

# 基于信号均值滤波+k-means+WKNN 的 Wifi 指纹定位算法研究

汪伦杰<sup>1</sup>, 廖兴宇<sup>2</sup>, 潘伟杰<sup>1</sup>, 吕 健<sup>1</sup>

(1 贵州大学 现代制造技术教育部重点实验室, 贵州 贵阳 550000 ; 2 中南大学 信息科学与工程学院, 湖南 长沙 410006)

**摘 要:** 首先利用平滑均值滤波方法对采集信号的 RSSI 值进行均值平滑处理, 提高信号采集的精度; 其次利用 k-means 聚类算法对指纹库进行优化提高指纹库的精度; 最后在 KNN 算法的基础上采用加权的方法来进行改进, 形成新的指纹算法 WKNN, 并将新的指纹匹配算法 WKNN 应用到文本设计的 Wifi 指纹定位系统中. 实验结果表明: 该改进算法在定位精度和稳定性方面较传统 Wifi 指纹定位算法有大幅度的提高.

**关键词:** 信号均值滤波; k-means 聚类; 加权 KNN; Wifi 指纹定位算法

**中图分类号:** TP393

**文献标识码:** A

**文章编号:** 1000—7180(2017)03—0030—05

## Wifi Fingerprint Location Algorithm in Indoor Location Based on Signal Mean Filter+k-means+WKNN

WANG Lun-Jie<sup>1</sup>, LIAO Xing-Yu<sup>2</sup>, PAN Wei-Jie<sup>1</sup>, LV Jian<sup>1</sup>

(1 Key Laboratory of Advanced Manufacturing Technology, Ministry of Education, Guiyang 550000, China;  
2 College of Information Science & Engineering, Central South University, Changsha 410006, China)

**Abstract:** In this paper, we first make use of the smoothing mean filtering method to smooth the RSSI value of the signal in order to improve the accuracy of signal acquisition. Secondly, we use k-means clustering algorithm to optimize the fingerprint database in order to improve the accuracy of the fingerprint database. Finally, we used a weighted approach to improve the traditional KNN algorithm and form a new fingerprint matching algorithm WKNN. The experimental results show that: Improved algorithm proposed in positioning accuracy and stability than traditional fingerprint Wifi positioning considered to have greatly improved.

**Key words:** signal mean filter; k-means clustering; weighted KNN; fingerprint location algorithm

### 1 引言

基于 Wifi 的位置指纹定位技术具有成本低、易部署、不受视线约束等优点, 已经成为当前室内定位研究的热点. 在基于 Wifi 的室内定位算法的研究中, Alonson J M, Ocana M<sup>[1]</sup> 等提出了在室内环境中利用空间机器人进行无线定位, 并设计一种基于软计算的位置估计系统, 估计待测目标的位置. Galvan-Tejada I<sup>[2]</sup> 等提出了一种 Wifi 与蓝牙相结合的室内定位算法, 利用三角测量的定位方法计算可测到的基站信号, 建立系统方程, 进而估计出位置信

息. Zandbergen P A<sup>[3]</sup> 等人将 Wifi 与 A-GPS、蜂窝定位进行联合定位, 提高了室内定位的精度. Laoudias<sup>[4]</sup> 等利用指纹库匹配的方法开发了 Airplace 系统, 定位精度能够达到 2~4 米. Yang<sup>[5]</sup> 等提出通过监听物理层信道的状态来提高 Wifi 指纹定位系统的精度. Fang<sup>[6]</sup> 等则通过从已测得的 RSSI 信号值中取特征性较强的信号, 从而降低室内环境中无线信号受到的多径效应的影响, 使得精度提高了 40%.

Wifi 指纹定位算法的主要思路是利用指纹库中存储的信号强度值与当前扫描到的信号值进行匹配

**收稿日期:** 2016—07—20; **修回日期:** 2016—08—30

**基金项目:** 国家科技支撑计划项目(2014BAH05F01, 2014BAH05F02, 2014BAH05F03); 贵州省科技计划项目(黔科合 J 字[2013]2108, Z 字[2013]4005); 黔发改投资([2012]2747, [2012]2484)

定位.在信号采集的过程中,如果不进行降噪处理会极大的影响定位精度.而现有的部分定位算法并没有考虑降噪.在Wifi指纹定位算法中,第二个影响定位精度的因素是指纹库的优劣,而现有的定位算法很少考虑指纹库的优化.第三个影响定位精度的因素是指纹匹配算法的优劣,传统指纹匹配采用的是KNN和NN算法,这两种算法都存在参考指纹单一、定位精度不高、稳定性差等弊端.针对上述三个主要的问题,本文提出了一种基于信号均值滤波+k-means+WKNN的Wifi指纹定位算法.首先采用信号均值滤波方法对采集信号过滤,保留精度高的信号强度值,以减小信号采集误差.

