An improved indoor localization algorithm based on SOM and WKNN

Jiuqi Luo, Changgeng Li*
School of Physics and Electronics, Central South University, Changsha, 410012, China

ABSTRACT

This paper focuses on WiFi indoor positioning based on received signal strength, and Weighed K-Nearest Neighbor (WKNN) algorithm is the most classic position estimation strategy. However, locating across the entire fingerprint database takes a lot of time. In this paper, a clustering algorithm based on Self-Organization Map (SOM) is proposed to shorten the positioning time. Meanwhile, an improved WKNN algorithm is proposed to further increase the positioning accuracy. The experiment results show that the positioning time is effectively cut down after clustering and the average positioning error of the proposed algorithm is 1.18m, which can achieve high accuracy in indoor environment.

Keywords: Indoor localization, Wi-Fi fingerprinting, clustering algorithm, WKNN algorithm

1. INTRODUCTION

With the rise of smartphone, Location Based Service (LBS) plays an important role in daily living, such as emergency rescue, personalized information delivery, shopping navigation and logistics industry[1]. Mobile users desire for more adequate personal location information under certain circumstance. For instance, it can be a real-time shopping guide in an entertainment center or shopping mall[2].

The wireless localization systems can be deployed for outdoor and indoor environment based on various key technologies, including satellites, Wi-Fi, RFID, bluetooth and cellular[3]. Mature outdoor real-time positioning application includes GPS, BeiDou, Galileo and GLONASS, which provides location services for users anywhere on the earth[4].

However, compared to maturely developed outdoor localization, research on indoor localization is still in a nascent phase, and several systems widely employed in outdoor environment can't meet the accuracy requirements of indoor localization. Localization system based on Wi-Fi, due to its low cost and convenience, is considered to be the most potential to realize LBS in indoor environment[5]. Wi-Fi-based indoor localization techniques can be divided into two categories: range-based (arrival-based) and fingerprint-based. The basic idea of range-based technology is to obtain and calculate the distance of the transmitter and the receiver, the time of arrival (TOA), the angle of arrival (AOA) and the time difference of arrival (TDOA) are three typical methods[6]. However, these methods require strict time synchronization, which will cost much in equipment configuration. Therefore, on some occasions requiring low-cost and low-complexity, fingerprint-based localization technology is more suitable[7].

However, in the complex indoor environment, there exists multiple access points (APs), and it costs much time to realize localization using fingerprint-based algorithm[8]. To shorten the positioning time and reduce the complexity of whole process, Li et al. proposed an AP selecting strategy based on variation, and it effectively reduces the scale of fingerprint database[9]. While Guo et al. focused on the online positioning stage and proposed improved K-means (K-means++) algorithm[10]. By partitioning the whole positioning zone, the procedure of matching algorithm was simplified. Inspired by these algorithms, a clustering algorithm based on Self-Organization Map (SOM) is proposed. By realizing partition localization, the complexity of online positioning is lowered. Meanwhile, an improved Weighted K-Nearest Neighbor (WKNN) algorithm is proposed to increase the positioning accuracy.

The rest of this paper is organized as follows. Section 2 describes new localization model in detail. Section 3 presents the experiment results and analysis. Finally, Section 4 draws conclusion and outlines future work.

*lcgeng@csu.edu.cn

International Conference on Computer Application and Information Security (ICCAIS 2021),
edited by Yingfa Lu, Changbo Cheng, Proc. of SPIE Vol. 12260, 1226009 · © The Authors.
Published under a Creative Commons Attribution CC-BY 3.0 License · doi: 10.1117/12.2637408

