EKF–GPR-Based Fingerprint Renovation for Subset-Based Indoor Localization with Adjusted Cosine Similarity

Junhua Yang *, Yong Li, Wei Cheng, Yang Liu and Chenxi Liu

School of Electronic and Information, Northwestern Polytechnical University, Xi'an 710072, China;
ruikel@nwpu.edu.cn (Y.L.); pupil_119@nwpu.edu.cn (W.C.); HNliuyang90@163.com or
hnliuyang90@mail.nwpu.edu.cn (Y.L.); pollux@mail.nwpu.edu.cn (C.L.)
* Correspondence: yjh5392@126.com or yjh5392@mail.nwpu.edu.cn; Tel.: +86-186-2925-0634

Abstract: Received Signal Strength Indicator (RSSI) localization using fingerprint has become a prevailing approach for indoor localization. However, the fingerprint-collecting work is repetitive and time-consuming. After the original fingerprint radio map is built, it is laborious to upgrade the radio map. In this paper, we describe a Fingerprint Renovation System (FRS) based on crowdsourcing, which avoids the use of manual labour to obtain the up-to-date fingerprint status. Extended Kalman Filter (EKF) and Gaussian Process Regression (GPR) in FRS are combined to calculate the current state based on the original fingerprinting radio map. In this system, a method of subset acquisition also makes an immediate impression to reduce the huge computation caused by too many reference points (RPs). Meanwhile, adjusted cosine similarity (ACS) is employed in the online phase to solve the issue of outliers produced by cosine similarity. Both experiments and analytical simulation in a real Wireless Fidelity (Wi-Fi) environment indicate the usefulness of our system to significant performance improvements. The results show that FRS improves the accuracy by 19.6% in the surveyed area compared to the radio map un-renovated. Moreover, the proposed subset algorithm can bring less computation.

1. Introduction

Localization and tracking play an important role in many military, industry, transportation, and tourism applications, including tracking airplanes, positioning a new product in the workshop, tracking the behavior of wild, protected animals, positioning firefighters inside a burning building, or guiding visitors to find the right track. The most popular and familiar localization method, as we all know, is the Global Positioning System (GPS). With GPS chips being embedded in many personal devices such as smartphones, sports watches, and traffic navigators, GPS localization has been proven very suitable for consumers. Through the Global Navigation Satellite System (GNSS), GPS system provides timing data to localization devices from multiple satellites. The device can trilaterate its position on the earth by using this data. The GPS system works well in clear view environments or open spaces, but encounters a problem in indoor environments where the signal suffers from non-line-of-sight (NLOS) propagation. In this case, indoor localization technology gradually developed.

An indoor localization system is a framework consisting of a network of customized or commercial devices that are used to wirelessly locate people or objects inside an indoor environment. For localization purposes, some technologies require extensive and additional deployment of

infrastructure covering the service area individually [1–4]. They may take much manpower and material resources on specialized client devices or tags attached to an object. Relatively speaking, the superiority of Received Signal Strength Indicator (RSSI)-based Wireless Local Area Networks (WLAN) positioning is embodied incisively and vividly [5,6]. There are two options for RSSI-based positioning: one option is to exploit the Access Points (APs) locations and the radio propagation path-loss models; another is fingerprinting algorithms with building floor plans. However, since the AP locations are not usually accessible and the path-loss model is uncertain, it makes this approach less attractive. The fingerprinting method has been shown to achieve more accurate positioning results than the method based on path-loss models. It does not require any extra investment cost in deploying APs by employing schemes of Wi-Fi fingerprinting [7–9], i.e., widespread APs installed in the building can be used for location estimation. Not only that, but the mobile devices are also widely available nowadays and are equipped with WLAN interface so there is no need for additional hardware to change or attach.

The RSSI fingerprint-based technique, however, requires an offline phase in which RSSI measurements are collected at known locations to initial training or calibration fingerprint data. Then, the position of a target device (TD) is estimated by matching an immediate RSSI measurement with the training data in the online phase [10,11]. In the offline phase, fingerprint collection is usually labor- and time-intensive, which eventually reduces localization efficiency and increases the cost. In addition, the measured RSSI values depend on fingerprint density, positions of access points, and locations of indoor objects. The cost of these manual calibrations holds back the widespread adoption of fingerprinting indoor localization. What is more serious is that the RSSI of AP is a variable varying with time, as shown in Figure 1, the signal heat map of an AP changes (dBm) on 27 September and 27 October, 2017. The test position is in indoor environment comprised of walls, doors, furniture, and people. It is obvious that the heat maps are different in a fixed area. Consequently, a new fingerprinting updated system is required whenever the environment alters in the indoor area [12].
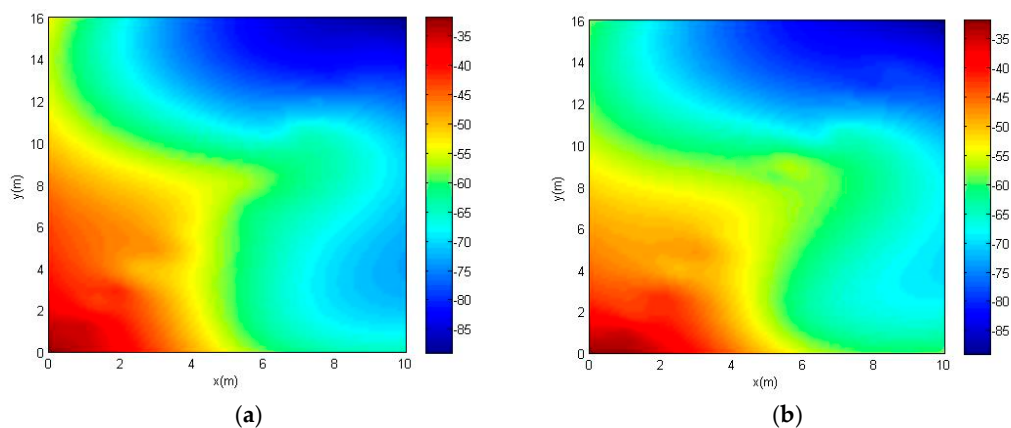

Figure 1. Fingerprint heat maps of one access point (AP; Media Access Control (MAC): B8-F8-83-CF-F8-96) on two different days. (**a**) 27 September 2017; (**b**) 27 October 2017.

To address this problem, several approaches for reducing fingerprinting load or updating the fingerprinting survey radio map have been proposed. One solution is proposed in [13], which considers having an incomplete fingerprint radio map with realistic coverage gaps, and studies the performance of several interpolation methods for recovering the missing fingerprint data. In the localization framework using unsupervised manifold alignment [14], it requires very little of the fingerprinting load, some crowdsourcing-based RPs, and plan coordinates of the indoor area. Further, without building a full fingerprinted radio map of the indoor environment, a scheme tries to construct the radio map with limited calibration [15]. However, all these works just attach importance to reduce the effort of

creating the radio map, but ignore the localization accuracy. There is a lack of methods to achieve an acceptable accuracy as well as reduction of the signal acquisition.

In this paper, we propose a Fingerprint Renovation System (FRS) to further reduce the fingerprinting load while still maintaining high exactness in performance. In the offline phase, we first built a fingerprinting radio map, as done traditionally. After that, the system is renovated, using a modified hybrid EKF–GPR algorithm, combining Extended Kalman Filter (EKF) and Gaussian Process Regression (GPR), where crowdsourcing-based points [16] add the localization region. In our research, a crowdsourcing-based point is defined as RP in a certain location when the original radio map needs an upgrade, and a crowdsourcing-based signal is defined as the RSSI value at the crowdsourcing-based point. In the online phase, a dynamic localization subset is detected to reduce computational cost. Then, the Weighted K Nearest Neighbor (WKNN) [17] employing ACS methods will be used for positing the TD. In order to make FRS fully limpid, we show our novel contributions as follows:

- As altered environment may be frequent in the radio map maintenance phase, we propose a novel and fast effective algorithm, which renovates the radio map by detecting the crowdsourcing-based signal. EKF–GPR is used to estimate the renovated radio map. The initial radio map is acted as state variable, and the crowdsourcing-based signal is treated as observed variable. The renovated radio map is also the state variable in the next renovation phase. In a real experiment environment, data collection works of 1030 fingerprint points are replaced by 50 crowdsourcing-based points, while the localization error is only increased by 19.2%.
- In the online phase, the calculation of matching algorithm using each fingerprint of the entire radio map is too large. In our proposed system, a specified amount of fingerprints is picked up randomly to execute the matching algorithm. The randomly subset of radio map is *S1*. Then, a subset *S2* in the entire radio map is produced around fingerprints acquired by matching algorithm, and an accurate position can be acquired.
- When choosing the matching algorithm, the ACS method is presented to replace the Cosine Similarity [18]. Errors in localization results are caused by using Cosine Similarity when the target point (TP) and chosen RPs are on the same line, because Cosine Similarity only spots the differences from the vector direction. In order to correct this error, Cosine Similarity is adjusted to improve the localization accuracy in this paper.

