms the quality criteria *good*, *medium* and *bad* were assigned to the membership functions. In practice, most of the rules contain a combination of the linguistic terms. The rules were further connected to five output membership functions with the criteria *very good*, *good*, *satisfying*, *moderate* and *bad.* The output membership functions are triangular functions of equal shape isochronously distributed on the output scale which also ranges from 0 to 1.

Quality criteria →	good	medium	bad
↓Linguistic term			
snrmax	0 0.06 0.16	0.06 0.16 0.28 0.38	0.28 0.38 1
snrstd	0 0.12 0.22	0.12 0.22 0.45 0.55	0.45 0.55 1
snrmean	0 0.10 0.20	0.10 0.20 0.39 0.49	0.39 0.49 1
snrsum	0 0.12 0.22	0.12 0.22 0.45 0.55	0.45 0.55 1

Tab. 1 Quality and quantity criteria of the membership functions in order to achieve an equally distributed input.

The next step was to assign the recorded data to the membership functions. Initially, there had been no empirical data on the magnitudes of the Euclidean distances. For this reason the data of a whole set of records in an environment was analyzed in post processing. All individual Euclidean distances were divided by the maximum value of all computed Euclidean distances in order to normalize them. In other words, all computed Euclidean distances were now assigned to a scale from 0 to 1. Furthermore, the overall amount of computed Euclidean distances was divided into three parts in a way that they were equally distributed to the three membership functions. With this information the membership functions could also be defined quantitatively. Tab. 1 presents an overview of the quality and quantity criteria of the membership functions.

The edges of the trapezoidal functions were spread over a range of 0.1 on the normalized scale by definition. Besides, the sum of all membership functions was always 1.

In the next step the rules were tested with respect to any collected input vector. The input vector contains the normalized Euclidean distances of the four linguistic variables and can be seen in Fig. 2 in the lower left box (values 0.17 0.1 0.1 0.1). The Rule Viewer of the MATLAB Fuzzy Logic Toolbox visualizes whether or not a rule can be applied. Usually, a number of 12 to 15 rules had to be tested and the very majority of them were *and*-rules which means that every part of a rule had to be fulfilled cumulatively in order to accept the result. The processing of the rules in order to get an output is called aggregation. The example in Fig. 2 shows that only rule #3 is met completely. This rule expresses that if only the linguistic terms `snrstd` and `snrmean` are *good*, the output will be *very good*. In fact, the input for these two terms is 0.1 each and meets or falls below the required values of 0.12 and 0.1 (Tab. 1) for "very goodness", respectively. In this rule #3 there is no requirement for the other two linguistic terms. Furthermore, it can be seen that some of the rules are met at least partially. This case occurs when the red line of the input hits the trapezoidal function at its edge.

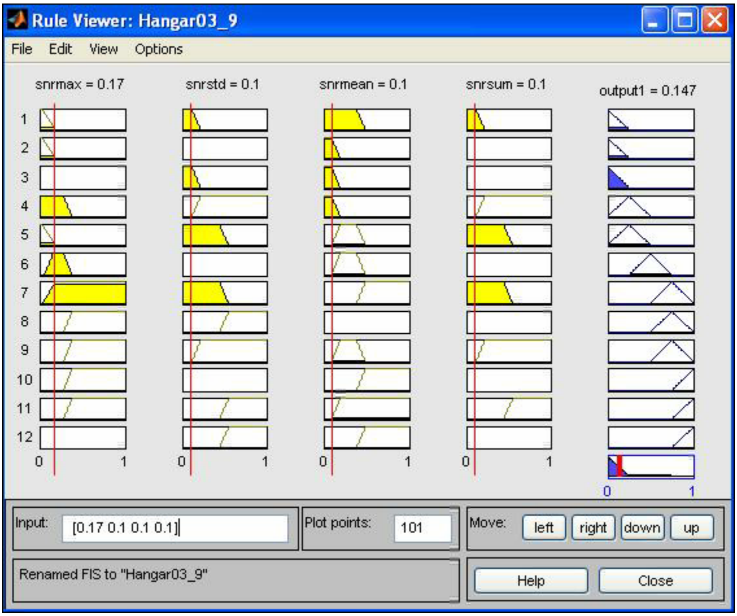

Fig. 2 Rule Viewer of the MATLAB Fuzzy Logic Toolbox

The individual outputs have to be summarized to a single value. This process is called defuzzification. The applied defuzzification method is calculation of the centroid. For every observation point the procedure of testing all rules is done with every available calibration point. The result is a ranking of calibration points. It is again conceivable to adopt the coordinates of that calibration point of the smallest output value after defuzzification. Just as well it is possible to take a set of the best calibration points again and to weight them with the reciprocal of the belonging output values and to average them. The first trials have been undertaken in a rectangular part of an abandoned airport hangar. The computation of an average out of a set of calibration point coordinates would always lead to coordinates of a point surrounded by the used calibration points. However, if a calibration point is located at one of the edges of the rectangular structure one provokes a systematic error by this kind of averaging. For this reason it was preferred to add a second stage of a fuzzy system which takes some basic topological issues into account.

