Advances in Computer Science and its Applications (ACSA)
Vol. 1, No. 3, 2012, ISSN 2166-2924

The Research of NN-KNN Hybrid Location Algorithm Based on Background Cloud Computing

[1]Guanqing Xu, [1]Ping Zhao, [1]Xiaoliang Zhang, [1]Riming Lin, [1]Pengcheng Liang

[1]School of Electronics and Information, Northwestern Polytechnical University, Xi' an, P. R. China

Email: guanqing616@sina.com

Abstract –In this paper, a hybrid location algorithm for fingerprint-based indoor position systems is proposed based on background cloud computing. A novel computing method is applied to the hybrid algorithm for reducing computational complexity. The performance of the hybrid algorithm is simulated with both the example of nearest neighbors (NN) and k-nearest neighbors (KNN) algorithm, and the computational complexity is analyzed theoretically. Simulation results indicate that the better location performance can be achieved and the computational complexity is reduced by proposed NN-KNN hybrid algorithm, and the NN-KNN hybrid algorithm is also suitable for other location systems.

Keywords –indoor location; location fingerprint; cloud computing; NN; KNN

1. Introduction

Recent years, applications based on distributed systems are concerned extensively. With the progress of the cloud computing, more and more traditional areas are combined with the new platform. For example, the cryptography theory is applied on cloud computing [1], and a distributed testing architecture for Service-Oriented architecture is proposed in [2]; meanwhile, some tools for distributed platform are also investigated [3]. More and more applications based on cloud computing are proposed, but little attention was paid to the indoor location based on cloud computing. Due to the huge computational complexity, indoor location algorithm, especially fingerprint based indoor location algorithm, can be improved combined with cloud computing.

At present, indoor location includes optical tracking location [4], assisted global position system (A-GPS) [5], radio waves and ultrasonic integrated location [6] and received signal strength (RSS)-based location [7], etc. Recent years, with the development of wireless local area network (WLAN), more and more related location technology and applications were proposed and fingerprint-based indoor position systems based on RSS were paid extensive attention.

Fingerprint-based indoor position systems mainly include nearest neighbors (NN) algorithm, k-nearest neighbors (KNN) algorithm, etc. The location process consists of two phases: off-line phase and on-line phase. In off-line phase, the tasks are sampling and building radio-map; in on-line phase, the task is location matching. In the whole process, it needs vast cost for storing data and computation, which limits the application of the two algorithms.

In this study, a fingerprint matching simplified algorithm and NN-KNN hybrid algorithm are proposed based on background cloud computing for reducing the computational complexity and improving the location performance. NN-KNN denotes that different algorithms are disposed and the final location result is got by synthesizing the result of each algorithm. In this paper, the reduced computational complexity is analyzed theoretically and the NN-KNN hybrid algorithm is simulated with the example of NN algorithm and KNN algorithm.

2. NN, KNN, and NN-KNN hybrid algorithm

The location process using the proposed algorithms is based on the background cloud computing. As shown in figure 1, when location request is received, radio-map is first optimized with the fingerprint matching simplified algorithm and the algorithms for NN-KNN hybrid algorithm are chosen as well, then the algorithms will be fused to MapReduce tasks in background cloud computing center based on Hadoop Architechture for computation, and the location result will be got by synthesizing each algorithm result.