Figure 2 shows the framework of the proposed algorithms, which consists of the offline phase and online phase. The rest of the paper is organized as follows. We first introduce studies that relate to updating the radio map, including some major symbols, in Section 2. Section 3 proposes EKF–GPR algorithms for automatically updating the indoor positioning model. After that, we present the method of acquiring the subset in the whole fingerprinting radio map in Section 4, and ACS method is used to actualize the localization of TD in Section 5. In order to test and verify the FRS, we evaluate our methods with Wi-Fi data obtained in a real indoor environment in Section 6. Finally, we conclude this article. Though the study is in the context of Wi-Fi fingerprinting, our FRS is an independent and general method. The FRS, therefore, may be used with any fingerprint signal, such as Bluetooth, Near Field Communication (NFC), cellular networks, and Radio Frequency Identification (RFID).
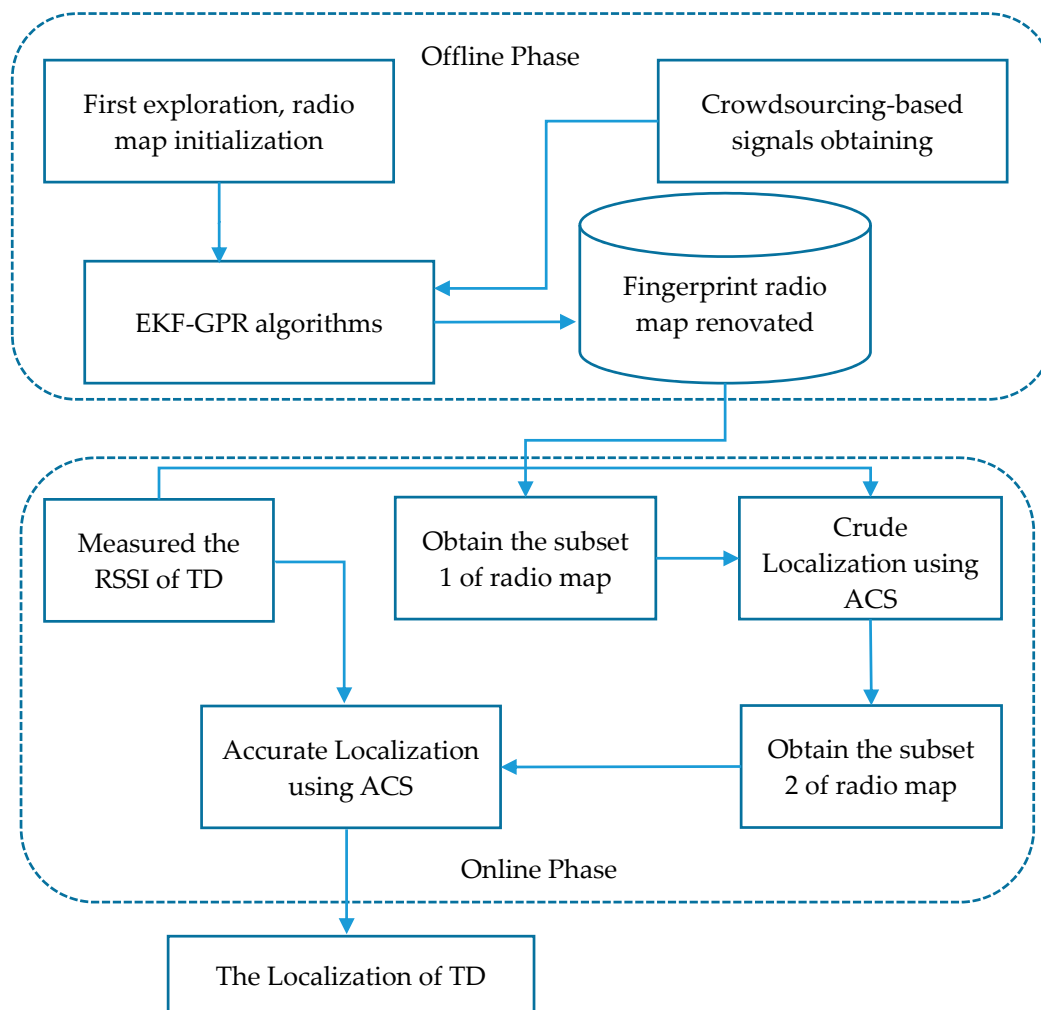
Figure 2. System overview of the Fingerprint Renovation System (FRS). EKF, Extended Kalman Filter; GPR, Gaussian Process Regression; RSSI, received signal strength indicator; TD, target device; ACS, Adjusted Cosine Similarity.

2. Related Work

The RSSI fingerprint matching method is used as the basic scheme of many indoor localization systems at present. It is an infrastructure-free approach without the requirement of utilizing expensive hardware. A common practice to construct an initialization radio map is to manually collect fingerprints at numerous known locations in the entire localization site. The radio map requires to update the area of interest to provide the accuracy that meets the requirements of daily life. The fingerprinting radio map can be updated automatically using crowdsourcing and machine learning methods [19,20]. As in previous studies [21], it is assumed that volunteers turn on their WLAN modules to contribute their traces of RSSI measurements while carrying wireless devices, i.e., mobile phones, in a localization building. First of all, a fingerprinting radio map should be built as the initialization data map.

It is required to generate a RSSI radio map during the offline phase of fingerprint-based localization. The construction of the fingerprinting radio map begins by dividing the site of interest into grids, which needs to know the floor plan in advance. RSSI values of the wireless signals transmitted by APs are gathered in calibration points inside the grids and stored into the radio map. This radio map is used for estimating the user's location during the online phase [22].

In the offline phase of two-dimensional (2-D) model, we divide the certain physical area into R known small cells. $L = \{l\_1, l\_2, \ldots, l\_j, \ldots, l\_R\}$ is defined as the center coordinates of these cells and:

$$l\_j = (x_j, y_j). \tag{1}$$

l_j is modelled as reference points (RPs). RSSI values from all access points (APs) within a certain range are measured and stored in the fingerprint data base, which is a measurement of data vectors that can be shown as:

$$P(l_j) = (p_1(l_j), p_2(l_j), \ldots, p_i(l_j), \ldots, p_A(l_j)), \tag{2}$$

where $j = 1, 2, \ldots, R$, R is the number of reference points(RPs), A is the number of APs selected in the localization area, $p_i(l_j)$ is the RSSI value from the i-th AP at RP l_j. The unique Media Access Control (MAC) address is used to discriminate different APs. Regularly, RSSI from many APs are detectable somewhere in a particular area, for those APs too weak to detect or under a certain value (e.g., -100 dBm) are set as $p_i = 0$. Then, all fingerprint signals in the radio map are:

$$P = (P(l_1), P(l_2), \ldots, P(l_j), \ldots, P(l_R)). \tag{3}$$

In the online phase, the RSSI of the target is surveyed as:

$$P^T = (p_1^T, p_2^T, \ldots, p_i^T, \ldots p_A^T). \tag{4}$$

The important notations are listed in Table 1. The radio map can be modified or updated before applying it in the online phase. In our system, it will be freshened with the help of crowdsourcing and EKF–GPR algorithms.

Table 1. Major Symbols Used in Radio Map.

Symbol	Meaning
L	Cartesian coordinates of RPs
$P(l_j)$	RSSI values from all APs at l_j RP
P	RSSI values of all the RPs
P_T	RSSI values from all APs at target point
$p_i(l_j)$	RSSI value from the i-th AP at RP l_j
$\hat{P}$	The prediction RSSI values
K	The Kalman gain
k	The index for the discrete time sequence
$\overline{P(l_j)}$	Average RSSI value of $P(l_j)$
$\overline{P^T}$	Average RSSI value of P^T
$S1$, $S2$	The subset of radio map

3. EKF–GPR Algorithms

The Kalman Filter has been associated with indoor localization in numerous researches. However, the function of the Kalman Filter in these research is, more often than not, concentrated upon path prediction and navigation [23–26]. There are few studies about using the Kalman Filter in fingerprint radio map updating, especially in combination with GPR. For all we know, GPR is used to update the radio map in [27] and estimate the virtual RPs' RSSI values in [28], but it is only GPR. On the other hand, the predictive capabilities are limited by the noise covariance and uncertainty of system model in classical filter algorithms. To solve this problem, GPR is used to provide the uncertainty of predictive value in combining EKF. The EKF–GPR algorithms do not depend on observation model and parametric prediction. Using non-parametric regression, all their parameters can be learned from training data.