## 2 Wifi指纹定位算法原理描述

离线阶段和在线阶段是Wifi指纹定位算法的两个重要阶段.位置指纹信息的提取是离线阶段的主要任务如式(1)所示,根据已有位置指纹信息推算待测节点的位置是在线阶段的主要任务如式(2)所示.算法原理描述如图1所示.

$$FP_i = (x_i, y_i, Wifi_i^1, Wifi_i^2, \dots, Wifi_i^s) \quad (1)$$

式中,  $(x_i, y_i)$  表示虚拟网格中的第  $i$  个采样点的实际坐标.  $Wifi_i^j$  表示采样点  $i$  处采集到来自信号源  $j$  的信号强度值.  $FP_i$  采样点  $i$  处的位置指纹.

$$(\hat{x}, \hat{y}) = \frac{1}{k} \sum_{i=1}^k (x_i, y_i) \quad (k \geq 2) \quad (2)$$

式中,  $(x_i, y_i)$  是单次KNN算法匹配得到的定位位置坐标.  $K$  次KNN算法匹配后求平均值将得到最终的估计位置  $(\hat{x}, \hat{y})$ .

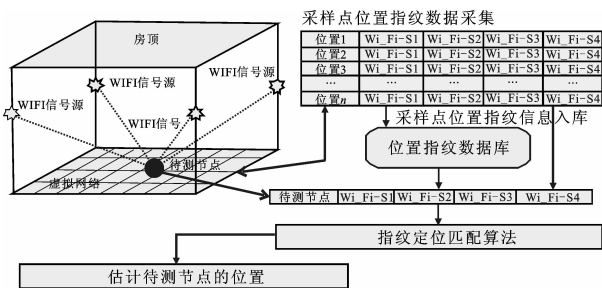


图1 WIFI指纹定位算法原理

### 2.1 离线阶段

①将监测区域划分成若干虚拟网格,将网格顶点定义为采用点.

②提取各采样点处的位置指纹信息存入指纹数据库中.

### 2.2 在线阶段

①当待测节点进入监测区域后,首先实时采集

$N$  组Wifi信号强度值.

②利用一定的指纹匹配算法,依据指纹库中已有的位置指纹信息,推算出待测节点的坐标.

### 2.3 位置指纹库

在选定的区域中,假设有  $n$  个采样点,可以获得  $s$  个AP的RSSI信号值  $Wifi_i^j$ ,那么在每一个采样点可以采集到的  $s$  的RSSI值作为一个指纹,遍历所有采样点便可以得到  $n$  个指纹,保存入库FP,如式(3)所示.

$$FP = \begin{bmatrix} Wifi_1^1 & Wifi_1^2 & \dots & Wifi_1^s \\ Wifi_2^1 & Wifi_2^2 & \dots & Wifi_2^s \\ \dots & \dots & \dots & \dots \\ Wifi_n^1 & Wifi_n^2 & \dots & Wifi_n^s \end{bmatrix} \quad (3)$$

式中,  $Wifi_i^j$  表示在第  $i$  个采样点处测得的第  $j$  个AP的RSSI值.  $FP_i = (Wifi_i^1, Wifi_i^2, \dots, Wifi_i^s)$  是指纹库的一个指纹.每个指纹对应一个唯一的位置  $p(x, y)$ .

