



Manifold Alignment-Based Radio Map Construction in Indoor Localization

Ping Ji, Danyang Qin^(✉), Pan Feng, and Yan Zhang

Key Laboratory of Electronic and Communication Engineering,
Heilongjiang University, No. 74 Xuefu Road,
Harbin, People's Republic of China
qindanyang@hlju.edu.cn

Abstract. In recent years, Wireless Access Point (WAP)-based Received Signal Strength Indication (RSSI) indoor localization technology has been of intriguing interest to deduce the coordinates of an object or an observer in the scene with RSSI being collected by various WAPs in a Range of Interest (ROI). The Radio Map construction by fingerprints is of great importance for indoor localization. Existing methods of Radio Map construction have encountered bottlenecks in this area, which will limit the application of indoor localization technology due to the deployment is massive and cumbersome. The spatial correlation between RSSI observations is adopted and the manifold alignment algorithm will be adopted to locate the user's current location without a complete Radio Map so as to reduce the requirements of the calibration fingerprints. Simulated Radio Map (SRM) scheme and Plan Coordinate (PC) scheme will be proposed and simulated separately to verify the correctness and efficiency of the proposed scheme.

Keywords: Indoor localization · Radio map construction · Spatial correlation
Manifold alignment

1 Introduction

Where is the nearest metro station? Is this product I want to buy in the supermarket? With the advent of mobile Internet, there will be more expectations for localization services. For airports, libraries, shopping malls and parking lots, localization services are expected to be more accurate and faster in such scenes.

RSSI-based localization technology can be divided into two main stages. The first stage is offline acquisition stage. It is necessary to identify a few landmarks (LM) inside the building and collect the location fingerprints of each LM in the building by the equipment. Next, the spatial coordinate data of the LM should be correlated to establish a localization fingerprint database, which is to be called Radio Map. The second stage is online localization stage. One user enters the localization area holding a mobile terminal to measure the RSSI data through the receiver equipped with WiFi. The RSSI data will be compared with the data in the manifold alignment algorithm to obtain the coordinates of the user according to the determination result.

The deployment of WAP is time consuming and labor intensive when put into use in larger environments such as airports and mega-stores although the Radio Map, built by fingerprinting, covers a very comprehensive set of information. Therefore, a solution that can reduce the deployment cost and workload is the key to solve the problem. The scheme proposed in this paper greatly reduces the deployment cost and users can be located without build complete Radio Map for the indoor environment. The program has taken advantage of the inherent spatial correlation reflected by the RSSI data to reduce the number of calibration fingerprints. We recall that there is a high spatial correlation between neighboring positions in the Radio Map. If a DS to reflect the mapping between each position in the target region can be obtained, the approximate location of the RSSI data source can be determined. Moreover, manifold alignment algorithm [1] will locate the collected RSSI data in real time.

The rest of the paper will be organized as follows. The current state of research will be described in Sect. 2. The basics of the manifold alignment algorithm will be introduced in Sect. 3. The SRM scheme and the PC scheme will be introduced respectively in Sect. 4. Section 5 will discuss performance evaluation and the full text will be summarized in Sect. 6.

2 Motivation

There are few ways to combine the Radio Map establishment and localization methods, the typical ones are multi-sensor Particle Filtering (PF) based on intelligent platform [2] and WLAN indoor localization based on Compressive Sensing (CS) [3]. The former achieved the Radio Map construction by a variety of smart phones sensor devices, simultaneously, achieved localization by PF. It is found that, however, this is a probabilistic localization method, which has complicated localization process and poor localization accuracy. The latter achieved localization by CS, which included Radio Map reconstruction under the condition of sparse Radio Map.

The unsupervised Radio Map is one of the algorithms to increase the efficiency of the Radio Map construction. Kim et al. proposed an unsupervised Radio Map construction algorithm based on the improved RSSI indoor propagation model in [4]. The measurement of RSSI adopted Client-Assistant (CA) proposed in [5]. There was high efficiency in this measurement. Nevertheless, the problems caused by a variety of different devices to collect RSSI data are inconsistent, especially due to the different sensitivity of the hardware.