3.1. EKF in Fingerprint Radio Map Renovation

In our proposed system, the primary method used in renovating the fingerprint is EKF. The EKF iteratively estimates the fingerprint radio map and updates the estimate with crowdsourcing-based points. In the renovation process, the crowdsourcing-based signal measurement equation is represented as a nonlinear model, and therefore, linearization should be performed to derive a linear equation. The EKF deals with the real-time linearization of the system function at the previous state estimate, and also includes the real-time linearization of the observation function at the corresponding predicted RSSI.

We model the localization state equation as a noisy measurement given by

$$l_j^k = F^{k-1} l_j^{k-1} + w^{k-1} , \tag{5}$$

where the superscript k is the index for the discrete time sequence, w^k is assumed to be white Gaussian noise (AWGN) with a normal probability distribution of mean 0 and variance Q^k, $w^k \sim \mathcal{N}(0, Q^k)$.

When we get the crowdsourcing-based signals, the model is defined as:

$$p_i(l_j) = f_i(l_j) + v , \tag{6}$$

where $f_i(l_j)$ is the noiseless RSSI value at location l_j from AP i, and v is the measurement noise which is modeled as additive Gaussian noise and $v \sim \mathcal{N}(0, d_j^2)$. In crowdsourcing-based schemes, the input locations l_j of clients also contain uncertainty due to localization errors. Therefore, we consider beyond the standard Equation (6) the input location with error n_j:

$$l_j = \widetilde{l_j} + n_j, \tag{7}$$

where $\widetilde{l_j}$ is the actual 2-D location of crowdsourcing-based point, $n_j \sim \mathcal{N}(0, M_j)$ The diagonal matrix M_j is 2-by-2 size assuming each dimension is independent, that is, $M_j[j, j] = s_j^2$ where s_j is the uncertainty of input location l_j, and all the non-diagonal elements of matrix M_j are zero. Then, we can achieve the equation from (6) and (7):

$$p_i(l_j) = f_i(\widetilde{l_j} + n_j) + v. \tag{8}$$

Taylor approximation [29] using noisy input l_j:

$$p_i(l_j) = f_i(l_j) + n_j^T \partial f_i + v, \tag{9}$$

where $\partial f_i = \partial f_i(l_j)/\partial l_j$ is the derivative of function $f_i(\cdot)$ with regard to l_j. The measurement equation in time sequence k, therefore, can be modeled as:

$$p_j^k = H^k l_j^k + n_j^T \partial f_i^k + v^k. \tag{10}$$

The measurement noise is additive Gaussian noise and $v^k \sim \mathcal{N}(0, D^k)$. Measurement covariance matrix D^k is diagonal. Table 2 summarizes the EKF processing.

Table 2. A summary of the EKF processing.

State Model:	$l_j^k = \mathbf{F}^{k-1} l_j^{k-1} + \mathbf{w}^{k-1}$
Measurement Model:	$p_j^k = H^k l_j^k + n_j^T \partial f_i^k + v^k$
(1) Initial filter:	$\hat{l}_j^{0+} = E(l_j^0)$ $P^{0+} = E[(l_j^0 - \hat{l}_j^{0+})(l_j^0 - \hat{l}_j^{0+})^T]$
(2) $k = 1$, k ++, calculate the partial differential matrix:	$F^{k-1} = \left.\dfrac{\partial l_j^{k-1}}{\partial l_j}\right\|_{\hat{l}^{(k-1)+}}$ $N^{k-1} = \left.\dfrac{\partial l_j^{k-1}}{\partial w}\right\|_{\hat{l}^{(k-1)+}}$
(3) The update of state estimation and estimation error covariance:	$P^{k-} = F^{k-1} P^{(k-1)+} F^{(k-1)T} + N^{k-1} Q^{k-1} N^{(k-1)T}$ $\hat{l}_j^{k-} = F^{k-1} \hat{l}_j^{(k-1)+}$
(4) Calculate the partial differential matrix:	$H^k = \left.\dfrac{\partial p_j^k}{\partial l_j}\right\|_{\hat{l}^{k-}}$ $M^k = \left.\dfrac{\partial p_j^k}{\partial v}\right\|_{\hat{l}^{k-}}$
(5) The measurement update of state estimation and the update estimation error covariance:	$K^k = P^{k-} H^{kT} (H^k P^{k-} H^{kT} + M^k D^k M^{kT})^{-1}$ $\hat{l}_j^{k+} = \hat{l}_j^{k-} + K^k[p_j^k - H^k \hat{l}_j^{k-} + n_j^T \partial f_i^k]$ $P^{k+} = (I - K^k H^k) P^{k-}$
(6) Go to step (2)	

The key components of an EKF are the state prediction and measurement models, which probabilistically describe the fingerprint evolution and the measurements returned by the crowdsourcing-based signals, respectively. As show in Table 2, these models are parametric descriptions of the radio map updating. The noise components and parameters of the models can be estimated from more than one crowdsourcing-based signal. Even though such EKF models are very efficient, their updating capabilities are limited because they often ignore the model aspects of the radio map renovation process.

3.2. GPR in Fingerprint Radio Map Renovation

GPR is an impressive, non-parametric method for learning regression functions from sample data. The fundamental advantages of GPR are their ability to learn noise and smoothness parameters from training data, their ability to provide uncertainty estimates, and their modeling flexibility [30].

In our proposed system, GPR is designed as an enhancement of measurement models and parametric prediction for EKF. In this section, we discuss the formulation of GP. The crowdsourcing-based measurement models is show in (9), then the RSSI values $p_i(l_j)$ can be rewritten as:

$$p_i(l_j) = f_i(l_j) + v_e, \tag{11}$$

where v_e is additive Gaussian noise and $v_e \sim \mathcal{N}(0, d_j^2 + \partial f_i^T M_j \partial f_i)$. We define the covariance function $c(l_*, l_j)$, which represents the correlation of two RSSIs at input locations l_* and the RP l_j. In that way, the transfer function between location l_* of the crowdsourcing-based point and its RSSI is defined as a Gaussian process:

$$f(l_*) \sim \mathcal{G}(m(l_*), c(l_*, l_j)), \tag{12}$$

where $m(l_*)$ is mean value, and $c(l_*, l_j)$ is the covariance. The covariance of RSSI can be expressed from [31]:

$$cov(p_a, p_b) = c(l_a, l_b) + d_j^2 \beta_{ab}, \tag{13}$$

where l_a and l_b denote any two localizations, $\beta_{ab} = 1$ if $a = b$, and 0 otherwise. As shown in Section 2, L is an R-by-2 matrix. Corresponding to the locations L, p_c is the vector of crowdsourcing-based RSSIs. Then the covariance matrix of p_c is shown:

$$cov(p_c) = P_f + d_j^2 I. \tag{14}$$

Here, P_f is the R-by-R covariance matrix over all R input RSSIs and I is the identity matrix of size R. All input RSSI values are jointly Gaussian process, $p(l) \sim \mathcal{N}(m(L), P_f + d_j^2 I)$, where $m(L)$ is a vector of R mean RSSIs about all locations in L. Based on (11) and (13), then the predicted mean RSSI $m(l^{k+1})$ is given by

$$m(l^{k+1}) = m(l^k) + c(l^k, L)^T (p(l) - m(L))[P_f + d_j^2 I + diag\{\Gamma_f M_j \Gamma_f^T\}]^{-1}, \tag{15}$$

where $diag\{\cdot\}$ is a diagonal matrix expression, Γ_f is an R-by-2 matrix, it contains R derivative function values about ∂f_i. In that way, the predictive variance of the RSSI is given by

$$(d_j^2)^{k+1} = c(l^k, l^k) - c(l^k, L)^T [P_f + d_j^2 I + diag\{\Gamma_f M_j \Gamma_f^T\}]^{-1} c(l^k, L). \tag{16}$$

So far, the evolution of the dynamic fingerprinting radio map is represented by the Gaussian process alone. GPR expresses relationship between the crowdsourcing-based signals and next state of radio map.

