American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)
ISSN (Print) 2313-4410, ISSN (Online) 2313-4402

http://asrjetsjournal.org/

RFID Localization System Based on K-Nearest Neighbor Algorithm and Extreme Learning Machine Algorithm with Virtual Reference Tags

Changsheng Zhang[a*], Yaoxian Li[b]

[b,c]*Tianjin Polytechnic University,Binshuixi Road Xiqing ,Tianjin,China,300387*

[a]*Email: 15122982172@163.com*

[b]*Email: 1362440896@qq.com*

Abstract

Aiming at the complex indoor environment, a new indoor localization algorithm was proposed based on the K-nearest neighbour algorithm (KNN) and extreme learning machine (ELM) by using virtual reference tags. In this paper, LANDMARC location algorithm is used in regional location during the online phase, and the ELM with virtual reference tags was introduced in the locked area. The design scheme of indoor positioning system based on Intel R1000 platform is proposed, and the system was realized by C++ language. The experimental data show that average error of the indoor positioning system is 0.3m and the reduction in estimation error is 38% over LANDMARC and 19% over ELM. The system can effectively improve the indoor localization accuracy in low tag density environments.

Keywords: Radio frequency identification(RFID); localization system; landmark; extreme learning machine; virtual reference.

1. Introduction

The global positioning system (GPS) is most widely used in the outdoor location. However, due to the signal path loss and other technical restrictions, GPS is not suitable for complex indoor environment. Using arrival angle (AOA), arrival (TOA) and arrival (TDOA) information of the time difference, many solutions have been proposed.

----

* Corresponding author.

Although these solutions provide high accuracy, they usually require advanced hardware, which will increase the cost and complexity of the system. In many types of indoor positioning system based on RSSI, LANDMARC is widely used due to the simple principle, high accuracy [1]. But its accuracy depends on the density of reference tags [2], and is conditioned by hardware features.

After the tag reaches a certain range of density, adding tags cannot improve the positioning accuracy, so the positioning performance is limited [2]. Due to the widespread deployment and low cost of Wireless Local Area Network (WLAN), WLAN based indoor localization solutions have gained more attention[3], and the Received Signal Strength (RSS) based fingerprinting approach is recognized as the most popular one among them[4]. From the RFID technology, this paper has carried on the research of fingerprint method based on the received signal strength.

Extreme learning machine (ELM) is a kind of new algorithm for single hidden-layer feed forward neural networks (SLFNs), the algorithm randomly generated input layer connection with the hidden layer weights and thresholds for hidden layer neurons, and no adjustments in the training process, you need to set the number of hidden layer neurons, and Only optimal solutions can be obtained [5]. Compared with the traditional training methods, the method has the advantages of fast learning and generalization performance. In [6], two localization algorithms: Weighted Path Loss (WPL) and Extreme Learning Machine (ELM) are proposed. WPL can provide a faster estimation while ELM can provide a higher localization accuracy. In this paper, based on RSSI localization algorithm is proposed on the basis of a hybrid algorithm of KNN-ELM with virtual reference tags, which is used to increase accuracy of the ELM model. The VIRE uses the concept of virtual reference tags which are locations with known distance between real reference tags [7]. The RSSI of Each virtual reference tag is calculated by its surrounding real reference tags RSSI [8]. Compared with LANDMARC algorithm and ELM algorithm, this algorithm can improve the position precision and efficiency under the condition without increasing the tag density.

2. Algorithm

2.1 Algorithm of RSSI Location

In RF system, the main attenuation form of electromagnetic wave is large scale fading, which related to many factors, including the time, the distance between the tag and the reader, the carrier, the carrier frequency and so on. Based on the Path Loss Model defined in [6], the signal strength $\Gamma(\mathrm{d})$ can be expressed as:

$$\Gamma(d) = \Gamma(d_0) + 10n\log(\frac{d}{d_0}) \qquad (1)$$

Where n is the path loss exponent, showing the rate of path loss with increasing distance, $\mathrm{d}_0$is known as the near distance, as the reference distance, d is the distance between the tag and the reader.

