

doi:10.3969/j.issn.1001-2400.2015.06.015

室内定位系统中指纹库的优化方法

马鑫迪, 马建峰, 高 胜

(西安电子科技大学 计算机学院, 陕西 西安 710071)

摘要: 指纹库的好坏直接影响到室内定位的定位精度,而现有的室内定位算法大都是建立在原始指纹库的基础之上的,缺乏对指纹库的优化处理.笔者利用 k -means 算法优化指纹库的构建过程,通过聚类的思想剔除已采集指纹库中的“噪点”,从而为室内定位算法提供较好的数据样本源.实验结果表明,与现有的室内定位系统相比,利用优化后的指纹库进行定位能够明显提高室内定位的精度.

关键词: 智能终端; k -means 方法; 指纹库优化; 去噪

中图分类号: TP302.7 **文献标识码:** A **文章编号:** 1001-2400(2015)06-0081-07

Fingerprint optimization method for the indoor localization system

MA Xindi, MA Jianfeng, GAO Sheng

(School of Computer Science and Technology, Xidian Univ., Xi'an 710071, China)

Abstract: The accuracy of the indoor localization is influenced directly by the quality of the fingerprint. But the indoor localization algorithms in existence are almost conducted based on the original fingerprint which is not optimized. The k -means is introduced to optimize the fingerprint in this paper. And deleting the collected bad-points through the theory of cluster could make the fingerprint more accurate for the indoor localization algorithm. Compared with the indoor localization systems in existence, the result of experiments shows that the optimized fingerprint can increase the accurate of indoor localization with a higher probability.

Key Words: smartphone; k -means method; optimize fingerprint; delete bad-point

近年来,基于位置的服务(Location-Based Services, LBSs)得到广泛的应用.作为 LBSs 的核心支撑技术,移动定位技术受到广泛的关注.现有的移动定位技术可分为:室外定位和室内定位,其中室外定位已有成熟的系统,如全球定位系统(Global Position System, GPS)、北斗等,并且已经达到较高的定位精度,能够有效地满足用户在室外环境中对精确性的需求.在室内环境下,用户通常也需要定位自己所在的位置,例如,在大型商场中,用户利用定位系统定位当前的位置,然后按照地图寻找自己需要的商品.然而,在室内环境中,卫星信号穿透建筑物墙壁后,信号衰减非常严重,并且室内障碍物较多,环境比较复杂,进一步引起信号的衰落.因此,室外定位技术无法应用于室内环境中,所以,如何实现高效精确的室内定位已成为研究热点和难点之一.

目前,研究人员致力于部署更加灵敏且精确的室内定位系统^[1-3].与此同时,用于定位过程中的定位感知技术也在不断丰富,如蓝牙技术、超声波定位感知技术、射频识别技术及基于无线局域网(Wireless Local Area Network, WLAN)的室内定位技术等. Laoudias 等^[4]利用指纹库匹配的方法开发了 Airplace 系统,其定位精度可达到 2~4 m; Yang 等^[5]总结了现有的室内定位算法,并利用终端上的传感器开发了基于 Wi-Fi

收稿日期:2014-07-10 网络出版时间:2015-03-13

基金项目:国家自然科学基金委员会-广东联合基金重点基金资助项目(U1135002);国家自然科学基金资助项目(61303221, 61272398);中央高校基本科研业务费专项资金资助项目(JY10000903001)

作者简介:马鑫迪(1989-),男,西安电子科技大学博士研究生, E-mail: xdma1989@163.com.

