A STATISTICAL MODELLING BASED LOCATION DETERMINATION METHOD USING FUSION TECHNIQUE IN WLAN

Reetu Singh, Luca Macchi and Carlo. S. Regazzoni[*]

D.I.B.E
University of Genova
Opera Pia 11a, Genova, Italy

Kostas.N. Plataniotis

Edward S. Rogers department
University of Toronto
Ontario, Canada, MSS3G4

ABSTRACT

Location information is of paramount importance in context aware Ambient Intelligence (AmI), Smart Space, traffic monitoring, surveillance network and cooperative communications services. This paper describes a Positioning determination solution based on wireless local area network (WLAN) signals. Position determination is based on the statistical modeling of the received signal at any position. This paper presents a probabilistic based statistical modeling approach for location estimation which incorporates fusion strategy in final step to combine efficiently the location individually reported by each WLAN transmitter. The system builds a radio map of the environment. The presented system is easier to implement and provide sufficiently good performance under all conditions. The accuracy with the 90% probability is reported to be 1.85 meters where as average error is reported to be 2.1 meters.

1. INTRODUCTION

Advancement in mobile world is catapulted by incorporating the mobile users position information. The position of a hand held device is an important information needed to enhance the communication efficiency. Basically, the positioning has been indigenously put apart into two parts, Indoors and Outdoors [1]. Since there are many ways to categorize the positioning systems, in this paper we will just refer to indoors positioning systems based on WLAN. This paper presents a probabilistic based statistical modeling approach for location estimation which incorporates fusion strategy in final step to combine efficiently the location individually reported by each WLAN transmitter. The system builds a radio map of the environment. The concept of radio map is based on collection of signal strength over a set of strategically selected coordinates in a known referenced environment. The set of signal strengths collected over known positions is treated as its signature. The issues related to duration of time in which the signal strength should be collected is still needed to be cleared. However on average, in the current state of art, typical time spent on collecting survey data set is 5 minutes with a sampling period of 1 second . Most of the indoor positioning systems are based on the radio map [1-12].

The presented system is easier to implement and provide sufficiently good performance. The system is efficient and works well. The accuracy with the 90% probability is reported to be 1.85 meters where as average error is reported to be 2.1 meters. The paper is divided into the following parts: section II explains the formalization of the problem and the maximum likelihood approach utilized. Section III talks about the experimental setup used for algorithm's performance verification. The experimental results are shown in section IV followed by the conclusion in the end.

2. STATISTICAL METHOD FOR LOCATION ESTIMATION

Statistical methods are generally based on knowledge of statistical distribution of the propagation model. Using statistical parameters of the propagation model, location can be estimated by utilizing a probabilistic approach. A model is said to be statistic if it gives a probability distribution of a signal in certain conditions. The statistical modeling of the environment should be the good choice where deterministic methods fails to imitate the environment. However, propagation parameters used for modeling should be derived from the environment itself. The best way to know these parameter is doing actual measurements. The parameters are computed by incorporating real measurements and the technique applied herein is maximum likelihood approach. Having estimated the propagation parameters, the location estimation simply reduces to an inference problem. The bayesian approach has been considered in the presented problem. It utilizes a priori knowledge of signal distribution at known locations. The location with the highest likelihood is chosen to be the estimate of user's position from individual WLAN

---

[*] This work was performed under co-financing of the MIUR within the project FIRB-VICOM.

transmitter. Finally, the exact location is approximated by using the weighting fusion approach on the set of estimated location. The weights are assigned on the basis of the ratio of the highest likelihood value occurring for each transmitter to the sum of all highest likelihood values. The detailed description is given in the following sections.

2.1. Formalization of the problem

The statistical propagation model describes a relationship between two physical quantities, location of receiver and access points [13][14][15][16]. Inside this model is present an error term defined by a probability distribution. In the application treated in this paper the physical quantities are the Received Signal Strength (RSS) and the distance between user and each access point. The proposed propagation model considering the i^{th} access point is shown in equation (1):

$$P_r{}^i(l) = P_0{}^i - \beta_1(l)^i \log(d^i(l, l_i)) - \beta_2(l)^i + \xi^i \qquad (1)$$

Where $P_r{}^i$ is the received power from the i^{th} access point at any location l. $P_0{}^i$ is the power received at d=1 meters, $\beta_1{}^i(l)$ an environmental parameter and $d^i(l, l_i)$ is the distance between user at location l and l_i,i^{th} access point's location. $\beta_2{}^i(l)$ is a parameter to take into account for Non Line of Sight effect and $\xi^i(0, \sigma^i(l))$ a gaussian variables with zero mean and variance equal to $(\sigma^i(l))^2$. It accounts for multi-path effect due to shadowing.

