Article

Gaussian Process Regression Plus Method for Localization Reliability Improvement

Kehan Liu [1], Zhaopeng Meng [2] and Chung-Ming Own [1],*

[1] School of Computer Software, Tianjin University, Tianjin 300350, China; kehanliu@tju.edu.cn
[2] School of Computer Software, Tianjin University of Traditional Chinese Medicine, Tianjin 300193, China; mengzp@tju.edu.cn
* Correspondence: chungming.own@tju.edu.cn; Tel.: +86-22-8740-1530

Abstract: Location data are among the most widely used context data in context-aware and ubiquitous computing applications. Many systems with distinct deployment costs and positioning accuracies have been developed over the past decade for indoor positioning. The most useful method is focused on the received signal strength and provides a set of signal transmission access points. However, compiling a manual measuring Received Signal Strength (RSS) fingerprint database involves high costs and thus is impractical in an online prediction environment. The system used in this study relied on the Gaussian process method, which is a nonparametric model that can be characterized completely by using the mean function and the covariance matrix. In addition, the Naive Bayes method was used to verify and simplify the computation of precise predictions. The authors conducted several experiments on simulated and real environments at Tianjin University. The experiments examined distinct data size, different kernels, and accuracy. The results showed that the proposed method not only can retain positioning accuracy but also can save computation time in location predictions.

Keywords: location estimation; RSS fingerprinting; Gaussian Process Regression; Naive Bayes

1. Introduction

Accurate, reliable, and real-time indoor positioning and position-based protocols and services are required in future social applications. Through a mobile device, a positioning system can help to determine its position and makes the position of the device available for position-based services such as navigation and tracking or monitoring. The location information of devices or users could considerably improve the performance of wireless network for network planning, network adaptation, and load balancing. Generally, the indoor localization problem is resolved using standard, low-cost, and already deployed infrastructure such as the strength database of received signals [1,2]. Instead of spending resources in deploying dedicated infrastructure and collecting data procedure, the purpose of indoor positioning study is to design and implement data fusion methods and algorithms using existing infrastructure.

Generally, received signal strength (RSS)-based location fingerprinting is based on the principle that each position has a unique set of signal values. When a system boots up, mobile devices receive the unique RSS value of the system, and they search the fingerprinting database and identify the entry that is most similar to the system unique RSS value as the estimated location. The main problem of a typical RSS fingerprinting system is that the real RSS value at any location is easily affected by the object (human or furniture inside the building) and multipath fading effects [3]. In other words, the RSS fingerprint obtained at different periods of time need not match the previous fingerprint stored in the database, leading to incorrect estimation results. Previous researchers have shown that the fingerprinting localization method can achieve meter-level accuracy if the RSS signature map

is not outdated. However, because of the dynamic nature of the radio channel and changes in the surrounding environment, the RSS signature maps must be updated multiple times daily. That is, the time and effort required to build the RSS signature map during the offline phase are the major drawbacks of fingerprinting-based localization.

In addition, the challenges in deploying reliable and scalable fingerprinting localization systems and facing user limits are also the barriers to reaching the accuracy required for use in commercial applications. Many studies have prompted in the elimination of the times at gathering RSS signatures. In [4], Haeberlen et al. demonstrated a system that allows for remarkably accurate localization across an entire office building over 12,000 square meters in area; most notably, regarding the scale, they localized a device to one of 510 cells in the building within seconds, yielding a success rate of greater than 95%. Moreover, Bisio et al. designed an indoor localization scheme of targets by using electromagnetic waves; this system was replaced by an offline phase in which the fingerprints in each point of the area of interest were estimated by means of electromagnetic waves [5]. In [6], the authors computed a position by considering the RSS measurements as a part of a random process, exploiting the information present in the acquired signals. With the efforts to build an RSS signature map having been reduced, the median error remains unacceptable for most practical indoor applications [7]. Furthermore, Zheng et al. utilized both the raw RSS values and their relation to construct a new stable and robust fingerprint for indoor position, their results indicated that the RSS variance problem can be solved without any manual calibration [8]. In this study, we developed a probabilistic framework for handling sparse training data in fingerprinting localization. Specifically, the calibration effort and costs of building the RSS signature map were reduced by modeling signal strength by using the advanced Gaussian process (GP), and the troublesome of RSS signature map was enhanced in the online phase.

Supervised learning in the form of regression and classification is an important constituent of statistics and machine learning, either for the analysis of data sets or as a subgoal of a more complex problem [9]. Traditionally, parametric models have been used for this purpose. A possible advantage also exists in the ease of interpretability; however, for complex data sets, simple parametric models may lack expressive power. A GP is a nonparametric model that is characterized completely by its mean function and covariance matrix [10]. A GP depends on several hyper parameters that can be estimated using training measurements. In our study, we used the properties of the conditional probability of a GP. The GP prior distribution was used for regression and predicting the RSS at locations with no prior measurements, and the Naive Bayes algorithm was also used to derive the obtained conditional Gaussian probability. We named the computation process Gaussian Process Plus. A major advantage of using a GP-based predicator is that in addition to the mean estimate of the RSS, an estimation of the variance is also produced, providing an indication of the uncertainty of the estimation [11]. The aforementioned process can be performed offline and generally must be performed only once.

Yiu et al. used Gaussian process regression (GPR) for received signal strength indication (RSSI) prediction in order to solve indoor location problems [12]. Partial measurements were first taken from the area of interest. They then used the Firefly algorithm to train and categorize the prior results derived using GPR. The trained results were used to build a refined fingerprinting database for the entire area of interest. The experiment showed that the GPR model could achieve a satisfactory result in predicting the RSS of an area with no prior measures. However, according to Yiu et al., the ability to expand the GPR model is restricted by the need to rebuild the fingerprinting database; in addition, the flexibility in the computation of the GPR models was limited in their study. Thus, to retain the flexibility of GP and increase the efficiency of indoor positioning, we proposed the Gaussian Process Regression Plus (GPRP) method to solve the aforementioned limitations, and we tested the experiments in both simulated and real environments.

The remainder of this manuscript is organized as follows: Section 2 reviews related studies investigating the properties of RSSI for signal transmission, and the definitions of GP. Section 3 details the GPRP measurement system used in the current study and describes the data analysis method.

Section 4 presents the simulated and real experiments and discussion, and Section 5 offers conclusions and recommendations for future research.

2. Preliminaries

2.1. RSS Fingerprint

RSS properties, which facilitate location fingerprinting, have been determined by many studies examining indoor positioning systems. This research demonstrates that user orientation could cause a variation of up to 5 dBm in the RSSI level [13,14]. At any location, the different orientations of users and mobile devices with respect to the transmitter could cause the mean RSS value to change. The modeling of the RSS-based location fingerprinting is essential for location determination algorithms; examples of RSS-based location fingerprinting models are the probabilistic approach model and preliminary analytical model [15]. Gaussian or lognormal distributions were used to model the randomness of RSS. For example, [16] summarized in a large-scale measurement that most RSS histograms could be fitted appropriately with Gaussian distributions but that few histograms could be fitted with bimodal Gaussian distributions.