2. LOCALIZATION MODEL

2.1 SOM algorithm

Traditional WKNN algorithm compares the characteristic distance between reference point (RP) and test point (TP) to find optimal nearest RPs[11-12]. However, it usually traverses the whole fingerprint database and costs much time. In order to lower the complexity of positioning, we choose clustering algorithm based on Self-Organization Map (SOM), which provides precise classification and effectively reduce the positioning time. The flow of SOM is depicted as Figure 1. Assuming that a set of Received Signal Strength (RSS) for one access point (AP) is $R = \{r_1, r_2, \cdots, r_m\}$, where m is the number of AP. The training and testing data are firstly normalized by

$$R' = \frac{R - \mathrm{ave}(R)}{\max(R) - \min(R)} \tag{1}$$

where $max(R)$ and $min(R)$ represents the maximum and minimum value of the RSS in R respectively, and $ave(R)$ represents the average of the RSS values in R. Next the node (clustering center) information is initialized. The node number can be set as 2-5 usually, and the initial clustering center can be randomly generated with the area of RPs. On finishing initializing the clustering system, the Euclidean distance D between RP and each clustering center is calculated by

$$D = \left\| R'_{RP} - R'_{C} \right\|^2 \tag{2}$$

where R'_{RP} and R'_C represents the RSS of RP and clustering center. With regard to the nearest center, it is defined as "excited node (EN)", while others are defined as "inhibited node (IN)". Reasonably, the EN are supposed to be more approaching to the RP, while the IN should be more remote to the RP. In other words, the RP belongs to the cluster whose node is excited. Depending on whether the node is excited or inhibited, the RSS of each node is updated, which is described by

$$R'_{C1} = \begin{cases} R'_C + k \cdot \left(R'_{RP} - R'_C\right), \ for\ EN \\ R'_C - k \cdot \left(R'_{RP} - R'_C\right), \ for\ IN \end{cases} \tag{3}$$

where R'_{C1} represents RSS of node after updating and k represents the learning rate. Here, an adaptive learning rate is adopted. It gradually reduces as the iteration lasts, which is described by

$$k = \exp(-0.5t) \tag{4}$$

where t represents the iteration times. Once the iteration meets the end condition, the clustering produces are accomplished. The fingerprint database is divided into several parts, which can be represented by the clustering center of corresponding cluster.

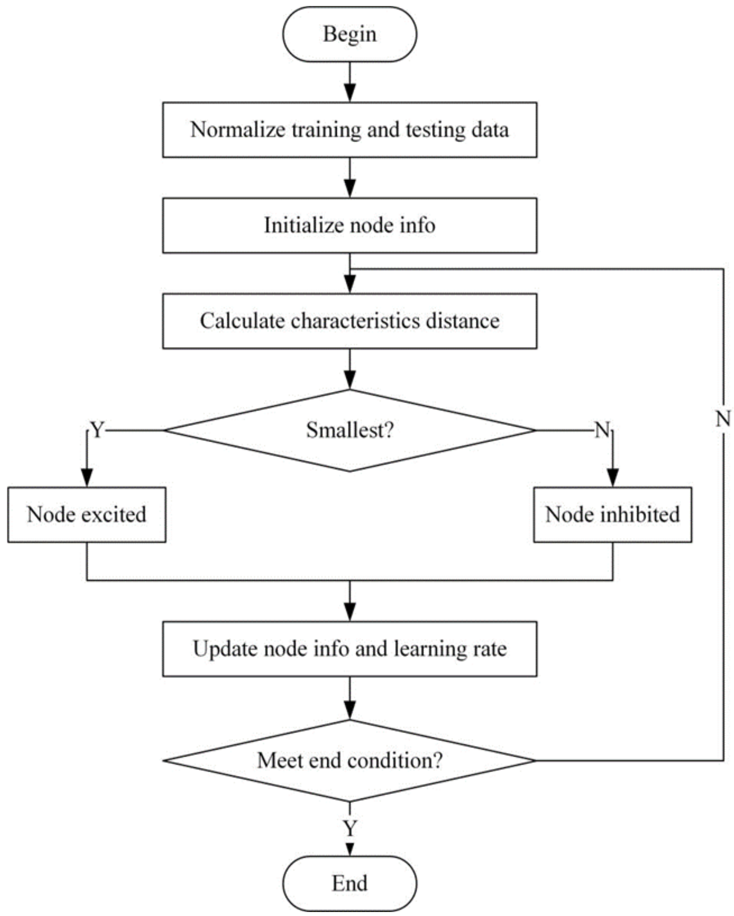

Figure 1. The flow of SOM.