In order to reduce the deployment workload and ensure the accuracy of Radio Map, the semi-supervised Radio Map construction algorithm has been proposed. Considering the deployment workload, the literatures started with reducing the number of LMs and RSSI resampled per LM. The difference is that [6] adopted Kernel Based Interpolation method to reconstruct Radio Map, and [7] adopted Linear Interpolation method to reconstruct Radio Map. However, the accuracy of the Radio Map obtained by the interpolation method was low. The other paper [8] adopted Label Propagation to learn the mapping between RSSI and unlabeled RSSI to implement Radio Map reconstruction. Although all the above algorithms can reconstruct the Radio Map, these algorithms were not used for indoor localization.

3 Manifold Alignment

The application of manifold alignment learning method should satisfy two conditions. First of all, it is required that the correlation between each pair of adjacent data points in both DSs, that is the neighboring correlation, will be strong. Even if the two DSs show different distributions or shapes in high-dimensional space, a common low-dimensional correlation between them should be ensured.

3.1 Neighborhood Weight

Manifold alignment requires two DSs to preserve neighborhood correlation in low-dimensional space. Accordingly, we select local linear embedding techniques [9].

Any high-dimensional space data point $z^{(i)}$ with N adjacent data points will be described as a neighborhood DS, notes as $\mathcal{N}(i) = [z^{(\mathcal{N}(i,1))}, \dots, z^{(\mathcal{N}(i,N))}]$. And the neighborhood weight of $z^{(i)}$ is $W_z^{(i)}$ satisfying Eq. (1).

$$W_z^{(i)} = \arg \min_{W_{ij}} \left\{ \left| z^{(i)} - \sum_{j \in \mathcal{N}(i)} W_{ij} z^{(\mathcal{N}(i,j))} \right|^2 \right\} \tag{1}$$

s.t. $\sum_{j \in \mathcal{N}(i)} W_{ij} = 1$

It is easy to understand that the closer $z^{(\mathcal{N}(i,j))}$ is to $z^{(i)}$, the higher its weight W_{ij} will be. For the data point j not belonging to $\mathcal{N}(i)$, there will be $W_{ij} = 0$. The computing method of the neighborhood weight by the closed form solution [10] can be optimized. For any data point i , its distance matrix D_i will be:

$$D_i = \left[z^{(i)} - z^{(\mathcal{N}(i,1))}, z^{(i)} - z^{(\mathcal{N}(i,2))}, \dots, z^{(i)} - z^{(\mathcal{N}(i,N))} \right]^T \tag{2}$$

Therefore, the optimized formula for W_{ij} is:

$$W_{ij} = \frac{\sum_{k=1}^N \left\{ (D_i D_i^T)^{-1} \right\}_{jk}}{\sum_{m=1}^N \sum_{n=1}^N \left\{ (D_i D_i^T)^{-1} \right\}_{mn}} \tag{3}$$

where $\{(D_i D_i^T)^{-1}\}$ denotes the element of the inverse matrix $(D_i D_i^T)^{-1}$ at the intersection of the u th row and the v th column.

3.2 Manifold Alignment Structure

The source DS \mathcal{X} consists of the X points in \mathbb{R}^h , the target DS \mathcal{Y} consists of the Y points in \mathbb{R}^h , and the manifold alignment F between the two DSs is expressed as:

$$F = \arg \min_{\mathbf{f}, \mathbf{g}} \{ \lambda^x \mathbf{f}^T L^x \mathbf{f} + \lambda^y \mathbf{g}^T L^y \mathbf{g} + \mu (\mathbf{f} - \mathbf{g})^T (\mathbf{f} - \mathbf{g}) \} \quad (4)$$