The equation 1 also known as log-power/distance propagation model is termed as a statistical model since it contains an assumed statistically distributed error term $(\sigma^i(l))^2$ [16]. This model holds good for any terminal receiving the signal from any access point. The equation consists of different parameters and statistical features of variables ξ. It is known [13] that if ξ is gaussian distributed, the probability distribution of $P_r(l)$ is also considered to be gaussian with $(\sigma^i(l))^2$ variance [16], expressed as:

$$f(P_r^i(l)|\beta_1{}^i, \beta_2{}^i, d^i(l, l_i), P_0{}^i) = \frac{1}{\sqrt{2\pi}\sigma^i(l)} e^{-\frac{1}{2}(Y)^2} (2)$$

$Y = \frac{P_r^i(l) - E(P_r^i(l)|\beta_1(l)^i, \beta_2(l)^i, d^i(l,l_i), P_0{}^i)}{\sigma^i(l)}$

where mean is given by

$$\begin{aligned} &E(P_r^i(l)|\beta_1(l)^i, \beta_2^i(l), d^i(l, l_i), P_0{}^i) \\ &= P_0{}^i - \beta_1(l)^i \log(d^i(l, l_i)) - \beta_2(l)^i. \end{aligned} \qquad (3)$$

In the next section maximum likelihood method is used to estimate the unknown parameters from collected measurements.

2.2. Maximum likelihood estimation using empirical analysis

In a problem like the one described in the previous paragraph the main difficulty is to define the propagation parameters because they are not obtainable from theoretical assumptions [15][16]. They depend upon the propagation environment and the best way to perform their estimation is to realize a wide number of measurements. Lets consider that an identically independent distributed ("*i.i.d*") sample of *n* random vectors of $P_r{}^{i,j}/l$ is collected on given reference coordinate positions. For a given random sample, the idea of statistical inference is to analyze the properties of the variable received power. The measurements consist in the choice of n "*good*" calibration points where "*good*" means that these points have to cover adequately all the environment involved in propagation model. Propagation parameters can be estimated from empirical data.

Starting from knowledge collected at the calibration points, it is possible to use the following procedure to perform the parameters estimation:

- distances $d^i(l, l_i)_{i=1,2,..,m}$ between each access point's location (l_i) and any user's location *l*.
- Received power $P_r{}^{i,j}(l)_{i=1,2,..,m, j=1,2,...,n}$ (user is at location *l*) where *i* is number of access points and *j* is number of survey data set collected at each location. The maximum number of access points is *m* and total number of measurements taken at each location is *n*.
- The power $P_o{}^i$ received at *d=1* meters from access point i^{th}.

In the previous paragraph we have already modelled probability distribution of the received power. If the parameters have been estimated, the probability function provides a measure of how likely presented signal strength sample outcomes are at a given position *l*, but the likelihood will depend upon the parameters utilized to model it though. However, if we know the real measurements and probability distribution model, i.e radio map, the likelihood of propagation parameters can be estimated by using equation (4). In the present case maximum likelihood method has been used. A typical method to perform a maximum likelihood estimation for the proposed model is to depict the likelihood function. Now assuming that observed samples are *i.i.d* from a distribution with probability density function (p.d.f) given in equation (2), it is possible to write the likelihood function $\Phi_i(\overline{x}(l))$ for i^{th} access point at any location *l*, as shown in equation (4). Mathematically speaking, likelihood is nothing but product of the conditional probability density function of each observation. The detailed information can be found in [17].

$$\Phi_i(\overline{x}(l)) = \prod_{j=1}^{n} f(P_r^{i,j}|l) =$$
$$= \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}\sigma^i} e^{-\frac{1}{2}(\frac{{P_r}^{i,j}(l) - E(P_r^i(l)|\beta_1(l)^i, \beta_2^i(l), d^i(l), {P_0}^i)}{\sigma^i(l)})^2} \tag{4}$$