2.2 Improved WKNN algorithm

WKNN algorithm is the most commonly applied algorithm in indoor positioning[13-14]. To further improve the positioning accuracy, an improved WKNN algorithm (AG-WKNN) is proposed. The weight of each RP is calculated and obtained by Gaussian distribution function. The flow of this algorithm is presented in Algorithm 1.

Algorithm 1: AG-WKNN

Input: $R_T = \{r_1, r_2, \cdots, r_m\}$, the RSS sequence of TP
Output: Estimated TP's position (x_T, y_T)

1: Obtain the RSS sequence of TP $R_T = \{r_1, r_2, \cdots, r_m\}$
2: Calculate the distance l_i between R_T and i-th clustering center R'_{Ci} by

$$l_i = \|R_T - R'_{Ci}\|^2$$

3: Classify TP where the distance is smallest
4: Among the cluster: Calculate the distance l_j between R_T and the partial fingerprint database by

$$l_j = \left\|R_T - R'_{pj}\right\|^2$$

5: Obtain the distance array $\varphi = [l_1, l_2, \cdots, l_j]$
6: Sort the array φ (From small to large)
7: Three TPs (x_1, y_1), (x_2, y_2), (x_3, y_3) with the least distance values are selected
8: Calculate the weight w_k by Gaussian distribution function
9: Obtain the TP's position by

$$(x_T, y_T) = w_1 \cdot (x_1, y_1) + w_2 \cdot (x_2, y_2) + w_3 \cdot (x_3, y_3)$$

10: **Return** TP's position (x_T, y_T)

Specifically, the Gaussian distribution function can be described by

$$f_k = a \cdot \exp\left\{-\frac{(l_k - b)^2}{2c^2}\right\} \tag{5}$$

where a, b, c can be set according to different systems. w_k is the normalized f_k, which is depicted by

$$w_k = \frac{f_k}{\sum f_k} \tag{6}$$

On this point, the estimated position of TPs can be obtained and the process of positioning is accomplished.

3. THE EXPERIMENT RESULTS AND ANALYSIS

In order to verify the performance of proposed scheme, experiments were conducted on the sixth floor of Physics Building in Central South University, Changsha city, Hunan province, China. The layout of the environment is shown in Figure 2.

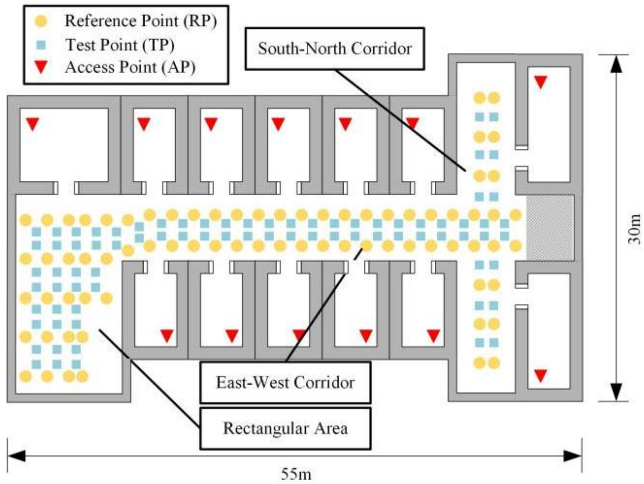

Figure 2. Layout of the environment.

The environment includes a rectangular area, an east-west corridor and a south-north corridor which cover 15m×9m, 55m×2m and 30m×2m respectively. 73 RPs are 2.4m far from their neighbor and are shown in yellow. 215 TPs are chosen randomly in the area and are shown in blue.

In clustering algorithm, the value of k (the number of node) has great influence on region division. The smaller the number of clusters, the smaller the differentiation will be. However, if the number of clusters is too large, the complexity of subsequent localization will increase. In order to find the optimal k, the experimental range is set as 2-5. Figure 3 shows the clustering results of 73 RPs for this environment, with different categories represented in different colors.