$\mathbf{f} = [f_1, \dots, f_x]^T$ and $\mathbf{g} = [g_1, \dots, g_y]^T$ are two vectors belonging to \mathbb{R}^x and \mathbb{R}^y , respectively, and \mathcal{P} is a set of point of pairs in \mathcal{X} and \mathcal{Y} . The pairing points are points where the source and target DSs are exactly or similarly falling in low-dimensional space. λ^x , λ^y and μ are the weight factors of different components in Eq. (4). The first term takes the minimum value to ensure that the $f_i - f_j$ is smaller, when the weight W_{ij} is higher, which can preserve the neighborhood relation in \mathcal{X} . And it is the same with the second one. The last term is the treatment of the difference between the pairing points of \mathbf{f} and \mathbf{g} . Then Eq. (4) can be rewritten as:

$$F = \arg \min_{\mathbf{f}, \mathbf{g}} \{ \lambda^x \mathbf{f}^T L^x \mathbf{f} + \lambda^y \mathbf{g}^T L^y \mathbf{g} + \mu (\mathbf{f} - \mathbf{g})^T (\mathbf{f} - \mathbf{g}) \} \quad (5)$$

where $L^x = [L^x_{i,j}] \forall i, j \in \mathcal{X}$ and satisfies:

$$L^x_{i,j} = \begin{cases} \sum_j W_{ij}^x, & i = j \\ -W_{ij}^x, & j \in \mathcal{N}_i \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

In addition, since $L^y = [L^y_{i,j}] \forall i, j \in \mathcal{Y}$, in order to obtain $L^y_{i,j}$, all of W_{ij}^x in Eq. (6) will be replaced by W_{ij}^y . Equation (5) needs to be subjected to a strict constraint of $f_i = g_i \forall i \in P$ (i.e. as $\mu \rightarrow \infty$), and it is defined as $Q^x = \mathcal{X} \setminus \mathcal{P}$ and $Q^y = \mathcal{Y} \setminus \mathcal{P}$. Then the problem in Eq. (5) will be translated into solving the following eigenvalues:

$$F = \arg \min_{\mathbf{h}} \left\{ \frac{\mathbf{h}^T L^z \mathbf{h}}{\mathbf{h}^T \mathbf{h}} \right\} \quad (7)$$

s.t. $\mathbf{h}^T \mathbf{1} = 0$

where \mathbf{h} and L^z will satisfy:

$$\mathbf{h} = \begin{bmatrix} \mathbf{f}_{\mathcal{P}} = \mathbf{g}_{\mathcal{P}} \\ \mathbf{f}_{Q^x} \\ \mathbf{g}_{Q^y} \end{bmatrix} \quad (8)$$

$$L^z = \begin{bmatrix} \lambda^x L^x_{\mathcal{P}\mathcal{P}} + \lambda^y L^y_{\mathcal{P}\mathcal{P}} & \lambda^x L^x_{\mathcal{P}Q^x} & \lambda^y L^y_{\mathcal{P}Q^y} \\ \lambda^x L^x_{Q^x\mathcal{P}} & \lambda^x L^x_{Q^xQ^x} & 0 \\ \lambda^y L^y_{Q^y\mathcal{P}} & 0 & \lambda^y L^y_{Q^yQ^y} \end{bmatrix} \quad (9)$$

$L^x_{I,J}$ ($L^y_{I,J}$) is a submatrix of L^x (L^y) consisting of the intersection of the row of element indices in vector I and the column of element indices in vector J . According to the structure of L^z , the structure of \mathbf{h} begins with the aligned element \mathcal{P} in \mathbf{f} and \mathbf{g} , then follows with the rest data points in \mathbf{f} , and finally ends up with the rest data points in \mathbf{g} .

Now, the final structure of embedded ε as shown in Eq. (10), because of the DSs need to be embedded in l -dimensional ($l < h$), which consists of the eigenvector $[\mathbf{h}^{(1)}, \dots, \mathbf{h}^{(l)}]$.

$$\varepsilon = \begin{bmatrix} \mathbf{f}_{\mathcal{P}}^{(1)} & \mathbf{f}_{\mathcal{P}}^{(2)} & \dots & \mathbf{f}_{\mathcal{P}}^{(l)} \\ \mathbf{f}_{\mathcal{Q}^x}^{(1)} & \mathbf{f}_{\mathcal{Q}^x}^{(2)} & \dots & \mathbf{f}_{\mathcal{Q}^x}^{(l)} \\ \mathbf{g}_{\mathcal{Q}^y}^{(1)} & \mathbf{g}_{\mathcal{Q}^y}^{(2)} & \dots & \mathbf{g}_{\mathcal{Q}^y}^{(l)} \end{bmatrix} \quad (10)$$

4 SRM Scheme and PC Scheme

SMP and PC schemes will be introduced in detail in this section, and the spatial correlation between RSSI values will be moved to calibration fingerprints with a limited number and online RSSI observations to enable user localization.

