

基于多元高斯混合模型的离线指纹数据库

秦宁宁^{*①②} 王超^① 杨乐^③ 孙顺远^①

^①(江南大学轻工过程先进控制教育部重点实验室 无锡 214122)

^②(南京航空航天大学电磁频谱空间认知动态系统工信部重点实验室 南京 211106)

^③(坎特伯雷大学电气与计算机工程系 克赖斯特彻奇 8011)

摘要: 针对室内环境下单次采样测量值的波动变化及信号间的相互干扰, 该文提出一种基于分区多元高斯混合模型(MVGMM)的室内定位系统。根据信号接入点(AP)铺设位置与空间结构, 系统采用一对多支持向量机算法对目标区域做分区操作, 以精确信号变化的区域范围。利用狭小分区内信号间的耦合关系, 建立基于信号间相互干扰的多元高斯混合模型, 以改善信号波动所造成的定位精度下降。当室内环境发生变化时, 基于分区多元高斯混合模型的自适应更新算法可对各分区指纹数据的可信度做出判断, 并以自适应算法更新信号波动较大分区的模型参数, 提高模型与现有环境间的耦合程度。实验结果表明, 该文算法可利用相对少量样本数据, 构建稳定可维护的室内信号分布模型, 相较于其他算法, 其定位精度也有一定程度提高。

关键词: 室内定位; 多元高斯混合模型; 分区; 自适应更新

中图分类号: TN911.7; TP391

文献标识码: A

文章编号: 1009-5896(2021)06-1772-09

DOI: [10.11999/JEIT200226](https://doi.org/10.11999/JEIT200226)

Off Line Fingerprint Database Based on Multivariate Gaussian Mixture Model

QIN Ningning^{*①②} WANG Chao^① YANG Le^③ SUN Shunyuan^①

^①(Key Laboratory of Advanced Process Control for Light Industry of Ministry of Education, Jiangnan University, Wuxi 214122, China)

^②(Key Laboratory of Dynamic Cognitive System of Electromagnetic Spectrum Space, Ministry of Industry and Information Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

^③(Department of Electrical and Computer Engineering, University of Canterbury, Christchurch 8011, New Zealand)

Abstract: For the fluctuation of single sampling measurement value and the mutual interference between signals in indoor environment, this paper proposes an indoor positioning system based on the partition MultiVariate Gaussian Mixture Model(MVGMM). According to the Access Point (AP) position and indoor spatial structure, the system uses SVM classification in “one-against-all” form to partition the target area in order to predict the subarea with signal changes. A MVGMM based on the mutual interference between signals is established by using the coupling relationship between multiple communication devices in the partition. It is important to improve the positioning accuracy which is affected by signal fluctuation. When the indoor environment changes, the adaptive updating algorithm based on the partition MVGMM can test the reliability of fingerprint data in each segmentation. Moreover, it can update the model parameters in the partition with large

收稿日期: 2020-03-31; 改回日期: 2020-08-24; 网络出版: 2020-09-03

*通信作者: 秦宁宁 ningning801108@163.com

基金项目: 国家自然科学基金(61702228, 61803183), 江苏省自然科学基金(BK20170198, BK20180591), 电磁频谱空间认知动态系统工信部重点实验室开放研究基金(KF20202104)

Foundation Items: The National Natural Science Foundation of China (61702228, 61803183), The Natural Science Foundation of Jiangsu Province (BK20170198, BK20180591), The Open Fund of Key Laboratory of Dynamic Cognitive System of Electromagnetic Spectrum Space of Ministry of Industry and Information Technology(KF20202104)

signal fluctuation by the adaptive algorithm to strengthen the coupling relationship between the model and the existing environment. Experimental result demonstrates that the proposed algorithm can build a stable and maintainable indoor signal distribution model by using a relatively small number of sample data. Its positioning accuracy is also improved to a certain extent compared to other algorithms.