As shown in Figure 2, when $k = 2$ or $k = 3$, there are hardly any distinctions in the region of east-west corridor. When $k = 5$, the number of clusters is a bit of large, which may increase the complexity of positioning. Therefore, k is set to 4 in this paper.

In order to verify the performance of the proposed SOM algorithm, traditional WKNN algorithm is firstly adopted. The localization results are shown in Table 1.

Table 1 shows the localization results improve after clustering. Specifically, the mean error has an improvement of 5.4% and the time of positioning reduce by 45.7%, which indicates that the proposed

SOM algorithm can effectively lower the positioning error and shorten the positioning time.

Table 1. Performance of SOM algorithm.

	TP number	**Error (m)**		**Time (s)**
		Mean	**Max**	
Cluster 1	134	0.98	6.17	0.0021
Cluster 2	14	0.67	2.18	0.000399
Cluster 3	13	0.80	2.16	0.000371
Cluster 4	54	2.09	7.59	0.000953
All	215	1.23	7.59	0.003823
Before clustering	215	1.30	8.36	0.00704

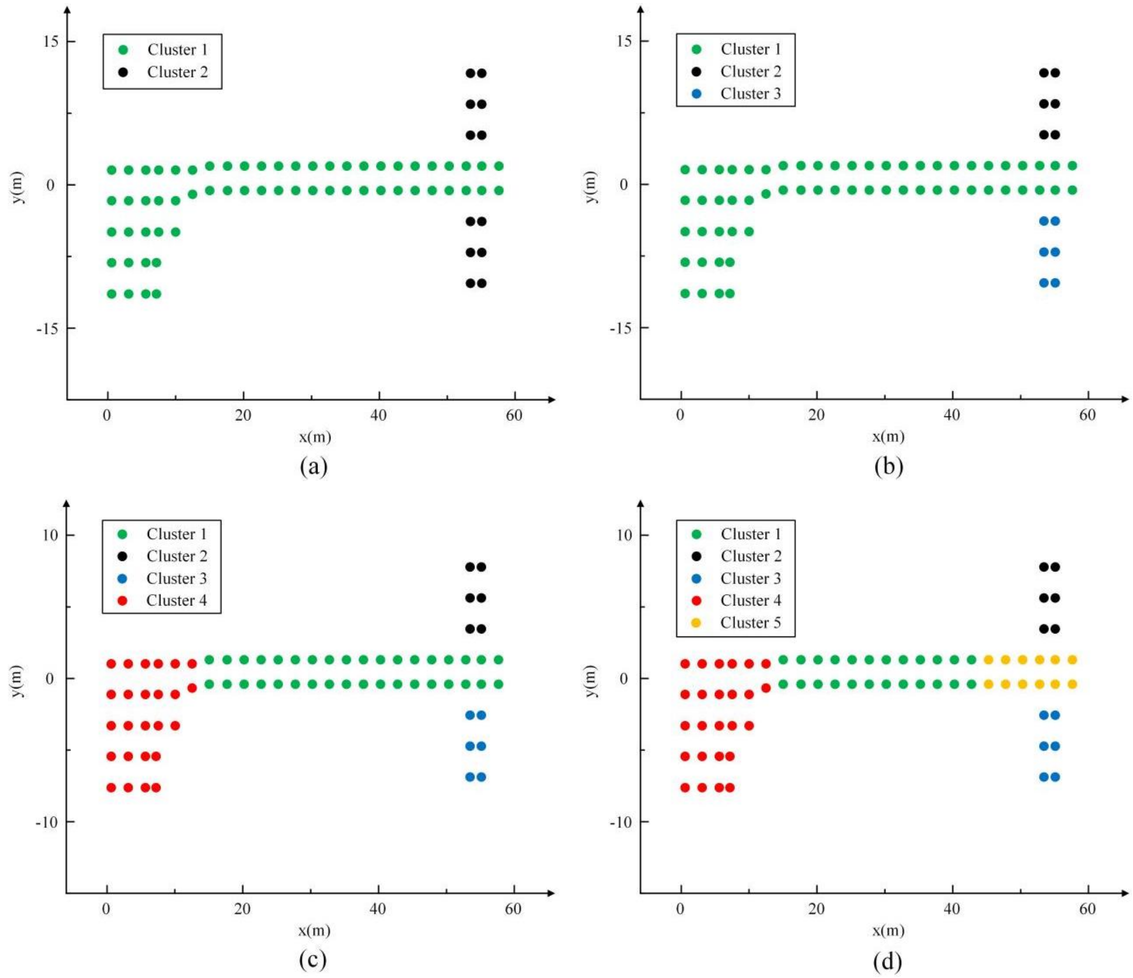

Figure 3. Clustering results under different k: (a) $k = 2$, (b) $k = 3$, (c) $k = 4$, (d) $k = 5$.

To further improve the accuracy, AG-WKNN is conducted in the environment. In equation (5), a and c are set to 1 in this paper. The value of b determines the weight w directly, and impacts on positioning accuracy. l_k represents the $k-th$ distance between RP and TP, b can be set near the first distance l_1, which is depicted by

$$b = n \cdot l_1 \tag{7}$$

where n is the coefficient. In order to find the optimal n, different values are tested, and the results are shown in Figure 4. Figure 4 shows that the positioning error goes up first, and declines later on as n increases. Specially, the positioning error minimizes as n equals to 1.03, and the minimum value goes to 1.18m.