4.1 SRM Offline Stage

The offline deployment stage requires the following operations:

- (1) The indoor environment information is input, such as the CAD file of the building, the location and height of WAPs, and RF simulator Volcano Lab can generate the simulated Radio Map $\mathcal{S} = [(s^{(1)}, p^{(1)}), \dots, (s^{(S)}, p^{(S)})]$, which contains all the grid points divided in the area.
 - $s^{(i)} = [s_1^{(i)}, \dots, s_K^{(i)}]$ is the simulated RSSI vector for K WAPs at the i th grid point in the area.
 - $p^{(i)} = [x^{(i)}, y^{(i)}]$ is the plane coordinate of the i th grid point.
- (2) The SRM is regarded as a source DS to figure W_{ij}^x and L^x of LLE by Eq. (3) and Eq. (6) respectively. The complexity of this step is $O(S^3)$.
- (3) The subset $\mathcal{C} = [(c^{(1)}, p_c^{(1)}), \dots, (c^{(C)}, p_c^{(C)})]$ of \mathcal{S} is taken as a calibration measure, and all grid points in \mathcal{C} are the paired data points in manifold alignment.
 - $c^{(i)} = [c_1^{(i)}, \dots, c_K^{(i)}]$ is the simulated RSSI vector of K WAPs at the i th grid point in the area.
 - $p_c^{(i)} = [x_c^{(i)}, \dots, y_c^{(i)}]$ is the plane coordinate at the i th grid point.

4.2 SRM Online Stage

During the online positioning stage, the following actions will be performed by the positioning server:

- (1) The server will be received O online RSSI values $\mathcal{O} = [o^{(1)}, \dots, o^{(O)}]$ sent by the user with O localization requests.

(2) The following sets will be defined:

- The data points in set \mathcal{P} are the data point of S that is paired with C .
- The data points in set \mathcal{Q}^x are the remaining parts of S that cannot be paired with C .
- The data points in set \mathcal{Q}^y are the online RSSI observations in \mathcal{Y} .
- $\mathcal{X} = [U_{i \in \mathcal{P}} s^{(i)} | U_{j \in \mathcal{Q}^x} s^{(j)}]$ is the DS formed by rearranging the SRM vectors, the accuracy of the RSSI data will be improved at the pairing position, which is called the *source DS*.
- $\mathcal{Y} = [c^{(1)}, \dots, c^{(C)} | o^{(1)}, \dots, o^{(O)}]$ is a new DS formed by the offline calibration fingerprint vector and the online observation RSSI vector connection, which is called the *target DS*.

The formation of the *source DS* and the *target DS* are shown in Fig. 1.

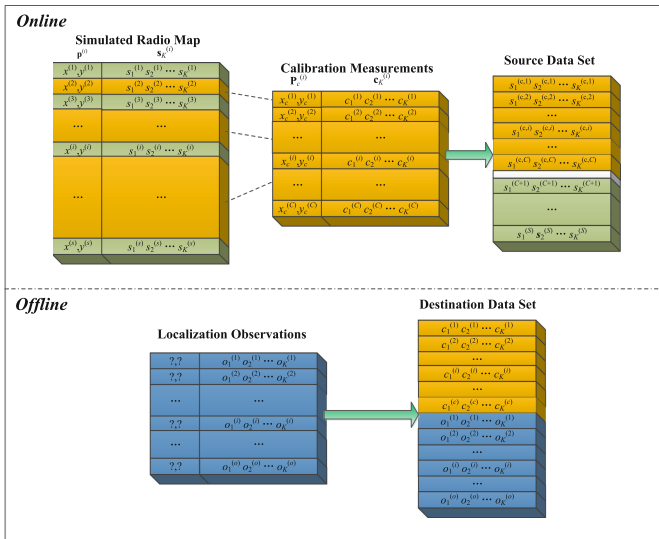


Fig. 1. The formation of the *source DS* and the *target DS* in SRM

(3) W_{ij}^y and L^y of the *target DS* are figured by Eqs. (3) and (6), separately. The complexity of this step is $O((C + O)^3)$.

