Journal of Engineering and Applied Sciences 3 (4): 307-311, 2008
ISSN: 1816-949X

A Regression-Model-Based Approach to Indoor Location Estimation

[1]M.C. Su, [1]C.Y. Li, [1]D.Y. Huang, [1]S.C. Lin, [1]G.D. Chen, [2]C.C. Hsieh and [3]P.C. Wang
[1]Department of Computer Science and Information Engineering,
National Central University, Taiwan, R.O.C.
[2]Department of Computer Science and Engineering,
Tatung University, Taiwan, R.O.C.
[3]Quality Management Center, Cathay General Hospital, Taiwan

Abstract: Recently, context-aware or location-aware computing has become an interesting research field and has many practical applications in commerce, tourism, public safety, entertainment, military environments, hospital management, etc. Many different approaches have been proposed to tackle the problem of determining the location of a user or a mobile device. In an outdoor environment, the Global Positioning System (GPS) is the most popular solution. However, due to the poor indoor coverage, the GPS cannot provide a satisfactory solution to the problem of indoor location estimation. Many different approaches have been proposed to tackle the indoor location estimation problem. In this study, by use of the Received Signal Strength Indication (RSSI) measurements, a simple approaches to indoor location estimation are introduced to provide a simple but effective solution to the indoor localization problem based on existing wireless LAN infrastructures. The approach is based on regression models. The performance of the proposed approaches is demonstrated by testing 2 data sets acquired from real-world environments.

Key words: Regression, model, indoor location, estimation

INTRODUCTION

Recently, the growth of interest in location-based applications provides a strong motivation to develop location estimation techniques (Brunato and Battiti, 2004; Gwon *et al*., 2004; Hashemi, 1993; Li *et al*., 2000; Mauve *et al*., 2001; Niculescu, 2004; Orr and Abowd, 2000; Pahlavan *et al*., 2000; Patwari *et al*., 2005). Accurate knowledge about the user's position has many applications in civil, public safety, commercial, military, entertainment applications, etc. For example, an emergency caller's location can be identified and then an emergency assistance can be provided on time. The location-based service can help the staff in hospitals or nursing rooms track the sick or the elderly who are away from visual supervision (Orr and Abowd, 2000). In a museum, location estimation techniques can help a tourist effectively through the museum.

While, the Global Position System (GPS) is the most reliable and widespread positioning system for outdoor location-based services, it is not a good choice for an indoor environment due to its poor indoor coverage. Therefore, many approaches to indoor location estimation rely on dedicated sensor networks and/or existing wireless Local Area Network (LAN) infrastructures. The advantage of the methods using dedicated sensor networks is that physical specification and quality of the location sensing results is under control of the designer (Orr and Abowd, 2000). However, the methods based on existing wireless infrastructures are more attractive and cost-effective than the ones using dedicated sensor networks because the former ones can be easily integrated into existing wireless environments without significant changes in both mobile platforms and network infrastructures (Li *et al*., 2004). Basically, the intensities of radio signals emitted from wireless networks can be used to detect the position of a device because of a functional dependence between the signal strength from an access point and the physical position of the device. However, the propagation patterns of radio signals are extremely complex and difficult to be mathematically modeled (Brunato and Battiti, 2004).

Some popular signal metrics related to the estimation of distance are the Received Signal Strength (RSS), carrier signal phase of arrival (POA) and time of arrival (TOA) of the received signal. While, we need major hardware

Corresponding Author: M.C. Su, Department of Computer Science and Information Engineering, National Central University, Taiwan, R.O.C.

modifications and dedicated drivers to acquire the measurements of POA and TOA, it will be relatively inexpensive and simple for us to get the Received Signal Strength Indication (RSSI) value via common wireless adapters. Due to the factors of cost, dedicated driver and maintenance complexity, this study focuses on the development of 2 simple indoor location estimation methods based on the use of the RSSI values. The first approach is based on radial-basis-function (RBF) networks and the second one is based on regression models.

The proposed indoor location estimation method: The goal of an indoor location algorithm is to determine the location of a mobile device from some signal matrices measured from a set of access points. Some partial reviews on localization algorithms can be found in (Niculescu, 2004; Savvides *et al*., 2004; Sun *et al*., 2003). Each proposed algorithm has its own considerations, advantages and limitations. In this study, a simple but effective methods are proposed to provide a solution to the indoor localization problem based on raw RSSI measurements.