3.3. EKF–GPR Algorithm in FRS

Extended Kalman Filters and Gaussian Process Regression are used in conjunction to create the EKF–GPR algorithm. The block diagram of the EKF–GPR algorithm is shown in Figure 3. The EKF–GPR algorithm first employs EKF to predict the renovation result of the radio map and then uses GPR to correct and estimate the error of the EKF prediction. As we all know, there is a certain prediction error in the EKF because of the uncertainty. The prediction error is associated with state estimation to excavate the correlation in EKF–GPR algorithm.
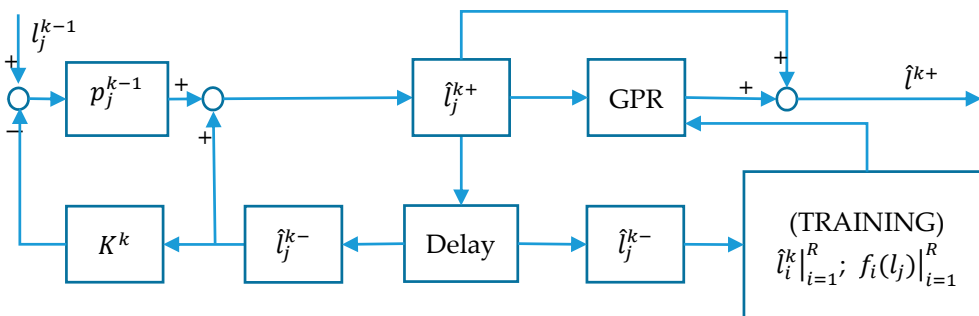

Figure 3. Block diagram of EKF–GPR algorithm.

We first use the GPR prediction model to generate the predicted mean:

$$\hat{m}^k = m^{k-1} + GPR_m(m^{k-1}, \hat{P}). \tag{17}$$

Conditioned on observation training data P, as shown in (3), the $GPR_m(\cdot)$ defines a Gaussian predictive distribution, and $\hat{P}$ is prediction data. Then, the input location error n_j in (7) can be transformed:

$$n_j = GPR_m(m^{k-1}, \hat{P}), \tag{18}$$

where n_j corresponds directly to the *GPR* uncertainty. The linearization of the prediction model is shown as:

$$\boldsymbol{\Gamma}^k = \boldsymbol{I} + \frac{\partial GPR_m(m^{k-1}, \hat{\boldsymbol{P}})}{\partial l_j^{k-1}}. \tag{19}$$

It is the sum of the identity matrix and the partial derivative of the GPR mean function. A change in any of the fingerprinting radio maps of the previous state has a direct effect on the renovation state. Meanwhile the GPR mean function only represents the change in time domain, and therefore the identity matrix is necessary. As shown step (3) of Table 2, estimation error covariance is achieved using these matrices:

$$\hat{c}^k = \boldsymbol{\Gamma}^k c^{k-1} (\boldsymbol{\Gamma}^k)^T + n_j. \tag{20}$$

According the GPR model, the predicted observation and the error covariance are shown as

$$p_j^k = GPR_m(m^k, \boldsymbol{P}) \tag{21}$$

$$R^k = GPR_c(m^k, \boldsymbol{P}). \tag{22}$$

Then the linearization of the observation model is:

$$H^k = \frac{\partial GPR_m(m^k, \boldsymbol{P})}{\partial l_j^k}. \tag{23}$$

From above, we can obtain the Kalman gain:

$$K^k = \hat{c}^k (H^k)^T (H^k \hat{c}^k (H^k)^T + R^k)^{-1}. \tag{24}$$

The update of the mean and estimation error covariance:

$$m^k = \hat{m}^k + K^k (p_j^k - \hat{p}_j^k) \tag{25}$$

$$c^k = (\boldsymbol{I} - K^k H^k) \hat{c}^k. \tag{26}$$

Given the verified EKF model, for each crowdsourcing-based signal, EKF–GPR calculates at each RP *j* the predicted signal mean (25) and RSSI estimation error covariance (26). The EKF–GPR algorithm employs the prediction model $\boldsymbol{\Gamma}^k$ as the input vector and the prediction errors mean m^k as the corresponding output in each training signal. As shown in Figure 3, we train the GPR over a set of training data $\hat{l}_i^k$ and $f_i(l_j)$. Note that EKF–GPR utilizes a model-based EKF and a standard GPR by combining them in a cascade form.

4. Subset of Radio Map

Radio map is generally built in a special server, with enough hardware resource that makes it unhindered to renovate itself. In the online phase, the resource of handheld devices is constrained. The complexity of the localization algorithm is usually proportionate to the localization accuracy. On condition that the localization algorithm is confirmed, the subset of radio map using for locating should be as concise as possible. On the other hand, for the sake of preserving user privacy and to make the location system scalable, the localization algorithm should be run on the mobile unit. Hence, the scale of the radio map need to be considered. In this section, a way of subset, which runs on the final localization unit, is presented to reduce the complexity of algorithm.

In order to reduce the computation complexity, choosing only a subset of the radio map for the positioning is a convenient way. However, the subset selection usually comes at the expense of a complexity algorithm, such as machine learning techniques [32] or principal component analysis [33].

These methods bring difficulties to the localization themselves. Two stages of subset are proposed in our method to solve this issue.

In the first stage, a random subset *S1* is obtained from the whole radio map. Before that, the elements of matrix *P* in (3) is randomly sequenced to form a new matrix P_r:

$$P_r = (P(l_{r1}), P(l_{r2}), \ldots, P(l_{rR})).\tag{27}$$

The idea of *S1* selection can be expressed in a mathematical form as follows:

$$\begin{pmatrix} P_{S1}(l_{r1}) \\ P_{S1}(l_{r2}) \\ \vdots \\ P_{S1}(l_{r\mathcal{M}}) \end{pmatrix}_{\mathcal{M}*1} = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & \cdots & 0 \end{pmatrix}_{\mathcal{M}*R} \begin{pmatrix} P(l_{r1}) \\ P(l_{r2}) \\ \vdots \\ P(l_{rR}) \end{pmatrix}_{R*1}\tag{28}$$

where $S1 = [P_{S1}(l_{r1}), P_{S1}(l_{r2}), \ldots, P_{S1}(l_{r\mathcal{M}})]^T$ is subset 1 of radio map, $\mathcal{M}$ is the number of access points in subset 1, and $\mathcal{M} \leq R$, R is still the number of RPs in initializing radio map. In (28), the first $\mathcal{M}$ RPs among out-of-order array $\{P(l_{r1}), P(l_{r2}), \cdots, P(l_{rR})\}$ are chosen, hence,

$$\{P_{S1}(l_{r1}), P_{S1}(l_{r2}), \ldots, P_{S1}(l_{r\mathcal{M}})\} = \{P(l_{r1}), P(l_{r2}), \cdots, P(l_{r\mathcal{M}})\}\tag{29}$$

can be obtained.

An appropriate localization algorithm, ACS in our method, is run based on *S1*. The result is a certain amount RPs from *S1*, which can be expressed as

$$\mathbb{S} = (P(l_{\mathbb{S}1}), P(l_{\mathbb{S}2}), \ldots, P(l_{\mathbb{S}\mathcal{N}})),\tag{30}$$

where $\mathcal{N}$, $\mathcal{N} \leq \mathcal{M}$, is the number of the nearest neighbor RPs in *S1* and it is decided by different scenes and algorithms. Then, we revive more intensive RPs in P_r by judging which ones are around the elements of $\mathbb{S}$, which is subset 2 (*S2*). In order to illustrate the whole subset choosing process, we show it in an unsophisticated indoor environment in Figure 4. The area filled by red triangles in (d) of Figure 4 is *S2* and it is the final subset which is used to acquire accurate localization.
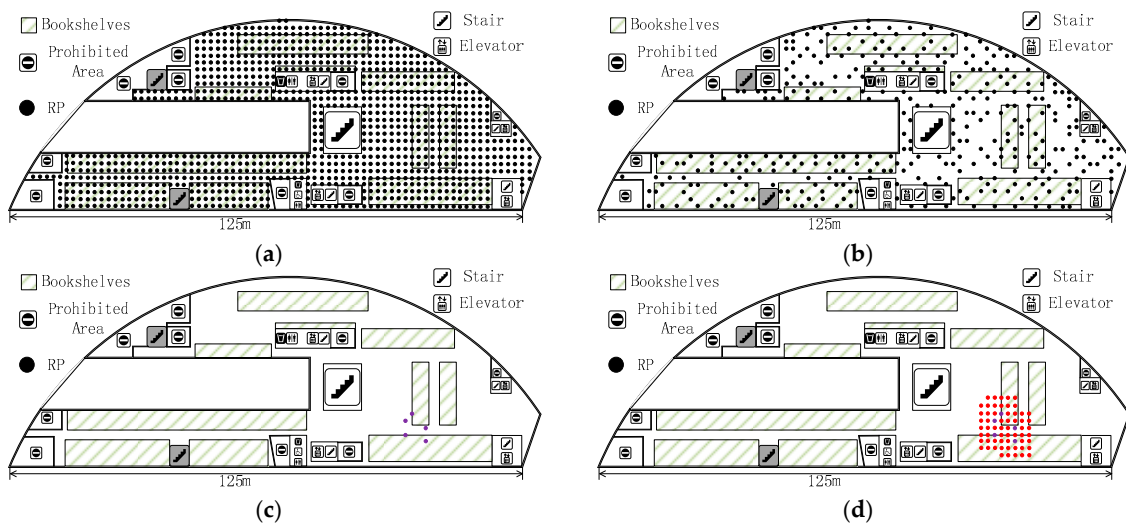

Figure 4. The process of choosing radio map subset. (**a**) The original reference points (RPs) in an indoor environment, which are shown as black circle dots; (**b**) Subset *S1* is shown as purple circle dots; (**c**) In the online phrase, five localization RPs in *S1*; (**d**) Subset *S2* is shown as red triangles.