(4) L^z can be calculated by Eq. (9), and there is:

$$\lambda^x = \frac{C + O}{S + C + O}, \quad \lambda^y = \frac{S}{S + C + O} \quad (11)$$

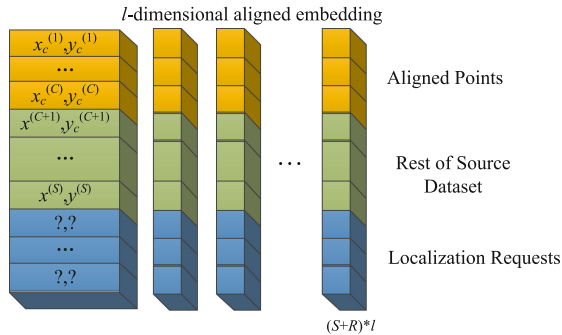


Fig. 2. Alignment embedded structure

(5) The eigenvalues of L^z will be calculated and the low dimensional embedding matrix ε can be constructed. The alignment embedded structure is showed in Fig. 2.

- The first C lines $\varepsilon_{\mathcal{P}} = [\mathbf{f}_{\mathcal{P}}^{(1)}, \dots, \mathbf{f}_{\mathcal{P}}^{(l)}]$ correspond to the calibration fingerprint C .
- The next $(S - C)$ lines $\varepsilon_{Q^x} = [\mathbf{f}_{Q^x}^{(1)}, \dots, \mathbf{f}_{Q^x}^{(l)}]$ are the rest of S that cannot be paired with C .
- The last O line $\varepsilon_{\mathcal{O}} = [\mathbf{g}_{Q^y}^{(1)}, \dots, \mathbf{g}_{Q^y}^{(l)}]$ indicates the online observations.

(6) The distance will be calculated from each line in $\varepsilon_{\mathcal{O}}$ to all the lines in $\varepsilon_{\mathcal{O}}$ and ε_{Q^x} , the closest line will be find. Attach its plan coordinate to the observation, then sent to the user. The iterations will be $O((C + O)^3 + (S + O)^3)$.

4.3 PC Scheme

It is difficult to get the location and height of WAPs in large scale indoor scenes. Less environment detail will be required if plane coordinates serve as the source DS. The scheme will project the WAPs in space onto the ground to generate the plane coordinate DS as well as the spatial correlation between adjacent data points will not be changed. Plane coordinate DSs still reflect the neighborhood relationships between data points in the *source DS*, although they cannot reflect the spread of RF in space. The plane coordinate DS $\mathcal{S} = [p^{(1)}, \dots, p^{(S)}]$ has been determined during the offline deployment stage. The other steps are the same as the SRM.

- In the *source DS*, the first $1-K$ elements of $\hat{p}^{(i)}$ are the same as $p^{(i)}$, and the k th element becomes x/y .
- The set $\mathcal{X} = [U_{i \in \mathcal{P}} \hat{p}^{(i)} | U_{j \in \mathcal{Q}^y} \hat{p}^{(j)}]$ will be modified.

The changed *source DS* and *target DS* are shown in Fig. 3.