Table 2 lists the performance using the proposed algorithm. The mean error and max error both reduce. Furthermore, the mean error has a reduction of 5.4% and 9.2%. Figure 5 shows the cumulative

distribution errors using proposed algorithm. The positioning accuracy improves when using SOM and AG-WKNN algorithm.

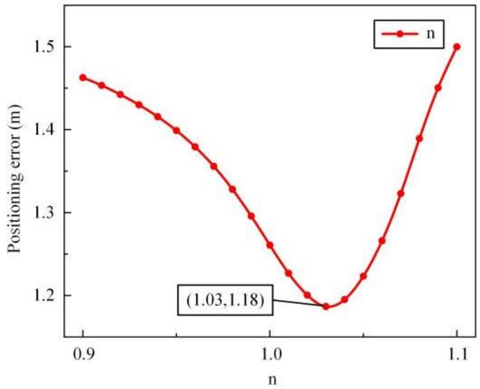

Figure 4. The relationship between n and positioning error.

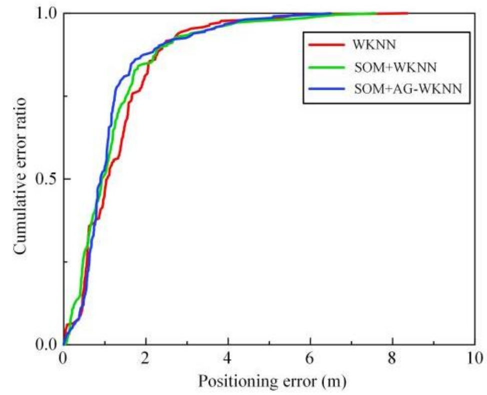

Figure 5. Cumulative distribution errors of proposed algorithm.

Table 2. The performance using the proposed algorithm.

Algorithm	Mean error (m)	Max error (m)
WKNN	1.30	8.36
SOM+WKNN	1.23	7.59
SOM+AG-WKNN	1.18	5.97

To make our results more convincing, the proposed algorithm is compared with existed algorithms. Zhong et al. proposed K-means algorithm based on neighborhood density[15]. Guo et al. proposed a hybrid algorithm, in which used K-means++ algorithm[10]. The performance of these algorithms is shown in Table 3.

Table 3. The comparison of proposed and existed algorithms.

Algorithm proposed by	Mean error (m)	Clustering time (s)
Zhong et al.	1.81	0.084822
Guo et al.	1.70	0.082227
This paper	1.18	0.081859

Table 3 shows the clustering time of three algorithms is nearly the same, but the mean error of the proposed algorithm outperforms the other two algorithms, with improvement of 34.8% and 30.6%. In conclusion, the results indicate that the proposed algorithm performs well and it will have a good prospect in real positioning scenarios.