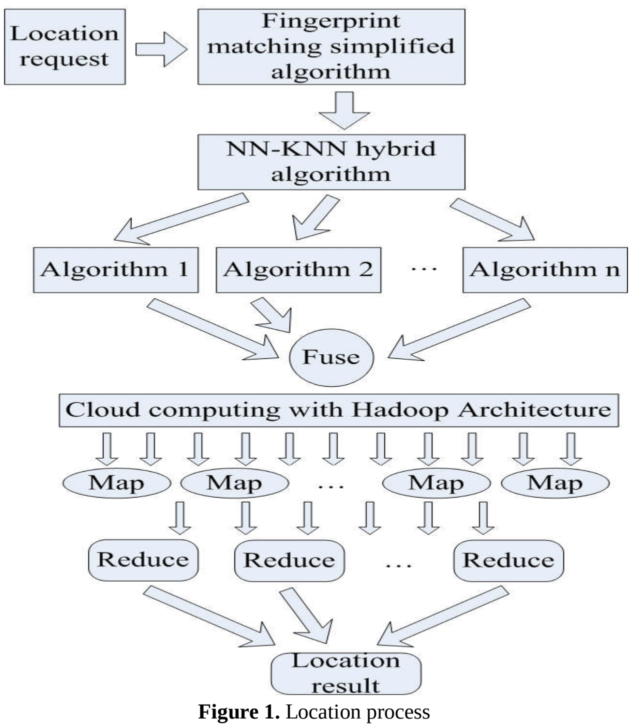

Figure 1. Location process

2.1. NN algorithm and KNN algorithm

In KNN algorithm, the distances between measured RSS samples in on-line phase and fingerprints in radio-map are computed. Assuming the numbers of reference point (RP) and access point (AP) are b and c respectively， the distances are defined as

$$L_{qi} = \left(\sum_{j=1}^{c} \left| s_j - S_{ij} \right|^q \right)^{1/q}, i = 1, 2, ..., b, \quad (1)$$

where S_{ij} is the RSS sample from *j*th AP at *i*th RP in radio-map, and s_j is the RSS sample from *j*th AP measured in on-line phase [8].

There are *k* RPs that are chosen according to the first *k* minimum distances defined by (1), and the estimated location is calculated by

$$\left(x_t, y_t\right) = \frac{1}{k} \sum_{i=1}^{k} \left(x_i, y_i\right), \quad (2)$$

where (x_i, y_i) is coordinate of *i*th RP, and (x_t, y_t) is the estimated coordinate of the test point (TP) [9].

In NN algorithm, the parameter *k* in (2) equals 1.

2.2. Fingerprint matching simplified algorithm

RSS decays exponentially as the distance increases because of the propagation loss of electromagnetic wave. RSS decays fast when the transmission distance is short, and RSS decays slowly as the transmission distance increases; RSS will be relatively stable when the transmission distance is around 8 m [10]. On the basis of the above theory, the RSS received by TP from AP that is over 8 m tends to be stable in certain location area, which indicates that there are much redundant computations in the process of fingerprint matching computation. The algorithm is designed for ignoring the RSS that is over 8 m to TP in the process of fingerprint matching computation to reduce computational complexity. The whole algorithm is described below.

Assuming there are *n* APs in certain location area and the RSS is set at critical value *α* when TP is 8 m to AP, the RSS that one certain TP receives from APs is X = (X1, X2, ⋯ , Xn), then let the elements of the vector X subtract α one by one. Xi is reserved when Xi is equal or greater than α and is discarded when lesser than α. Assuming that there are *m* elements of the vector X reserved, then, a new vector x = (x1, x2, ⋯ , xm) (*m is equal or lesser than n.*) is formed with the reserved elements of the vector X sequentially.

In the process of fingerprint matching computation, assuming the vector of one certain RP in the radio-map is Y = (Y1, Y2, ⋯ , Yn), a new vector y = (y1, y2, ⋯ , ym) is formed with the selected elements of the vector Y in accordance with the vector x, and then the distance between the vector x and y is calculated as

$$d = \sqrt{\left(x1 - y1\right)^2 + \left(x2 - y2\right)^2 + ... + \left(xm - ym\right)^2}. \quad (3)$$

The traditional distance between vectors is replaced by d calculated with fingerprint matching simplified algorithm. The same calculation is done on other RPs.

In table 1, the reduced computational complexity in one RP is first given, and the total reduced computational complexity is also given with the assumption that there are *p* RPs in the location area.

Table 1. Analysis of computational complexity

Algorithms	Arithmetic	
	addition/subtration	multiplication
Traditional algorithm	2n	n
Fingerprint matching simplified algorithm	3m	m
Reduced computational complexity	2n-3m	n-m
Total reduced computational complexity	p(2n-3m)	p(n-m)

2.3. NN-KNN hybrid algorithm

The proposed NN-KNN hybrid algorithm adopts diversified algorithms for joint location. The location tasks in on-line phase are allocated to background cloud computing center based on Hadoop Architecture for dispose, and the location result of each algorithm will be summated in weight to draw the final location result.