Where $\overline{x}(l)$ is the vector of propagation parameters to estimate,

$$\overline{x}(l) = (\beta_1(l), \beta_2(l))$$

It is known that the likelihood function for the i^{th} access point $\Phi_i(\overline{x}(l))$ is not depending on $\sigma^i(l)$ and maximum likelihood estimates of the parameter are those values that maximize the likelihood, see equation (5):

$$\Phi_i(\overline{x}(l)) = (\frac{1}{\sqrt{2\pi}\sigma^i(l)})^n e^{-\frac{SSE^{i,j}(l)}{2(\sigma^i(l))^2}} \tag{5}$$

$$SSE^i(l) = \sum_{j=1}^{n}(P_r^{i,j}(l) - E(P_r^i(l)|\beta_1(l)^i, \beta_2^i(l), d^i(l), {P_0}^i)^2 \tag{6}$$

Maximizing the likelihood also means minimizing eq.(6). The best estimation is reached by minimizing SSE^i [16][17]. To obtain maximum likelihood solution, equation 6 is expanded by substituting eq.(3) and can be expressed in matrix form after satisfying minimizing condition. All the unknown parameters are transported on left side of the equality sign and known parameters on right side. The maximum likelihood solutions are given by the least square estimator $\widehat{\overline{x}}(l)$.

$$\widehat{\overline{x}}(l) = (Y^T Y)^{-1} Y^T S \tag{7}$$

where Y and S are defined as:

$$S = \begin{bmatrix} {P^1}_r - P_o \\ \cdot \\ \cdot \\ {P^n}_r - P_o \end{bmatrix}, Y = \begin{bmatrix} logd^i & 1 \\ \cdot & \\ \cdot & \\ logd^i & 1 \end{bmatrix} \tag{8}$$

Above derivation can be found in [17]. By substituting $\widehat{\beta}(l)$ parameters in equation 6, the MLE's of $\sigma^i(l)$ can by estimated as

$$\widehat{\sigma^i}(l) = \sqrt{\frac{SSE^i(l)}{n}} \tag{9}$$

2.3. Location estimation

Bayes theorem, starting from an a priori knowledge, achieves to find out a posterior probability density function. In this paper, a priori knowledge is the estimation of propagation parameters performed on known locations, as explained in previous paragraphs. Application of Bayes theorem in WLAN location estimation problem with individual access points can be useful to write p.d.f. of received power on a given location *l*, as shown in equation (10)[17]. On observing the current signal strength value, likelihood of that data at each stored location can be computed from $f(\overline{P_r^i}|l, \widehat{\overline{x}})$.

$$p(l|\overline{P_r}^i) = \frac{f(\overline{P_r^i}|l, \widehat{\overline{x}})\pi(l)}{\int f(\overline{P_r^i}|L, \widehat{\overline{x}})\pi(L)dL} \tag{10}$$

where $\overline{P_r^i}$ is the vector of received power from each access point and L is the vector of stored location, where $l = 1, .., L$. $\pi(l)$ is the prior probability of location which is fixed to maximum, i.e one in present case as the probability of finding all positions are assumed to be equal. The denominator sums up to one due to equal value of $\pi(L)$ at every location *l*. Knowledge of probability of observation for each possible location *l* provides information about dependency of the observed signal properties on the location variable. Henceforth Bayes theorem used as an inference tool provides a probability of location variable given the observation. When a signal is received, probability density for all stored location is calculated. The location with the highest likelihood is chosen to be the estimate of user's position, l_{est} from individual WLAN transmitter.

$$l_{est} = argmax(p(l_i|\overline{P_r}^i)) \tag{11}$$

The location estimates are combined by using weighting fusion approach. The weights *w* are assigned on the basis of the ratio of the highest likelihood value occurring for each transmitter to the sum of all highest likelihood values i.e location with high probability value will have larger weights. In the end, it is possible to write the final estimate of the location $l_{est_{Fin}}$ in the following way:

$$l_{est_{Fin}} = l_{est_1} w_1 + l_{est_2} w_2 + ... + l_{est_i} w_i \tag{12}$$

Where *i* is the number of access points and w_i is the weight associated to l_{est_i}.