Key words: Indoor positioning; MultiVariate Gaussian Mixture Model(MVGMM); Subarea; Adaptive algorithm

1 引言

全球卫星导航系统在室外环境下已被广泛应用于为人们提供位置服务,但信号的缺失也导致该系统无法在复杂的室内环境下发挥作用^[1,2]。WiFi设施的广泛铺设和智能手机的普及,使得基于接收信号强度(Received Signal Strength, RSS)值的室内定位系统,得到了大批研究人员的密切关注。然而,由于无线通讯设施的设计初衷并非为人们提供室内导航,因此如何降低环境波动对无线信号的不确定干扰导致的定位影响,已成为现有研究不得不面对和解决的难点。

基于已有无线设施的定位系统,其常见商用装置不具备自主可编辑功能,仅能提供室内通用RSS测量值,这使得传统基于到达时间差^[3]和到达距离差^[4]等方法,无法直接平移应用。利用测量信号与实际位置间匹配运算的指纹定位算法,弥补了无线设施所发送信号在时间和空间特性上的缺失,可有效实现信号环境与实际场景的映射,为基于RSS值的室内定位提供了可能。

离线指纹通过对目标区域做网格划分,在选定参考点处做多次RSS信号采集,以实现对室内RSS信号分布的拟合。由于墙壁对电磁信号的衰减和反射导致天线的辐射模型并非定向均匀^[5,6],基于辐射全向所构建的信号分布模型如PL模型^[7]仅适用于相对开阔的室外环境,面对复杂室内环境模型无法准确地刻画信号强度在室内空间中的分布状况。针对此问题,传统方法常假设信号室内环境下的分布符合高斯模型^[8],但据文献^[9]分析可知,所得信号分布更趋于左倾模型,故此类假设在相对复杂环境下常导致定位失效。文献^[10]则通过核函数,改进原有关于信号分布的估计算法,以提升定位精度。基于箱型模型的概率方法无需假设信号分布模型,仅通过离散化连续测量值便可实现对信号分布的近似拟合^[11]。然而,该方法过度依赖于箱数目与箱宽度,且当测量数据维度过大时,算法的运行过程将变得极为低效。Zhao等人^[12,13]则利用高斯过程(Gaussian Processes, GP)模型拟合室内环境下的信号分布,但模型需要假设不依赖于测量数据的高斯噪声。

面对大型室内场景,上述方法将耗费大量人力物力,且易受环境因素影响,所构建的离线指纹库

与实际场景中信号分布的映射关系也会因时间变化而减弱,需不断修正离线指纹库,以降低时间积累所造成的映射误差累计。针对大型室内场景下采样数据量大及维护成本高等问题,本文通过分区操作精确维护区域,并根据分区内信号间的耦合关系提出一种分区多元高斯混合模型(MultiVariate Gaussian Mixture Model, MVGMM)以提高对信号分布的拟合程度。模型根据信号接入点(Access Point, AP)位置与物理连通结构对目标区域进行划分,并通过1对多支持向量机模型实现分区操作。在相对狭小的分区区域内,利用信号间存在的相互干扰分别建立多元高斯混合模型,以强化信号的拟合程度,最终达到改善分区定位精度的效果。当环境发生变化时,算法以信息熵作为分区数据更新判据,以及时响应分区变化对指纹库的影响,降低维护成本。从而在室内定位应用中,实现少量数据支撑高效可维护指纹库的构建。

2 构建离线指纹库

2.1 指纹收集方案

随着目标区域的增大,区域内参考点(RP)数量随之增加,这将极大影响在线定位速度,且由信号衰减模型可知,信号RSS值随与AP距离的增大而减小。目标区域过大常产生极端信号值,也会导致在线定位结果出现大偏移现象。根据墙壁对AP信号的遮挡影响以及信号的有效辐射距离,划分区域可对应于几个房间或一段走廊,本文将目标楼层区域分为3类:房间,走廊与楼梯口,走廊区域根据物理连接距离分割为几段,以表征相应区域内AP放置的不同。

区域划分后,使用训练阶段收集的样本数据为每个区域创建分区指纹库,区域训练数据包括区域标记、RP位置及其采集各AP的RSS值。

目标区域由 K 个划分区域 Ω_k , $k \in \{1, 2, \dots, K\}$ 组成,并将区域 Ω_k 划分为 N_k 个网格,取每个网格的几何中心作为参考点 \mathbf{x}_n^k ,其中 $n \in \{1, 2, \dots, N_k\}$, $\mathbf{x}^k = [\mathbf{x}_1^k, \mathbf{x}_2^k, \dots, \mathbf{x}_{N_k}^k]$ 为 $2 \times N_k$ 维位置矩阵表示RP位置。对于各RP位置 \mathbf{x}_n^k ,相应的区域标志为 $\mathbf{y}_n = [y_n^1, y_n^2, \dots, y_n^K]$,其中 $y_n^k = 1$ 且 $y_n^i = 0$, $i \neq k$,表示参考点位于 k 分区。在 \mathbf{x}_n^k 收集到的来自 M_k 个AP的RSS样本值为 $\mathbf{r}_n^k = [r_{n1}^k, r_{n2}^k, \dots, r_{nM_k}^k]$,其中

r_{nm}^k 表示在 \mathbf{x}_n^k 处收集到的来自第 m 个AP的RSS样本值, $m \in \{1, 2, \dots, M_k\}$ 。