Current signal-based RSS location systems have two problems. First, a considerable manual calibration effort is required to construct a radio map in the offline training phase; second, the positioning accuracy changes with the environmental dynamics. Three dynamic factors observed to change frequently over time in the environment are proposed in [17], including the presence of people, relative humidity, and movement. These factors easily affect the radio signal propagating from access points (APs) to mobile devices and are responsible for changes in positioning accuracy. The results of the experiments conducted in a cafeteria showed that an RSS distribution at a fixed location with a specified beacons was not constant with time (Figure 1). The RSS values calibrated previously in the radio map may be outdated; this condition degrades positioning accuracy because of the presence of people.
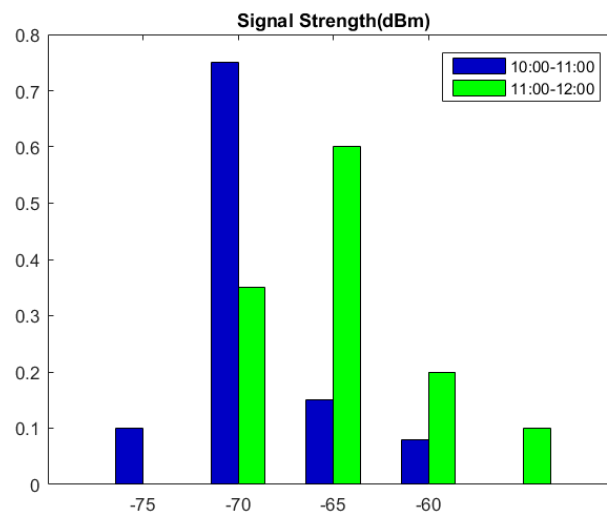

Figure 1. RSS distributions during different time periods in the cafeteria.

Traditionally, an average RSS has been considered to be lognormally distributed according to a large-scale fading model. Such RSS is generally predictable and follows several standardized path loss models.

The wall attenuation factor (WAF) model is useful for describing the slow-fading phenomenon and attenuation in signal propagation in indoor environments [18]. In this model, the attenuation

factor is used to predict signal propagation behavior when walls are the main obstacle. The following equation shows how attenuation influences RSS:

$$P\left(d\right)_{dbm} = P\left(d_o\right)_{dbm} - 10 \times n \times \log\left(\frac{d}{d_o}\right) - nW \times WAF \tag{1}$$

where n indicates the rate of increase in signal attenuation with the propagation distance, $P\left(d_o\right)$ is the RSS at a distance of reference point d_o, and d is the distance between the transmitter and the receiver. Furthermore, nW is the number of obstacles (walls) between the transmitter and the receiver, and WAF is the attenuation value resulting from the obstacles. In this equation, if nW is greater than a certain constant C, the value of nW is considered to represent the number of walls at which the attenuation factor stops influencing the signal. We can then use the constant value instead of nW. When we select a subset of APs satisfying a certain property in our alternative set of fingerprint definitions, we retrieve the corresponding position from the offline fingerprinting database by using the location estimation algorithm. Generally, several methods are used for determining the nearest neighbor location. For example, we can use the Euclidean distance:

$$\text{Eucdist}\left(S, R\right) = \sqrt{\sum_{i=1}^{n}\left(s - r_i\right)^2}$$

or the Mahalanobis distance:

$$\text{Mahaldist}\left(S, R\right) = \sqrt{\left(S - R\right)^T S^{-1}\left(S - R\right)}$$

where S is the RSS value for the target location and R is the value closest to S in the fingerprinting database.

2.2. Gaussian Process

The GPR is a new machine-learning method based on Bayesian theory and statistical learning theory. It provides a flexible framework for probabilistic regression and is widely used to solve high-dimension, small-sample, or nonlinear regression problems. From the view of the function space, the GP defines a distribution over functions. GPs are the extension of multivariate Gaussian model to the infinite-sized vector of real-valued variables. The GP is fully specified by a mean function and a covariance function such as:

$$\begin{cases} m\left(x\right) = E\left[f\left(x\right)\right] \\ k\left(x|x'\right) = E\left[\left(f\left(x\right) - m\left(x\right)\right)\left(f\left(x'\right) - m\left(x'\right)\right)\right] \end{cases} \tag{2}$$

where x and $x' \epsilon R$ are random variables [19]. The GP equation is $f\left(x\right) \sim gp\left(m\left(x\right), k\left(x, x'\right)\right)$. For simplification, the mean function is usually set to zero in the data preprocessing stage.

Because the key assumption in GP modeling is that the data can be represented as a sample from a multivariate Gaussian distribution, we can infer that:

$$\begin{bmatrix} x \\ x^* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K & K_*^T \\ K_* & K_{**} \end{bmatrix}\right) \tag{3}$$

where T indicates matrix transposition. The function probability follows a Gaussian distribution, and the conditional probability of $x_*|x$ can be computed as:

$$\text{p}\left(x_*|x\right) \sim N\left(K_* K^{-1} y, K_{**} K^{-1} K_*^T\right) \tag{4}$$

The optimal estimation for x_* is the mean of this distribution $\overline{x_*} = K_* K^{-1} x$ [20,21], and the uncertainty in the estimation is the variance $\text{var}\left(x_*\right) = K_{**} - K_* K^{-1} K_*^T$. The aforementioned

description indicates that the GPR is suitable for predicting unknown data with limited known data. GPR is also the central idea used in this study.

Some works have used the GP to help generate a fingerprint database in the offline phase [22,23]. Ferris et al. demonstrated how to use the GP to generate likelihoods at locations for which no calibration data were available, and proved that Gaussian regression could be applied successfully to various localization problems [22]. Atia et al. [24]. considered using the GPR technique to bridge the GPS outage; they ultimately obtained an 80% improvement in the position root square mean error (RMSE). Richter et al. analyzed different GPR models for WLAN fingerprinting and provided useful advice regarding GPR [22]. Cho et al. created a GPR-based radio map construction method for surveying data that could obtain high accuracy and availability, though realistic data were rare [25]. These studies have proved that GPR is a viable means of improving positioning accuracy and releasing the labor at collecting fingerprinting data.

2.3. Weight-RSS Propagation Model

For estimating and tracking the parameters of RSS propagation model in the indoor environment, many studies employ adaptive Bayesian framework to cope with unpredictable, inter-calibration and fading radio effects. However, these methods cannot drive the acceptable performance due to the complex propagation phenomena in the radio transmission. In [8], Zheng et al. proposed weight-RSS, a calibration-free solution to solve the RSS variance caused by device heterogeneity and complex environmental factors. In Zheng's system, their location can be estimated by matching all the RSS values and the relation from the test device with all the entries of the fingerprint mapping results.