w_i is defined by the equation (13):

$$w_i = \frac{argmax(p(l_{est_i}|\overline{P_r}^i))}{\sum_{i=1,..,m} argmax(p(l_{est_i}|\overline{P_r}^i))} \tag{13}$$

3. EXPERIMENTAL TEST BED

A typical room layout at the university of Genova is elected as the test bed for the proposed systems. The size of the building is 40×30 meters, three access points, one of type Cisco 350 , and two Cisco 1100 are installed which operate in 2.4 Ghz frequency and cover 40 meters at 11Mbps data rate, as shown in figure 1. The characteristic detail can be found herein [18]. The access points are based on 802.11b IEEE standard WLAN. They have omnidirectional antenna with gain 2.2dbi. The transmitter work in 2.4 GHz band. The experiments were performed in North-West part of the building involving laboratory and a corridor. To receive wireless signals integrated NIC card on IBM pentium IV laptop has been used. The test site is grided into the size of one square meters. In figure 1, black dots are the positions where set of measurements coming from three access points are recorded. The collected signal strength is reported in units of negative decibel (-dbm) according to standard industrial practice. At each position approximately 300 samples are collected by keeping the sampling rate of one second and receiver was kept oriented towards west direction throughout the experiments.

4. EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

The algorithms is implemented in Matlab 6.5 environment. The developed method was tested on the experimental data. Test data was collected at various location, throughout the test site, in different day and time. First at the same position measurements was collected and position algorithm was used to computes user's location. The computed position was compared with the real position by mean of Euclidean distance that gives the shortest distance between real and computed position. The result is shown in figure 2. The average error comes out to be 2.1 meters. More elaborate experiment was done to find out how positioning error varies in space. For this purpose set of location coordinates distributed in space was chosen and measurements was collected on every position. The error obtained for these location is shown in figure 3. The error shown is distributed in space and time. The rms obtained on same data set can be seen in figure 4 which shows maximum error obtained at time between 11-12 am and minimum rms is obtained between 3-4 pm. The rms varies from 2.3 to 2.8 meters.

5. CONCLUSION

The paper presents statistical modelling based location determination method that applies the fusion techniques to obtained position. The system builds a radio map of the environment. The bayesian approach has been considered in the presented problem. It utilizes a priori knowledge of signal distribution at known locations. The estimate of the location estimate are combined by using weighage fusion approach.

6. REFERENCES

[1] J. Hightower and G. Borriello - Location Systems for Ubiquitous Computing - IEEE Computer, pp. 57-66, IRS-TR-01-001, August 1 2001.

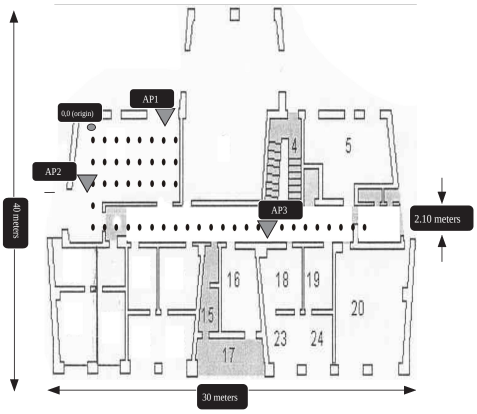

Fig. 1. Experiment Test site

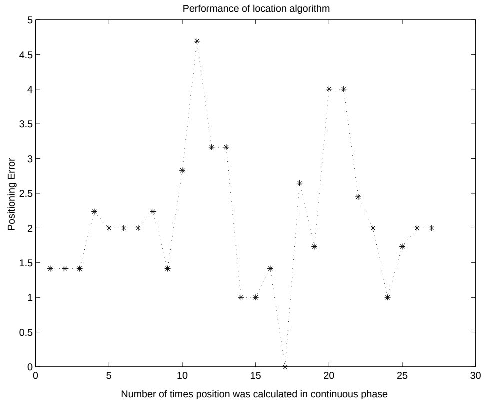

Fig. 2. Positioning error obtained for a particular position continuously in time