The RSS value is indicated as the voltage measured by a receiver's received strength indicator (RSSI) circuit. RSS is often equivalently reported as measured power. In free space, signal power decays proportional to d^{-2}, where d is the distance between the transmitter and the receiver. In real-world channels, the ensemble mean received power decays proportional to d^{-np}, where the path-loss exponent, n_p, is typically between 2 and four due to the 2 major sources of environment dependence in the measured RSS, multi-path signals and shadowing (Patwari *et al*., 2005). It is a very demanding but extremely difficult challenge to derive a functional relationship between the position of a mobile device and raw RSSI measurements because of the complexity of indoor radio propagation, severe multi-path problem and variety of obstructions (e.g., furniture, walls, building, etc). Based on a wide variety of measurement results and analytical evidence, (Hashemi, 1993; Patwari *et al*., 2005; Roos *et al*., 2002), the difference between a measured received power and its ensemble average is modeled as log-normal (i.e., Gaussian if expresses in decibels). A critical problem associated with this statistical-based approach is that the parameters are somehow related to the environment and there are no universally good values for them (Roos *et al*., 2002); therefore, a large number of observations are required to have a precise estimate for the corresponding parameters.

In this study, we propose an alternative but simple approach to building a functional relationship between the position of a mobile device and raw RSSI measurements. The proposed regression-model-based method involves the following steps.

Step 1: Collect a set of observations. Each observation consists of a pair $(\underline{s}_i, \underline{p}_i)$, i = 1,..., N, where N is the number of observations,

$$\underline{s}_i = (s_{i1}, \cdots, s_{in})^T \in \Re^n$$

and

$$\underline{p}_i = (p_{i1}, p_{i2})^T \in \Re^2.$$

The parameter n is the number of access points (APs) placed in an indoor environment. The vector, $\underline{s}_i$, is an n-dimensional vector containing the n raw RSSI signal strength values and the vector, $\underline{p}_i$, is a 2-dimensional vector containing the physical coordinates of the mobile device. For the jth AP, the N pairs, $(\underline{s}_i, \underline{p}_i)$, can be transformed into N pairs, (s_{ji}, d_{ji}), where s_{ji} is the signal strength value emitted from the jth access point and d_{ji} is the distance from the coordinates of the mobile device, $\underline{p}_i$, to the jth AP.

Step 2: For each AP, build a second-order regression model with three regression coefficients,

$$R_j^d(s_j) = c_{j0}^d + c_{j1}^d s_j + c_{j2}^d s_j^2,$$

from the N pairs, (s_{ji}, d_{ji}), i = 1, ..., N, where s_j represents the signal strength value emitted from the jth access point. In addition to the regression model, D^d_j, we build a 2nd-order regression model,

$$R_j^{sd}(s_j) = c_{j0}^{sd} + c_{j1}^{sd} s_j + c_{j2}^{sd} s_j^2,$$

for the standard deviation curve computed from the N pairs, (s_{ji}, d_{ji}). The standard deviation curve is computed as follows. We dichotomize the strength interval into several sub-intervals and then compute the standard deviation of the samples in each sub-interval. An example is shown in Fig. 1, where the strength interval is divided into 6 sub-intervals in this case. The curve in green color corresponds to the regression model R^d_j and the curve in pink color corresponds to the regression model R^{sd}_j. Therefore, we need to store 6 parameters for each AP.

Step 3: Discretize the environment into squares of fixed size, say L × M squares. Compute the distance between the center point of each square and each AP. Let R^{cp}_{ij} denote the distance between the center point of the ith square and the jth AP. The vector

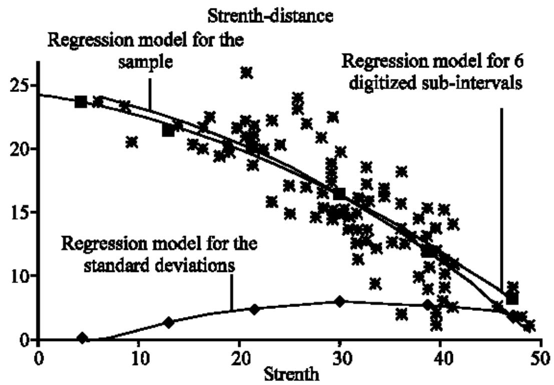

Fig. 1: The regression models built for each AP from the N pairs, (s_{ji}, d_{ji})

$$\underline{s} = (s_1, \ldots, s_n)^T$$

represents the present measured signal strength values. According to the regression model, the mobile device is probably with the distance,

$$R_j^d(s_j) \pm R_j^{sd}(s_j),$$

far away from the jth AP. We choose the winning square i* that minimizes the following distance measure to provide an estimate about which square the device is located at:

$$i^* = \operatorname*{Arg\,min}_{i=1,\ldots L\times M} \sum_{j=1}^{n} \frac{(d_{ij}^{cp} - R_j^d(s_j))^2}{(R_j^{sd}(s_j))^2} \tag{3}$$