## 3 k-means+WKNN的Wifi指纹定位算法

### 3.1 基于均值滤波的Wifi信号采集误差修正

在实际的定位环境中,由于多径效应、绕射、障碍物等不确定性因素的影响,Wifi信号的RSSI值波动比较严重<sup>[7]</sup>.为了有效减小采集数据瞬时波动造成的影响,本文采用均值平滑滤波技术对指纹库(FDB)中的指纹数据进行平滑滤波处理.指纹库均值平滑滤波处理的原理是:

①选择  $k$  个与本参考点  $P_i(x_i, y_i)$  之间的实际距离  $d < \delta$  的参考点  $P_j(x_j, y_j)$  如式(4)所示.

$$P_j(x_j, y_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \leq \delta \quad (j = 1, 2, \dots, k) \quad (4)$$

②取这  $k$  个距离最近的参考点的Wifi信号RSSI值的均值,并将该平均值赋值给  $P_i(x_i, y_i)$  作为RSSI值,如式(5)所示.

$$RSSI(P_i) = \sum_{j=1}^k RSSI(P_j) \quad (5)$$

式中,  $\delta$  表示距离的常数,用来控制均值平滑滤波的平滑程度,通常只能设置成略大于参考点之间的平均距离.  $K$  表示满足要求的参考点数量.

### 3.2 基于k-means聚类算法对指纹库进行优化

聚类算法的思想是将一组特征数据分为一个一个的类,在一个类之内的数据相似度比较高,而类与类之间的相似度比较低.采用k-means聚类算法对指纹库进行优化处理过程中,可以把每个采样点处采集到的来自各个Wifi信号源  $AP_i$  的一组Wifi信

号值作为一个聚类,假如有  $m$  个采用点,  $n$  个 Wifi 信号源,则总共会有  $m \times n$  个聚类。

**定义**  $X$  为  $m \times n$  个聚类的集合.  $X$ .size 为聚类集合  $X$  的大小.  $RSSI_i/\text{cluster}_i$  表示第  $i$  个聚类;  $RSSI_{ij}$  是第  $i$  个聚类中,第  $j$  个信号强度值;  $RSSI_i$ .size 为第  $i$  个聚类的大小.  $\Psi_i$  表示第  $i$  个聚类的中心.  $\Omega$  为信号噪声阈值,当  $|RSSI_{ij} - \Psi_i| > \Omega$  时表示当前的信号强度值  $RSSI_{ij}$  为噪声, k-means 聚类降噪原理如图 2 所示. 则可以将算法原理描述为伪代码如下:

```

①given:  $X$ 
②for  $i=1$  to  $X$ .size
     $i=X$ .size,  $j=RSSI_i$ .size
③ $\phi_i = \sum_{i=1, j=1}^{i=X, \text{size}, j=RSSI_i, \text{size}} RSSI_{ij} / RSSI_i$ .size
④Check: if  $|RSSI_{ij} - \Psi_i| > \Omega$ 
    then Delete( $RSSI_{ij}$ ) from  $RSSI_i$ .
⑤if exist( $|RSSI_{ij} - \Psi_i| > \Omega$ ) == false
    then choose a new cluster  $RSSI_i$  to check.
⑥if all the clusters  $RSSI_i$  ( $i=1, \dots, X$ .size)
has been processed then stop the Algorithm.
  
```

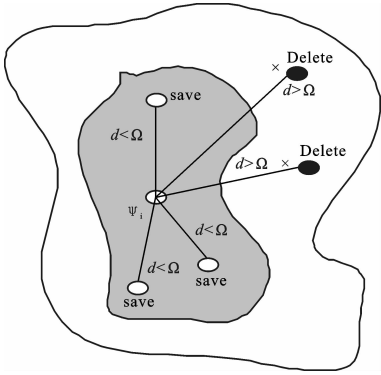


图 2 k-means 聚类降噪原理

### 3.3 基于 WKNN 的指纹匹配算法

KNN 算法(K 近邻算法)是 NN 算法(最近邻算法)的改进算法. 由于 KNN 算法易于实现,算法简单,所以在 Wifi 指纹定位算法中位置估计阶段通常采用 KNN 算法作为位置估计算法. KNN 算法在实际定位阶段会将实时采集到的信号强度值与指纹库进行匹配,寻求最为相似的指纹,最后用这条最为相似指纹的实际位置作为采样点的估计位置. KNN 算法选择的参考指纹较为单一,定位结果不稳定,容易产生较大的定位误差<sup>[8]</sup>. 为了有效解决 KNN 算法存在的参考指纹单一,定位结果不稳定的问题,本文提出了一种加权 KNN 算法(WKNN). 本文在 WKNN 算法中引入了权重系数  $\delta$ ,权重序数  $\delta$  用于

