INDOOR LOCATION USING WIRELESS NETWORKS BASED ON BAYESIAN REASONING

Jesús F. Rodríguez-Aragón, Vidal Moreno Rodilla, Belén Curto Diego
Fco. Javier Serrano Rodríguez, Raúl Alves Santos and M. José Polo Martín
Computer Science and Automatic Department, University of Salamanca, Pza. Los Caídos S/N, Salamanca, Spain

Keywords: Indoor location sensing, Bayesian networks, Positioning algorithm, Wifi location.

Abstract: This paper describes a solution for the indoor location in the context of wireless local networks. Firstly, the processes of sampling and training are done by off-line scene analysis. Secondly, the mobile entity can be localized in a self-positioning fashion according to the Bayesian Network based method.

1 INTRODUCTION

At present, indoor location systems have created expectations in certain environments. To mention some examples, we make reference to locating people in hospitals or nursing homes (both medical staff and patients), locating products in a warehouse or guidance systems in museums (Segvic et al., 2009). The overwhelming growth of the use of wireless technology in recent years has been due mainly to the appearance of low cost products on the market that let us create an indoor wireless network easily. Almost every public building has a wireless network that can be used as the basis of a tracking system.

Wireless location techniques can be classified according to the method used for signal measurements: Time of Arrival (TOA), Angle of Arrival (AOA) and Received Signal Strength (RSS) (Liu et al., 2007; Kanaan and Pahlavan, 2005).

The first of the three techniques has been the most investigated up to date. However, problems are still unsolved due to the difficulties of predicting the signal propagation indoors. This same happens with AOA, since AOA is based on TDOA (Time Difference of Arrival) which also depends on the signal propagation. Radio propagation suffers from multipath effect (Alavi and Pahlavan, 2003). The method described in the present paper is found among the ones that use RSS to calculate current position based on a prior scene analysis, since RSS method doesn't depends on multipath effect.

This work focuses on developing a self-positioning indoor location algorithm based on the ratio of *RSS loss/position* from different visible access points (AP) in a specific area. In the present case, the environment is a public building with a wireless network structure, so that issuers (Wifi access points) already exist, i.e. it's not necessary to build a new infrastructure, and the detector (simple or multiple devices with a wireless network card) is the one that moves on the environment.

Each position has a range of possible signal strength losses of near visible APs. Using statistical techniques (Ladd et al., 2004; Wang et al., 2003), we can infer a number of possible positions from a measurement of signal strength loss of visible APs. The more APs providing information, the greater reliability of the prediction will be achieved.

In this sense, we make a proposal for uncertainty treatment similar to (Ladd et al., 2004). In order to improve the accuracy of the algorithm, we perform several data processing prior to infer a possible result.

Each of the processes composing the indoor location system based on wireless networks is described in this paper. Firstly, the sampling method is detailed and then the training and location processes are defined. Finally, a detailed description of the particular implementation of the algorithm is shown, as well as the results obtained.

2 WLAN IEEE 802.11 STRUCTURE

An 802.11 WLAN is based on a cellular architecture (the system is subdivided into cells). Each cell, called

F. RodrÃguez-AragÃşn J., Moreno Rodilla V., Curto Diego B., Javier Serrano RodrÃguez F., Alves Santos R. and JosÃľ Polo MartÃn M.
INDOOR LOCATION USING WIRELESS NETWORKS BASED ON BAYESIAN REASONING.
DOI: 10.5220/0002808001070112
In *Proceedings of the 6th International Conference on Web Information Systems and Technology (WEBIST 2010)*, page
ISBN: 978-989-674-025-2

BSS (*Basic Service Set*), is controlled by a base station, called AP (*Access Point*). An 802.11 WLAN can be composed of a single AP, and therefore, for a single cell.

The whole interconnected WLAN, including the different cells with their respective APs, is considered as a 802 single network by the upper levels of the OSI model, and it is called ESS (*Extended Service Set*).

Beacon frames are one of the management frames in the 802.11 standard. An access point periodically sends a Beacon frame to broadcast its presence and network information to the client stations in its environment. Stations can get a list of available access points looking for Beacon frames continuously in all 802.11 channels. Beacon frames contain the necessary information to identify network characteristics and to be able to connect to the desired access point. Some of the fields contained are: AP MAC address, SSID (*Service Set IDentifier*), RSSI (*Received Signal Strenght Indicator*), supported rates, timestamp (used in synchronization), etc.