5. Localization Based on Adjusted Cosine Similarity

During the online phase, the RSSI value in real time at the TP is compared with radio map using certain matching algorithm, such as Euclidean Distance and Cosine Similarity [34]. The algorithm of Euclidean Distance could not deliver a high localization accuracy because some statistical regularities are ignored in Euclidean Distance, which can be estimated from the TP and fingerprint radio map. On the other hand, the two algorithms can also be combined with K Nearest Neighbor (KNN) [35] and WKNN methods which focus on the calibration or calibration-free method to deal with the RSSI value difference between the RP and the TP. The KNN algorithm uses the RSSI of TP to search for K closest matches of RPs in fingerprint, then taking the mean value of these K location coordinates. WKNN is expanded on the basis of KNN but gives different weights to these K-referenced coordinates. In this paper, Cosine Similarity and WKNN are chosen to achieve TP localization.

Traditionally, Cosine Similarity investigates the correlation of RP and TP from the view point of similarity. The expression of Cosine Similarity is shown as

$$CS(x,y) = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2} \sqrt{\sum_i y_i^2}} = \frac{\langle x, y \rangle}{\|x\| \, \|y\|}. \tag{31}$$

However, there are deficiencies about Cosine Similarity, in that it only spots the differences from the vector direction, but ignores the absolute value. Errors in the localization result are made when using Cosine Similarity when the TP and *K* RPs are on the same line. In order to address this concern, the ACS method is presented in this paper. The correlation between RSSI vectors P^j and P^T is based on ACS, i.e.,

$$A_j = ACS(P^j, P^T) = \frac{\sum_{a=1}^{A} (P_a^j - \overline{P^T})(P_a^T - \overline{P^T})}{\sqrt{\sum_{a=1}^{A} (P_a^j - \overline{P^T})^2} \sqrt{\sum_{a=1}^{A} (P_a^T - \overline{P^T})^2}}, \tag{32}$$

where $\overline{P^T}$ is the average RSSI value from the all APs at the TP, and:

$$\overline{P^T} = \frac{1}{A} \sum_{a=1}^{A} P_a^T. \tag{33}$$

In the KNN method, *K* coordinates of most similar RPs to TP are employed to calculate the mean value. The expression is given as follows,

$$(x_T, y_T) = \frac{1}{K} \sum_{i=1}^{K} (x_i, y_i), \tag{34}$$

where (x_T, y_T) is the measurement localization coordinate of TP, is the coordinates of *K* RPs. Without assigning weight, KNN is the basic Nearest Neighbor algorithm. Meanwhile, it is assigned a weight to every coordinate according to the value of similarity in the WKNN method. Normally, w_j is the weight of the *j*-th selected RP, which can be calculated as

$$w_j = \frac{1}{E_j \sum_{i=1}^{k} \left(\frac{1}{E_i}\right)}, \ j = 1, 2, \ldots, K. \tag{35}$$

In this way, the coordinate of TP using WKNN can be shown as

$$(x_T, y_T) = \sum_{i=1}^{K} w_j(x_i, y_i). \tag{36}$$

In order to illustrate the improvement of ACS, we carry out an experiment in a real indoor environment. We performed the tests in the second floor of School of Electronics and Information

in Northwestern Polytechnical University (NWPU). There are three APs (AP1-5c:dd:70:b6:89:71, AP2-5c:dd:70:b6:92:51, AP3-5c:dd:70:b6:75:71), that are formed in a straight line in the corridor, as shown in Figure 5. The RSSIs of T1 and T2 are shown in Table 3.
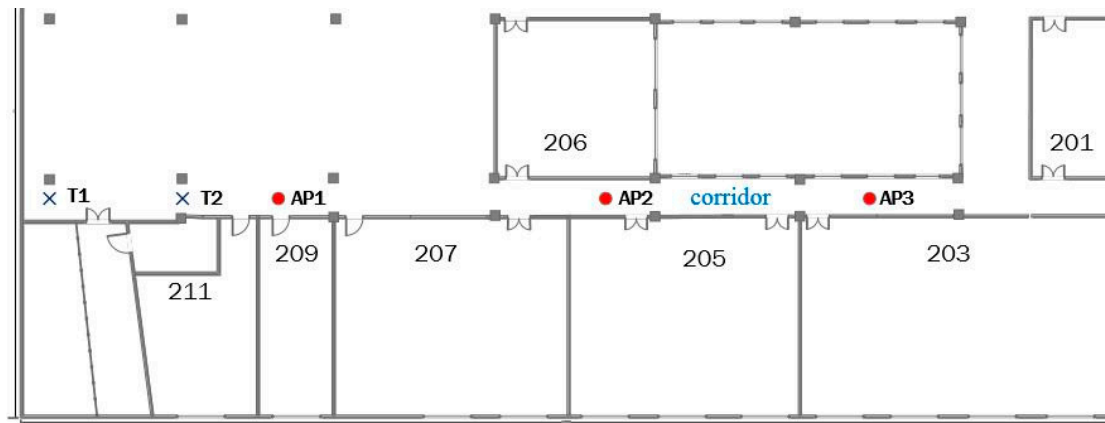

Figure 5. Experiment of ACS in the real environment, APs are shown as red circle dots, two test points is shown as blue crosses, and the five points are in one straight line.

Table 3. The RSSIs of ACS test points.

	AP1	AP2	AP3
T1	−49 dBm	−55 dBm	−66 dBm
T2	−36 dBm	−49 dBm	−64 dBm

The similarity between T1 and T2 using (31) is 0.9945, and there is a remarkable homogeneity between them. If this situation is emerged in localization phrase, T1 and T2 will be judged as two very close points. However, we can see that there is a certain interval between them. Unlike the Cosine Similarity, the similarity value is 0.703 using (32). T1 and T2 are no longer that close, which meets the actual circs better.

6. Experimental Evaluation

In this section, we consider the experimental testbed and evaluate the performance of the proposed FRS method comparing it to other methods in the area of WLAN location determination. All systems were implemented in the NWPU library for fair comparison.

6.1. Experimental Settings

The experiment is carried out at the fourth floor of the NWPU campus library. Except for the prohibited area, there is still about 4100 m^2 left, as shown in Figure 6. The experimental area is mainly focused on bookshelves and a reading area. It has an immense space, including bookshelves, desks, and frequent flow of people. We conducted experiments on two different periods, the RPs collecting date (Database_1) was from 16 October 2017 to 22 October 2017, and the crowdsourcing-based signals were collected on 16 November 2017. The renovated radio map using crowdsourcing-based signals is marked as Database_2. In the online phase, four TPs (Figure 7) were used to detect the accuracy of localization using Database_1 and Database_2, respectively.
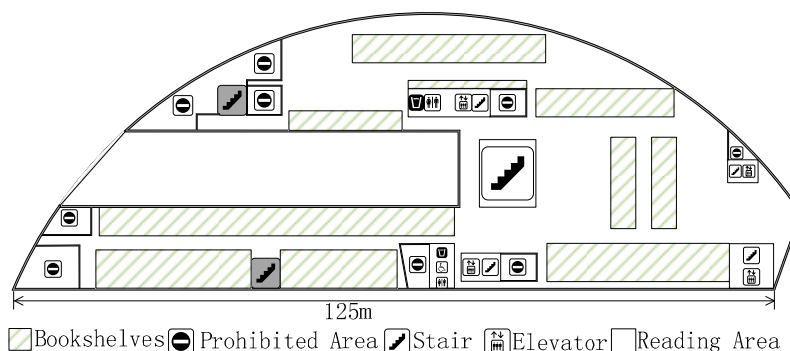
Figure 6. Plane graph of the fourth floor in the Northwestern Polytechnical University (NWPU) campus library.
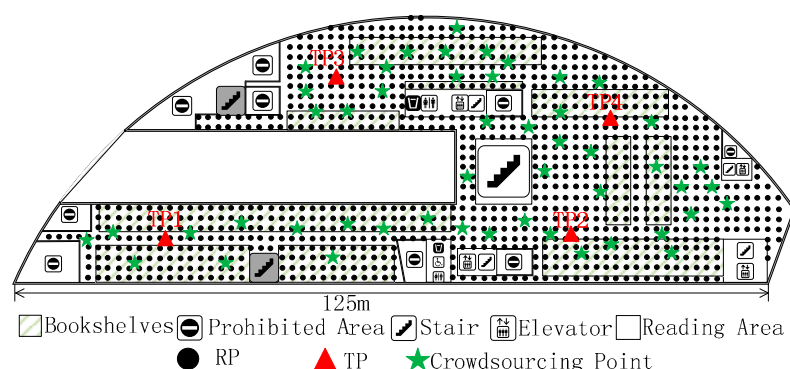

Figure 7. The distribution of RPs, target points (TPs), and crowdsourcing-based points in the experimental area.