网络出版地址: <http://www.cnki.net/kcms/detail/61.1076.TN.20150313.1719.015.html>

(Wireless-Fidelity)信号且可在三维空间上实时定位的 LIFS 定位系统; Ozdenizci 等^[6]则提出采用 NFC 标签辅助定位的方法,利用终端具有的 NFC 扫描功能来扫描附近的 NFC 标签进行辅助定位; Yang 等^[7]提出通过监听物理层信道的状态来提高精度; Liu 等^[8]采用多用户协同定位的方法,使用户定位更准确; Fang 等^[9]则通过从已测得的 RSS 信号值中提取特征性较强的信号,从而降低室内环境中无线网络的多径效应,使得定位精度提高约 40%。

另外,在某些研究工作中,部分学者利用 k -means 聚类算法^[10]和主成分分析(Principal Component Analysis, PCA)^[11]的思想对指纹库进行处理,从而降低计算的复杂度并提高定位的精度;而 Fang 等^[12]通过研究,提出动态混合投影(Dynamic Hybrid Projection, DHP)方法对主成分分析和多重判别分析(Multiple Discriminant Analysis, MDA)进行优势互补,使定位精度得到进一步提升。

利用指纹库中存储的信号值与当前扫描到的信号值进行匹配定位是目前室内定位算法的主要思路之一。但是,在采集指纹数据的过程中,由于信号源功率不稳定或者室内环境的变化,会使终端采集到的信号数据波动比较大,笔者将这些波动比较大的数据称为信号“噪点”。现有的室内定位算法在进行定位时大都没有考虑到对指纹库的优化,而指纹库的好坏又直接影响到定位的精度,所以,采用未优化的指纹库会在一定程度上影响到定位精度。因此,如何优化指纹库,剔除指纹库中的信号“噪点”,提高定位精度是要解决的主要问题。

文中采用 k -means 聚类算法^[13]对指纹库进行优化处理,把在每一个位置采集到的每个 AP 的一组信号值作为一个类进行计算,将这组信号值中不属于该类的数据剔除掉,即将采集的信号“噪点”剔除。通过部署实验环境,采集实验数据,并将优化后的指纹库和未优化的指纹库用于同一定位算法进行定位结果对比,结果表明对指纹库进行优化后,定位误差小于 1 m 的概率相比于优化前有明显提高。

1 k -means 聚类算法

聚类思想^[13]是指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,而类之间的数据相似度比较低。

笔者采用 k -means 聚类算法^[13]实现对指纹库的优化过程。在聚类问题中,给定训练样本 $\{x^{(1)}, x^{(2)}, \dots, x^{(n)}\}$ 。然后,利用 k -means 算法将样本聚类成 k 个簇。例如,将太空中的星星进行聚类计算,最终将所有的星星聚集成 k 个星团。首先,需要随机地选取太空中的 k 个点(或者 k 个星星)作为 k 个星团的质心,并对每颗星星计算其到每一个星团质心的距离,选择距离最近的那个星团作为这颗星星所属的类。经计算后,每一颗星星都有了属于自己的星团。最后,对于每个星团,重新计算其质心。重复以上步骤,直到质心不变或者变化很小。

在利用 k -means 算法进行指纹库优化时,可以把指纹库里的信号值看作太空中的星星进行聚类计算,首先,计算每个类的质心,然后,判断该类中的每一个元素是否满足聚类条件属于这个类,如果不属于,则删除元素并重新计算其质心;重复以上步骤直到所有的元素均满足条件为止。

2 指纹库优化算法及定位算法

首先设计了指纹库的优化算法;然后,对定位系统中的定位算法进行了描述,并解释了指纹库优化算法在定位算法中的执行过程。

2.1 指纹库优化算法

采用 k -means 聚类算法对指纹库进行优化处理。在优化处理过程中,可以把每个位置采集到的每个无线接入点(Access Point, AP)的一组信号值作为一个聚类,共有 mn 个聚类。其中, m 为采集的位置个数, n 为采集数据过程中采集到的所有 AP 数。

利用 k -means 聚类算法剔除信号“噪点”的具体实现算法描述如下:

① given:

$$X = \{[x_{11}, x_{12}, \dots, x_{1j}, \dots], [x_{21}, x_{22}, \dots, x_{2j}, \dots], \dots, [x_{i1}, x_{i2}, \dots, x_{ij}, \dots], \dots, [x_{m1}, x_{m2}, \dots,$$

```

 $x_{mj}, \dots \}$ 
② flag=0
③ for  $i = 1$  to  $mn$ 
④ while true
⑤   flag=0
⑥    $\mu_i = \frac{\sum_{j=1}^{x[i][].length} x[i][j]}{x[i][].length}$ 
⑦   for  $j = 1$  to  $x[i][].length$ 
⑧     if  $|x[i][j] - \mu_i| > \Phi$ 
⑨       delete  $x[i][j]$ 
⑩     flag=1
⑪   if flag=0
⑫     break
    
```

X 集合为 mn 组信号值, x_{ij} 为第 i 个聚类中第 j 个信号值, $x[i][].length$ 为第 i 个聚类中元素的个数, $i \in 1, 2, \dots, mn$, 质心 μ_i 表示第 i 个聚类的中心, Φ 为阈值, 当信号值满足 $|x[i][j] - \mu_i| > \Phi$ 时, 认为信号值 x_{ij} 为“噪点”, 将其删除。

假设 $x[i][]$ 集合为终端在位置 (a, b) 处采集到的某个 AP 的信号集合, 则首先计算集合 $x[i][]$ 的平均信号值, 作为第 1 轮的中心 μ_i , 然后再判断集合 $x[i][]$ 中的每个元素是否满足条件属于这个聚类, 如果不属于, 则直接作为坏点删除, 否则会对构建的指纹库影响比较大, 最终会影响到定位的精度。重复以上步骤, 直到集合 $x[i][]$ 中所有的元素均满足条件属于这个聚类为止。利用优化算法对数据库中的数据进行优化处理, 最终计算出 mn 个聚类的中心作为构建指纹库的元素。

2.2 定位算法

将优化后的指纹库应用于加权 k 最邻域 (Weighted K-Nearest Neighbor, WKNN) 算法^[4] 进行定位测试。WKNN 算法的执行流程如图 1 所示。

$$S_i = \frac{1}{((R_{i1} - R'_1)^2 + (R_{i2} - R'_2)^2 + \dots + (R_{in} - R'_n)^2)^{1/2}}, \quad 0 < i \leq m \quad (1)$$

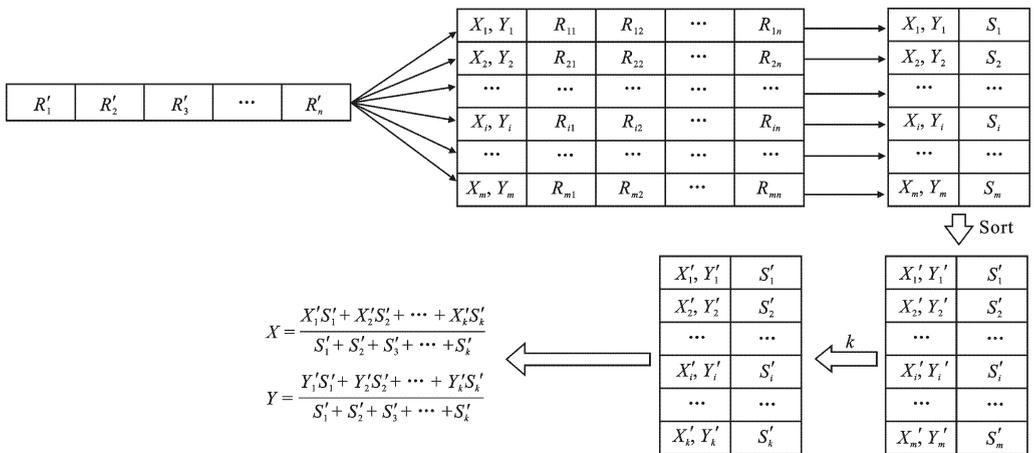


图 1 WKNN 算法执行流程

WKNN 算法执行过程中, 首先需要根据式(1)计算出 $S_i, (X_i, Y_i), R_{i1}, R_{i2}, R_{i3}, \dots, R_{in}$ 为存储在指纹库中第 i 个样本点对应的坐标及 n 个 AP 的信号强度, $R'_1, R'_2, R'_3, \dots, R'_n$ 为用户在当前位置采集到的对应 AP 的信号强度, 这样就会求出集合 $S = \{S_1, S_2, S_3, \dots, S_m\}$ 。集合 S 与指纹库中的 m 个样本点坐标相对应, 然后, 系统将集合 S 按从大到小的顺序排序为集合 S' , 根据参数 k 从集合 S' 中选择出前 k 个 S'_i 对应的坐标进行定