Assume that m APs are deployed in an indoor environment and the physical space of the indoor environment is modeled as a finite space $L = \{l_1(x_1, y_1), l_2(x_2, y_2), \ldots, l_n(x_n, y_n)\}$ [8]. RSS values of all APs at the location l_i can be represented as follows:

$$R_i = \left(r_i^1, \, r_i^2, \, \ldots, \, r_i^q, \ldots, \, r_i^m \right)$$

and the weight-RSS of location l_i is defined as:

$$D_j = \left\{ \left(r_i^1, s\left(r_i^1 \right) \right), \, \left(r_i^2, s\left(r_i^2 \right) \right), \, \ldots, (r_i^m, s(r_i^m)) \right\}$$

where $s(r_i^j)$ is the index of r_i^j after R_i was sorted as descending order. Accordingly, Zheng et al. assume that the distance from the fingerprint map is D_{tr} (off line), and the distance from a test device is D_{te} (on line). The weighting factor is computed as follows:

$$F_{\text{tr,te}} = \{ f_i | \, 1 \leqslant i \leqslant m \}$$

and:

$$f_i = 1 - \frac{\left| s\left(r_{\text{tr}}^j \right) - s\left(r_{\text{te}}^j \right) \right|}{\max(s\left(r_{\text{tr}}^j \right) - s\left(r_{\text{te}}^j \right))}$$

Then the authors can derive the difference between distance from fingerprint map D_{tr} and from a test device D_{te} is:

$$Dist\left(D_{\text{tr}}, \, D_{\text{te}} \right) = \sqrt{\sum_j^m f_i (r_{\text{tr}}^j - r_{\text{te}}^j)^2}$$

3. System Design

In [12], Yiu et al. used the trained GPR model to estimate the signature map of an area. For optimizing hyper parameters in GPR, they proposed that the signature map be rebuilt; however, inefficient computation restricted the flexibility of the Gaussian model. Thus, in our study, we focused

on GPR system improvement and used the Naive Bayes algorithm to reduce the computation complexity of the indoor positioning system, and retain the flexibility of the models of GP.

In Figure 2, we illustrate the proposed system framework and compares it with that of Yiu et al. Firstly, our GPRP method and Yiu's method created a discrete and partial RSS fingerprinting database in the testing area. Accordingly, we all derived and tested several GP models for system computation. The different GP models had distinct distribution functions and hyper parameters. However, Yiu et al. proposed a method based on categorized and GP algorithms, and finished with the rebuilding of the RSS fingerprinting database; by contrast, our GPRP method employed the Naive Bayes method to select all the predicted location from the GP models, and the computation was more easily derived and was less time-consuming. Furthermore, to stress the efficiency of our proposed method, in Figure 2, the solid line indicates the training phase, and the blue dashed line indicates the testing phase. FL_i was derived as the final estimated location for each method. The less computation steps are needed in our GPRP method.
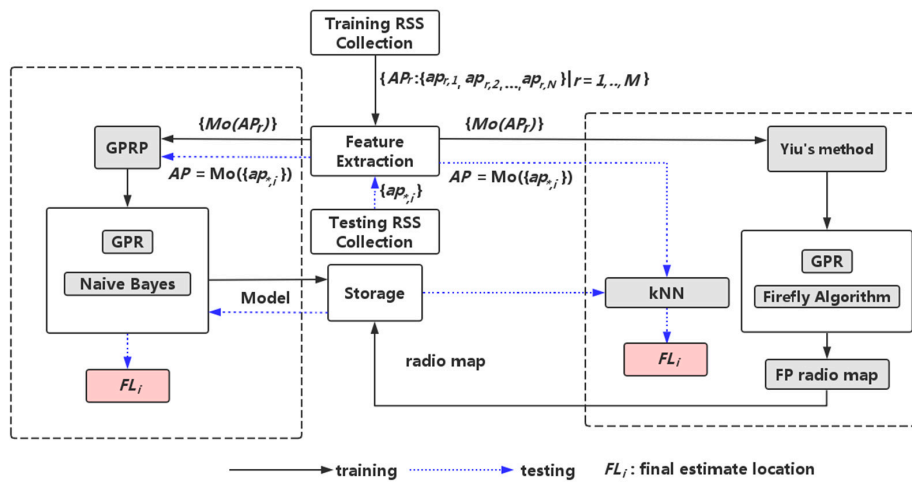

Figure 2. Framework of GPRP's method compared to Yiu's method.

3.1. The Method of Gaussian Process Regression Plus

In our method, an offline RSS fingerprinting database must be built first. We assume that the testing area has M reference points in locations L_r ($r \in \{1, 2, \cdots, M\}$). In the sampling phase, we collect the RSS value at the rth reference point of N APs from 1 to jth times; that is:

$$AP_r = \{ap_{r,1}, ap_{r,2}, \cdots, ap_{r,N}\}$$

where:

$$ap_{r,1} = Mo(\{ap_{r,1}^1, ap_{r,1}^2, \ldots, ap_{r,1}^j\})$$

$Mo(.)$ is the mod function, which is used to derive the remainder of the division. $ap_{r,N}^j$ is denoted as the odd moment data at the rth reference point of N APs from 1 to jth times; by contrast, AP_r is the filtering set of signal strength received at the rth reference point.

When we attempt to predict the unknown position at the input of RSS values, we obtain a series of RSS values AP_* as follows:

$$AP_* = \{ap_{*,1}, ap_{*,2}, \cdots, ap_{*,N}\}$$

Because of the noises in our environment, we can coincide with the regression problem as $L(AP_*) = f(AP_*) + v$ fort the simplification. The latent function $f(.)$ describes the relationship between the RSS and spatial coordinates. v denotes the physical noise; the value would follow the Gaussian distribution.

Thus, the GP is defined as the function of the mean and covariance, which represent the principal characteristic and structure of the model. In addition, the RSS value in location positioning can be derived to the absolute distance; in our study, the authors derived the log-distance path-loss model as the mean function; that is:

$$m\left(AP_*\right) = \frac{\sum_{j=1}^{N} \alpha_j \times \left(C_j - \zeta_j \times log\left(\parallel AP_{*,j} - AP_{*,j}^S \parallel\right) + \varepsilon_j\right)}{\sum_{j=1}^{N} \alpha_j} \tag{5}$$

where C_j is a constant value, $AP_{*,j}$ is the collected online RSS value of unknown position $*$ at jth AP, and $AP_{*,j}^S$ is the RSS value of jth AP at the source of unknown position $*$. α_j represents the weight of $AP_{*,j}$, which indicates the trustable degree of the AP. ζ_j represents the path-loss exponent of jth AP, which is computed as the noise degree of jth AP. In our method, this value follows the Gaussian distribution N $\left(0, \sigma_i^2\right)$; σ^i is the covariance of noises in our experiment. Accordingly, the covariance function is defined as follows:

$$k\left(AP_*, AP_r\right) = exp\left(\frac{\sum_{j=1}^{N} \beta_j \left\|AP_{*,j} - AP_j\right\|}{\sum_{j=1}^{N} \beta_j}\right) \tag{6}$$

where β_j is computed as the weight of AP_j. To reduce the computation complexity, we derive the mod function; that is, we use AP_r^m instead of AP_r in Equation (6).