The main idea of Eq. (3) is as follows. Basically, the smaller the Euclidean distance between the vectors

$$(d_{i1}^{cp}, \ldots, d_{in}^{cp})^T$$

and

$$(R_1^d(s_1), \ldots, R_n^d(s_n))^T$$

is, the more probable the device is at the ith square. However, the smaller the standard deviation, $R^{sd}_j(s_j)$, the more important the distance difference,

$$(d_{ij}^{cp} - R_j^d(s_j))$$

Therefore, we use the inverse of the standard deviation as the weighting factor of the corresponding distance difference. Basically, the larger the number of squares the more precise the location estimate. However, simulations showed that the precision saturated when the number of squares reached a certain value. In addition, larger number of squares requires higher computational cost. To make a tradeoff between the degree of precision and the computational cost we adopt a 2-pass winning selection scheme. In the first pass, we select a candidate square from a smaller number of squares by Eq. (1). In the following, a truly winning square is selected from a small region surrounded the candidate square in the second selection pass.

RESULTS AND DISCUSSION

Two data sets were used to test the proposed methods. The first data set, Location Fingerprinting Measurements, was downloaded from the Web site: http://ardent.unitn.it/software/data/. The data set consists of 257 measurements throughout a target environment with a size of roughly 30×25 m as shown in Fig. 2a. In this environment, a wireless LAN using IEEE802.11b standard is composed of 6 AVAYA WP-IIE APs. This data set was chosen because the authors in (Brunato and Battiti, 2004) had used four different machine learning methods, such as support vector machine, weighted k Nearest Neighbors (kNN) method, Bayesian approach and multi-layer perceptron, to test the data set and the data set was freely available on the Internet for comparisons.

The second data set was collected at a corridor shown in Fig. 2(b) which was about 19.2 × 30.9 m. With 1.8 m separation between adjacent points, total 66 measurement locations along the corridor were measured. At each measurement location, 50 complete measurements were taken by a person with an IPAQ HP6300 palmtop computer in hand and oriented towards to different directions. The average of the 50 measurements was computed and used as the representative measurement of that measurement location. There were 5 APs (2 D-Link DWL-G700AP, 1 D-Link DWL-7100AP, 1 Buffalo WHR-G54S and 1 ASUS Spacelink AP) were located at the corridor.

Each data set was split into a training data set and a testing data set as shown in Table 1. For comparisons, these data sets were tested against the k Nearest Neighbors (kNN) method with k = 3, the RBF networks and the regression-model-based method with the 2-pass selection scheme. The comparisons were conducted based on the distance error in meters, the number of parameters required for each method and the processing time in mini-seconds. The performance achieved by the three methods was tabulated in Table 2. While the computational cost of the kNN method was proportional to the number of training data, the computational cost of

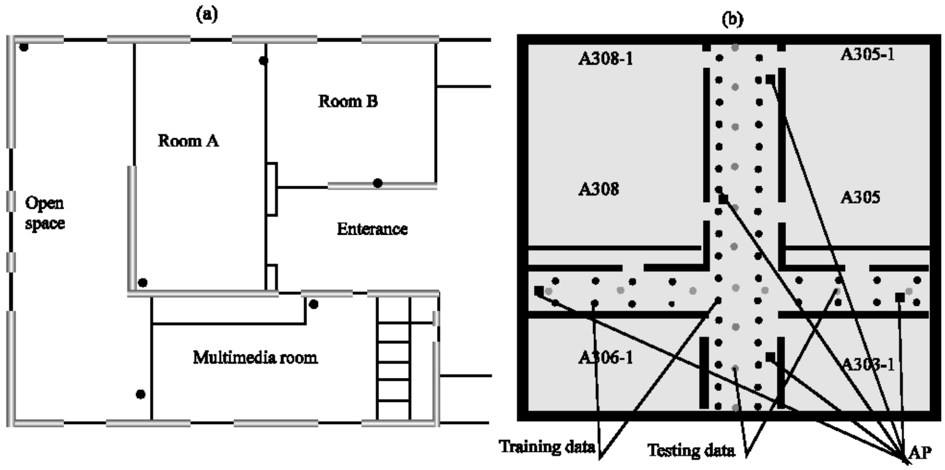

Fig. 2: The testing environments. (a) The environment at the University of Trento (Brunato and Battiti, 2004). (b) The corridor at National central University