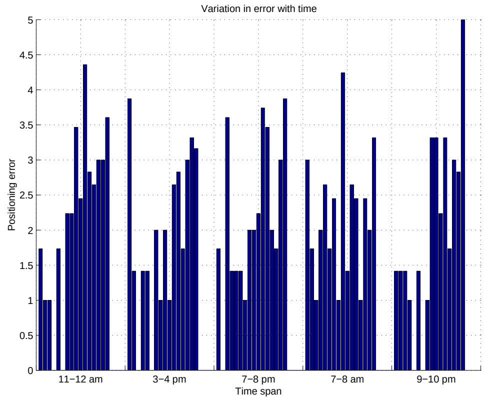

Fig. 3. Positioning error distributed in time and space

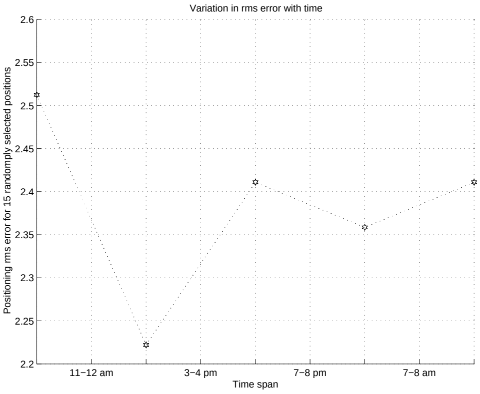

Fig. 4. RMS error distributed in time

[2] K. Kaemarungsi, P. Krishnamurthy - Modeling of Indoor Positioning Systems Based on Location Fingerprinting - IEEE Infocom, Hong Kong, March 2004.

[3] Y. Gwon, R. Jain, T. Kawahara - Robust Indoor Location Estimation of Stationary and Mobile Users - IEEE Infocom, Hong Kong, March 2004.

[4] M. Youssef A. Agrawala - On the Optimality of WLAN Location Determination Systems - Communication Networks and Distributed Systems Modeling and Simulation Conference, San Diego, California, January 18-24 2004.

[5] R. Battiti, M.Brunato, A. Villani - Statistical learning theory for location fingerprintings in wireless LANs - Technical report DIT-02-0086, 2002.

[6] T. Roos, P. Myllymaki, H. Tirri - A statistical modelling approach to location estimation - IEEE Transaction on mobile computing, vol.1, no.1, pp.59-69, January 2002.

[7] P. Castro, P. Chiu, T. Kremenec, R. Muntz - A probabilistic location service for wireless network environments - In proceedings of Ubicomp 2001, pp.18-24, Springer Verlag, September 2001.

[8] M. Youssef, A. Agrawala, A. U. Shankar - WLAN location determination via clustering and probability distribution - IEEE PerCom 2003, March 2003.

[9] P. Millymaki, T. Roos, H. Tirri, P. Misikangas, J. Sievanen - A probabilistic approach to WLAN user location estimation - The third IEEE workshop on wireless LANs, 2001.

[10] M. Youssef, A. Agrawala - Small-scale compensation for WLAN location determination systems - IEEE WCNC 2003, March 2003.

[11] P. Castro, P. Chiu, T. Kremenek, R. Muntz - A probabilistic location service for wireless network environments - Ubiquitous Computing 2001, Atlanta, September 2001.

[12] P. Bahl and Venkata N. Padmanabhan - Radar: An In-building RF-based User Location and Tracking System - IEEE Infocom 2000, volume 2, pages 775-784, March 2000.

[13] G.L. Stuber - Principle of Mobile Communication - Kluwer Academic Publishers, The Netherlands.

[14] B. Sklar - Rayleigh Fading Channels in Mobile Digital Communication Systems Part1: Characterization - IEEE Communication Magazine, pp. 90-100, July 1997, .

[15] K. Pahlavan and P. Krishnamurthy - Principles of Wireless Networks: A Unified Approach - Prentice Hall PTR, 2002.

[16] T. S. Rappaport - Wireless Communications: Principles and Practice - Prentice-Hall Inc., New Jersey, 1996.

[17] M.H. DeGroot - Probabilistic and Statistics - second ed. Reading,Mass:Addison-Wesley,1986.

[18] http://www.cisco.com/en/US/products/hw/wireless