The objective of our system is to predict the RSS value as $f_n\left(AP_*\right)$ for all two-dimensional inputs x and y. To achieve this, the system models are defined as:

$$f_n\left(AP_*\right) \sim GP\left(\text{m}\left(AP_*\right), k\left(AP_*, AP_r\right)\right)$$

That is, the GP models for the indoor positioning are divided into parts: one part is based on the x and y coordinates of all training locations. Thus, we let $X_n \triangleq [x_1, x_2, \ldots, x_S]$ be a set of training samples, where S represents the number of training coordinate data; then:

$$S_x = \{(x_1, \, y_n\left(x_1\right)), (x_2, \, y_n\left(x_2\right)), \, \ldots (x_S, \, y_n\left(x_S\right))\}$$

and:

$$S_y = \{(x_n\left(y_1\right), \, y_1), (x_n\left(y_2\right), \, y_2), \, \ldots (x_n\left(y_S\right), \, y_S)\}$$

are the S-dimensional column vectors containing the RSS measurements from all training locations.

Accordingly, we define the GP model to predict location as follows:

$$f_x\left(AP_*\right) = \frac{\sum_{j=1}^{N} \alpha_j \times \left(C_j - \zeta_j \times log\left(\left\|AP_{*,j} - AP_{*,jS}\right\|\right) + \varepsilon_j\right)}{\sum_{j=1}^{N} \alpha_j} \\ + exp\left(\frac{\sum_{j=1}^{N} \beta_j \left\|AP_{*,j} - AP_j\right\|}{\sum_{j=1}^{N} \beta_j}\right) : x\epsilon\{1, 2, \cdots, w\} \tag{7}$$

and:

$$f_y\left(AP_*\right) = \frac{\sum_{j=1}^{N} \alpha_j \times \left(C_j - \zeta_j \times log\left(\left\|AP_{*,j} - AP_{*,jS}\right\|\right) + \varepsilon_j\right)}{\sum_{j=1}^{N} \alpha_j} \\ + exp\left(\frac{\sum_{j=1}^{N} \beta_j \left\|AP_{*,j} - AP_j\right\|}{\sum_{j=1}^{N} \beta_j}\right) : y\epsilon\{1, 2, \cdots, h\} \tag{8}$$

We use the hyper parameter vector $\theta \triangleq [C, \alpha, \varepsilon, \beta]$ to optimize our GP model; different parameters would influence the estimated results. In our experiments, we trained hyper parameters by training the data S_x and S_y.

The GPR has many kernels, some of which are listed in Tables 1 and 2. The performance of the GPR was validated using the different sets of mean and covariance kernel functions. The variances were examined in our experiments.

Table 1. Summary of several commonly-used mean functions [26].

Mean Function	Equation Expression
Zero	0
Constant	$\mathbb{C}; \left(x \times x' + \sigma_0^2\right)^p$
Linear	$\alpha x + b$
Poly	$\sum_i^N \alpha^i x^{N-i}$

Table 2. Summary of several commonly-used covariance functions [26].

Covariance Function	Equation Expression
Constant	σ_0^2
Linear	$\sum_{d=1}^{D} \sigma_d^2 x_d x_d'; \left(x \times x' + \sigma_0^2\right)^p$
Polynomial	$\exp\left(-r^2/(2l^2)\right)$
Exponential	$\exp\left(-r/l\right)$
Rational quadratic	$\left(1 + r^2/(2\alpha l^2)\right)^{(-\alpha)}$

Figure 3 presents the results from the GPR model of Equations (7) and (8); the red line reflects the GPR model with six values of S_y, which was the location predicted by the model with the fixed y coordinate. These two red start lines indicated the barriers of the upper and lower covariance values, one cycles red line tracked the change path of the mean values. Besides, the blue line reflects the GPR model with S_x, and with the fixed x coordinate. Two blue start lines and one cycles blue line are represented the same things as red lines. This figure reveals that possible predicting results are barrier inside the covariance lines, and the x and y locations with those derived positions are intersected inside these areas.
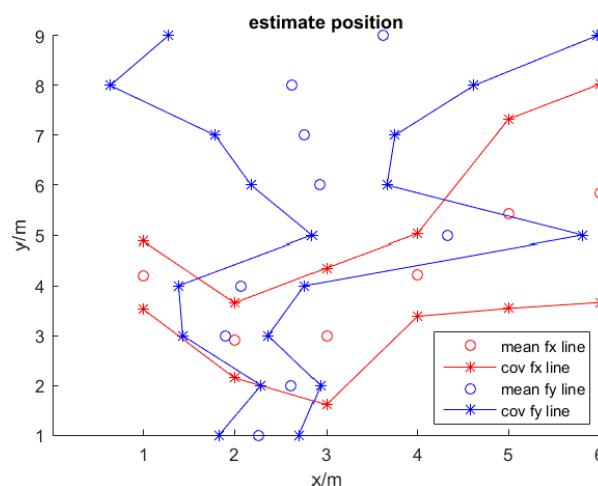

Figure 3. The demonstration of position predicting with GPR, red line is computed with the fixed *y* coordinate, blue line is computed with the fixed *x* coordinate. The results reveal the barriers of the upper and lower covariance values.

3.2. Naive Bayesian Location Model

Our proposed system derived the positions *x* and *y* with S_x and S_y individually, sometimes, S_x and S_y have multiple solutions. To combine these results, our system employed the Naive Bayesian model.

Bayes' theorem is a simple mathematical formula used for calculating conditional probabilities. If E is a random experiment and $B, A_1, A_2, \cdots, A_n$ are events in E, then the following conditions are met:

(1) $P(A_i) > 0$, $i = 1, 2, \cdots, n$,
(2) Events $A_1, A_2, \cdots, A_n$ are partitioned by the sample space; their values are independent of each other,
(3) $P(B) > 0$.

Therefore, $P(A_i | B) = \frac{P(A_i \times B)}{P(B)} = \frac{P(A_i) P(B|A_i)}{\sum_{j=1}^{n} P(A_i) P(B|A_j)}$, $i = 1, 2, \cdots, n$. This equation can be proved using the conditional probability theorem and total probability theorem [27].