位计算. 由于 S_i 表示当前位置扫描到的信号值与指纹库中处理后的信号值的欧氏距离的倒数, 所以, 可以将集合 S 的元素作为每个坐标的权重, 并利用选择出的 k 个点进行用户位置的计算.

文中所采用的 k -means 算法将用于计算指纹库中的坐标 (X_i, Y_i) 处的元素 $R_{i1}, R_{i2}, R_{i3}, \dots, R_{in}$, 其计算流程如图 2 所示.

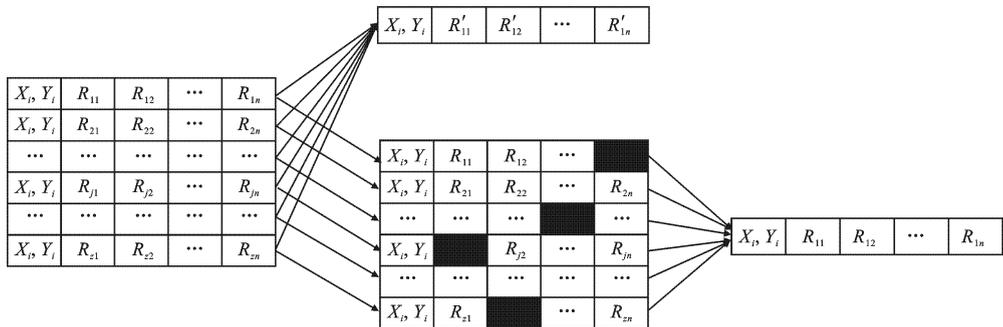


图 2 k -means 算法流程图

假设系统在一个坐标处采集了 z 组数据, 并且每组数据有 n 个 AP 信号值. 首先, 系统从数据库中导出采集的原始数据 $(X_i, Y_i), R_{j1}, R_{j2}, R_{j3}, \dots, R_{jn}$, 然后利用 k -means 算法进行优化计算, 将筛选出的“噪点”删除, 如图 2 中黑色方格所示, 最后将删选后的数据进行求中心得到 $R_{i1}, R_{i2}, R_{i3}, \dots, R_{in}$ 作为终端在坐标 (X_i, Y_i) 处的特征值. 而没有通过 k -means 算法优化的数据, 如图 2 中 $(X_i, Y_i), R'_{i1}, R'_{i2}, \dots, R'_{in}$ 数据所示, 由于存在信号“噪点”, 所以, 得出的 $R'_{i1}, R'_{i2}, \dots, R'_{in}$ 特征值与正常值存在明显偏差, 最终导致定位不准确.

3 实验及评估

首先描述系统的开发环境; 然后利用系统采集数据, 并对数据集进行分析, 观察指纹库是否需要进行分析; 最后对系统进行测试, 并对测试结果进行定位精度分析.

3.1 实验环境及数据分析

由于 Android 终端设备处理能力有限, 并且存在功耗问题, 所以 Laoudias 等开发的 AirPlace 系统存在系统缺陷. 笔者在 AirPlace 系统的基础上进行改进, 将主要计算中心从 Android 终端迁移到服务器端, 使得 Android 终端计算量减小, 从而降低功耗. 系统开发环境如表 1 所示.

表 1 系统开发环境

硬件环境	ASUS F83vf, HTC new Desire T328d
软件环境	Windows 7, Eclipse, Android SDK, Java JDK, MySQL

基于实验室环境采集数据并进行系统测试. 首先, 将系统设定为每个位置采集 30 组数据, 每隔 0.5 s 采集一次, 然后, 选择当前所处的位置, 并采集数据. 采集结果如图 3 所示.