4.2 Second Stage

The main objective of the second stage is to find an appropriate weighting of the output candidates of the first stage. All calibration points which have not been selected as candidates in the first stage, are already excluded for the second stage. The latter consists again of four linguistic variables, however, neither of them contains signal-to-noise information in a direct manner. They are:

1. the absolute value of the output of the first fuzzy stage (`fuz1-absolute`)
2. the quotient between the best and second best output of the first fuzzy stage (`fuz1-quotient`)
3. the number of Delaunay neighbours (`Delaunay-Neighbourhood`)
4. the periphery flag (`periphery`)

The first two linguistic terms which deal with the output parameters of the first fuzzy stage play an important role. If the absolute value of the output of the first stage or the quotient between the first and the second best output is extraordinary small it will be assumed that the observation point is very close to the calibration point. As a consequence, the latter will get a very heavy weighting in a way that the resulting coordinates will only slightly differ from that of the calibration point. Linguistic terms #1 and #2 are connected to three membership functions called high, medium and low. In contrast to the first fuzzy stage the parameters of the membership functions were set manually.

In order to avoid the shift of observation point coordinates located at the periphery of an environment relatively to the centre by averaging the coordinates of the surrounding calibration points, some topological basis data of the calibration points have to be acquired. For this reason a Delaunay triangulation of all calibration points of the environment was conducted. By means of this all Delaunay neighbours could be determined for each calibration point (Fig. 3). It is easy to understand that calibration points located closer to the centre of the environment have more Delaunay neighbours than those located more in the periphery. Therefore, the linguistic term 3 simply expresses the number of Delaunay neighbours.

In order to get a deeper insight of a peripheral location of a calibration point a further investigation was conducted. For each calibration point the azimuths to all Delaunay neighbours were determined and sorted in ascending order. Then the maximal angle between two neighboured azimuths was evaluated. If this value is small, e.g. below 120 deg, the point will definitely not be located at the periphery of a rectangular environment. However, if the angle is rather high, e.g. 180 deg, the calibration point will probably be located in the periphery at one side of the environment. If the angle is even higher, e.g. 270 deg, the calibration point will probably be located in a corner of the shape of the environment. If a calibration point detected to be at the edge of the environment is the best candidate of the first fuzzy stage, its coordinates will get a significantly higher weight in the second fuzzy stage in order to avoid the effect of centralization. The maximal angle of each calibration point is also visualized in Fig. 3. The linguistic terms #3 and #4 are connected to four membership functions each which were called *high*, *above average*, *moderate* and *low*. Like terms #1 and #2 they were set manually, too.

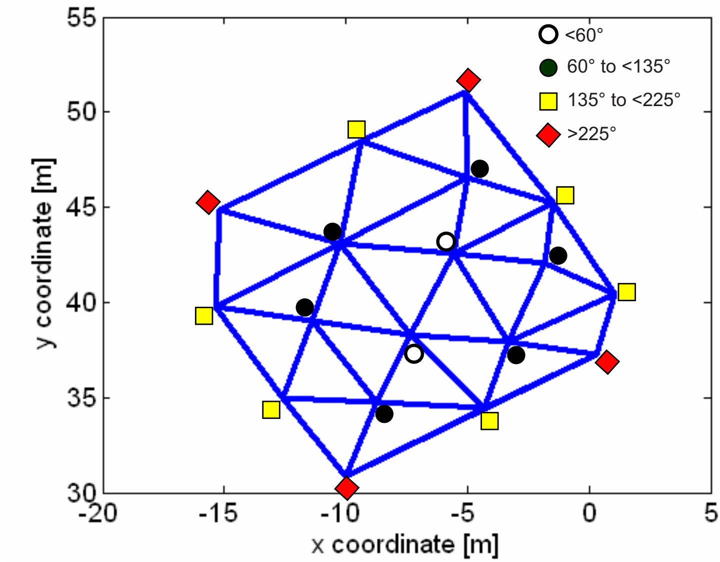

Fig. 3 Delaunay triangulation of calibration points and classification of maximal angles

The second fuzzy system runs in a sequential manner with respect to the calibration point hierarchy of the first fuzzy stage. Finally, the normalized reciprocals of the outputs of the second fuzzy stage are directly used as weighting factors for the desired coordinates of the observation point.