Bayesian Decision Theory is built on Bayesian probabilities, which enable us to assert a prior belief in a data point coming from a certain class [28]. This theory is used as a classifier such as the Naive Bayesian classifier, which can classify rapidly and accurately [29,30]. According to our proposed method, RSS values transmitted from APs were all independent, and our proposed GPR model derived some predicted candidate sets of S_x and S_y. To reduce the possible sets, the Naive Bayesian positioning method was used to combine and derive the most satisfactory results. The Naive Bayesian positioning method is detailed as follows:

(1) The sample space is divided by S_x and S_y, and events in the sample spaces are represented as $L_i = (x_i, y_i)$, which is the place derived from the proposed GPR method.
(2) The prior probability $P(L_i)$ is computed as:

$$P(AP_* | L_i) = P(AP_1, AP_2, \cdots, AP_n | L_i) = \sum_{j=1}^{n} P(AP_j | L_i) \tag{9}$$

(3) The posterior probability is computed as:

$$P(L_i | AP_*) = \frac{P(AP_* | L_i)}{P(AP_*)} = \frac{P(AP_* | L_i) P(L_i)}{\sum_{j=1}^{n} P(AP_* | L_j) P(L_j)} \tag{10}$$

Accordingly, the final estimation location is defined as:

$$L_i : P(L_i | AP_*) = \max(P(L_i | AP_*))$$

and if $P(L_i | AP_*)$ is equal to zero, x and y are selected as follows:

$$P(AP_* | x_i) = P(AP_1, AP_1, \cdots, AP_n | x_i) = \prod_{j=1}^{n} P(AP_j | x_i) \tag{11}$$

and:

$$P(x_i | AP_*) = \frac{P(AP_* | x_i)}{P(AP_*)} = \frac{P(AP_* | x_i) P(x_i)}{\sum_{j=1}^{n} P(AP_* | x_j) P(x_j)} \tag{12}$$

then:

$$P(y_i | AP_*) = \frac{P(AP_* | y_i)}{P(AP_*)} = \frac{P(AP_* | y) P(y_i)}{\sum_{j=1}^{n} P(AP_* | y_j) P(y_j)} \tag{13}$$

(4) The final location and error between the estimate positions are computed as follows:

$$P(x_i | AP_*) = \max(P(x_i | AP_*)), P(y_i | AP_*) = \max(P(y_i | AP_*))$$

and:

$$P(y_i | AP_*) = \max(P(y_i | AP_*)), P(x_i | AP_*) = \max(P(x_i | AP_*))$$

We then define the real location as TL_i and $FL_i = \{(x_i, y_i)\}$, and the RMSE is derived as:

$$\text{Error} = \sqrt[2]{\frac{\sum_{i=1}^{N}(TL_i - FL_i)^2}{N}} \tag{14}$$

4. Experiments and Discussion

4.1. Experiment Environment Initialization

In this section, we conducted experiments in two environments. One involved simulated experiments programmed using Matlab 2015. The other environment was the physical experiment, the area of which included four beacons deployed on the corner of a square area that was 30 m × 30 m. In the simulation experiment, the signal was smooth; thus, the RSS energy declined gradually from the source; thus, the received energy could be predicted easily according to the distance from the source. However, the physical experiment was performed on the second floor in the 55th office building of Tianjin University. It was approximately a 7 m × 15 m area (Figure 3). The area was a typical building hall with some pillars in each corner of the area and was relatively open in the center, with people able to walk around freely. Four Bluetooth transmitters were placed in the corner of the area to ensure the signal full coverage of the testing area.

The testing area was divided into 54 locations for area measurements. The locations are marked by black dots in Figure 4. The four Bluetooth transmitters are marked by the red dots. For receiving online RSS data, we employed a Samsung Note 3 as the data collector in the testing area. Each Bluetooth transmitter was designed to transmit 100 signals per second, and our system collected 20 samples in each location; 15 signals were for system training, and five signals were for localization testing. In addition, we applied our GPRP method with the mod function to avoid potential overfitting problems through a single signal measurement.
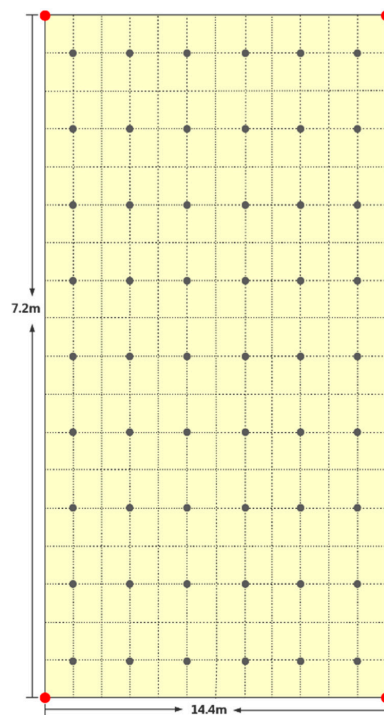

Figure 4. Physical testing area.

4.2. The Effect of the Different Training Data Size

For the purpose of reducing the burden of the RSS data collection in the offline phase in this experiment, the authors attempted to find a balance between using limited training data and increasing the localization accuracy in the emulation environment. Figure 5 shows the cumulative distribution function (CDF) of the localization error for different training dataset sizes. We tested distinct 30%, 60%, and 90% data sizes to evaluate the performance of our proposed method. The results showed that larger training datasets achieve more satisfactory results than do smaller training datasets. The average error rates were 2.98, 2.69, and 2.45 m for the 30%, 60%, and 90% size of datasets, respectively (Figure 5). Comparing to the 2.41 m error rate of whole dataset, we can tell that the performance is acceptable at the saving almost half training time, but only increasing 0.11 times of the error rate.
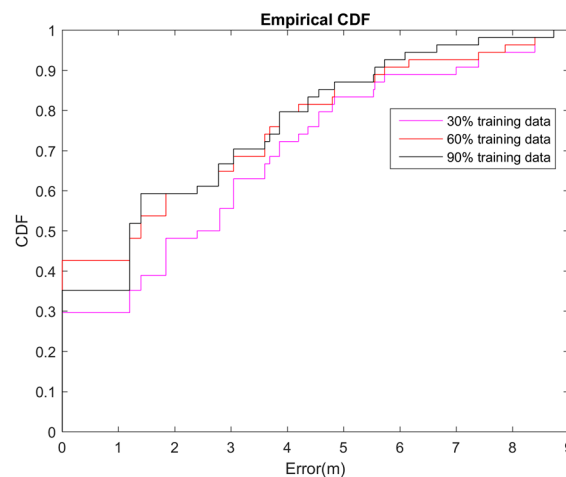
Figure 5. CDF of localization error for different training dataset sizes.