The process of NN-KNN hybrid algorithm is described in the following paragraphs and figure 2.

- First, in off-line phase, RSS of each RP from APs is measured and the radio-map will be built according to the measured RSS.
- Second, in on-line phase, the RSS of TP will be first measured, and according to the radio-map, algorithms for NN-KNN hybrid algorithm will be selected and fingerprint matching of each selected algorithm using the fingerprint matching simplified algorithm proposed in this paper will be made.
- Finally, the result of each algorithm got from background cloud computing center will be summated in weight to get the coordinate of TP.

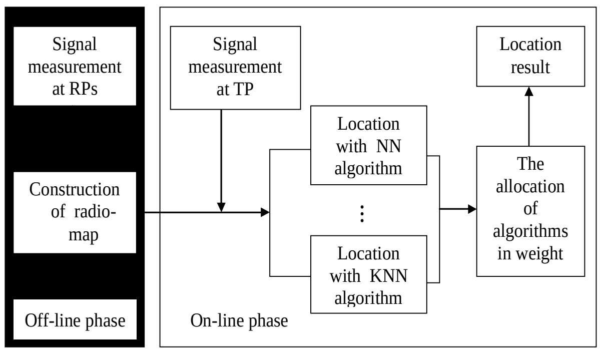

Figure 2. The process of NN-KNN hybrid algorithm

3. Algorithm simulation design

The performances of the proposed algorithms were studied through computer simulation, which was all done based on the platform of MATLAB 7.0. We chose a real area in the campus as the simulation environment, and its environmental feature was considered as representative. As to the algorithms, we first studied the influences of different k values of KNN algorithm on the mean location error, which might have some connections with the performances of NN-KNN hybrid algorithm. Moreover, for the performances display of NN-KNN hybrid algorithm, different situations were taken into consideration, and some representative ones were simulated, which include comparison of NN, KNN, and NN-KNN hybrid algorithms with two different k values, and comparison of NN-KNN hybrid algorithms on different weightings with two different k values. At last, we also studied the performances of NN-KNN hybrid algorithm with over two k values, which is important for designing parallel algorithms based on cloud computing.

3.1. Experimental environment

Experimental environment is one floor of a building which consists of two rectangles with dimensions of 44m by 32m and 14m by 12m respectively. As shown in figure 3, the grey part is outdoor area and the white part is location area. In the location area represented by the white part, there are some rooms comparted by walls and 9 APs represented by red five-pointed stars. As for the allocation of RPs, we virtualize the location area to grids with side length of 2 m, and RPs are set at cross points of the grids. TPs for simulations are set randomly in the location area.

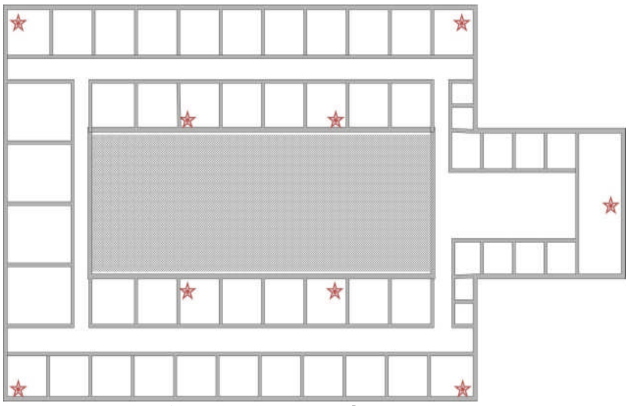

Figure 3. Experimental environment

3.2. Simulation parameters

In the simulation experiment, the RSS of the radio-map and TP are measured according to the Motley-Keenan (MK) propagation model [11], which can reflect the loss of the radio signal propagating in indoor environment accurately, because the signal being decayed by the wall is taken into account. The MK model is described as

$$L_{pico} = 37 + 20 * \log(\lambda) + 3 * N_w, \quad (4)$$

where L_{pico} denotes RSS; λ denotes the distance between transmitting end and receiving end; N_w denotes the numbers of walls being passed through by the radio signal.