2.2 Algorithm of ELM

ELM is a kind of machine learning algorithm based on Single-hidden Layer Feed forward neural Networks

(SLFNs) architecture. It has been proved to provide good generalization performance at extremely fast learning speed [5]. Following is the brief description about ELM given by Huang [5]. For SLFNs, the outputs with L hidden nodes can be represented as:

$$y_N(\boldsymbol{x}) = \sum_{i=1}^{L} \boldsymbol{\beta}_i g_i(\boldsymbol{x}) = \sum_{i=1}^{L} \boldsymbol{\beta}_i G(\boldsymbol{a}_i, b_i, \boldsymbol{x}) \qquad (2)$$

Where a_i, b_i are the weights and bias connecting the input nodes and the ith hidden node, β_i are the output weights connecting the ith hidden node and the output nodes, and $G(a_i, b_i, x)$ is the activation function which gives the output of the ith hidden node with respect to the input vector x.

Suppose we have N arbitrary distinct training samples $(x_i, t_j), j = 1,2, \dots, N$, we can represent the SLFN for each sample as equation:

$$y_N(\boldsymbol{x}_j) = \sum_{i=1}^{L} \boldsymbol{\beta}_i G(\boldsymbol{a}_i, b_i, \boldsymbol{x}_j), j = 1,2, \dots, N. \qquad (3)$$

The above N equations can be written compactly as:

$$\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{T} \qquad (4)$$

Where

$$\boldsymbol{H} = \begin{bmatrix} G(\boldsymbol{a}_1, b_1, \boldsymbol{x}_1) & \cdots & G(\boldsymbol{a}_L, b_L, \boldsymbol{x}_L) \\ \vdots & \dots & \vdots \\ G(\boldsymbol{a}_1, b_1, \boldsymbol{x}_N) & \cdots & G(\boldsymbol{a}_L, b_L, \boldsymbol{x}_N) \end{bmatrix}_{N \times L},$$

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1^T \\ \vdots \\ \boldsymbol{\beta}_L^T \end{bmatrix}_{L \times m}, \boldsymbol{T} = \begin{bmatrix} \boldsymbol{t}_1^T \\ \vdots \\ \boldsymbol{t}_L^T \end{bmatrix}_{N \times m}$$

$\boldsymbol{H}$ is called the hidden layer output matrix of ELM; $\boldsymbol{T}$ is called expected output.

Unlike the traditional training algorithms for neural network, which needs to adjust the input weights and hidden layer biases, Huang in [5] has proved that these parameters of SLFN can be randomly assigned if only the activation function is infinitely differentiable. Thus, the hidden layer output matrix H remains to be fixed once these parameters are randomly initialized. To train a SLFN is simply equivalent to find a least solution of equation (5), Training neural network with a single hidden layer can be converted into solving the optimal solution of a linear system which gives:

$$\|\boldsymbol{\beta}_{LS} - \boldsymbol{T}\| = \min_{\beta} \|\boldsymbol{H}\boldsymbol{\beta} - \boldsymbol{T}\| \qquad (5)$$

Solving a linear system equation (4) and output weights can be determined:

$$\boldsymbol{\beta} = \boldsymbol{H}^{+}\boldsymbol{L} \qquad (6)$$

Where $\boldsymbol{H}^{+}$ is the Moor-Penrose generalized inverse of $\boldsymbol{H}$.

3. KNN-ELM Algorithm Analysis

The KNN-ELM hybrid location algorithm proposed in this paper combines the recent K algorithm and the machine learning algorithm of extreme learning machine, which can be divided into two stages. First, we construct the LANDMARC region localization space, namely the zone lockdown phase. Because the computation of LANDMARC algorithm is smaller and the location is more accurate, it can be used to quickly search the area to be located. Second, the selected area is introduced into the ELM training model, which is used to calculate the exact location of the target by the ELM training model, namely lock stage. Before getting the ELM training model, the RSSI of the reference tag is used to add the virtual tags in the area of each partition, and then the machine is trained. Algorithm flow chart is shown in Figure 1.