在处理样本集构建指纹库之前, 先对样本集数据进行观察, 随机抽取终端在某个位置扫描到的某个 AP 的一组信号值, 观察其波动是否明显. 表 2 描述了随机抽取数据的波动大小.

由表 2 可得, 在 14 s 和 25 s 处, 数据存在明显波动, 而且如此大的波动幅度会影响系统定位的精度. 所以, 可以认为 14 s 和 25 s 处的数据为信号“噪点”, 若不删除, 必定会影响定位的精度. 因此可以得出, 进行信号数据“去噪”是非常有必要的.



图 3 数据采集

3.2 实验结果评估

为了验证经 k -means 算法优化后的指纹库可以提高定位的准确率, 笔者在实验室环境中多次使用设计的定位系统, 分别采用优化后的指纹库和未优化的指纹库利用相同的定位算法进行定位并记录定位结果, 然

表 2 随机抽取数据波动大小

时间/s	1	2	3	4	5	6	7	8	9	10
信号值/dB	-45	-45	-45	-45	-45	-45	-45	-45	-45	-45
时间/s	11	12	13	14	15	16	17	18	19	20
信号值/dB	-45	-45	-45	-94	-45	-45	-45	-45	-45	-45
时间/s	21	22	23	24	25	26	27	28	29	30
信号值/dB	-45	-45	-45	-45	-82	-45	-45	-45	-45	-45

后对结果数据进行定位误差分析,其中一组定位结果如图 4 所示(黑色圆点表示当前位置,内圆表示离当前位置 1 m 的距离,外圆表示离当前位置 2 m 的距离)。

从多次试验中随机选取 3 组定位结果进行分析,并从每组结果中分别抽取 30 个,50 个,100 个误差点进行统计分析,3 组统计结果分别如图 5~7 所示。

图 5~7 中横坐标表示误差分别为 1 m,2 m,⋯,5 m,纵坐标表示误差在 1 m,2 m,⋯,5 m 内的概率。因为 3 组数据分别采集于不同的时间点,信号干扰程度也不同,所以完全可以表现出优化后指纹库的性能。



