RESEARCH OF INDOOR LOCATION METHOD BASED ON THE RFID TECHNOLOGY

Feifei Guo, Chunkai Zhang, Min Wang, Xiaofei Xu

Department of Computer Science and Technology
Harbin Institute of Technology Shenzhen Graduate School
Shenzhen, China, 518055
ckzhang@hotmail.com

Abstract

Growing convergence among mobile computing devices and embedded technology promotes the development and the deployment of "context-aware" applications, where location is the most essential context. In this paper we present an improved probabilistic method based on RFID (Radio Frequency Identification) for locating objects inside buildings. The major advantage of this algorithm is to improve the probabilistic method based on Bayesian theory by introducing reference tags to amend the initial location estimation. Based on the experiment, we prove that we can get a higher accuracy in indoor location estimation with this method.

Keywords: indoor location, RFID, probabilistic method, reference tag.

1. Introduction

In general, there are two methods to get the location estimation, distance related method and distance unrelated method. The former should calculate the distance from the access points (AP) to the mobile terminal (MT), and then estimate the position by Trilateration or Triangulation algorithm. TOA(Time of Arrival), AOA(Arrival of Angle), TDOA(Time Difference of Arrival) and RSSI-based techniques are the four principal techniques to get the distance estimation. The latter, the distance unrelated method, however, doesn't need distance information and it mainly includes proximity based algorithm and distance in hops based algorithm. A prime example of this method is the centroid algorithm.

The TOA-based technique employs the signal transmission time from the APs to the MT to get the distance estimation [1]. But it must be implemented by synchronizing the wireless network which is formed through installing the GPS or atomic clock on APs. The AOA-based technique is based on the capability of network nodes to sense the direction from which a signal is received [2]. Luckily, it can be used in the phonetic channel and doesn't need accurate system timer. But AOA sensing requires either an antenna array or several ultrasound receivers so that it is easy to appear location blind spot and has much cost. The TDOA-based technique measures the delay time of the signal arrival, that the AP clocks must be accurately synchronized to calculate the time differences of received packets [3], that we often needs to rebuild the APs for timer accuracy, which may lead to many problems in the complex real situation. The RSSI-based technique estimates the distance by the RF propagation loss model which is a simple mathe-

matical expression representing the relationship between the RSSI and the distance between the sender and the receiver. However, RSSI is influenced by so many parameters and establishing an appropriate RF propagation loss model for off-the-shelf APs is not possible. The centroid algorithm is based on the network connectivity completely and it locates the object by regarding the geometrical center of surrounding reference points as the centroid which is the predicted position [4]. The centroid algorithm has easy implementation and fewer calculate amounts, but it needs more reference points so that its cost is high relatively in the indoor environment.

Due to the complexity of the indoor environment, such as the noisy, multipath effect and NLOS (Non-Line-of-Sight) properties, AOA, TOA and TDOA techniques cannot get a desirable accuracy in indoor context. Besides, AOA needs antenna array which has too high cost to fit for the indoor location, TOA and TDOA both require synchronization setting, it is infeasible for the inexpensive indoor hardware to provide fine-grain time synchronization. Also, the centroid algorithm needs a high cost to get a desirable accuracy as it needs a large number of references nodes. Compared with the above methods, the RSSI-based technique is based only on the RSSI [5], which need not to add additional sensor/actuator infrastructure, so we work on the RSSI-based location technique, we know that to estimate the distance by the RF propagation loss model is not reliable in the complex indoor context, so here we confine the RSSI-based location techniques to the machine learning methods.

In general, RSSI-based location techniques can be divided into the deterministic and the probabilistic techniques. Although both of them require a long and human labor-intensive training phase, the deterministic technique provides less precision than the probabilistic technique. For the deterministic technique [6], the area is subdivided into some small cells and data are obtained in these cells from APs, and in the positioning phase the collected objective data is then matched against the signal strength vectors mapped onto position in the sample tables and at last the closest match is returned. For the probabilistic positioning techniques [8-11], a probability distribution of the user's position is defined over the area of the movement. The goal of the positioning is to get a most probable position of the tracked user. A prime example of this method is the Bayesian learning method. T. Roos builds a Bayesian model with a preset number of discrete cells and gets the most probable position estimation by calculating a posteriori distribution over the location area [9]. Bayesian learning method is one of the machine learning methods used in the RSSI-based location estimation. An alternative machine learning method is the K-nearest neighbor (KNN) [12-14] which is a case-based method and can also be used to solve RSSI-based location estimation problem.