4.3 Results

The first test run of the system was conducted in one part of approximately 20 m by 20 m of the abandoned airport hangar which had a total dimension of 40 m by 60 m. In the 400 m² part of the hangar five access points of the IEEE 802.11b standard were installed, one in each corner and one in the centre. Three of the access points were Netgear ME 102 and the two remaining SMC 2671W. 16 calibration points were uniformly distributed in the test area. This adjustment was not done exactly but according to visual judgement. The average distance between two calibration points was approximately 5 m. Another 16 observation points were located in a manner that they alternate with the calibration points. Again, the fixing of the sites was done by visual judgement. Consequently, the approximate distance between two observation points was also 5 m. Each calibration and observation point was referenced to millimetre accuracy to a local coordinate system by an electro-optical tachymeter. Fig. 4 provides a schematic overview of the arrangement of access points, calibration points and observation points.

Signal-to-noise ratio measurements were then conducted on the calibration points in order to establish a database. In order to provide the required four parameters for the first fuzzy stage a sample of about ten epochs was collected on each point. Subsequently, the signal-to-noise measurements on the observation points were conducted in exactly the same manner.

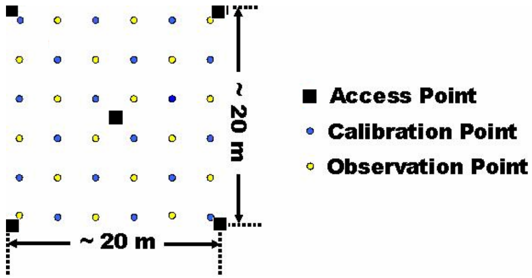

Fig. 4 Arrangement of access points, calibration points and observation points in the hangar environment

The results presented in the following were obtained with the settings of the fuzzy systems described above whereas the best four outputs of the first fuzzy stage were selected for the second stage. Of course, it is possible to vary this number. For simplicity, as parameter for error consideration and comparison of the results the average 2D position error of all individual results is given here. After the first stage the average 2D position error was 3.59 m. This is significantly better than a simple weighted computation of Euclidean distances without application of the fuzzy approach with an average 2D position error of 4.47 m. After the second stage the error decreased to 3.27 m. The best individual result is a 2D error of 0.35 m, the worst case is 6.58 m. There is no significant difference between observation points at the periphery and closer to the centre. The main results are presented in Tab. 2.

	Average 2D position error [m]			
Number of calibr. points	Euclidean Distance	Fuzzy Stage 1	Fuzzy Stage 2	EPE
16	4.47	3.59	3.27	No data
32	4.47	2.48	1.61	4.40

Tab. 2 Accuracy of the different indoor positioning approaches in the hangar environment

The results of the fuzzy approach have also been compared to the commercial Ekahau Positioning Engine (EPE). It is well-known that the EPE yields better results in heterogeneous environments where partition walls support the idea of the defined allowed paths and the user position history. In fact, with the EPE it is not possible to achieve the results of the fuzzy logic approach, the average 2D position error is 4.40 m whereas the individual errors vary from 0.96 m to 9.07 m. This is even more surprising since the results are not directly comparable. This is due to the fact that with the EPE on every observation point a calibration record was conducted and every 36 points were calculated. If this method was applied to the fuzzy logic approach, the error would decrease to 2.48 m after the first fuzzy stage and even 1.61 m after the second stage.

There has been only slight modification to the standard settings of the fuzzy system so far. A thorough review of the applied rules could be a key to a further improvement of the result. However, the extension to a higher number of membership functions was evaluated and led to a deterioration of the position accuracy since the user is in danger to lose the understanding of the whole system. As a consequence, the user is no longer able to express the necessary number of evident rules.

5 Modifications for a Heterogenous Environment (Office Building)

Both the fuzzy logic and the EPE system have also been tested in a combined office and laboratory building. This building has a similar dimension of approximately 20 m by 20 m. However, it is separated into ten offices and laboratories of different sizes, all of them separated to each other by massive, but relatively thin partition walls. A former investigation with the EPE led to excellent 2D position results of 1 m and better [2]. This was due to the two basic techniques of the EPE: first, the path design which requests the user to ascertain the allowed user paths in a building in advance on a digital map in order to avoid trajectories crossing walls (thick, continuous lines in Fig. 5 right); second, the consideration of the user history and a predefined maximum user velocity which rules out sudden user movements from one epoch to the next. The good results in [2] had been achieved since all observation points were located more or less exactly on the predefined paths. For the comparison with the fuzzy approach the observation points were located on all accessible points, regardless of their distance to the paths. Similar to the hangar, the office environment was equipped with five access points. The main features of the test design can be gathered from Fig. 5.