(a) 未优化定位结果 (b) 优化后定位结果

图 4 定位结果

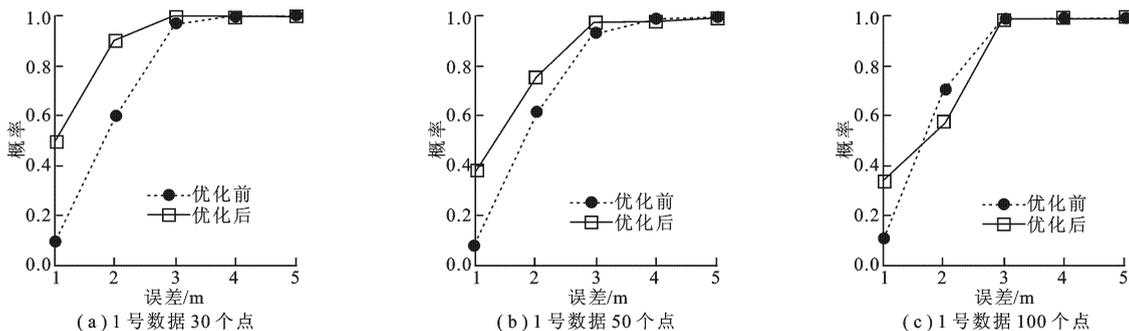


图 5 1 号数据结果分析

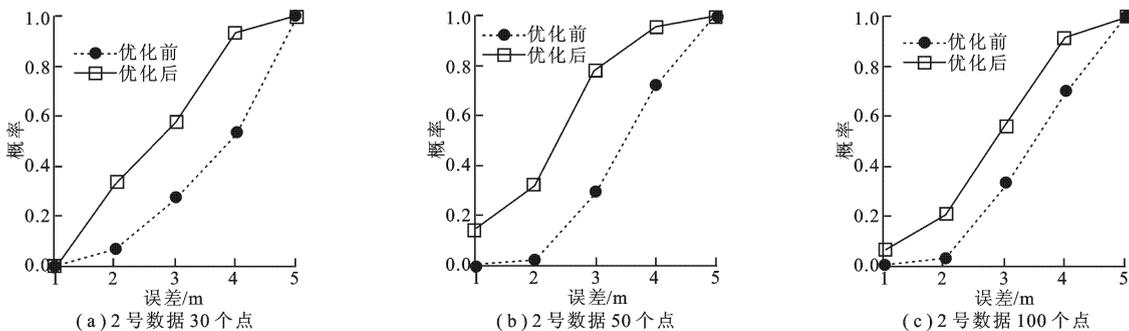


图 6 2 号数据结果分析

对图 5 进行分析,可以得出,当抽取 30 个点进行对比时,采用未优化的指纹库进行定位,误差在 1 m 之内的概率只有 10%,而采用优化后的指纹库进行定位,误差在 1 m 之内的概率达到了 50%;当抽取 50 个点进行对比时,采用未优化的指纹库进行定位的误差在 1 m 之内的概率只有 8%,而采用优化后的指纹库进行定位的误差在 1 m 之内的概率为 38%;当抽取 100 个点进行对比时,采用未优化的指纹库进行定位的误差在 1 m 之内的概率为 11%,而采用优化后的指纹库进行定位的误差在 1 m 之内的概率为 34%。因此,可以得出采用优化后的指纹库进行定位的误差在 1 m 之内的概率更大。

对图 6 进行分析,可以得出,在采集 2 号数据时,信号干扰比较大,导致定位误差比较大,但是也可以从

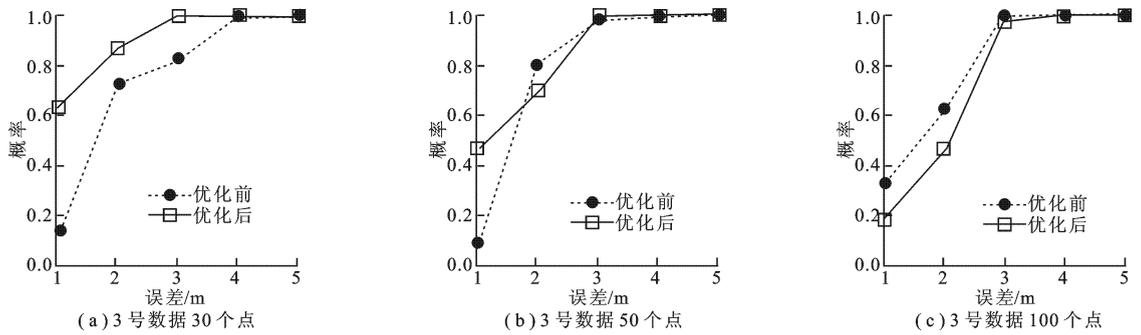


图 7 3 号数据结果分析

统计的结果分析出优化后的指纹库能使定位更精确. 分析结果显示采用未优化的指纹库进行定位的误差在 1 m 之内的概率均为 0, 在 2 m 之内的概率分别为 0.07, 0.02, 0.03, 而采用优化后的指纹库的定位误差在 1 m 之内的概率分别为 0, 0.14, 0.06, 在 2 m 之内的概率分别为 0.33, 0.32, 0.21. 因此, 可同样说明采用优化后的指纹库进行定位能够提高定位的准确率.

对图 7 的分析同图 5 分析, 也能得出同样的结论.

因此得出结论, 采用 k -means 算法进行优化后的指纹库能使定位误差以更大的概率保持在较小的范围内, 也就是说, 优化后的指纹库与未优化的指纹库相比, 可以大大提高定位的精度.