2.2 分区模型构建

综合考量区域分类的精确度与效率, 以1对多方式设置支持向量机概率分类模型, 可有效解决根据AP位置与物理连通所设置分区的分类问题^[14]。对于预先设置的各分区, 以目标是否位于本分区进行二分类标识, 通过训练数据构建各分区SVM模型。对于给定 K 个分区, 设立 K 个SVM模型, 取各分区内所有参考AP的测量信号组成当前观测数据 $\mathbf{r} = [r_1, r_2, \dots, r_M]$, 其中 M 为目标所接收到区域内AP数量, 对于未接收到的信号值取为-100 dB。针对当前观测数据, 各分区SVM模型可给出目标是否位于相应区域内的分布概率 $p(y^k = 1|\mathbf{r})$, 其中 y^k 为分区标识, 表示目标位于分区 Ω_k 内, $k \in \{1, 2, \dots, K\}$ 。通过各分区SVM模型给出的分布概率 $p(y^k = 1|\mathbf{r})$, 可对目标所在分区做初步判断, 并作为1级判断依据。

算法采用基于概率SVM的分区操作, 将离线阶段所获取参考点观测数据划分为训练集与测试集, 对分区判断模型进行训练。通过 K 个SVM模型可获取目标位于相应分区的概率值, 但分区交界处信号分布复杂, 易造成分区模型的1级判断失误, 且当目标位于分区交界处时, 其受两分区的影响程度相当, 1级判据无法给出符合精度要求的判别结果。由此, 算法通过设置2级判断依据^[15], 以克服对于分区交界附近的测试点误判断问题。选取被判定区域内2个最大概率的分区区域, 即 $p(y^i = 1|\mathbf{r})$ 与 $p(y^j = 1|\mathbf{r})$, $i, j \in \{1, 2, \dots, K\}$, 且 $p(y^i = 1|\mathbf{r}) > p(y^j = 1|\mathbf{r})$, 其差值可表示为

$$\Delta y_p = p(y^i = 1|\mathbf{r}) - p(y^j = 1|\mathbf{r}) \quad (1)$$

当 $\Delta y_p > \Delta y$ 时, 说明 i 分区对测试点影响力远大于 j 分区, 可将参考点判定于 i 分区, 其中 Δy 为2级判断阈值。对于 $\Delta y_p < \Delta y$, 则将两区域均判定为目标区域, 可分别做相应的区域匹配运算, 并将各自分区所得目标位置做概率平均, 以求取最终位置估计。

2.3 多元高斯混合模型构建

在给定分区内, 可通过分区指纹库与实时观测数据的匹配结果, 确定目标位置。传统指纹库构建方法通常需在参考点处进行过饱和和信息采集, 以频率分布近似区域内各AP信号强度值的概率分布, 分布模型假设与实际分布的映射偏差, 会直接导致离线指纹库的整体偏移。为解决这种苛刻的“精准”化分布模型假设要求, 论文跳脱RP与所获取RSS信号间的单一映射关系, 将目标区域内的RP位置与其RSS值做联合分布假设。

考虑到划分后狭小分区内信号间的相互干扰, 可利用不同AP信号的相关性建立多元高斯混合模型(MVGMM), 通过不断增加高斯元素个数, 以利用不同参数的概率密度函数加权和近似分区内RP位置与所获取各AP信号间的联合概率密度分布^[16], 弥补常见工作中对AP信号间耦合关系的忽略。多元高斯混合模型的概率分布函数可表示为

$$p(\mathbf{x}) = \sum_{c=1}^C w_c p_N(\mathbf{x}|\boldsymbol{\mu}_c, \mathbf{P}_c) \quad (2)$$