4.3. Effect on Different GPR Kernel

To test location estimation accuracy with different GPR kernels, 30% of the 54 reference locations in the training area (black dots in Figure 2) were selected in this experiment; resets were used for the testing. The GPR was the kernel we adopted in our proposed GPRP method described in Section 3.1. In addition, we compared the derived results with different GPR kernels, it is the CL kernel, the CL kernel was found to be in Section 3.1 too. The CL kernel is based on the constant mean and linear covariance kernel; Tables 1 and 2 show the equation. The popular reason of using CL kernel is because of the simplified computation. The results show that our proposed GPRP method outperformed the CL kernel (Figure 6). Table 3 summarizes the overall results, which reveal that the estimation accuracy of the GPRP method was acceptable, especially inside the 4-m area. The average computation times were 42.80 and 43.81 s for the CL and our proposed method, respectively. The accumulative accuracy of our proposed model was much higher than that of the CL method, though the computation time was only increased by approximately 2%.

Table 3. Localization results in testing area.

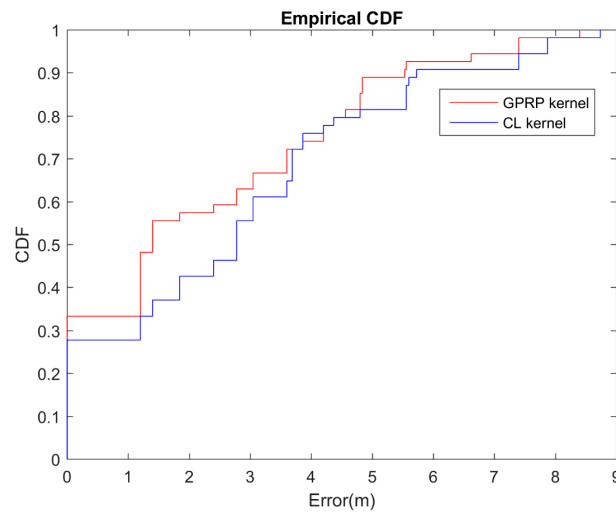|	RMSE (m)	1 m	2 m	3 m	6 m
GPRP method	2.28	33.33%	57.41%	62.96%	92.59%
CL method	2.78	27.79%	42.59%	55.56%	90.74%

Figure 6. CDF of localization error for different kernels comparison.

4.4. Performance Evaluation of GPRP

To check the obstacle effects on the signal transmission in these simulations, we tested the signal transmission on several situations; one situation was tested using the same environment with different obstacles, and the other situation was tested robustly by the different obstacles.

4.4.1. GPRP Robustness on Location Accuracy Testing

According to Equation (1), the WAF is defined as the obstacle factor, which is used to represent the obstacle effects on the RSS. Figure 7a shows the optimal situation with receiving signals; the distribution is displayed as the cyclic shape. In addition, Figure 7b–d indicate strength mapping, with WAF of 1.1, 3.2, and 6.4, respectively These values were indicated as light, medium, and heavy crowds in the environment. According to these pictures, we can observe that the WAF affects the simulation environment.
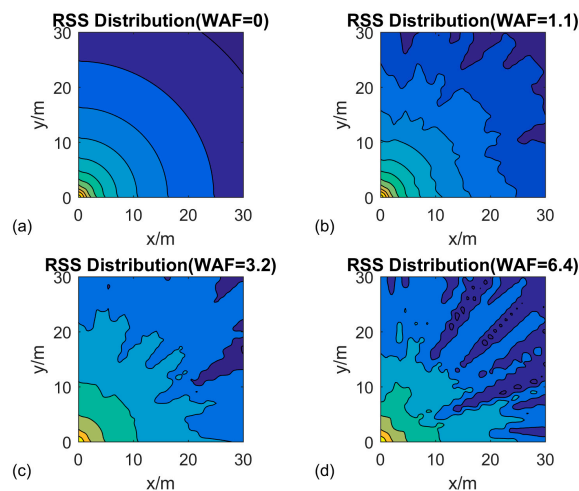

Figure 7. RSS testing with different obstacles.

To test the robustness, the validation was performed according to the changing environment. Thus, the WAF value was changed on the offline training and online testing. We defined four different environments in Table 4. These four environments were represented by the same offline training

and dynamic online testing with light, medium, heavy, and super heavy crowds. Figure 8 shows that the GPRP method could adjust to the dynamic environments in terms of acceptable accuracy. For example, we can observe that the accuracy decrease to less than 10%, even in the light-crowd training to heavy-crowd testing.
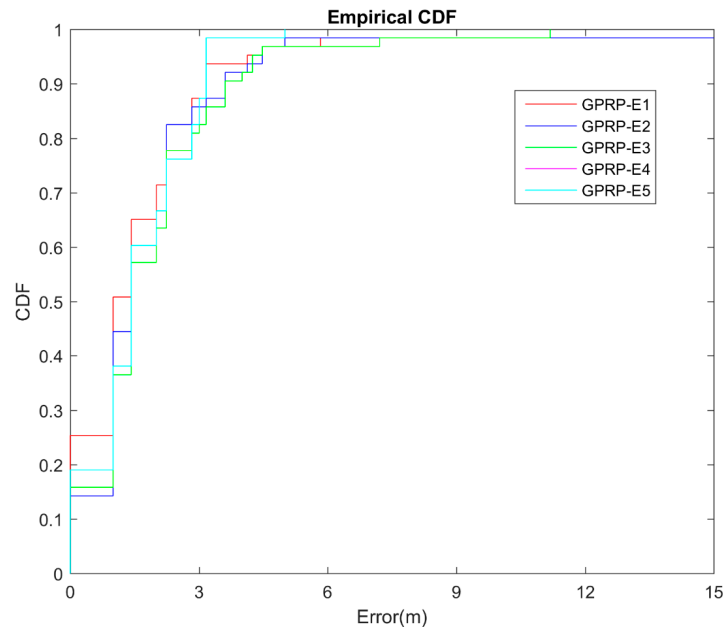

Figure 8. CDF of localization error for different obstacle environments.

Table 4. Robustness testing for four different environment.

ID for Testing Environment	WAF for Off-Line Training	WAF for On-Line Testing
E1	0.2	0.2
E2	0.2	0.3
E3	0.2	0.5
E4	0.2	1
E5	0.2	2

4.5. Localization Accuracy on Different Environments

In these experiments, we evaluated our system performance in two different environments: simulated and field. In the environment of the field experiment, the location examined was the second floor of the 55th building of Tianjin University. For all the experiments, we adopted 30% data for training and 70% data for testing. Firstly, we compared three methods for the general fingerprinting database: our proposed GPRP, the method of Yiu [12], general fingerprinting (FP), and Weight-RSS by Zheng et al. [8]. The FP and Zheng methods are all based on the RSS location system, which is required to build an RSS mapping database of an entire area. Both the computation time and the database preparation are bottlenecks in the system execution.