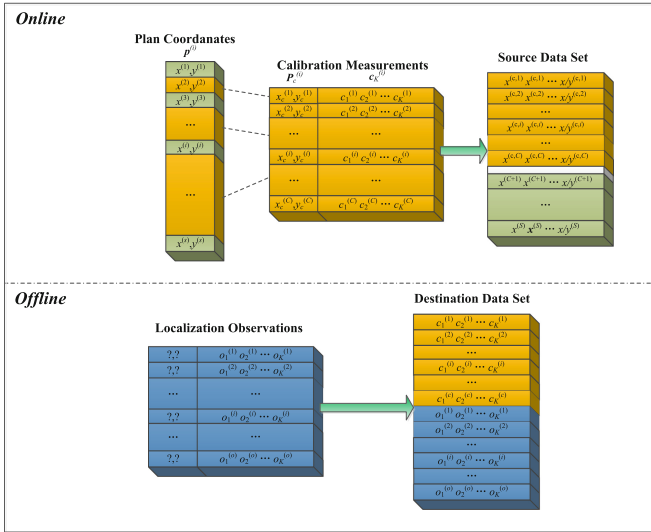


Fig. 3. The formation of the *source DS* and the *target DS* in SRM

5 Performance Analysis

Floor-7 of Physical Laboratory Building in Heilongjiang University is taken as the indoor localization test environment. As shown in Fig. 4, five WAPs are deployed along the corridor. The testing area will be divided into 219 mesh points with 1 m spacing between adjacent mesh points. In this section, the neighborhood set size, localization observations number and WAPs number will be performed based on MATLAB. The simulations reveal that the superior performance of PC to SRM.

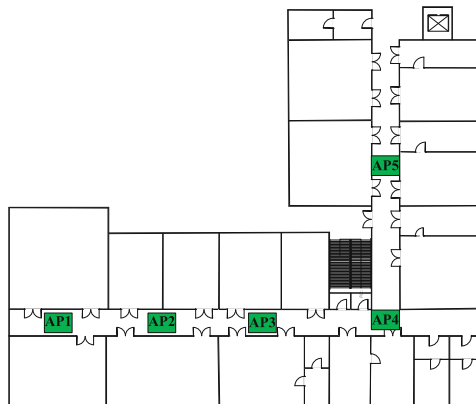


Fig. 4. Indoor structure plan and WAP deployment

5.1 Influence of Neighborhood Set Size

As shown in Fig. 5, the number of neighbors affects the performance of SRM and PC in some degree. The percentages of localization observations and calibrated fingerprints are 5% and 25%, respectively. From Fig. 5, we can find that the number of neighbors is an important factor for the localization accuracy. It is clear that when the number of neighbors is from 10 to 15 (a small number), the localization error level is greater than it is from 20 to 25 (a large number). When there are fewer neighbors, the outlier value of neighborhood has a greater influence on the result of weight. However, the localization error rises again if the neighbors number reaches 35–50 (a very large number), which indicates that each point owns many adjacent points as the number of neighbors is too large. At this point, the localization accuracy will be reduced. The concept of neighborhood dilution [11] can be used to explain this issue.

5.2 Influence of Localization Observations Number

The effect of the localization observations number on the average localization error performance is depicted in Fig. 6 when there are 25 neighbors and the calibration fingerprints number is 25%. The localization errors of the two schemes decrease slightly as the localization observations number increases. It is found that the average localization error is less than 3.5 m as the localization observations number increases from 1 to 50. For most indoor localization applications, the localization observations number does not have a significant impact on the localization accuracy.

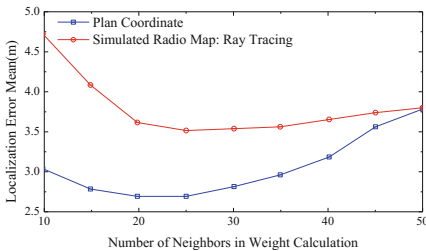


Fig. 5. Mean positioning error against the number of neighbors

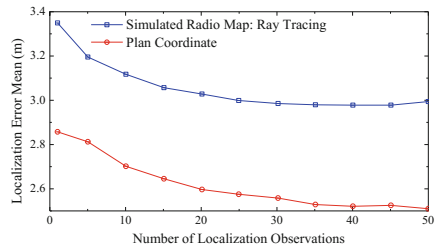


Fig. 6. Mean positioning error against the percentage of localization observations

5.3 Influence of the Number of WAPs

The effect of the WAPs number on average localization error under different percentages of calibration fingerprints is depicted in Fig. 7 when the localization observations number is 11 and the number of neighbors is 20. SRM and PC refer to simulated radio map and plan coordinate respectively. The percent in parentheses is the

calibration load. As expected, both algorithms will achieve high localization accuracy as the number of WAPs within a certain range of the number of WAPs is growing. In fact, the performance of all indoor localization scheme based on RSSI will be better as the WAPs number increases. The localization errors of the PC scheme are 5.7 m and 3.8 m, respectively, when the calibration load is 20% and the WAPs number is 3 and 4. It can be observed that, for any calibration load, the localization error of the SMP scheme is higher than PC scheme, which shows that the superior performance of PC to SRM.