4 结束语

文中引入 k -means 算法能较好的对指纹库进行优化. 实验数据分析表明, 优化指纹库是有必要的. 优化后的指纹库在干扰较小时, 能明显提高定位误差在 1 m 之内的概率; 在干扰较大时, 可以同样使定位误差在 2 m 之内的概率得到明显提高. 因此可得出结论, 与未优化的指纹库相比, 优化后的指纹库可以显著提高定位精度.

参考文献:

- [1] Want R, Hopper A, Falcao V, et al. The Active Badge Location System [J]. ACM Transactions on Information Systems, 1992, 10(1): 91-102.
- [2] Ward A, Jones A, Hopper A. A New Location Technique for the Active Office [J]. IEEE Personal Communications, 1997, 4(5): 42-47.
- [3] Priyantha N B. The Cricket Indoor Location System [D]. Cambridge: Massachusetts Institute of Technology, 2005.
- [4] Laoudias C, Constantinou G, Constantinides M, et al. The Airplace Indoor Positioning Platform for Android Smartphones [C]//Proceedings of the IEEE 13th International Conference on Mobile Data Management. Piscataway: IEEE, 2012: 312-315.
- [5] Yang Z, Wu C, Liu Y. Locating in Fingerprint Space: Wireless Indoor Localization with Little Human Intervention [C]//Proceedings of the 18th Annual International Conference on Mobile Computing and Networking. New York: ACM, 2012: 269-280.
- [6] Ozdenizci B, Ok K, Coskun V, et al. Development of an Indoor Navigation System Using NFC Technology [C]//Proceedings of the 4th International Conference on Information and Computing. Piscataway: IEEE, 2011: 11-14.
- [7] Yang Z, Zhou Z, Liu Y. From RSSI to CSI: Indoor Localization via Channel Response [J]. ACM Computing Surveys, 2013, 46(2): 25.
- [8] Liu H, Gan Y, Yang J, et al. Push the Limit of WiFi Based Localization for Smartphones [C]//Proceedings of the 18th Annual International Conference on Mobile Computing and Networking. New York: ACM, 2012: 305-316.
- [9] Fang S H, Lin 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.

- [10] Bai S, Wu T. Analysis of k -Means Algorithm on Fingerprint Based Indoor Localization System [C]//Proceedings of the IEEE 5th International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications. Piscataway: IEEE, 2013: 44-48.
- [11] Fang S H, Lin T N. Principal Component Localization in Indoor WLAN Environments [J]. IEEE Transactions on Mobile Computing, 2012, 11(1): 100-110.
- [12] Fang S H, Wang C H. A Dynamic Hybrid Projection Approach for Improved Wi-Fi Location Fingerprinting [J]. IEEE Transactions on Vehicular Technology, 2011, 60(3): 1037-1044.
- [13] Wagstaff K, Cardie C, Rogers S, et al. Constrained K-Means Clustering with Background Knowledge [C]//Proceedings of the Eighteenth International Conference on Machine Learning. Williamstown: IMLS, 2001: 577-584.

(编辑: 王 瑞)

(上接第 65 页)

- [14] Cao T V, Aunet S, Ytterdal T. A 9-bit 50MS/s Asynchronous SAR ADC in 28nm CMOS [C]//2012 NORCHIP Conference. Washington: IEEE Computer Society, 2012: 6403105.
- [15] Mueller J H, Strache S, Busch L, et al. A Calibratable Capacitance Array Based Approach for High Resolution CR SAR ADCs [C]//Proceedings of the 19th International Conference on Mixed Design of Integrated Circuits and Systems. Piscataway: IEEE, 2012: 183-188.
- [16] Li Z, Lu Y X, Mo T T. Calibration for Split Capacitor DAC in SAR ADC [C]//Proceedings of the 10th International Conference on ASIC. Washington: IEEE, 2013: 1-4.
- [17] Sun L, Dai Q Y, Qiao G S. Analysis on Capacitor Mismatch and Parasitic Capacitors Effect of Improved Segmented-capacitor Array in SAR ADC [C]//Proceedings of the 3rd International Symposium on Intelligent Information Technology Application. Piscataway: IEEE Computer Society, 2009: 280-283.

(编辑: 郭 华)