Accordingly, we employed our proposed method in the actual environment. The area was divided into 6 × 9 blocks. Every block was approximately 1.4 m long and 1.2 m wide. All the experiments were carried out by using 30% data for training and 70% data for training. About the FP and Zheng method, we need to build the fingerprint database, it took 54 × 10 s = 540 s to accomplish, FP method took another 1 s to predict the online location, and Zheng method took more time (2 s) to predict. Furthermore, Yiu method and our proposed method were selected 0.3 percent samples to build the training database, the computation time was 54 × 0.3 × 10 s = 160 s to construct the basement. However, according to our description in Section 2, Yiu's method cost average 25 s to train the GPR model, 75 s to

train parameters in the model, and 46 s to rebuild the fingerprinting database on the average. Overall, Yiu's method would spend about 307 s to finish the prediction. Besides, when our proposed method built the fingerprint database, we only needed 25 s to train the GPRP model, after that, the offline prediction would cost about 125 s to complete this procedure. Overall, our proposed method only cost 205 s to predict the online position on the average. Table 5 showed all the data for these four methods. According to the comparing results in the Table 5, the results reveal that the computation time would be extended with the sampling size of fingerprint database, because the procedure of RSS value re-mapping to the position was changeable. Our proposed method not only can save the computation time of building the RSS fingerprint database, but also can remain the accuracy of online prediction.

Table 5. Total performance of four comparison methods.

	Simulated Experiments		Field Experiments	
	Time Cost (s)	Average Error (m)	Time Cost (s)	Average Error (m)
FingerPrint (FP)	4201	1.46	541	2.34
Yiu method [12]	1385	4.88	307	3.09
Zheng method [8]	4203	1.6	542	2.93
GPRP	1395	1.82	205	2.28

In this experiment of physical area, we compared four methods for the general fingerprinting database: our method GPRP, the method of Yiu, the method of Zheng, and fingerprinting (FP). The graph (Figure 9) below showed all the comparison. The average errors were 2.28, 2.34, 3.09 and 2.93 m for our proposed GPRP method, the FP method, the Yiu method, and Zheng method respectively. The combination of graph and table showed that our proposed method can retain accuracy and reduce computation time in the simulated localization system.
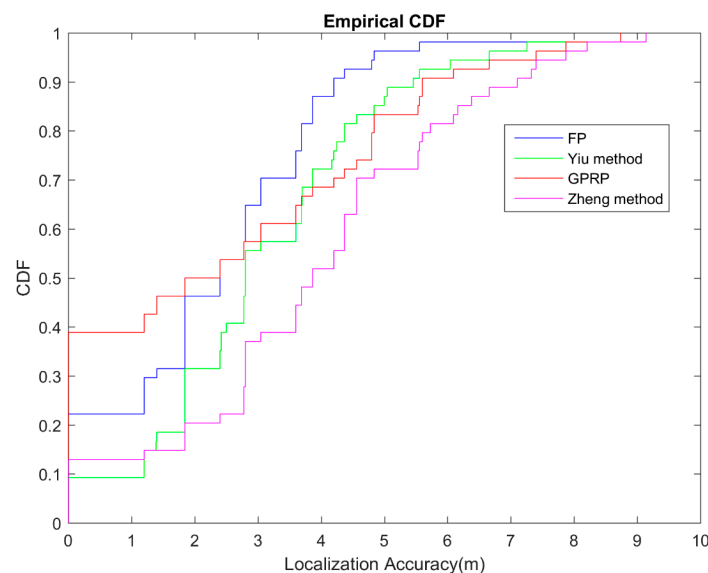

Figure 9. CDF of physical localization error for different algorithms comparison.

In the simulation area, when we compared our method to the FP method, the results showed that the performance of our method was almost the same as that of the FP method (Figure 10); however, our method required only half the computation time of the FP method. This advantage was based on the saving of comparing process from the lookup table. In addition, the Yiu method had the partial advantage of flexibility, though it had the longest computation time of all three methods. The average

errors were 1.82, 1.46, 4.88, 7.2 and 1.6 m for our proposed GPRP method, the FP method, the Yiu method, and Zheng method respectively. The cumulative error rate at 3 m was 0.87, 0.90, 0.34 and 0.85 for the GPRP method, the FP method, the Yiu method and Zheng method, respectively. Overall, the current results still revealed that our proposed method can retain accuracy and reduce computation time in the simulated localization system.
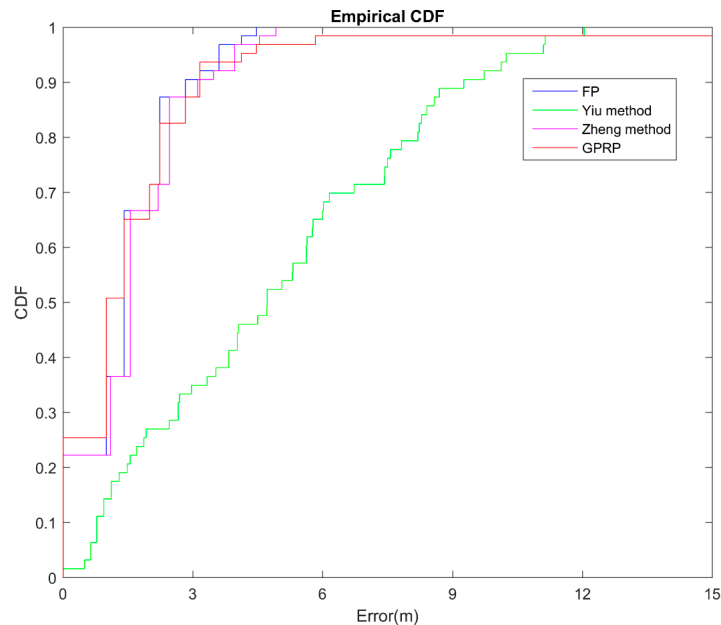

Figure 10. CDF of simulated localization error for different algorithms comparison.

5. Conclusions

Generally, the main challenge in RSS-based location positioning is the high sensitivity of the technique to environmental changes. Variations in RSS measurement reduce estimation accuracy. In other words, if the radio propagation signal strength were correlated with the distance between the transmitter and the receiver, location determination would be a trivial problem. However, the relationship between these two parameters is dynamic rather than straightforward. Hence, many methods have been proposed for obtaining location predictions when the position of the receiver is changing. One such basic method is using a fingerprinting database, which involves straight computation for RSS signal collection. Both time consumption and searching from the lookup table database are bottlenecks in this method.

In this study, we proposed the GPRP method for adapting the flexibility of the GP and determining the location by using the simplified computation of the Naive Bayes method. Both computation time and efficiency can be improved using this method. In our experiments, we tested our method by examining the 55th building of Tianjin University in both a simulated and physical environment. The experiments considered distinct data size, different kernels, and accuracy testing. The results showed that our proposed method can ensure accuracy and reduce computation time in location prediction.