We collect RPs on the fourth floor of the NWPU campus library, an irregular plane, and the total number of RPs is 1030, as shown in Figure 7. These RPs were taken at two-meter intervals. All of RSSIs were taken in the middle of grids using Wi-Fi Analyzer in a Nubia Mobile Phone. Overall, 105 physical APs were detected for testing. Except for infrastructure APs, fixed on the library ceiling, there are about 10 temporary APs left. There are 10 APs between two phases and are also one part of the alternative radio map. At each RP, a client could measure 15 distinctive physical APs on average. These APs are ordered according to the RSSI value, and we choose the eight largest APs as $P(l_j)$. In order to obtain the exact result, we face north, west, south, and east, respectively, to sample the RSSI at every RP. In the same way, the crowdsourcing-based signal at a certain point is obtain for renovating the radio map. As we stated before, $\widetilde{l_j}$ is the actual 2-D location of crowdsourcing-based point. In this experiment, the number of this unit is 50 and it is chosen randomly, as shown in Figure 7.

We use a number of baseline parameters in each phase. When using the WKNN localization method, the parameter $k = 5$. The number of APs at each RP or TP, A = 10, and it includes the APs with the 10 largest RSSI value. As we stated before, the total number of RPs is 1030, $R = 1030$, and $\mathcal{M} = 100$ in *S1*. In (30), $\mathcal{N} = 5$ is the same as k. The nearest 24 RPs are revived pointing at each of the five RP, hence there are 125 revived RPs of *S2* at most, because these RPs may coincide with each other.

6.2. The Effect of Radio Map Renovation

Here, we use four TPs during the experimental period to investigate the performance of our renovation method. We conduct the ACS localization method on targets with two fingerprint radio maps, Database_1 and Database_2.

We then evaluate localization result with the two radio maps respectively. The localization experiment is carried out on 17 November 2017, one day after radio map renovation. The four TPs

are distributed in the experimental area as the red triangle, as shown in Figure 7. At each TP, the ACS localization algorithm is run 30 times with 1 min intervals. Figure 8 shows the real-time localization errors versus the 30 times location. Before renovating the radio map, as shown in Figure 8a, four TPs have similar localization errors. Given Database_2 in Figure 8b, each localization error of TP is reduced. The mean error are 3.72 m and 2.99 m, corresponding to Database_1 and Database_2, respectively.
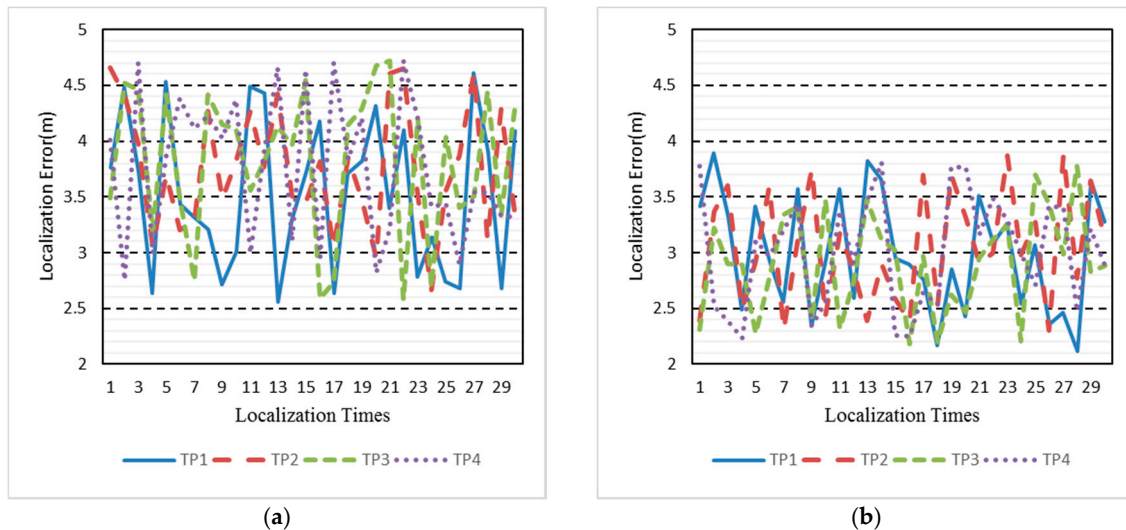

(**a**)   (**b**)

Figure 8. Localization error before (**a**) and after (**b**) radio map renovation.

We also compare FRS with interpolation and extrapolation methods (IEM) [13] based localization algorithms. The experimental RPs collecting date was from 28 November 2017 to 3 December 2017 and the RPs location is same as in Figure 7, which is called Database_3. Crowdsourcing-based signals are also collected during this phase to renovate Database_1. The renovated result here is Database_4, which is similar to Database_2. Then, the same four TPs shown in Figure 7 are used for localization. After that, some certain number percentages of Database_3 are removed to locate the TPs. Figure 9 shows the results of the localization using different methods and radio maps. In this Figure, *Original* is the localization result using Database_3, *FRS* represents the localization error using Database_4, *Partial* shows the localization result along with the removed percentages of Database_3, and *IEM* indicates the error using interpolation method based on different percentages of Database_3. The *IEM* method shows better localization results when the removed percentages of radio map are below 15%. However, the *FRS* method has a lower error when the removed percentages of radio map are above 15%. Although both of them reduce the labour force to update fingerprint radio map, *FRS* proposed in this paper has a greater scope of application. As the Percentage of Removed Fingerprints increases, the localization errors of *Partial* and *IEM* method grow larger. The *FRS* method is unaffected by the Removed Fingerprints and only needs 50 crowdsourcing-based points. Meanwhile, comparing *Original* and *FRS*, data collection works of 1030 fingerprint points are replaced by only 50 crowdsourcing-based points, while the mean localization error increases from 2.45 m to 2.92 m.
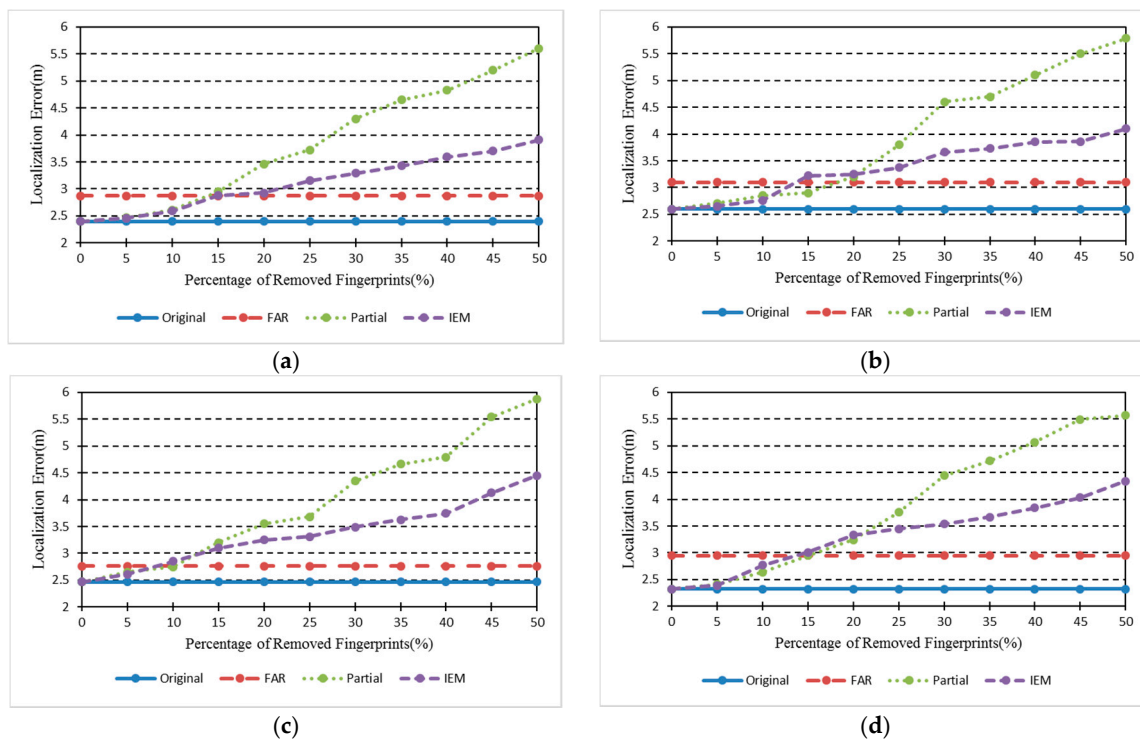
Figure 9. Localization results of four TPs. Original: The localization error obtained from Database_3; FAR: The localization error obtained from Database_4. Both of them are irrelative with Percentage of Removed Fingerprints. (**a**) TP1; (**b**) TP2; (**c**) TP3; (**d**) TP4.