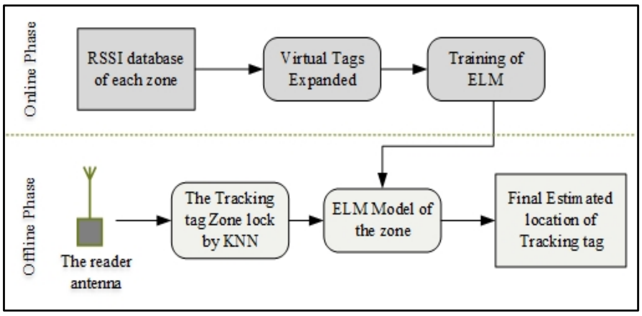

Figure1: Flowchart of integrated KNN-ELM using virtual reference tag approach.

3.1 LANDMARC Algorithm

Target region locking is mainly based on the LANDMARC algorithm of the "K -nearest neighbour". The following brief introduction is the LANDMARC algorithm [9, 10]. The arrangement of the reader and tag is shown in Figure 2.

It is assumed that the total of m readers and n reference tags and u target tags. The above field strength value is replaced by vector, the field strength vector of the target is denoted as $\boldsymbol{T} = (T_1, T_2, \cdots, T_m)$, which T_j represents the field strength from the jth reader to target; the field strength vector of the tag reference is denoted as $\boldsymbol{R}(i) = (R_1(i), R_2(i), \cdots, R_m(i))$, which $R_j(i)$ represents the field strength from the jth reader to reference tags. In order to determine the extent of the reference tag and the target neighbour, for each target tag $p(p \in [1, u])$, $\boldsymbol{E}_i = |\boldsymbol{T} - \boldsymbol{R}(i)| = \sqrt{\sum_{j=1}^{m} (T_j - R_j(i))^2}$ is defined [11], which represents the distance between the ith reference tags and the targets, the smaller the E_i is, the closer the distance is, $i \in [1, n]$.

By comparing the value of $\boldsymbol{E}_i$, we find the nearest reference tags as the nearest neighbour tag, and the coordinates of the k tags are known. The area will be a region which is surrounded by the k tags, the introduction of regional database, k equals to 4 is generally better [12], as shown in Figure 2.

3.2 ELM Training With Virtual Reference Tag

After the LANDMARC area is locked out, it needs to expand and construct the measurement matrix. The construction of measurement matrix is the key to complete the position locking, which has great influence on the positioning accuracy.

Through the fixed RFID reader reads signal strength value of the N reference tags as $RSSI_{nm}(1 \ll n \ll \mathrm{N}，1 \ll m \ll 3)$, and signal strength value for T as $\mathrm{RSSI}_{tm}(1 \ll t \ll \mathrm{T}，1 \ll m \ll 3)$, all of them are collected by 3 pairs of antennas.

To improve positioning accuracy we propose using virtual reference tags. So, we determine these virtual reference tags locations with specific coordinates between real reference tags and calculate their RSSI using the RSSI of real reference tags. According to the method presented in VIRE algorithm, $n-1$ virtual reference tags are assumed between each two real reference tags [7]. Therefore, there are $(n+1)^2-4$ virtual reference tags, 4 real reference tags and totally $(n+1)^2$reference tags in each section [7]. Therefore, the RSSI values of each virtual reference tag will be calculated according to what was said in VIRE approach. Since $n-1$ virtual reference tags are assumed between each two real reference tags, $n \times n$ subsections are generated in each section as shown in Figure 2. And then through the calculation of the virtual label RSSI worth to $RSSI_{vm}$ and virtual label coordinates $(i，j)$.

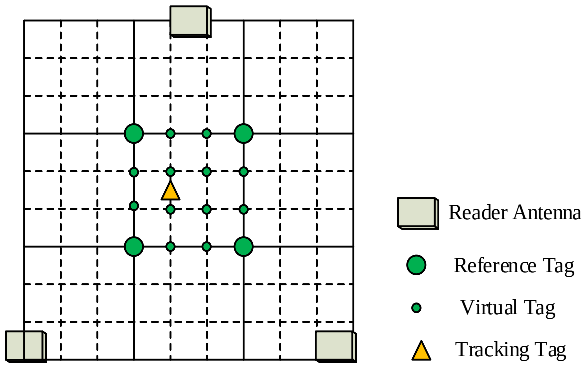

Figure 2: Placement of RIFD tracking tags, reference tag and reader.