3 BAYESIAN NETWORKS

Classical inference models avoid considering any *a priori* information. However, in certain occasions the use of *a priori* knowledge can be a very useful contribution to the process of inference. Models including *a priori* information are based on the Bayes Theorem (Equation 1).

$$P(H|E,c) = \frac{P(H|c) \cdot P(E|H,c)}{P(E|c)} \quad (1)$$

where $P(H|E,c)$ is the probability that an hypothesis H is fulfilled after considering the effect of evidence E in the context c (*a posteriori* probability); $P(H|c)$ is the probability that an hypothesis H is fulfilled in an isolated context c (*a priori* probability); $P(E|H,c)$ is the probability of occurrence of evidence E assuming the hypothesis H as true in the context c (conditional probability); $P(E|c)$ indicates the probability of occurrence of evidence E in the context c.

In our case, the hypothesis H is being in a particular cell; the evidence E is the current measurement; and the context c is the path previously done until the current measurement is taken and it provides information on the following possible positions.

All these are conditional probabilities, i.e., they specify the reliability of an hypothesis based on the assumption that some other initial hypothesis be true. Therefore, theory makes no sense if there are no *a priori* values of previous probabilities.

Signal strength values received from different visible APs are the main data source in this work. A signal strength value is linked to the position where the measurement has been obtained. Therefore, *a posteriori* probability is the probability of receiving a particular signal strength value for an AP conditioned on being at a particular cell.

4 SCENE ANALYSIS

In this work, we consider three main processes throughout the workflow for the location method: sampling, training and location as shown in Figure 1. During the sampling and the training process, an *offline* scene analysis is performed, and during location process, the mobile receptor is *online* localized from a particular measurement.

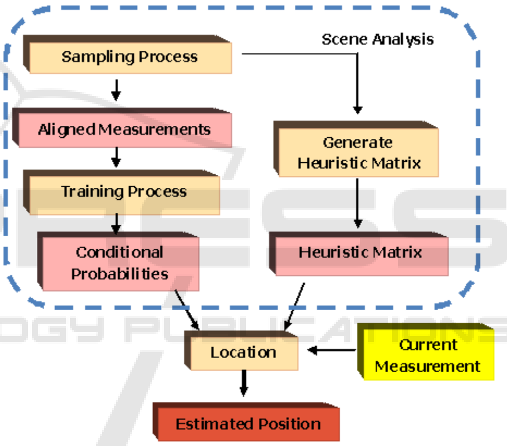

Figure 1: Location Workflow.

4.1 Signal Strength Map Creation

Initially, the sample area is represented by a grid map which is discretized into cells. Each position is identified by three values: (x, y, θ), where (x, y) are planar spatial coordinates of the cell and θ is the orientation with which the sample has been taken. Figure 2 shows graphically these three values. Signal strength of every visible AP is measured in each cell.

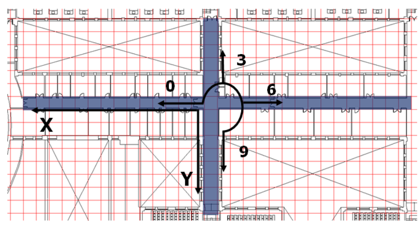

Figure 2: Sampling area.

Information about the different visible APs in the sampling area is collected and a study in order to remove those APs that provide unusefull information is performed.

Unlike (Ladd et al., 2004), sub-sampling is performed at lower resolution for each cell, i.e., the cell is divided into 9 possible positions in order to have a more complete probabilities distribution. Thus, measurements are taken at the center of the cell and they are also taken away from the center (in two configurations, $\times$ or $+$, as shown in Figure 3).

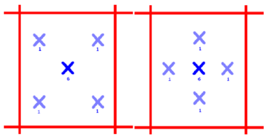

Figure 3: Possible sub-sampling configurations.

Therefore, the sampling process consists of sets of 10 measurements with $\times$ configuration and others with $+$ configuration. These sets of measurements are performed with the greatest possible randomness, taking place at different times in a day in order to compensate variations in the environment that may occur, as the difference of network users, greater number of obstacles in the measurements (due to more people walking through corridors, etc). Each of the sets are stored in a data structure for each position, in which the visible APs and measures (central and unfocused) are specified.

4.2 Generating an Heuristic Matrix

The analysis of the scene also includes the calculation of an heuristic in order to take into account the probability evolution depending on distance between two positions A and B. If position B is far away from position A, the probability of moving from one position to another will be less than if position B is adjacent to position A.