区分不同参考指纹在定位匹配中的不同的重要程度,权重序数  $\delta$  的推导过程如式(6)~(9)所示, WKNN 算法最终估计位置计算如式(10)所示.

$$\overline{RSSI}_i = \frac{1}{n} \sum_{j=1}^n RSSI_{ij} \quad (6)$$

$$\sigma_i = \sqrt{\frac{1}{n} \sum_{j=1}^n (RSSI_{ij} - \overline{RSSI}_i)^2} \quad (7)$$

$$R_i = \sigma_i / \overline{RSSI}_i \quad (8)$$

$$\delta_i = \frac{\sigma_i}{\sum_{i=1}^k R_i} \quad (9)$$

$$(\hat{x}, \hat{y}) = \frac{1}{k} \sum_{i=1}^k (\delta_i x_i, \delta_i y_i) \quad (10)$$

式中,  $RSSI_{ij}$  是第  $i$  个聚类中的第  $j$  个元素,  $\overline{RSSI}_i$  是第  $i$  个聚类的 RSSI 均值.  $n$  为聚类中元素的个数.  $\sigma_i$  为聚类  $i$  ( $RSSI_i$ ) 中 RSSI 值的标准差,  $\sigma_i$  越大表明聚类  $i$  ( $RSSI_i$ ) 中大部分 RSSI 值和其均值 ( $\overline{RSSI}_i$ ) 之间的差异较大,反之则差异较小.  $R_i$  为离散系数表征聚类  $i$  的 RSSI 值标准差  $\sigma_i$  与聚类  $i$  的 RSSI 均值之间的离散化程度.  $K$  为与待测节点最邻近的  $K$  个指纹参考点.

## 4 实验与数据分析

### 4.1 定位实验部署

本项目组以贵州大学机械楼 3 楼的两排实验室以及走道(总共包括:4 个实验室、两条走道)为定位区域进行了定位实验,定位区域地图如图 3 所示. 实验过程中我们采用基于 Android 平台开发出的 APP 采集监测区域中的 Wifi 信号,采用基于 J2EE 平台开发出的定位服务器负责接收客户端上传的数据并对其进行解析、匹配最终实现定位, J2EE 定位服务器端界面如图 4 所示;采用 MySQL 创建指纹数据库,并将其配置到定位服务器上,为服务器端应用程序提供数据支持.

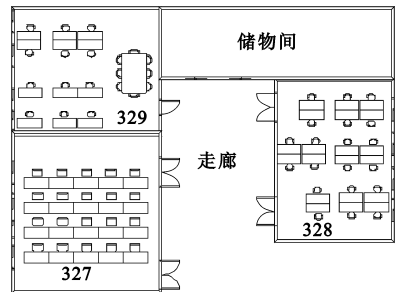


图 3 定位区域地图

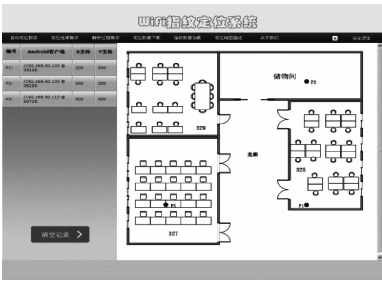


图4 J2EE定位服务器端界面图

4.2 定位实验结果分析

(1)均值滤波效果

针对室内无线信号震荡强烈的问题,本文采用平滑均值滤波方法来对信号强度值进行处理,以提高定位精度.如图5是均值滤波的效果,其中Y轴为RSSI强度值,X轴为距离信号源的距离,黑线表示原始的RSSI值,虚点表示平滑均值滤波后的RSSI值,虚线段表示理想的RSSI值.从图5中可以清晰看出均值滤波效果明显,滤波后的RSSI值比较接近理想RSSI值.