ELM consists of an offline phase and an online phase. During the offline phase, some RFID tags are adopted as reference tags in order to build up an empirical database. In addition, each RSSI sample is represented as$(RSSI_{nm}, (i，j))$ and$(RSSI_{vm}, (i，j))$.The vectors $RSSI_{nm}$and $RSSI_{vm}$ are the inputs of the ELM and the corresponding location vectors (i，j) are the training targets of ELM. The hard-limit transfer function$\mathrm{G(a,b,x)} = \mathrm{hardlim(ax + b)}$ is chosen as the activation function in this paper. ELM training process is introduced in the previous section. It is mainly the following three main steps: (1) randomly assign values to hidden node parameters; (2) Calculate hidden output matrix H; (3) Calculate the output weight β, by equation (6).

During the online phase, we need to enter the sample database $RSSI_{tm}$ to the model of ELM, The output given by ELM is the estimated location of the tracking tag. Due to the ELM, large amounts of data can increase accuracy of the model. If only by increasing the number of entity reference tags, the cost of system will be increased, and the measurement process is rather complicated, so we increased by using virtual reference tags in the above steps, which makes it possible to not only meet the requirement of ELM training data, but also increase the accuracy of training.

4. System Structure and Result Analysis

Location hardware system including the R1000 Intel radio frequency identification and development platform, PC, remote reader antenna, XCTF-8030a passive RFID tags was built in the open hall. The system can read out a total of 256 levels of 0~255 signal strength value and uses 3 pairs of antennas in time division multiplexing instead of multiple readers. The PC machine is the host of this system. Three pairs of antennas placed in three different locations, and tag is about 1.5 meters apart, which placed as shown in Figure 2. In this system, 16 reference tags were used, which means N=16 mentioned in section 3.2. Interface software is developed based on the above algorithm, has set up a positioning system for the hardware and software together with mixed programming by using VC++ gets MATLAB engine [13, 14]. It is not only easy to get ELM training model in MATLAB tools, but also simplifies source program codes. Software interface is shown in Figure 3.

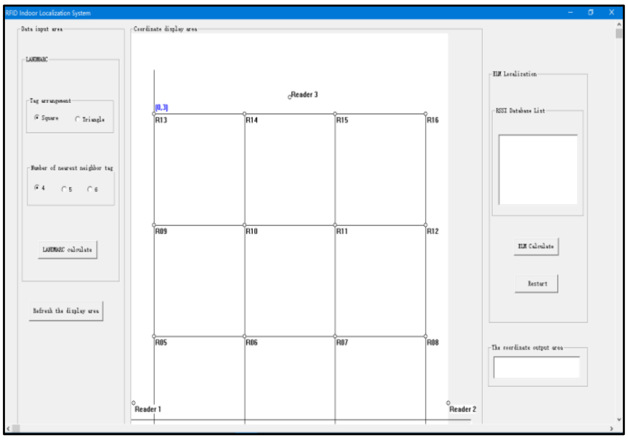

Figure3: Screenshot of the system running.

We define the location estimation error [9] $e = \sqrt{(x-\tilde{x})^2 + (y-\tilde{y})^2}$ to be the distance between the real location coordinates $(\tilde{x}, \tilde{y})$ and the system estimated location coordinates(x, y). With the method of statistical probability distribution function e is the location error in the actual measurement, L is the probability distribution function of the horizontal coordinates. $P(e < L)$ is a percentage that the number of measured results to the total number of times measurements, which e is smaller than the horizontal coordinate L. Such as $P(e < 0.5)$ is the result of all measurement errors, the number of e which is smaller than 0.5m to the total number of measurements.30 sets of data is measured continuously in our experiment, and the curve of $P(e < L)$ is draw according to the measurement results.

4.1 The Path Loss Exponent n

The signal strength value is measured at about 10 meters, and the measuring distance is 1 meters. At each location, 10 RSSI samples are collected in 1 day. Figure 4 shows the average signal strength of collected RSSI data at various locations. Based on the data we collected, we use a curve fitting method to construct the relationship between RSSI and distance, as:

$$\Gamma(d) = -39.497 + 10 \times 1.48 \times log(d) \qquad (9)$$

The pass loss exponent n is taken as 1.48 and the reference pass loss coefficient $\Gamma(d_0)$ as -39.497dBm. We assume that d and $\Gamma(d_0)$ remain unchanged in the entire test period.