其中, C 表示组成元素的个数, $p_N(\mathbf{x}|\boldsymbol{\mu}_c, \mathbf{P}_c)$ 表示均值为 $\boldsymbol{\mu}_c$, 协方差为 \mathbf{P}_c 的高斯组成成分, 权重 w_c 的加和为1。

基于式(2), 利用分区 Ω_k 内RP位置与RSS信号值联合分布的后验概率可将多元高斯混合模型表示为

$$p(\mathbf{x}, \mathbf{r}|y^k = 1) = \sum_{c=1}^{C_k} w_c^k p_N(\mathbf{z}|\boldsymbol{\mu}_c^k, \mathbf{P}_c^k) \quad (3)$$

其中, $y^k=1$ 表示目标处于第 k 个分区, \mathbf{r} 表示在参考点 \mathbf{x} 处接收到各AP信号的RSS值, $\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{r} \end{bmatrix}$, w_c^k 为多元高斯混合模型的组成元素权重, $\boldsymbol{\mu}_c^k$ 为元素均值, \mathbf{P}_c^k 为元素协方差, C_k 为组成元素数量。

为提高MVGMM模型对于目标区域内所采集样本数据的拟合效果, 采用EM算法估计模型参数。MVGMM模型选用 C_k 个均值为 $\boldsymbol{\mu}_c^k$, 协方差为 \mathbf{P}_c^k 的多元高斯函数, 拟合分区 Ω_k 内 N_k 个参考点 $\mathbf{x}^k = [\mathbf{x}_1^k, \mathbf{x}_2^k, \dots, \mathbf{x}_{N_k}^k]$ 与所获取各AP信号RSS值 $\mathbf{R}^k = [\mathbf{r}_1^k, \mathbf{r}_2^k, \dots, \mathbf{r}_{N_k}^k]$ 的联合概率分布。利用K-means算法将样本数据聚类为 C_k 个初始簇, 取各簇平均值与协方差用于初始化EM算法参数且各簇初始权重设置相同。联合概率分布的对数似然形式可表示为

$$\begin{aligned} L(\mathbf{z}^k, \mathbf{w}^k, \boldsymbol{\mu}^k, \mathbf{P}^k) &= \lg \prod_{n=1}^{N_k} \sum_{c=1}^{C_k} w_c p_N(\mathbf{z}_n^k, \gamma_{n,c}^k | \boldsymbol{\mu}_n^k, \mathbf{P}_n^k) \\ &= \sum_{n=1}^{N_k} \lg \sum_{c=1}^{C_k} w_c p_N(\mathbf{z}_n^k, \gamma_{n,c}^k | \boldsymbol{\mu}_n^k, \mathbf{P}_n^k) \end{aligned} \quad (4)$$

其中, $\mathbf{z}^k = [\mathbf{x}^k, \mathbf{R}^k]$, $\mathbf{w}^k = [w_1^k, w_2^k, \dots, w_{C_k}^k]$, $\boldsymbol{\mu}^k = [\boldsymbol{\mu}_1^k, \boldsymbol{\mu}_2^k, \dots, \boldsymbol{\mu}_{C_k}^k]$, $\mathbf{P}^k = [\mathbf{P}_1^k, \mathbf{P}_2^k, \dots, \mathbf{P}_{C_k}^k]$, 隐变量 $\gamma_{n,c}$ 表示第 n 个样本值属于第 c 个高斯组成元素的概率。

由此确定观测数据 \mathbf{z}_n^k 属于第 c 个高斯组成元素的概率可表示为

$$\gamma_{n,c}^k = \frac{w_c^k p_N(\mathbf{z}_n^k | \boldsymbol{\mu}_c^k, \mathbf{P}_c^k)}{\sum_l w_l^k p_N(\mathbf{z}_n^k | \boldsymbol{\mu}_l^k, \mathbf{P}_l^k)} \quad (5)$$

通过式(5)可计算多元高斯函数的权重 w_c^k 、均值 μ_c^k 和协方差 P_c^k 分别为

$$w_c^k = \frac{\sum_{n=1}^{N_k} \gamma_{n,c}^k}{N_k} \quad (6)$$

$$\mu_c^k = \frac{\sum_{n=1}^{N_k} \gamma_{n,c}^k z_n^k}{\sum_{n=1}^{N_k} \gamma_{n,c}^k} \quad (7)$$

$$P_c^k = \frac{\sum_{n=1}^{N_k} \gamma_{n,c}^k (z_n^k - \mu_c^k)(z_n^k - \mu_c^k)^2}{\sum_{n=1}^{N_k} \gamma_{n,c}^k} \quad (8)$$