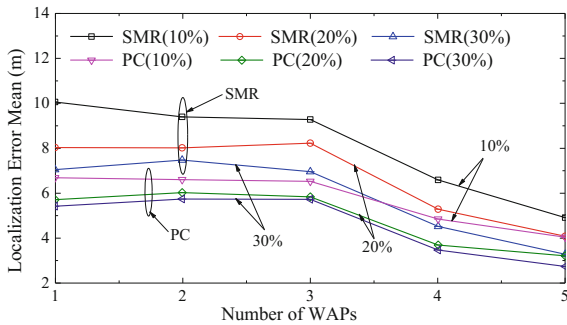


Fig. 7. Mean positioning error performance against the number of WAPs

6 Conclusion

This paper mainly solves the problem of Radio Map construction which will determine the accuracy and efficiency of the indoor localization. The RSSI values are collected by WAPs and a limited calibration load is assumed for practical deployment. The scheme proposed in this paper will reduce the number of calibration fingerprints by the spatial correlation between RSSI observations in the same indoor area. And the user's current location is located by the popular alignment algorithm without a complete Radio Map. Two schemes, SRM scheme and PC scheme, are proposed and simulated. The simulation results show that localization error can be limited from 0.4 to 0.8 compared with the complete Radio Map when the percentage of calibration fingerprints is selected from 70% to 80%. In the future, the algorithms will be studied by unsupervised manifold alignment so that reduce Calibration work of SRM and PC.

Acknowledgement. This work was supported by the National Natural Science Foundation of China (61771186), Postdoctoral Research Project of Heilongjiang Province (LBH-Q15121), University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province (UNPYSCT-2017125).

References

1. Pei, Y., Kim, T.K., Zha, H.C.: Unsupervised random forest manifold alignment for lipreading. In: IEEE International Conference on Computer Vision, pp. 129–136. IEEE Computer Society, Sydney (2013)
2. Cappello, F., Sabatini, R., Ramasamy, S.: Particle filter based multi-sensor data fusion techniques for RPAS navigation and guidance. In: Metrology for Aerospace, Benevento, pp. 395–400. IEEE (2015)
3. Feng, C., Au, W.S.A., Valaee, S.: Received-signal-strength-based indoor positioning using compressive sensing. *IEEE Trans. Mob. Comput.* **11**(12), 1983–1993 (2012)
4. Rusli, M.E., Ali, M., Jamil, N.: An improved indoor positioning algorithm based on RSSI-trilateration technique for internet of things (IOT). In: International Conference on Computer and Communication Engineering, Kuala Lumpur, pp. 72–77. IEEE (2017)
5. Liu, Y., Sheng, X., Marston, S.R.: The impact of client-side security restrictions on the competition of cloud computing services. *Int. J. Electron. Commer.* **19**(3), 90–117 (2015)
6. Ducru, P., Josey, C., Dibert, K.: Kernel reconstruction methods for Doppler broadening temperature interpolation by linear combination of reference cross sections at optimally chosen temperatures. *J. Comput. Phys.* **335**(2), 535–557 (2017)
7. Wu, Z., Guo, X., Huang, X.: A liver vessel skeleton line reconstruction method based on linear interpolation. In: International Conference on Virtual Reality and Visualization, Xi'an, pp. 257–260. IEEE (2013)
8. Gong, C., Tao, D., Liu, W.: Label propagation via teaching-to-learn and learning-to-teach. *IEEE Trans. Neural Netw. Learn. Syst.* **28**(6), 1452–1465 (2017)
9. Fakhr, M.W.: Sparse locally linear and neighbor embedding for nonlinear time series prediction. In: Tenth International Conference on Computer Engineering & Systems, Cairo, pp. 371–377. IEEE (2016)
10. Chen, L., Sun, J.Q.: The closed-form solution of the reduced Fokker–Planck–Kolmogorov equation for nonlinear systems. *Commun. Nonlinear Sci. Numer. Simul.* **41**, 1–10 (2016)
11. Barra, A., Agliari, E.: A statistical mechanics approach to autopoietic immune networks. *J. Stat. Mech: Theory Exp.* **51**(7), 165–169 (2010)