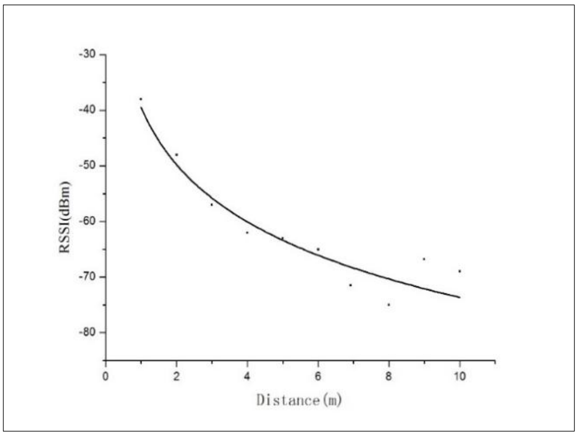

Figure 4: Relationship between RSSI and distance.

4.2 Experimental Results

1) Comparison between KNN and ELM: We use the VIRE method to increase the training database of ELM, which will increase the accuracy of the model training during the training process. Besides the number of input variables, another parameter that could affect the localization accuracy of ELM is the number of hidden nodes in ELM hidden layer. Figure 5 shows the performance comparison between the KNN algorithm and the ELM with different number of hidden nodes. It can be seen that as the number of hidden nodes in ELM hidden layer increases, the localization accuracy of ELM improves. Although the hidden nodes can increase the number of neurons and the scope of training, there is not the more the better. Besides, more hidden nodes will affect the speed of training.

In test, when the number of hidden nodes increases to 5000, the localization accuracy of ELM is 0.37m which becomes better than the performance of KNN (0,48m). But when the hidden layer node is increased to 6000, the effect is not good. In summary, Selection of hidden nodes number appropriately can improve the localization accuracy.

2) Comparison between KNN, ELM and KNN-ELM: In the open hall after testing, the results were calculated using the original KNN algorithm and the algorithm proposed in this paper. It can be seen that the new improved algorithm is better than the existing KNN algorithm on the same 256 level hardware system.

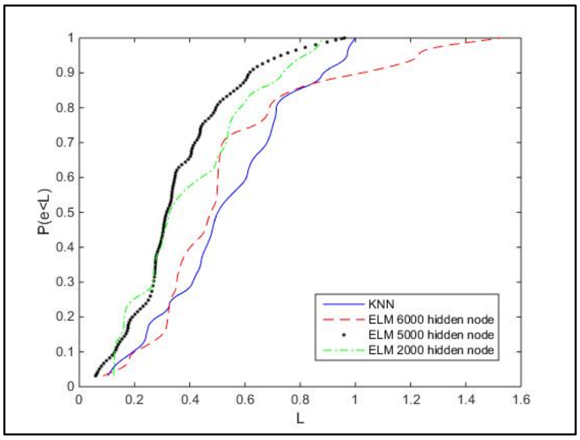

Figure 5: Cumulative percentile of error distance (L) of KNN and ELM using virtual reference tags with different number of hidden nodes.

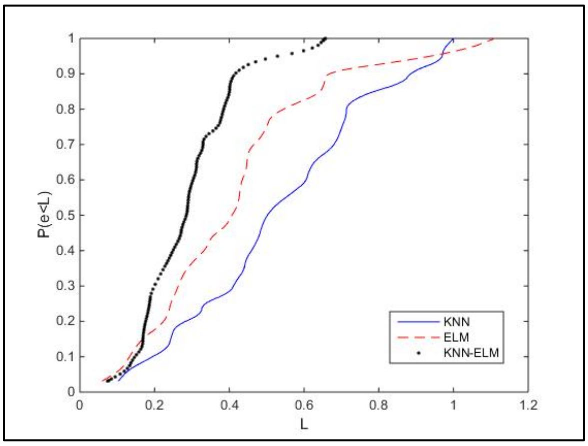

Figure 6: Cumulative percentile of error distance (L) for different methods.