The following simulation experiments all take (4) as the propagation model and White Gaussian Noise of 10dB SNR will be added, and 100 TPs are set randomly in the location area. In (1), q is set at the value 2 in accordance with the general standard of fingerprint-based indoor position systems, that is to say, Euclidean distance serves as the distance mode of the simulation.

4. Algorithm simulation results and analysis

The values of k have a great effect on KNN algorithm, so the variations of mean location error with k varying from 1 to 10 were first investigated and the simulation results were shown in figure 4. We can see that the mean location error of KNN algorithm varies with different k values and reaches minimum when k equals 2, so we chose k equals 2, 3, and 5 for the simulation of NN-KNN hybrid algorithm, with which the location errors of KNN algorithm are the minimum three. In the simulations, different weightings for NN-KNN hybrid algorithm were set. For example, “1:1” means that the location results of NN algorithm and KNN algorithm weighted in NN-KNN hybrid algorithm are equal, and analogously “3:7” means that the location results of NN algorithm and KNN algorithm weighted in NN-KNN hybrid algorithm are 30 percent and 70 percent respectively.

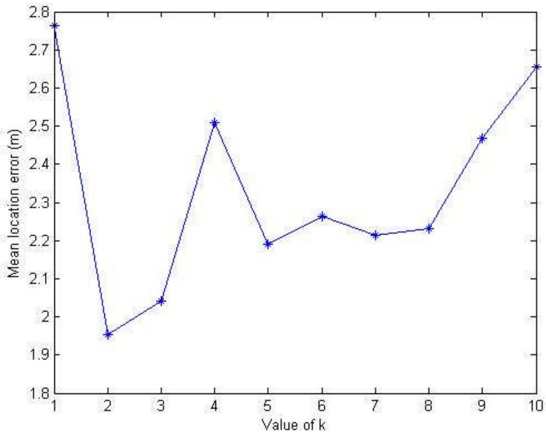

Figure 4. The variation of mean location error with different k

4.1. Simulation design and results

a) First part: Under the weighting of “1:1”, the performances of NN algorithm, KNN algorithm and NN-KNN hybrid algorithm were compared with k equaling 2 and 3 separately, which were shown in figure 5 and figure 6 respectively.

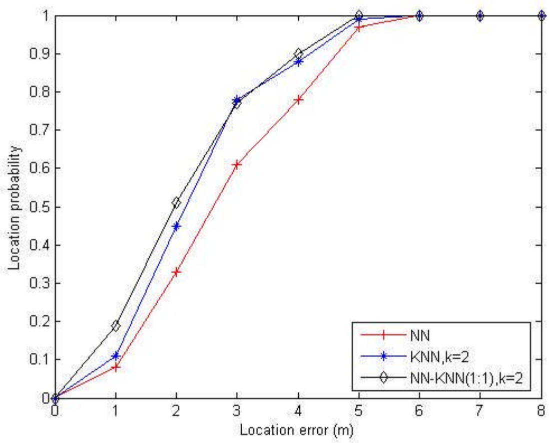

Figure 5. The performance comparison of NN algorithm, KNN algorithm, and NN-KNN hybrid algorithm (k equals 2)

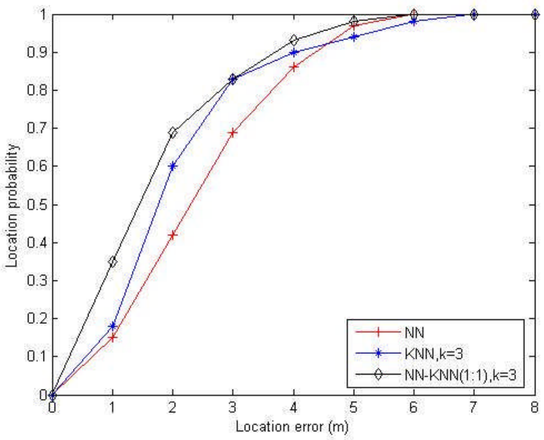

Figure 6. The performance comparison of NN algorithm, KNN algorithm, and NN-KNN hybrid algorithm (k equals 3)

b) Second part: The performances of NN-KNN hybrid algorithm in different weightings were compared with k equaling 2 and 3 separately, which were shown in figure 7 and figure 8 respectively, and the performance comparison of NN-KNN hybrid algorithm with over two values of k was also investigated and simulated with k equaling 2, 3 and 5, which was shown in figure 9.