To define the evolution of probability depending on the distance between two positions, an exponential function is used, since the probability of moving from one position to another decreases exponentially depending on the distance traveled (Equation 2)

$$P = e^{-\lambda d} \tag{2}$$

where d is the distance between positions. For determining the distance between two positions, we consider the number of cells of separation plus the number of changes in direction. λ is a constant that reflects the importance of distance between positions in the heuristic. When $d > 6$, we consider the probability is zero. The heuristic is implemented by a square matrix of dimensions $D \times D$, where D is the number of locations in the sampling area.

4.3 Training Process

Once all the measurements have been completed, there is a set of values of signal strength for each of the visible access points at each cell. The probability of being in a cell is conditioned to the signal strength measurement of an AP. For each AP in each cell conditional probabilities are calculated using

$$p = \frac{L_x}{N} \tag{3}$$

where L_x is the number of samples of a same signal strength value, and N is the number of total samples.

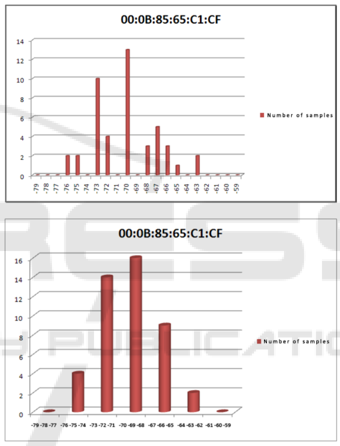

Figure 4: Grouping of probabilities.

To prevent the frequencies of certain signal strength values are zero around the maximum frequency zone, ranges for the evidences are defined. Each group contains n probabilities of signal strength values (Figure 4).

5 LOCATION

The schema of the location algorithm used is shown in Figure 5.

The location algorithm calculates an estimated position of the receiver agent from a vector R of signal strength measures that contains a measure for each of the visible APs taken in an unknown position.

$$R = \langle R_1, R_2, \ldots, R_n \rangle \tag{4}$$

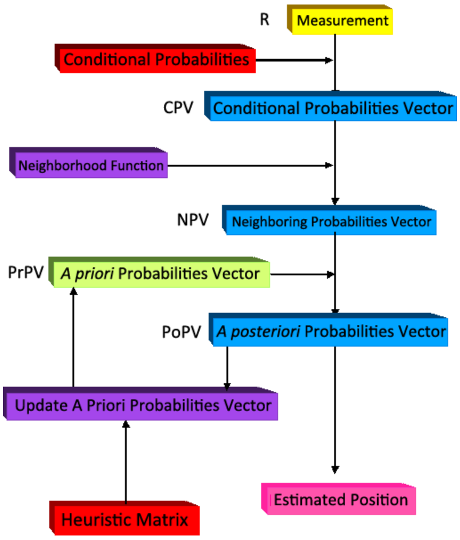

Figure 5: Schema of the Location Algorithm.

5.1 Calculating the Conditional Probabilities Vector

From the measures vector, we now proceed to calculate the conditional probabilities vector, CPV, which represents the location probability for each cell. This vector has as many components as cells considered. Each component of CPV is calculated by the following expression shown in Equation 5.

$$CPV_j = \prod_{i=1}^{N} P_i(AP_i, R_i) \tag{5}$$

where AP_i refers to the *ith* AP; R_i is the current received signal strength from AP_i; and $P_i(AP_i, R_i)$ is the probability of receiving the signal strength R_i from AP_i while being at the *jth* cell. This probability is found in the Bayesian network obtained during training process.

Once CPV is calculated, we can infer the current position based on the cell with the greatest probability within the vector. However, it is possible to process the data in order to obtain a more reliable result, incorporating the neighborhood function. What is sought is to select the candidate cell using a criterion different from only maximum probability. The new criterion takes into account that candidate cells will be found in those areas having a higher accumulated probability.

There were cases where the cell with the highest probability was most likely without context, i.e., there are more or less contiguous cells whose probability is high, but on the other hand, there is an isolated and remote in distance cell whose probability is the highest.

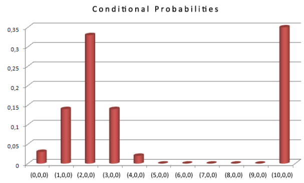

Figure 6: Neighborhood error.

est. Noting Figure 6, it suggests that the isolated cell is surely not the candidate cell.

5.2 Calculating the Neighboring Probabilities Vector

To avoid such errors, we recalculate the entire vector of conditional probabilities, influencing the probability of a cell with the probability of adjacent cells. Thus, the new vector, called neighboring probabilities vector, NPV, is calculated by Equation 6.

$$NPV_i = \alpha CPV_{i-1} + \beta CPV_i + \alpha CPV_{i+1} \tag{6}$$

where NPV_i is the *ith* component of the neighboring probabilities vector, i.e., the probability of the *ith* cell of being the current position; $CPV_{i-1}, CPV_i, CPV_{i+1}$ are the conditional probabilities of the immediately preceding, current and immediately following cells, respectively; and α and β are weighting constants. Figure 7 shows the neighboring probabilities vector for the example shown in Figure 6.