4. CONCLUSION

In this paper, clustering algorithm based on SOM and improved WKNN algorithm, called AG-WKNN, are proposed to increase the positioning accuracy and cut down the positioning time. The experiment results show that the positioning error and time both decline after using the proposed algorithm, and the experiments also show that the proposed algorithm has higher accuracy than existed algorithm. Better clustering algorithm is still desirable to enhance the indoor localization system and this is the future work.

REFERENCES

[1] Zhang, H., Liu, K., Jin, F., Feng, L., Lee, V. and Ng, J., "A scalable indoor localization algorithm based on distance fitting and fingerprint mapping in Wi-Fi environments," Neural Computing and Applications, 32, 5131-5145 (2019).

[2] Sotenga, P. Z., Djouani, K., Kurien, A. M. and Mwila, M., “Implementation of an indoor localisation algorithm for Internet of Things,” Future Generation Computer Systems, 107, 1037-1046 (2020).

[3] Yang, H., Zhang, Y., Huang, Y., Fu, H. and Wang, Z., “WKNN indoor location algorithm based on zone partition by spatial features and restriction of former location,” Pervasive and Mobile Computing, 60, 101085 (2019).

[4] Li, X. and Bharanidharan, M., “RSSI fingerprinting based iphone indoor localization system without Apple API,” Wireless Personal Communications, 112, 61-74 (2019).

[5] Zhao, M., Qin, D., Guo, R. and Xu, G., “Research on Crowdsourcing network indoor localization based on co-forest and Bayesian compressed sensing,” Ad Hoc Networks, 105, 102176 (2020).

[6] Zhang, C., Qin, N., Xue, Y. and Yang, L., “Received signal strength-based indoor localization using hierarchical classification,” Sensors (Basel), 20, 102176, (2020).

[7] Gan, H., Khir, M. H. B. M., Djaswadi, G. W. B. and Ramli, N., “A hybrid model based on constraint OSELM, adaptive weighted SRC and KNN for large-scale indoor localization,” IEEE Access, 7, 6971-6989 (2019).

[8] Yu, Y., Chen, R., Liu, Z., Guo, G., Ye, F. and Chen, L., “Wi-Fi fine time measurement: Data analysis and processing for indoor localization,” Journal of Navigation, 73, 1106-1128 (2020).

[9] Li, C., Huang, H. and Liao, B., “An improved fingerprint algorithm with access point selection and reference point selection strategies for indoor positioning,” Journal of Navigation, 73, 1182-1201 (2020).

[10] Guo, T., Chai, M., Xiao, J. and Li, C., “A hybrid indoor positioning algorithm for cellular and Wi-Fi networks,” Arabian Journal for Science and Engineering, (2021) <https://doi.org/10.1007/s13369-021-05925-9>

[11] Hoang, M. T., Yuen, B., Dong, X., Lu, T., Westendorp, R. and Reddy, K., “Recurrent neural networks for accurate RSSI indoor localization,” IEEE Internet of Things Journal, 6, 10639-10651 (2019).

[12] Zafari, F., Gkelias, A. and Leung, K. K., “A Survey of indoor localization systems and technologies,” IEEE Communications Surveys and Tutorials, 21, 2568-2599 (2019).

[13] Zekavat, S., Buehrer, R. M., Durgin, G. D., Lovisolo, L., Wang, Z., Goh, S. T. and Ghasemi, A., “An overview on position location: Past, present, future,” International Journal of Wireless Information Networks, 28, 45-76 (2021).

[14] Ninh, D. B., He, J., Trung, V. T. and Huy, D. P., “An effective random statistical method for indoor positioning system using WiFi fingerprinting,” Future Generation Computer Systems, 109, 238-248 (2020).

[15] Zhong, Y., Wu, F., Zhang, J. and Dong, B., “WiFi Indoor localization based on k-means,” Inter. Conf. on Audio, Language and Image Processing, 663-667 (2016).