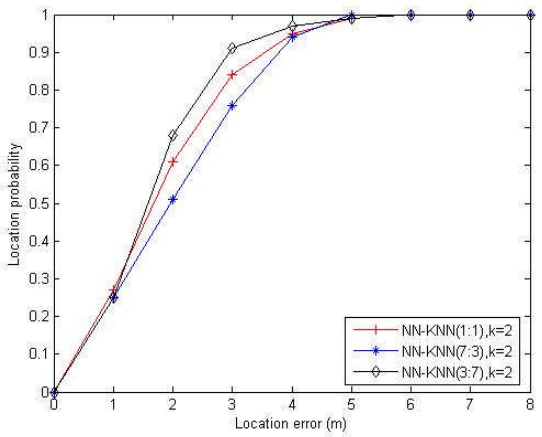

Figure 7. The performance comparison of NN-KNN hybrid algorithm in different weightings (k equals 2)

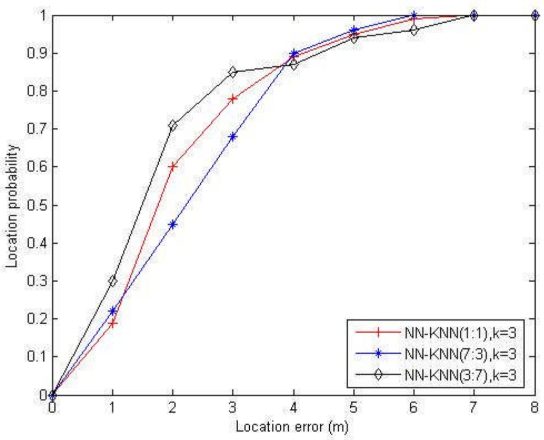

Figure 8. The performance comparison of NN-KNN hybrid algorithm in different weightings (k equals 3)

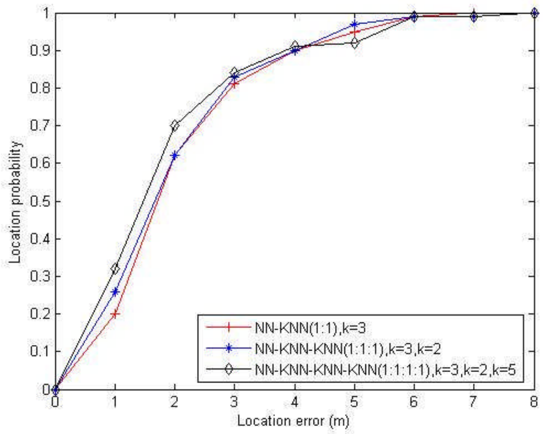

Figure 9. The performance comparison of NN-KNN hybrid algorithm with over two values of k

4.2. Analysis of the simulation results

As shown in figure 5 and figure 6, NN-KNN hybrid algorithm with the weighting of “1:1” outperforms NN algorithm and KNN algorithm when the location error is less than 2m, which is more valuable because the location accuracy within 2m is paid more attention. From figure 7 and figure 8 we can conclude that the performance of NN-KNN hybrid algorithm varies with different weightings and has a better performance when weightings are given more to the algorithm with a better performance. Figure 9 shows that the NN-KNN hybrid algorithm with more algorithms tends to have a better performance.

From the above simulation figures we can see that the NN-KNN hybrid algorithm with k equaling 2 outperforms the one with k equaling 3 and it is accordant to that the KNN algorithm with k equaling 2 outperforms the one with k equaling 3, which indicates that the value of k has an effect on NN-KNN hybrid algorithm and KNN algorithm with the same tendency.

5. Conclusion and future work

In this paper, a fingerprint matching simplified algorithm and a NN-KNN hybrid algorithm based on background cloud computing were presented. The fingerprint matching simplified algorithm was analyzed theoretically and the NN-KNN hybrid algorithm was analyzed with simulation experiments. Simulation results showed that the proposed algorithms could improve the performance of indoor location and the computational complexity was reduced.

In the future, the algorithm will be analyzed in other environments and an adaptive algorithm according to the environment will also be investigated [12-19].