Table 1: The data sets used for testing the indoor location estimation methods

Data set	# of training data	# of testing data	# of APs	Environment size
1st	100	157	6	30 ×25 m
2nd	50	16	5	19.2 ×30.9 m

Table 2: The performance achieved by the three methods

	Data set	kNN	RBF network	The regression model
1	Distance Err. (m)	3.621	4.254	4.249
	# of parameters	800	141 (15 hidden nodes)	36 (6 Aps)
	Processing time (ms)	37/157	200/157	16/157
2	Distance Err. (m)	3.56	3.74	4.13
	# of parameters	350	74 (9 hidden nodes)	30 (5 Aps)
	Processing time (ms)	1.0/16	31/16	1.0/16

the proposed method and the RBF-based method were not dependent on the number of training data. From Table 2, we found that the kNN method achieved the lowest distance error and the regression-model-based method required the least amount of parameters. In addition, the proposed method was faster than the other 2 methods. The price paid by the kNN method for achieving the lowest distance error is the storage of the large amount of the training data. For example, the kNN method had to store 8 parameters for each training data consisted of 6 raw RSSI signal strength values and a 2-dimensional physical coordinates for the first data set. As for the RBF-based method, it took the largest amount of processing time.

CONCLUSION

Accurate knowledge about the user's position has many applications. A regression-model-based indoor location estimation method was introduced in this study. Simulation results showed that the regression-model-based method requires the least amount of parameters and processing time. If the distance error is not the major consideration then the proposed method can be considered as a good method to be implemented at mobile devices with a small amount of computational resource and memory.

ACKNOWLEDGEMENT

This research was partly supported by the 96CGH-NCU-A3 and the National Science Council, Taiwan, R.O.C, under the NSC-96-2221-E-008-017-, the NSC 96-2422-H-008-001-and the NSC-96-2524-S-008-002-.

REFERENCES

Brunato, M. and R. Battiti, 2004. Statistical learning theory for location fingerprinting in wireless LANs. Elsevier Science.

Gwon, Y., R. Jain and T. Kawahara, 2004. Robust indoor location estimation of stationary and mobile users, in INFOCOM. Twenty-third Ann. Joint Conf. IEEE. Comput. Commun. Soc., 2: 1032-1043.

Hashemi, H., 1993. The indoor radio propagation channel. Proc. IEEE., 81 (7): 943-968.

Li, X., K. Pahlavan, M. Latva-Aho and M. Ylianttila, 2000. Indoor geolocation using OFDM Signals in HIPERLAN/2 Wireless LANs. The 11th IEEE. Int. Symp. Personal. Indoor and Mobile Radio Commun., 2: 1449-1453.

Mauve, M., A. Widmer and H. Hartenstein, 2001. A survey on position-based routing in mobile ad hoc networks. IEEE Network, 15 (6): 30-39.

Niculescu, D., 2004. Positioning in ad hoc sensor networks. IEEE Network, 18 (4): 24-29.

Orr, R.J. and G.D. Abowd, 2000. The smart floor: A mechanism for natural user identification and tracking. Proc. 2000 Conf. Human Factors in Computing Systems (CHI 2000), ACM Press, New York.

Pahlavan, K., X. Li and J.P. Make, 2002. Indoor geolocation science and technology. IEEE. Commun. Mag., 40: 112-118.

Patwari, N., J.N. Ash, S. Kyperountas, A.O. Hero III, R.L. Moses and N.S. Correal, 2005. Locating the nodes: Cooperative localization in wireless sensor networks. IEEE. Signal Processing Mag., 22 (4): 54-68.

Roos, T., P. Myllymaki and H. Tirri, 2002. A statistical modeling approach to location estimation. IEEE. Trans. Mobile Comput., 1 (1) 59-69.

Savvides, A., L. Girod, M.B. Srivastava and D. Estrin, 2004. Localization in Sensor Networks. In: Wireless Sensor Networks, Raghavendra, C.S., K.M. Sivalingam and T. Zanti (Eds.). Norwell, MA: Kluwer.

Sun, G., J. Chen, W. Guo and K.J.R. Liu, 2005. Signal processing techniques in network-aided positioning: A survey of state-of-the-art positioning designs. IEEE. Signal Processing Mag., 22 (4): 12-23.