The performance between three kinds of method comparison results as shown in the Figure 6, in the same environmental conditions, through the measurement of 30 different positions of the target label data, The average localization accuracy of all tracking tags by using LANDMARC, ELM and KNN-ELM using virtual reference tag is 0.48m, 0.37m and 0.30m. In this case, KNN-ELM using virtual reference tag enhances the precision of localization accuracy by 38% over LANDMARC, 19% over ELM respectively. In summary, the KNN and ELM hybrid algorithm by using virtual reference tags can provide better positioning accuracy than the LANDMARC algorithm.

5. Conclusion

In this paper, a combination algorithm based on the K-nearest neighbour and ELM with virtual reference tags is proposed. This algorithm effectively combines the characteristics of LANDMARC algorithm for the fast selection of the nearest neighbour tags and the high accuracy of the extreme learning machine algorithm. With the location hardware system with Intel R1000 RFID development platform build in the open hall, the experimental results show that the estimation error of ELM is increased by 23% compared to LANDMARC without the increase of label density, while the KNN-ELM algorithm with virtual reference tags is improved by 38% compared with the LANDMARC algorithm.

Due to the limitation on the number of tags and the transmit power of readers, the scale of system performance is limited. As the future work, there will be a much larger reference tag array in a much larger sensing area to study the effects of different spacing environment and the number of virtual reference tags.

Acknowledgments

This research is funded by Tianjin Research Program of Application Foundation and Advanced Technology (No. 15JCYBJC16300).

References

[1] Huang, Chih-Ning., and Chan,Chia-Tai., “ZigBee-based indoor location system by k-nearest neighbor algorithm with weighted RSSI” [J]. Procedia Computer Science, 2011, pp. 58-65.

[2] Bahl, P., and Padmanabhan, V., “Radar: An in-building RF based user location and tracking system”, in proceedings of IEEE Infocom, 2000, pp. 775-784.

[3] Xiao,W., Ni, W., and Toh, Y., “Integrated Wi-Fi Fingerprinting and Inertial Sensing for Indoor Positioning”. International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2011, pp.1-6.

[4] Meng, W., Xiao, W. Ni, W., and Xie, L., “Secure and Robust Wi-Fi Fingerprinting Indoor Localization”, 2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2011, pp.1-7.

[5] Huang, G.B., Zhu, Q.Y, and Siew, C.K., “Extreme Learning Machine: a New Learning Scheme of Feed forward Neural Networks” [C]. Proceedings of the International Joint Conference on Neural Networks. Piscataway: Institute of Electrical and Electronics Engineers Inc, 2004, pp.985-990.

[6] Zou, H., Wang, H., Xie, L., and Jia, Q.S., “An RFID indoor positioning system by using weighted path loss and extreme learning machine”, in Cyber-Physical Systems, Networks, and Applications (CPSNA),IEEE 1st International Conference on,2013, pp. 66-71.

[7] Zhao, Y.Z., Liu, Y.H., Lionel, M., and Ni, L.M., “Vire: active RFID-based localization using virtual reference elimination” [C].The 2007 international Conference on Parallel Processing (ICPC 2007), 2007, pp. 56-56.

[8] Li, J.H., Qi, R., Wang, Y., and Wang, F., “An RFID Location Model Based On Virtual Reference Tag Space”, Journal of Computational Information Systems, June 2011, pp. 2104-2111.

[9] Ni, L., Liu, Y., Lau, Y., and Patil, A., “LANDMARC: indoor location sensing using active RFID” [A].

Proceedings of the First IEEE International Conference on Pervasive Computing and Communications (PerCom2003) [C]. Dallas, Texas, USA, March 2003, pp. 407-415

[10] Bahl, P., Padmanabhan, V.N., and Balachandran, A., “Enhancements to the RADAR User Location and Tracking System” [R]. Microsoft Research Technical Report, 2009.

[11] Sun, Y., Fan, Z.P., “RFID technology and its application in indoor positioning” [J]. Journal of Computer Applications, 2005, (5), pp. 1205-1208.

[12] Ni, L., Liu, Y., Lau, Y., and Patil, A., “LANDMARC: Indoor Location Sensing Using Active RFID”, Wireless networks, vol. 10, no. 6, 2004, pp. 701–710.

[13] Liu, W., “Matlab and C/C++ mixed program design”, Beijing, Beihang University Press, 2005, pp.56-59.

[14] Ruan, S.Y., “MATLAB Program design”, Beijing, Publishing House of Electronics Industry, 2004, pp.65-6