对不同的 C_k 值重复聚类与EM估计过程，如表1所示，对比不同 C_k 值的拟合效果得到最优分量数的多元高斯混合模型。

采用EM算法更新高斯组成成分时，易产生具有奇异协方差矩阵的组成元素，为避免奇异矩阵的出现，算法通过在协方差矩阵中加入单位矩阵以保证其非奇异性。由于不同室内环境中AP数量与位置具有较大差异，MVGMM模型需根据具体环境选取合适 C_k 对相应区域内的信号分布情况进行拟合。考虑目标区域多为大型室内场景，样本数据复杂且在线匹配运算存在实时性要求，故算法采用贝叶斯信息准则优化MVGMM模型的过拟合状况，以确定最优高斯组成成分个数，避免出现维度灾难现象，数量选取规则可表示为

$$\text{BIC} = C_k \lg(N_k) - 2 \lg(L(z^k, w^k, \mu^k, P^k)) \quad (9)$$

基于上述分析可得，离线指纹库可表示为 $(w^k; \mu^k; P^k)$ ，其中 $k \in \{1, 2, \dots, K\}$ 。

3 在线阶段：目标定位

利用所采集样本数据构建关于目标区域的离线指纹库，通过该指纹库可对区域内目标位置进行估计。若当前时刻目标接受 M 个AP的RSS值为 $r^{\text{now}} = [r_1^{\text{now}}, r_2^{\text{now}}, \dots, r_M^{\text{now}}]$ ，通过分区判断模型可得目标在各分区的分布概率 $p(y_k = 1 | r^{\text{now}})$ ， $k \in \{1, 2, \dots, K\}$ 。

利用当前测量值 r^{now} 可选取对应分区 Ω_k 内对目标定位有效的AP信号测量值 $\bar{r}^{\text{now}} = [r_1^{\text{now}}, r_2^{\text{now}}, \dots, r_{M_k}^{\text{now}}]$ 。根据多元高斯分布的条件概率准则，可得给定观测数据下目标位置的后验概率分布^[17]

$$p(x_n^k | \bar{r}^{\text{now}}, y^k = 1) = \sum_{c=1}^{C_k} w_c^k p_N(x_n^k | \mu_{c,[x]}^k, P_{c,[x]}^k) \quad (10)$$

其中，基于表1所得离线指纹库 $(w^k; \mu^k; P^k)$ ，可得 $\mu_{c,[x]}^k = \mu_{c,[x]}^k + P_{c,[x,r]}^k (P_{c,[r,r]}^k)^{-1} (\bar{r}^{\text{now}} - \mu_{c,[r]}^k)$ (11)

$$P_{c,[x,x]}^k = P_{c,[x,x]}^k - P_{c,[x,r]}^k (P_{c,[r,r]}^k)^{-1} P_{c,[r,x]}^k \quad (12)$$

联合分区概率与式(6)，可得目标在分区 Ω_k 内各参考点的后验分布概率为

$$p(x_n^k, y^k = 1 | \bar{r}^{\text{now}}) \propto p(x_n^k | y^k = 1, \bar{r}^{\text{now}}) p(y^k = 1 | \bar{r}^{\text{now}}) \quad (13)$$