6.3. The Impact of Crowdsourcing-Based Points in Amount and Location

In the proposed FRS, the amount and position of crowdsourcing-based points directly impact the stabilization and accuracy of localization. There are 50 crowdsourcing-based points, as shown in Figure 7, to achieve all the result above. In this section, different numbers and positions of crowdsourcing-based points are set up to reveal the distinction of localization accuracy. The collection work of crowdsourcing-based signals is similar to normal RP. The only difference between the two is the randomness of crowdsourcing-based points.

First of all, 10 crowdsourcing-based points are set around TP1 and TP2, away from TP3 and TP4. Figure 10 shows the distribution of crowdsourcing-based points. After playing a role in the FRS system, crowdsourcing signals bring different localization accuracy as shown in Figure 11a. The mean localization error of TP1, TP2, TP3, and TP4 are 3.09 m, 2.98 m, 3.87 m, and 3.93 m, respectively. The localization errors of TP1 and TP2, which are around by crowdsourcing-based points, are reduced by 22.1% compared to TP3 and TP4. We can also see the otherness of root-mean-square error (RMSE) in Figure 11b. The localizations of TP1 and TP2 are more precise than TP3 and TP4.
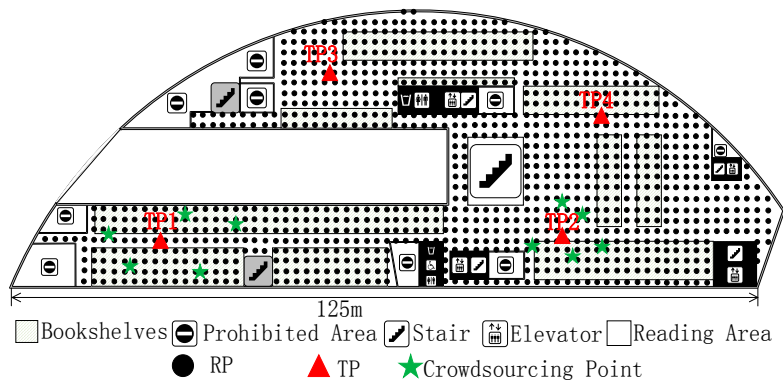
Figure 10. Ten crowdsourcing-based points around TP1 and TP2, away from TP3 and TP4.
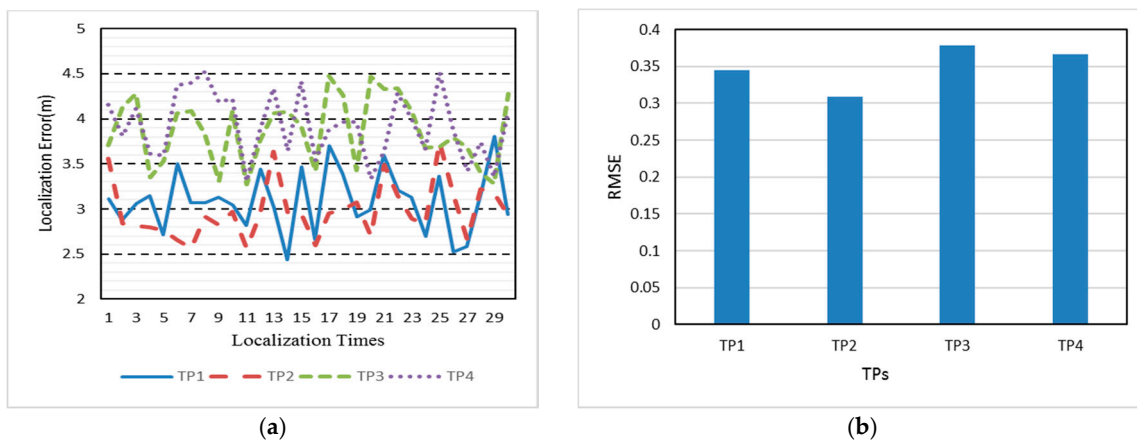

Figure 11. (**a**) Localization error of TP1 and TP2 compared to TP3 and TP4 with different situations about crowdsourcing-based points; (**b**) RMSEs of these four TPs.

Except for positions of crowdsourcing-based points, the amount also affects the localization accuracy. In the experimental area, we added the crowdsourcing-based points randomly, as shown in Figure 7. The interval is 10 and the scope is from 0 to 100. At each TP, the localization process is repeated 30 times with different amounts of crowdsourcing-based points. The localization errors of the four TPs are shown in Figure 12.
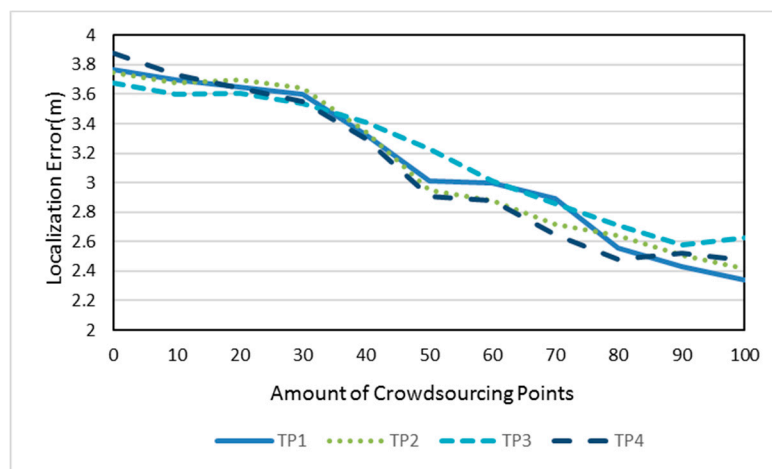

Figure 12. Localization error of TP1, TP2, TP3, and TP4 with different amounts of crowdsourcing-based points.

It indicates that the localization deviation is decreasing as the amount of crowdsourcing-based points increases. The promotion of localization accuracy is obvious from 30 to 80, while the trend is slowed down at others.

7. Conclusions

The RSSI fingerprint technique has the advantages of simplicity, deployment practicability, and supplying reasonable accuracy. Therefore, fingerprinting localization has attracted a lot of attention. However, it is hard to upgrade a fingerprint radio map due to harsh force and intensive labor requirements. In this paper, we have proposed and tested indoor localization with FRS, a subset of a radio map, and ACS, which achieves precise indoor localization.

We tested our proposed system in a real indoor environment. For altered RSSI in different date, FRS efficiently renovated the radio map, used new fingerprint with suitable algorithms and then found the location of the TP. Compared to the mean localization 3.72 m of un-renovated radio map, we achieved 2.99 m mean error using the renovated radio map. The localization accuracy is promoted about 19.6% and this value is raising as the amount of crowdsourcing-based points increases.

Moreover, the subset choosing of radio map reduces localization computation in the proposed FRS. In our experiment, it needs repeating ACS of a total of 1030 times without subset. This number is only 125 for each TP and the computation reduced about 87% in the online phase. We also consider that the FRS can use the ACS algorithm introduced to improve their positioning systems.

Author Contributions: Junhua Yang conceived the idea, designed the experiments and wrote the paper; Yong Li and Wei Cheng helped with the algorithm; Yang Liu and Chenxi Liu helped to analyze the experimental data.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

1. Ni, L.; Liu, Y.; Lau, Y.C.; Patil, A.P. LANDMARC: Indoor location sensing using active RFID. In Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, Fort Worth, TX, USA, 26 March 2003; pp. 407–415.