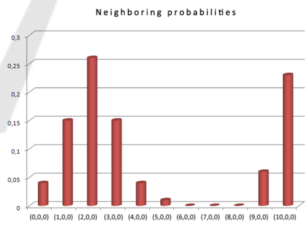

Figure 7: Neighboring Probabilities Vector.

Current position may again be inferred from this new vector. However, we apply an heuristic to obtain a more reliable result.

5.3 Calculating the *a Priori* Probabilities Vector

An *a priori* Probabilities Vector, PrPV, is generated with as many components as possible cells. Each component represents the probability of that cell being the candidate cell. In short, PrPV provides a prediction of the final result. This vector is updated each run. On the first run, all components have the same value, $\frac{1}{S}$, where S is the number of cells of the sample area. This means that the initial prediction is the same for all cells in the sample space. This is logical because there is no prior information.

5.4 Calculating the *a Posteriori* Probabilities Vector

The vector from which the final results are calculated is the *a posteriori* Probabilities Vector, PoPV, that is calculated from Neighboring Probabilities Vector, NPV, and *a priori* Probabilities Vector, PrPV, using the Equation 7.

$$PoPV_j = \frac{NPV_j \cdot PrPV_j}{\sum_{i=1}^{S} NPV_i \cdot PrPV_i} \tag{7}$$

where $PoPV_j$ is the *jth* component of the *a posteriori* probabilities vector; NPV_j is the *jth* component of the neighboring probabilities vector; $PrPV_j$ is the *jth* component of the *a priori* probabilities vector; S is the number of cells of the sample area; and $\sum_{i=1}^{S} NPV_i \cdot PrPV_i$ is used in order to normalize the result.

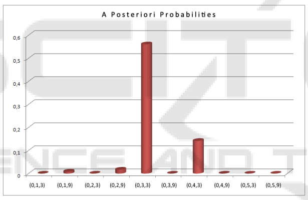

Figure 8: *A posteriori* Probabilities Vector.

Current position is now inferred based on the cell whose probability is the greatest (Figure 8).

5.5 Update the *a Priori* Probabilities Vector

Having obtained the estimated cell, the algorithm has to be prepared for the next execution. At this point, current position is known and it provides heuristic information about the next possible cell, assuming the continuity of the path. For that purpose, the heuristic matrix generated during the training process is used. This matrix provides information on the probability of moving from one cell to another.

Dismissing discontinuities in the path, cells that are found within a certain distance from current cell may be removed from the set of possible candidate cells. The update of the *a priori* probabilities vector is performed using Equation 8.

$$PrPV_i(t+1) = H \cdot PrPV_i(t) \tag{8}$$

where H is the heuristic matrix; $PrPV_i(t)$ is the *ith* component of the current *a priori* probabilities vector; and $PrPV_i(t+1)$ is the *ith* component of the *a priori* probabilities vector which is being calculated.

6 RESULTS

Using the wifi network infrastructure available in the Science Faculty of the University of Salamanca, the sample area was situated on the second floor and is compound by two perpendicular corridors and discretized in cells of $2.40m \times 2.40m$. Cells are identified with three values: (x, y, θ), as was shown in Figure 2. Sample space is compound by 43 cells with 2 possible orientations, and a central cell (located at the intersection of the two corridors) with 4 possible orientations.

Our experiments were conducted by a human operator carrying a HP iPAQ H3850 with a PCMCIA Compaq WL110 wireless Ethernet card. This particular card uses an Orinoco chipset. An Acer C500 PDA with Windows Mobile 5.0 was also used in our experiments. The access to the device was implemented in C/C++ after performing a study of the Orinoco Linux driver.

In our case, the number of visible APs used during the training and location process was 24. A prior study was done in order to remove APs that don't provide relevant information.

Having established the visible APs, we started with the sets of measurements. In this work, data used to train the Bayesian network consist of 10 sets of 10 measurements each. These measurements were performed at different times and on different days, so that the presence of extraneous factors were reflected in the data: in a college building, there are more people in the morning than in the afternoon, there are some rush hours between lectures, etc.