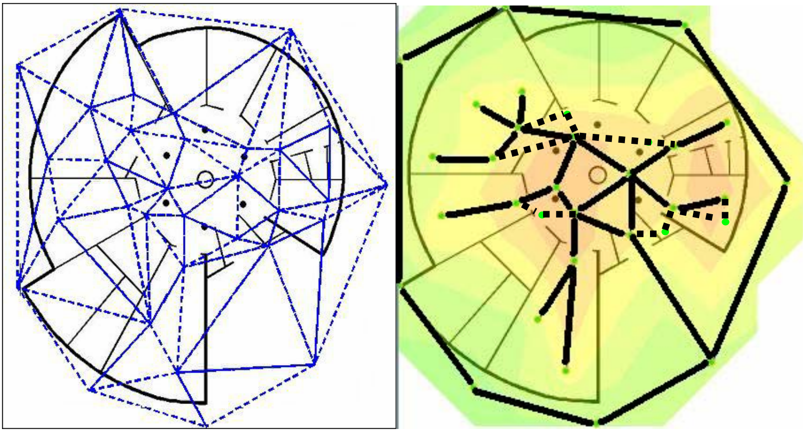

Fig. 5 Unfavourable Delaunay trinangulation in the office building (left) and the remedy (right): usage of the EPE paths combined with additional lines

Soon it became obvious that for the fuzzy system a Delaunay triangulation is fruitless since most of the triangle sides will cross the walls of the building (dashed lines in Fig. 5 left). For this reason the topological information had to be adapted to the specific environment. In the case of the office building the paths of the EPE calibration were used and supplemented by some more lines depicted in Fig. 5 right as dotted lines.

Despite of the adaptation of the topological information in the fuzzy system the EPE remained superior with respect to position accuracy. Again the total number of points was 32, 16 calibration points and 16 observation points. A total number of eleven points could be directly compared to the EPE. The average 2D position error of the fuzzy system is 3.14 m including one outlier of more than 10 m. After elimination of this outlier the average is 2.37 m, but still worse than the EPE result of 1.94 m.

It has to be pointed out that the presented results consider only position computations where the observation points do not coincide with the calibration points since this is the bigger challenge for systems based on pattern recognition algorithms.

6 Conclusion and Further Work

An approach of using fuzzy inference systems in order to determine positions on the basis of WLAN signal-to-noise ratio measurements and of pattern recognition was presented. The first results of the two-stage system are encouraging and show less variation with respect to different environments than the commercial EPE. If the position computation is limited to observation points which do not coincide with calibration points at all, the accuracy will be about 3 m for both environments. If observation points coincide with the calibration points, the accuracy will be significantly better. However, the processing of the topological information has to be adapted to the environment. It is conceivable that this adaptation could be realized automatically, possibly by another fuzzy logic system.

The investigation in the two presented environments are ongoing and will be extended to further environments. In the short term the investigations also comprise a better understanding of the fuzzy inference systems which could provide a (limited) improvement of accuracy. Furthermore, there are challenges with respect to the number of epochs necessary for an appropriate positioning. Of course, a moving user would prefer to determine his position by one single signal-to-noise ratio measurement. This is still contradictory to the principle of the first fuzzy stage where statistical information of the signal-to-noise ratio data are used as inputs.

7 References

[1] Ekahau, Ekahau Positioning Engine 2.0: 802.11-based Wireless LAN positioning system. *An Ekahau Technology Document*, November 2002.

[2] Eissfeller, B., D. Gänsch, S. Müller and A. Teuber, "Indoor Positioning Using Wireless LAN Radio Signals" in *Proceedings of ION-GNSS 2004*, Long Beach, California, 21-24 September 2004, pp. 1936-1947A. N. Editor and A. Nother (Eds.), *An Edited Book*, Wiley, Chichester, U.K, 2003.

[3] Teuber, A., G.W. Hein, "Analysis and Reduction of Systematic Errors in Wireless LAN Positioning" in *Workshop Proceedings of WPNC '05*, Hannover, Germany, 17 March 2005, pp. 87-92.

[4] Zadeh, L. A., Fuzzy Sets, Information and Control, 1965.

[5] Zadeh; L. A., Fuzzy Logic, Neural Networks, and Soft Computing, Communications of the ACM, Issue 37, No. 3, pp. 77-84, 1994.

[6] Haberler, M., Einsatz von Fuzzy Methoden zur Detektion konsistenter Punktbewegungen, *Geowissenschaftliche Mitteilungen Nr. 73, Schriftenreihe der Studienrichtung Vermessung und Geoinformation, Technische Universität Wien*, Institut für Geodäsie und Geophysik, 2005.