2. Fantuzzi, M.; Masotti, D.; Costanzo, A. A novel integrated UWB–UHF one-port antenna for localization and energy harvesting. *IEEE Trans. Antennas Propag.* **2015**, *63*, 3839–3848. [CrossRef]

3. Scherhäufl, M.; Pichler, M.; Stelzer, A. UHF RFID localization based on evaluation of backscattered tag signals. *IEEE Trans. Instrum. Meas.* **2015**, *64*, 2889–2899. [CrossRef]

4. Martinelli, F. A robot localization system combining RSSI and phase shift in UHF-RFID signals. *IEEE Trans. Control Syst. Technol.* **2015**, *23*, 1782–1796. [CrossRef]

5. Honkavirta, V.; Perala, T.; Ali-Loytty, S.; Piche, R. A comparative survey of WLAN location fingerprinting methods. In Proceedings of the 6th Workshop on Positioning, Navigation and Communication 2009 (WPNC'09), Hannover, Germany, 19 March 2009; pp. 243–251.

6. Kaemarungsi, K.; Krishnamurthy, P. Properties of indoor received signal strength for WLAN location fingerprinting. In Proceedings of the First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services (MobiQuitous'04), Boston, MA, USA, 26 August 2004.

7. Torres-Sospedra, J.; Montoliu, R.; Trilles, S.; Belmonte, O.; Huerta, J. Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems. *Expert Syst. Appl.* **2015**, *42*, 9263–9278. [CrossRef]

8. Fang, S.; Lin, T.; Lee, K. A novel algorithm for multipath fingerprinting in indoor WLAN environments. *IEEE Trans. Wirel. Commun.* **2008**, *7*, 3579–3588. [CrossRef]

9. Choi, T.; Chon, Y.; Kim, Y.; Cha, H. Enhancing WiFi-fingerprinting accuracy using RSS calibration in dual-band environments. *Pervasive Mob. Comput.* **2016**, *30*, 45–57. [CrossRef]

10. Han, D.; Jung, S.H.; Lee, M.; Yoon, G. Building a practical Wi-Fi-based indoor navigation system. *IEEE Pervasive Comput.* **2014**, *13*, 72–79.

11. Jung, S.; Moon, B.; Han, D. Unsupervised learning for crowdsourced indoor localization in wireless networks. *IEEE Trans. Mob. Comput.* **2016**, *15*, 2892–2906. [CrossRef]

12. Pajovic, M.; Orlik, P.; Koike-Akino, T.; Kim, K.J.; Aikawa, H.; Hori, T. An unsupervised indoor localization method based on received signal strength (RSS) measurements. In Proceedings of the 2015 IEEE Global Communications Conference (GLOBECOM), San Diego, CA, USA, 6–10 December 2015.

13. Talvitie, J.; Renfors, M.; Lohan, E.S. Distance-based interpolation and extrapolation methods for RSS-based localization with indoor wireless signals. *IEEE Trans. Veh. Technol.* **2015**, *64*, 1340–1353. [CrossRef]

14. Majeed, K.; Sorour, S.; Al-Naffouri, T.Y.; Valaee, S. Indoor localization and radio map estimation using unsupervised manifold alignment with geometry perturbation. *IEEE Trans. Mob. Comput.* **2016**, *15*, 2794–2808. [CrossRef]

15. Sorour, S.; Lostanlen, Y.; Valaee, S.; Majeed, K. Joint indoor localization and radio map construction with limited deployment load. *IEEE Trans. Mob. Comput.* **2015**, *14*, 1031–1043. [CrossRef]

16. Wang, B.; Chen, Q.; Yang, L.T.; Chao, H. Indoor smartphone localization via fingerprint crowdsourcing: Challenges and approaches. *IEEE Wirel. Commun.* **2016**, *23*, 82–89. [CrossRef]

17. Ko, J.; Klein, D.; Fox, D.; Haehnelt, D. GP-UKF: Unscented Kalman filters with Gaussian process prediction and observation models. In Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems, San Diego, CA, USA, 29 October–2 November 2007; pp. 1901–1907.

18. Yim, J.; Park, C.; Joo, J.; Jeong, S. Extended kalman filter for wireless LAN based indoor positioning. *Decis. Support Syst.* **2008**, *45*, 960–971. [CrossRef]

19. Wu, C.; Yang, Z.; Liu, Y.; Xi, W. Wireless indoor localization without site survey. In Proceedings of the IEEE INFOCOM, Orlando, FL, USA, 25–30 March 2012; pp. 64–72.

20. Rai, A.; Chintalapudi, K.K.; Padmanabhan, V.N.; Sen, R.; Chintalapudi, K. Zee: Zero-effort crowd-sourcing for indoor localization. In Proceedings of the 18th Annual International Conference on Mobile Computing and Networking, ACM, Istanbul, Turkey, 22–26 August 2012; pp. 293–304.

21. Zhuang, Y.; Syed, Z.; Li, Y.; El-Sheimy, N. Evaluation of two WiFi positioning systems based on autonomous crowdsourcing of handheld devices for indoor navigation. *IEEE Trans. Mob. Comput.* **2016**, *15*, 1982–1995. [CrossRef]

22. Kanaris, L.; Kokkinis, A.; Fortino, G.; Liotta, A.; Stavrou, S. Sample size determination algorithm for fingerprint-based indoor localization systems. *Comput. Netw.* **2016**, *101*, 169–177. [CrossRef]

23. Ali-Loytty, S.; Perala, T.; Honkavirta, V.; Piche, A.R. Fingerprint kalman filter in indoor positioning applications. In Proceedings of the 18th IEEE International Conference on Control Applications, Saint Petersburg, Russia, 8–10 July 2009; pp. 1678–1683.

24. Huerta, J.M.; Vidal, J.; Giremus, A.; Tourneret, J. Joint particle filter and UKF position tracking in severe non-line-of-sight situations. *IEEE J. Sel. Top. Signal Process.* **2009**, *3*, 874–888. [CrossRef]

25. Yang, H.; Zhang, R.; Bordoy, J.; Höflinger, F.; Li, W.; Schindelhauer, C.; Reindl, L. Smartphone-based indoor localization system using inertial sensor and acoustic transmitter/receiver. *IEEE Sens. J.* **2016**, *16*, 8051–8061. [CrossRef]

26. He, S.; Lin, W.; Chan, S.-G. Indoor localization and automatic fingerprint update with altered AP signals. *IEEE Trans. Mob. Comput.* **2017**, *16*, 1897–1910. [CrossRef]

27. Chang, Q.; Li, Q.; Shi, Z.; Chen, W.; Wang, W. Scalable indoor localization via mobile crowdsourcing and Gaussian process. *Sensors* **2016**, *16*, 381. [CrossRef] [PubMed]

28. Rasmussen, C.E.; Williams, C.K.I. *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, UK, 2005.

29. McHutchon, A.; Rasmussen, C.E. Gaussian process training with input noise. In Proceedings of the Advances Neural Information Processing Systems, Granada, Spain, 12–15 December 2011; pp. 1341–1349.

30. Rasmussen, C.E. Gaussian processes in machine learning. In Proceedings of the Machine Learning Summer School Conference, Tubingen, Germany, 4–16 August 2003; Volume 3176, pp. 63–71.

31. Chen, Y.; Yin, J.; Chai, X.; Yang, Q. Power-efficient access-point selection for indoor location estimation. *IEEE Trans. Knowl. Data Eng.* **2006**, *18*, 877–888. [CrossRef]

32. Fang, S.; Lin, T. Principal component localization in indoor WLAN environments. *IEEE Trans. Mob. Comput.* **2012**, *11*, 100–110. [CrossRef]

33. Li, C.; Qiu, Z.; Liu, C. An improved weighted k-nearest neighbor algorithm for indoor positioning. *Wirel. Pers. Commun.* **2017**, *96*, 2239–2251. [CrossRef]

34. Han, S.; Zhao, C.; Meng, W.; Li, C. Cosine similarity based fingerprinting algorithm in WLAN indoor positioning against device diversity. In Proceedings of the 2015 IEEE International Conference on Communications (ICC), London, UK, 8–12 June 2015; pp. 4313–4317.

35. Xie, Y.; Wang, Y.; Nallanathan, A.; Wang, L. An improved k-nearest-neighbor indoor localization method based on spearman distance. *IEEE Signal Process. Lett.* **2016**, *23*, 351–355. [CrossRef]