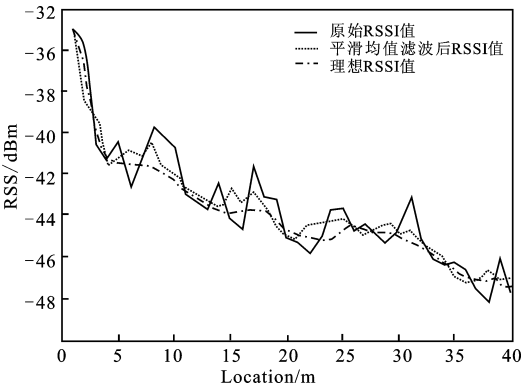


图5 均值滤波效果

(2)k-means 聚类算法对指纹库的优化效果

为了验证经过 k-means 算法优化后的指纹库可以提高定位准确性.本文在如图3所示的定位场景中进行了多次定位试验(在相同的定位场景中,采用相同的定位算法,采用经 k-means 算法优化后的指纹库和未经任何优化的指纹分别进行定位实验)并记录定位结果,然后从多次试验中随机选取2组进行比对分析.如图6是经单次 k-means 算法优化后的指纹库聚类效果,其中图6(a)是聚类前的样本空间,(b)是聚类后的样本空间.图7(a、b)是第1组指纹库优化效果,图7(c、d)是第2组指纹库优化效果.

从图6中可以发现 k-means 算法聚类效果明显,可以很好的将本来离散分布的采样点按照特征

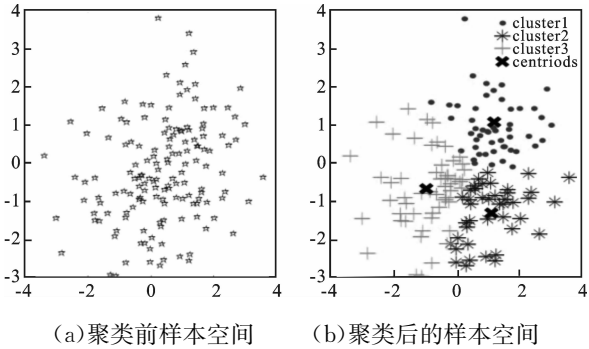


图6 k-means 算法对指纹库的优化效果

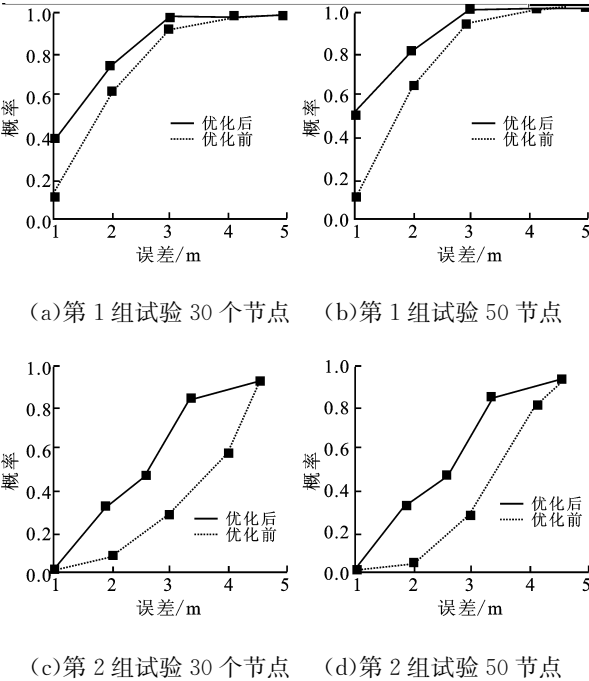


图7 指纹库优化前后的定位效果

信息分成有共性的3类,还可以发现经 k-means 算法优化后的指纹库与未经优化的指纹库相比较优势明显,经 k-means 算法优化后的指纹库定位精度更高.

(3)WKNN 算法对 KNN/NN 算法的优化效果

为了验证 WKNN 算法的性能,本文在如图3所示的定位场景中,采用相同的指纹库,然后采用 WKNN/KNN/NN,3 种不同的算法分别进行定位实验,图8中可以看出,WKNN 算法的收敛距离为2米左右,KNN 算法的收敛距离为3米左右,NN 算法的收敛距离为3.8米左右.还可以看出 NN 算法的稳定性较差,定位结果的精度波动较大,而 WKNN 算法的稳定性比 KNN 和 NN 算法都要好.从三种算法的收敛距离和定位精度稳定性两方面来比较,WKNN 算法的性能要明显优于 KNN 算法和 NN 算法.