Author Contributions: In this study, Chung-Ming Own, Zhaopeng Meng and Kehan Liu conceived and designed the researches; Kehan Liu performed the experiments; Chung-Ming Own and Kehan Liu analyzed the data; Chung-Ming Own and Zhaopeng Meng wrote the paper. All authors read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1.　Own, C.-M.; Meng, Z.; Liu, K. Handling neighbor discovery and rendezvous consistency with weighted quorum-based approach. *Sensors* **2015**, *15*, 22364–22377. [CrossRef] [PubMed]

2.　Skalar, B. Rayleigh fading channels in mobile digital communication system part 1: Characterization. *IEEE Commun. Mag.* **1997**, *35*, 136–146. [CrossRef]

3.　Yilmaz, H.B.; Tugcu, T. Location estimation-based radio environment map construction in fading channels. *Wirel. Commun. Mob. Comput.* **2015**, *15*, 561–570. [CrossRef]

4.　Haeberlen, A.; Flannery, E.; Ladd, A.M.; Rudys, A.; Wallach, D.S.; Kavraki, L.E. Practical robust localization over large-scale 802.11 wireless networks. *Int. Conf. Mob. Comput. Netw.* **2016**. [CrossRef]

5.　Bisio, I.; Cerruti, M.; Lavagetto, F.; Marchese, M.; Pastorino, M.; Randazzo, A. A trainingless wifi fingerprint positioning approach over mobile devices. *IEEE Antennas Wirel. Propag. Lett.* **2014**, *13*, 832–835. [CrossRef]

6.　Bisio, I.; Lavagetto, F.; Marchese, M.; Sciarrone, A. Smart probabilistic fingerprinting for WiFi-based indoor positioning with mobile devices. *Pervasive Mob. Comput.* **2016**. in press. [CrossRef]

7.　Hossain, A.K.M.; Soh, W.-S. A survey of calibration-free indoor positioning systems. *Comput. Commun.* **2015**, *66*, 1–13. [CrossRef]

8.　Zheng, Z.; Chen, Y.; He, T.; Sun, L.; Chen, D. Feature learning for fingerprint-based positioning in indoor environment. *Int. J. Distrib. Sens. Netw.* **2015**, *2015*, 1–11. [CrossRef]

9.　Xu, C.; Tao, W.; Feng, Z.; Meng, Z. Robust visual tracking via online multiple instance learning with fisher information. *Pattern Recognit.* **2015**, *48*, 3917–3926. [CrossRef]

10.　Xu, C.; Feng, Z.; Meng, Z. Affective experience modeling based on interactive synergetic dependence in big data. *Future Gener. Comput. Syst.* **2016**, *54*, 507–517. [CrossRef]

11.　Bekkali, A.; Masuo, T.; Tominaga, T. Gaussian processes for learning-based indoor localization. In Proceedings of the IEEE International Conference on Signal Processing, Communications and Computing, Xi'an China, 14–16 September 2011; pp. 1–6.

12.　Yiu, S.; Yang, K. Gaussian process assisted fingerprinting localization. *IEEE Internet Things J.* **2015**. [CrossRef]

13.　Youssef, M.; Agrawala, A. The Horus location determination system. *Wirel. Netw.* **2008**, *14*, 357–374. [CrossRef]

14.　Nurminen, H.; Talvitie, J.; Ali-Loytty, S.; Muller, P.; Lohan, E.S.; Piche, R.; Renfors, M. Statistical path loss parameter estimation and positioning using RSS measures in indoor wireless networks, positioning using RSS measurements in indoor wireless networks. In Proceedings of the International Conference on Indoor Positioning and Indoor Navigation, New South Wales Sydney, Australia, 13–15 November 2012; pp. 1–9.

15.　Small, J.; Smailagic, A.; Siewiorek, D.P. Determining User Location for Context Aware Computing through the Use of a Wireless Lan Infrastructure. Available online: http://www-2.cs.cmu.ed/~aura/docdir/small00.pdf (accessed on 26 March 2016).

16.　Zhou, J.; Chu, K.M.; Ng, J.K.-Y. Providing location services within a radio cellar network using ellipse propagation model. *Int. Conf. Adv. Intell. Netw. Appl.* **2005**, *1*, 559–564.

17.　Chen, Y.-C.; Chiang, J.-R.; Chu, H.-H.; Huang, P.; Tsui, A. Sensor-assisted wifi indoor location system for adapting to environmental dynamics. In Proceedings of the 8th ACM International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems, Montreal, QC, Canada, 10–13 October 2005; pp. 118–125.

18.　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]

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

20.　Larios, D.F.; Barbancho, J.; Molina, F.J. Locating sensors with fuzzy logic algorithms. In Proceedings of the IEEE Workshop on Computational Intelligence and Sensor Technology, Paris, France, 11–15 April 2011; pp. 57–64.

21.　Atia, M.M.; Noureldin, A.; Korenberg, M.J. Dynamic online-calibrated radio maps for indoor positioning in wireless local area networks. *IEEE Trans. Mob. Comput.* **2013**, *12*, 1774–1787. [CrossRef]

22.　Farid, Z.; Nordin, R.; Ismail, M. Recent advances in wireless indoor localization techniques and system. *J. Comput. Netw. Commun.* **2013**. [CrossRef]

23. Ciurana, M.; Cugno, S.; Barceló-Arroyo, F. WLAN indoor positioning based on TOA with two reference points. In Proceedings of the 2007 4th Workshop on Positioning, Navigation and Communication, Hannover, Germany, 22 March 2007; pp. 23–28.

24. Krishnakumar, A.S.; Krishnan, P. The theory and practice of signal strength-based location estimation. In Proceedings of the IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing, San Jose, CA, USA, 19–22 December 2005; pp. 1–10.

25. Diono, M.; Rachmana, N. Indoor positioning system based on received signal strength (RSS) fingerprinting. In Proceedings of the 8th International Conference on IEEE Telecommunication Systems Services and Applications (TSSA), Kuta, Indonesia, 23–24 October 2014.

26. Ounpraseuth, S.T. Gaussian Processes for Machine Learning. *J. Am. Stat. Assoc.* **2008**, *103*. [CrossRef]

27. Stephen, M. Thomas Bayes's bayesian inference. *J. R. Stat. Soc.* **1982**, *145*, 250–258.

28. Alpaydin, E. Bayesian decision theory. *Bayesian Anal. Uncertain. Econ. Theory* **2014**, *11*, 47–59.

29. Qiong, W.; Garrity, G.M.; Tiedje, J.M. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **2007**, *73*, 5261–5267.

30. Lin, Y.P.; Chen, Z.P.; Yang, X.L. Mail filtering based on the risk minimization Bayesian algorithm. In Proceedings of the 6th World Multi-Conference on Systemics, Cybernetics and Informatics, Orlando, FL, USA, 14–18 July 2002; pp. 282–285.