In this paper, we present an improved RSSI-based machine learning location algorithm. Here we adopt the RFID technology to obtain the RSSI [7]. In our algorithm we build a Bayesian model to get the primary location estimation, and then we introduce reference tags which are RFID passive tags and the KNN algorithm to amend the above estimation. Thus we can get higher location accuracy.

In the subsequent sections, this paper will make an introduction to Bayesian Learning location technique and KNN location technique in section 2; in section 3 we will describe our improved algorithm in detail; in section 4 the experiment results and analysis will be presented; lastly, we will give a conclusion clearly.

2. Location Method Based On RSSI

2.1. Bayesian Learning Method

Bayesian learning method [8-11] employs the Bayesian theorem to estimate the unknown position, given the signal measurements x in position i, then the position i is calculated by the posteriori as

$$i = \arg\max_i P(i \mid x) = \arg\max_i \frac{P(x \mid i)P(i)}{P(x)} \quad (1)$$

where $p(x \mid i)$ is the probability of receiving a sample x from position i and $p(i)$ is the probability of a MT being at this position i which initially can be considered as uniform in the location area, $p(x \mid i)$ can be calculated from the calibration information table (CIT) that is built ahead of time. Therefore, the location estimation problem becomes a standard maximization problem.

The main drawback of the above method is the large number of calibration samples necessary to construct the distribution $p(x \mid i)$ [8]. One possible approach to reduce the number of calibration samples is clustering [10].

2.2. K-Nearest Neighbor Method

KNN method is based on estimating the position i depending on the average (in physical space) of the coordinates of the k closest reference points to the received RSSI vector x. The generalized vector distance $d(x, y^i)$ can be described as

$$d(x, y^i) = \frac{1}{d}\left(\sum_{k=1}^{d} \frac{1}{w_k} \mid x_k - y_k^i \mid^p\right)^{\frac{1}{p}} \quad (2)$$

where p=2 denotes the Euclidean distance and p=1 the Manhattan one. The weight w_k can be used to bias the distance by a factor that indicates how reliable the calibration sample y_i is, but the improvement [13] is not very important.

The main problem of the KNN algorithm is the size of the CIT, which also makes the system slower due to the search times. One possible solution is to average the calibration points from every given position, thus reducing the CIT size.

In the weighted k-nearest neighbors (WKNN) [14] method, the average of the k-nearest coordinates is weighted by the distance in the RSSI space, that is

$$l_i = \frac{\sum_{j=1}^{k} (1/d(x, y^j) + d_0) l_j}{\sum_{j=1}^{k} (1/d(x, y^j) + d_0)} \quad (3)$$

where l_j is the physical coordinates of position j (with calibration vectors y) and d_0 is a small real constant to avoid division by zero. Traditional KNN is a special case of WKNN without using distance-dependent weights. Results show that WKNN achieves lower estimation error, the size of the CIT and the computation time being their main drawbacks.

3. Improved Algorithm

3.1. Notation Introduction

We denote random variables and their values by the same lowercase letters. In particular, l denotes location, and o denotes an observation variable or vector. We assume that the observation variable is a vector of received signal strength values for a set of RFID reader. The training data D consists of n samples, denoted by $(l_i\ o_i)$, $i \in \{1...n\}$, where n is the number of training samples. With a slight abuse of notation, we use the general notation $p(.)$ to denote all probability distributions. Conditional probabilities are denoted by $p(.|.)$. For any given location l we can obtain a distribution $p(o \mid l)$. By application of the Bayesian rule, we can then obtain the so-called *posterior distribution* of the location:

$$p(l \mid o) = \frac{p(o \mid l)\, p(l)}{p(o)}\, p(l \mid o) \qquad (4)$$

where $p(l)$ is the *prior probability* of being at location l before knowing the value of the observation variable. It gives a principled way to incorporate background information. Because the denominator $p(o)$ does not depend on the location variable l, it can be treated as a normalizing constant whenever only relative probabilities or probability ratios are required.