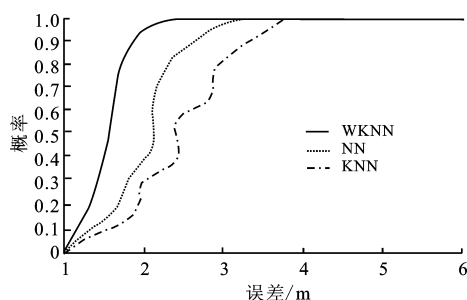


图 8 WKNN/KNN/NN 三种算法定位精度对比

## 6 结束语

本文针对传统 Wifi 指纹定位算法存在的三个突出的问题展开研究. 提出成新的指纹算法 WKNN, 并将新的指纹匹配算法 WKNN 应用到文本设计的 Wifi 指纹定位系统中. 为了验证算法改进的有效性, 我们在真实的场景中进行了定位实验, 实验结果表明: 本文提出的改进算法在定位精度和稳定性方面较传统 Wifi 指纹定位算法有大幅度的提高. 对于推动室内定位技术的发展具有重要的意义.

## 参考文献:

- [1] Alonso J M, Ocana M, Hernandez N, et al. Enhanced WiFi localization system based on soft computing techniques to deal with small-scale variations in wireless sensors[J]. Applied Soft Computing, 2011, 11(8): 4677-4691.
- [2] Galvan Tejada I, Sandoval E I, Brena R. Wifi bluetooth based combined positioning algorithm[J]. proceed Engineering, 2012, 7(4): 101-108.

- [3] Zandbergen P A. Accuracy of iPhone locations: A comparison of assisted GPS, WiFi and cellular positioning[J]. Transaction in GIS, 2009, 13(S1): 5-25.
- [4] Christos Laoudias, Robert Piche, Christos G. Device self-calibration in location systems using signal strength histograms[J]. Journal of Location Based Services, 2013, 7(3): 165-181.
- [5] Yang Z, Zhou Z, Liu Y. From RSSI to CSI: Indoor localization via channel response[J]. ACM Computing Surveys, 2013, 46(2): 1-25.
- [6] Fang S H, Liu T N, Lee K C. A novel algorithm for multipath fingerprinting in Indoor WLAN environments[J]. IEEE Transactions on Wireless Communications, 2008, 7(9): 3579-3588.
- [7] Huaiyu Li, Xiuwan Chen, Guifei Jing, et al. An indoor continuous positioning algorithm on the move by fusing sensors and WiFi on smart phones[J]. Sensor, 2015, 15(12): 31244-31267.
- [8] Wang J, Hu A, Liu C, et al. A floor-map-aided WiFi/pseudo-odometry integration algorithm for an indoor positioning system [J]. Sensors, 2015, 15(4): 7096-7124.

## 作者简介:

**汪杰伦** 女, (1992-), 硕士研究生. 研究方向为数字媒体设计. E-mail: 1530305469@qq.com.

**廖兴宇** 男, (1987-), 博士研究生. 研究方向为大数据与数据挖掘.

**潘伟杰** 男, (1983-), 博士, 副教授, 硕士生导师. 研究方向为文化遗产数字化、数字化艺术与设计、云计算与安全.

**吕健** 男, (1983-), 博士. 研究方向为创新设计、工业设计等.

(上接第 29 页)

- [7] 张国辉, 高亮, 李培根, 等. 改进遗传算法求解柔性作业车间调度问题[J]. 机械工程学报, 2009, 45(7): 145-151.
- [8] Demir Y, İsleyen S K. An effective genetic algorithm for flexible job-shop scheduling with overlapping in operations[J]. International Journal of Production Research, 2014, 52(13): 3905-3921.
- [9] Kacem I, Hammadi S, Borne P. Pareto-optimality approach for flexible job-shop scheduling problems: hybridization of evolutionary algorithms and fuzzy logic [J]. Mathematics and computers in simulation, 2002, 60(3): 245-276.

- [10] Xia W, Wu Z. An effective hybrid optimization approach for multi-objective flexible job-shop scheduling problems[J]. Computers & Industrial Engineering, 2005, 48(2): 409-425.

## 作者简介:

**姚妮** 女, (1978-), 土家族, 硕士, 实验师. 研究方向为智能信息处理、计算智能及其应用.

E-Mail: zzulijsjyn@163.com.

**李红婵** 女, (1983-), 硕士, 讲师. 研究方向为智能信息处理、计算智能及其应用.