References

[1] M. Sudha, M. Monica, Enhanced Security Framework to Ensure Data Security in Cloud Computing Using Cryptography, Advances in Computer Science and its Applications, Vol. 1, No. 1, (2012) 32-37.

[2] Youssef Bassil, Distributed Cross-Platform and Regression Testing Architecture for Service-Oriented Architecture, Advances in Computer Science and its Applications, Vol. 1, No. 1, (2012) 9-15.

[3] Senthil Murugan. B, Santhosh Kumar. V, A Survey of Workflow Management Tools for Grid Platform, Advances in Information Technology and Management, Vol. 1, No. 1, (2012) 1-3.

[4] A Smith, H Balakrishnan, M Goraczko, Tracking moving devices with the cricket location system, In Proc ACM Mo-biSYS, Massachusetts, USA, 2004, pp. 190-202.

[5] G L Sun, J Chen, W Guo, Signal processing techniques in network-aided positioning: a survey of state-of-the-art positioning designs, Signal Processing Magazine, vol. 22, July 2005, pp. 12-23.

[6] C Fretzagias, M Papadopouli, Cooperative location-sensing for wireless networks, In Proc PerCom'04, Orlando, 2004, pp. 121-131.

[7] H Marko, L Juha, I Hannu, Using calibration in RSSI-based location tracking System, In Proc CSCC, Crete, 2001, pp. 461-465.

[8] Ahmad Hatami, Kaveh Pahlavan, Comparative statistical analysis of indoor positioning using empirical data and indoor radio channel models, IEEE CCNC 2006 proceedings, Las Vegas, 2006, pp. 1018-1022.

[9] Yongliang Sun, Yubin Xu, Lin Ma, KNN-FCM hybrid algorithm for indoor location in WLAN, 2009 2nd International Conference on Power Electronics and Intelligent Transportation System, Shenzhen, 2009, pp. 251-254.

[10] Peigang Sun, Hai Zhao, Dingding Luo, Research on RSSI-based location in smart space, ACTA ELECTRONICA SINICA.China, Vol. 35, July 2007, pp. 1240-1245.

[11] Li Jiang, Research on the key technologies of indoor location based on WIFI, Dalian University of Technology, Dalian, 2010, pp. 15-16.

[12] Yudong Zhang, Lenan Wu, A Robust Hybrid Restarted Simulated Annealing Particle Swarm Optimization Technique, Advances in Computer Science and its Applications, vol.1, no.1, pp.5-8, 2012

[13] Satwik Motamarri, M. Satwik, G. Rajesh Chandra, A MODEL FOR IMPROVING GUILT PROBABILITIES TO IDENTIFY DATA LEAKAGES, Advances in Computer Science and its Applications, vol.1, no.2, pp.131-135, 2012

[14] Pothineni Lavanya, Manoj Kumar Tyagi, K. Poornima, N. Mahesh, Nymble: Intercepts Misbehaving Users, Advances in Computer Science and its Applications, vol.1, no.2, pp.155-163, 2012

[15] Yudong Zhang, Lenan Wu, Rigid Image Registration by PSOSQP Algorithm, Advances in Digital Multimedia, vol.1, no.1, pp.4-8, 2012

[16] Ruisong Ye, Huiliang Li, Self-similar Patterns in Three-dimensional Space Generated by Cellular Automata, Advances in Digital Multimedia, vol.1, no.1, pp.34-36, 2012

[17] Arezoo Mortazavian, Maryam Zekri, Bahram Karimi, Artificial Intelligence based Method for Adaptive Backstepping Control in a Class of Affine Nonlinear Systems, Advances in Digital Multimedia, vol.1, no.1, pp.29-33, 2012

[18] Yudong Zhang, Lenan Wu, Artificial Bee Colony for Two Dimensional Protein Folding, Advances in Electrical Engineering Systems, vol.1, no.1, pp.19-23, 2012

[19] Amir Hammouda, Noureddine Hidouri, A Voltage Drive Scheme for an Isolated Hybrid Diesel-Photovoltaic Water Pumping System based on DC machine, Advances in Electrical Engineering Systems, vol.1, no.2, pp.82-88, 2012