3.2. MLR Algorithm Description

We represent an improved algorithm—machine learning reference (**MLR**) which takes advantage of the Bayesian and WKNN idea and improves the overall location accuracy by introducing the reference tag concept. The final position estimation is obtained by followings:

(i) Input the objective vector to the Bayesian model, then the output which is two-dimensional coordinate is the rough position estimation of the target.

(ii) Find the nearest reference tags in accordance with the WKNN algorithm and compute the amending position.

(iii) Regard the average of above two positions as the final position of the target.

In the first phase, we take the following steps to derive the primary position:

(1) Build a Bayesian model of which the input is an observation vector of which the dimension is the reader number and the vector elements is the RSSI for each reader and the output is coordinate based on the *calibration data*.

(2) Input the objective vector and get the final result.

The critical issues are collecting of the *calibration data* and the training procedure. In accordance with the Bayesian rule, we only need to compute the product of $p(l)$ and $p(o \mid l)$. $p(l)$ can be considered as the occurrence frequency of samples which is obtained in l of all training samples. We adopt a probabilistic method—Kernel Method to compute $p(o \mid l)$:

$$p(o \mid l) = \frac{1}{n_1} \sum_{l_i = l} K(o; o_i) \qquad (5)$$

where n_1 is the number of training vectors in l and $K(o; o_i)$ denotes the *kernel function.*

$$K_{Gauss}(o; o_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(o - o_i)^2}{2\sigma^2} \right) \qquad (6)$$

where σ is an adjustable parameter that determines the width of the kernel. The result is the calibration point which has the largest posterior probability and of which the position coordinate is (x_1, y_1).

In the second phase, we define the signal strength vector of a tracking tag as $S = \{S_1, S_{2,\ldots} S_n\}$ where S_i denotes the signal strength received on reader i, where $i \in (1 \ldots n)$, and for reference tags, the corresponding signal strength vector as $\theta = (\theta_{1,} \theta_2 \ldots\ldots \theta_n)$. And the Euclidean distance between two kind tags is defined as:

$$E_j = \sqrt{\sum_{i=1}^{n} (\theta_i - S_i)^2} \qquad (7)$$

When there are m reference tags, a tracking tag has the vector $E = (E_{1,} E_2 \ldots\ldots E_m)$ .We select k reference tags which have lower E value. The unknown tracking tag's coordinate (x_2, y_2) is obtained by:

$$(x_2, y_2) = \sum_{i=1}^{k} w_i (x_i, y_i) \qquad (8)$$

where w_i is the weight of the ith neighboring reference tag. The computation of these weight values is very important to our research. Empirically, the weight is given by:

$$w_j = \frac{1/E_i^2}{\sum_{i=1}^{k} 1/E_i^2} \quad (9)$$

The final position estimation is computed as the average of (x_1, y_1) and (x_2, y_2) .

4. Experiment Result and Analysis

We simulate an actual rectangular scene (30×20 m) and each corner of the area is deployed with a reader and the area is divided by 2-meter grids on average. The training data is collected systematically on the center of each grid, which we call *calibration points*, 40 observations are recorded, each consisting of RSSI from all readers. We place reference tags on the four acmes which we call *reference points* of each grid. Similarly, the reference data is also consisting of RSSI value from all readers and they are collected on reference points. At each of the reference point, 20 observations are gathered. We select the junctions of each grid as the *test points*, at each of which we collect 20 observations.

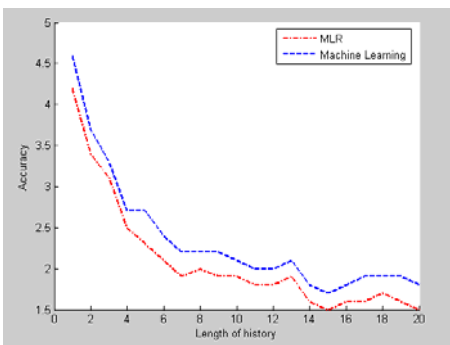

Fig.1 Relation between length of history and accuracy

From the above graph, we can see for the same test data, our algorithm has much higher location accuracy than the machine learning method in the whole level, and it appears higher precision with the increase of the number of test observations, so we can affirm that the location accuracy is related with the history data closely. Our observations are obtained in certain time interval, so if many observations are detected on certain location, it means that the target object is moving at low speed or it doesn't move at all. Similarly, if few observations are detected, we can think the target travels in so high speed that we can't record its observation value. However, the precision is related with other factors also. Such as the number of readers and training samples, the deploy way of readers and tags and the parameters of MLR method and so on. In view of the paper space, we only present the experiment conclusion which is when there are four readers and they are deployed in the four corners, 15 observations at each calibration point, we get much higher accuracy.