During the sampling process, measurements performed away from the center of the cell have been incorporated in order to achieve a more accurate distribution of the data. Proposed sub-sampling at a lower resolution was reached by dividing each cell in different positions as shown in Figure 3. Thus, 6 central measurements and 4 unfocused measurements were performed for each set. 5 sets, with 10 measurements each, were taken with $\times$ configuration and another 5, with 10 measurements each, with $+$ configuration.

The heuristic matrix used in the location algorithm is a matrix whose dimension is 90×90, since there are 90 possible positions in our sample space. Rows and columns are the different possible cells located in the same order. The function used to generate the heuristic matrix is defined by Equation 2. In our case, using $\lambda = 0.3$, the results are satisfactory.

During the bayesian network training process, signal strengths values are grouped, so that conditional probabilities are calculated in groups composed of 3 signal strengths values. Before calculating the location from a particular measurement is necessary to find the group of that measurement.

In the location process, a data post-treatment has been performed, as the usage of the neighborhood function and the actual usage of the heuristic matrix to update the *a priori* probabilities vector.

The neighborhood function used in our case is defined by Equation 6 taking as weighting factors $\alpha = \frac{1}{6}$ and $\beta = \frac{4}{6}$, so that neighboring probabilities vector is defined by Equation 9.

$$NPV_i = \frac{1}{6}CPV_{i-1} + \frac{4}{6}CPV_i + \frac{1}{6}CPV_{i+1} \qquad (9)$$

In order to check the accuracy of the proposed algorithm, we take a testbed of 100 different paths. The results improve considerably once traversed the first positions (5 - 6) of each path, when heuristic is being taken into account. Testbed results provide an algorithm accuracy close to 60%.

Figure 9 shows the implemented application during the location process of a person carrying a PDA and walking through the corridors.

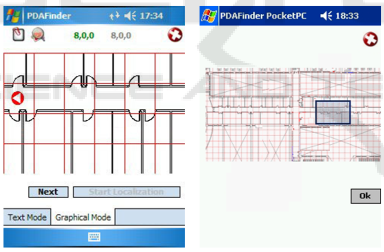

Figure 9: Application during the location process.

7 CONCLUSIONS

Due to the included optimizations and, after analyzing the results of the algorithm, that can be considered efficient for location with cells of $2.40m \times 2.40m$.

The sampling process is performed trying to include a more accurate data distribution, including different cases to add more information into the sampling. The location algorithm uses statistical techniques and, therefore, during sampling process, fullest possible information must be collected. Compared with the method proposed in (Ladd et al., 2004), data processing has been performed in order to achieve a better effectiveness for the position inference. Heuristic information is used, assuming that movements are continuous and great distance movements between two measurements are not real. Current position provides information about the next future position.

Working lines of research are open in the optimization of the sampling process collecting more information about the nature of the measures: independence of the environment, periodical behavior, etc. so that the result is based not only in terms of *signal strength loss/position*, but adding new information, date, time, temperature, etc. Improvements can also be achieved in the location algorithm itself, with heuristic optimizations, etc.

ACKNOWLEDGEMENTS

The work has been carried out within projects financed by the *Junta de Castilla y León* $SA030A-07$ and the *Spanish Ministry of Science and Innovation* $DPI2007-62267$. The main author has also worked under the support of a Junta de Castilla y León fellowship.

REFERENCES

Alavi, B. and Pahlavan, K. (2003). Modeling of the distance error for indoor geolocation. In *Proc. IEEE Wireless Communications and Networking Conference (WCNC), vol. 1, pp. 668 - 672*.

Kanaan, M. and Pahlavan, K. (2005). A comparison of wireless geolocation algorithms in the indoor environment. In *Proc. IEEE Wireless Communications and Network Conference (WCNC), vol. 1, pp. 177-182*.

Ladd, A., Bekris, K., Rudys, A., Wallach, D., and Kavraki, L. (2004). On the feasibility of using wireless ethernet for indoor localization. In *IEEE Transactions on Robotics and Automation, vol. 20-3, pp. 555 - 559*.

Liu, H., Darabi, H., and Liu, J. (2007). Survey of wireless indoor positioning techniques and systems. In *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 37, No. 6, pp. 1067 - 1078*.

Segvic, S., Remazeilles, A., Diosi, A., and Chaumette, F. (2009). A mapping and localization framework for scalable appearance-based navigation. In *Computer Vision and Image Understanding, vol. 11, Issue 2, pp. 172 - 187*.

Wang, Y., Jia, X., and Lee, H. (2003). An indoor wireless positioning system based on wireless local area network infrastructure. In *6th. International Symposium on Satellite Navigation Technology Including Mobile Positioning & Location Services, paper 54, pp. 22 - 25*.