5. Conclusion

In this paper, we research the indoor location method based on RFID technology and RSSI. The batch process and periodicity of RFID make us measure RSSI quickly and accurately. But the location accuracy of machine learning usually cannot meet the user's need. So we bring in the reference tag concept to amend the rough accuracy. The result of experiments shows that the improved idea is feasible in the actual location but amount of effort must be used in the procedure of collecting training data. The most key is our algorithm has much higher accuracy. It should be emphasized that our algorithm is fit for the objects which move in low speed.

6. References

[1] D.Niculescu, B.Nath. "Ad Hoc Posiitoning System(APS) Using AOA". *In Proceedings of the 20st Annual Joint Conference of the IEEE Computer*

and Communications Societies, Anchorage, AK, 2003,3:1734~1743.

[2] X.Li, K.Pahlavan. " Super-resolution TOA estimation with diversity for indoor geolocation". *IEEE Transactions on Wireless Communication* .2004, 3(1):224~234 .

[3] K.Pahlavan, Li. Xinrong and J. P. Makela, "Indoor geolocation science and technology," *IEEE Communications Magazine, vol. 40, issue 2*, pp 112-11, Feb.2002.

[4] N. Bulusu, J. Heidemann, D. Estrin. "GPS-Less low cost outdoor localization for very small devices". *IEEE Personal Communications.* 2000, 7(5):28~34.

[5] G. V. Z`aruba , M. Huber . F,A. Kamangar and I. Chlamtac, "Indoor location tracking using RSSI readings from a single Wi-Fi access point", *Springer Science + Business Media*, LLC 2007.

[6] A. Smailagic, D.P. Siewiorek, J. Anhalt, D. Kogan and Y. Wang, "Location sensing and privacy in a context-aware computing environment", *IEEE Wireless Communications Magazine* (Oct. 2002) 10–17.

[7] A.M. Ladd, K. Bekris, A. Rudys, G. Marceau, L.E. Kavraki and D.S. Wallach, Robotics-based location sensing using wireless ethernet, in: *Proceedings of the 8th ACM MobiCom*, Atlanta, GA(Sept. 2002) pp. 227–238.

[8] T. Roos, P. Myllyma¨ki, H. Tirri, P. Misikangas and J. Sieva¨nen. "A Probabilistic Approach to WLAN User Location Estimation". *International Journal of Wireless Information Networks*, Vol. 9, No. 3, July 2002.

[9] T. Roos, P. Myllym¨aki, and H. Tirri, "A statistical modeling approach to location estimation," IEEE Transactions on Mobile Computing, vol. 1, no. 1, pp. 59–69, 2002.

[10] M. Youssef, A. Agrawala, A.U. Shankar and S.H. Noh, "A probabilistic clustering-based indoor location determination System", Tech. Report, University of Maryland CS-TR 4350.

[11] A. Smailagic, D.P. Siewiorek, J. Anhalt, D. Kogan and Y. Wang, "Location sensing and privacy in a context-aware computing environment", *IEEEWireless Communications Magazine.*

[12] P. Bahl and V. Padmanabhan. "Enhancements to the RADAR User Location and Tracking System". *Technical Report* MSR-TR-2000-12, Microsoft Reasearch,2000.

[13] P. Prasithsangaree, P. Krishnamurthy, and P. K. Chrysanthis, "On indoor position location with wireless LANs," In Proceedings of 13th IEEE International Symposium on Personal Indoor and Mobile Radio Communi-cations (PIMRC '02), vol. 2, pp.720– 724, Lisboa, Portugal, September 2002.

[14] M.BrunatoandC.K.Kall´o, "Transparent location fingerprinting for wireless services," *Tech. Rep.* DIT-02-0071, September 2002.(Oct. 2002) 10–17.