目标位于分区 Ω_k 内各参考点的分布权重可更新为

$$q_n^k = \frac{p(x_n^k, y^k = 1 | \bar{r}^{\text{now}})}{\lambda} \quad (14)$$

表 1 MVGMM模型的参数估计

输入：高斯组成元素数量 C_k ，高斯组成元素的初始权重 \bar{w}_c^k ，初始均值 $\bar{\mu}_c^k$ ，初始协方差 \bar{P}_c^k ，分区 Ω_k 内参考点位置 $x^k = [x_1^k, x_2^k, \dots, x_{N_k}^k]$ 与相应采样值 $R^k = [r_1^k, r_2^k, \dots, r_{N_k}^k]$ ， $c \in \{1, 2, \dots, C_k\}$
输出：高斯组成元素权重 w_c^k ，均值 μ_c^k ，协方差 P_c^k
do
$w_c^k \leftarrow \bar{w}_c^k, \mu_c^k \leftarrow \bar{\mu}_c^k, P_c^k \leftarrow \bar{P}_c^k$
E步：
for $n \in \{1, 2, \dots, N_k\}, c \in \{1, 2, \dots, C_k\}$
计算 $\gamma_{n,c}^k$ //基于式(5)
end
M步：
for $c \in \{1, 2, \dots, C_k\}$
计算 w_c, μ_c^k, P_c^k //基于式(6)–式(8)
end
while $((L(\bar{z}^k, \bar{w}^k, \bar{\mu}^k, \bar{P}^k) - L(z^k, w^k, \mu^k, P^k)) > 0)$
$w_c^k \leftarrow \bar{w}_c^k, \mu_c^k \leftarrow \bar{\mu}_c^k, P_c^k \leftarrow \bar{P}_c^k$

其中, λ 为归一化因子, 则分区 Ω_k 内各参考点为目标位置的分布权重为 $\mathbf{q}^k = [q_1^k, q_2^k, \dots, q_{N_k}^k]$ 。

综上所述, 目标的位置估计为

$$\mathbf{x}^{\text{now}} = \sum_{k=1}^K \mathbf{x}_k^{\text{now}} = \sum_{k=1}^K \sum_{n=1}^{N_k} q_n^k \mathbf{x}_n^k \quad (15)$$

4 在线阶段: 指纹库更新

基于短期过饱和信息采集所构建的离线指纹库, 很难及时有效地跟随场景内信号变化, 从而导致系统需对时变区域进行必要的信息更新和指纹修正。对于非分区定位系统而言, 信息片段的局部变化都会导致指纹库的集体失效。本文给出的区域分类模型, 将目标区域按AP位置分布与物理连通原则进行划分, 使得小范围的区域异动仅对所在子分区的指纹库产生修正影响, 从而降低后期定位系统的维护成本。

4.1 分区判别标准

算法通过分区 Ω_k 内各RP点的分布权重计算其信息熵, 从而实现对分区所构建MVGMM模型可信度的衡量。当信息熵小于阈值时, 表明前期对分区 Ω_k 内各参考点的数据采集已无法表征现有环境中AP信号的RSS分布情况, 需要重新对相应分区进行离线采集, 以提供可表征现有环境因素的采样数据。通过实时观测数据 $\bar{\mathbf{r}}^{\text{now}} = [r_1^{\text{now}}, r_2^{\text{now}}, \dots, r_{M_k}^{\text{now}}]$ 所得参考点后验概率分布的信息熵可表示为

$$H(\mathbf{x}^k, y^k = 1) = - \sum_{n=1}^{N_k} p(\mathbf{x}_n^k, y^k = 1 | \bar{\mathbf{r}}^{\text{now}}) \lg p(\mathbf{x}_n^k, y^k = 1 | \bar{\mathbf{r}}^{\text{now}}) \quad (16)$$

对于 $H(\mathbf{x}^k, y^k = 1) < \Delta h_k$, 则需对分区 Ω_k 所构建模型进行参数更新, 其中 Δh_k 为相应分区信息熵阈值。

4.2 分区模型维护

当分区 Ω_k 中所得信息熵小于分区阈值时, 通过新采集 N_k 个数据 $\mathbf{z}^k = [\mathbf{z}_1^k, \mathbf{z}_2^k, \dots, \mathbf{z}_{N_k}^k]$ 对分区模型进行参数更新。基于现有多元高斯混合模型, 通过采集数据可对模型参数进行自适应调整, 以推导出与现有环境具有更加紧密耦合关系的MVGMM模型。与EM算法一致, MVGMM模型参数的自适应过程也是一个两步估计^[18]。首先, 通过K-means算法将新增观测数据 $\mathbf{z}^k = [\mathbf{z}_1^k, \mathbf{z}_2^k, \dots, \mathbf{z}_{N_k}^k]$ 聚类为 C_k 簇, 并通过聚类所得数据权值 \bar{w}_c , 均值 $\bar{\boldsymbol{\mu}}_c^k$, 协方差 $\bar{\mathbf{P}}_c^k$ 初始化各高斯组成成分。利用式(5)–式(8)可分别计算新增观测数据的统计权重 n_c^k , 值 $E_c^k(\mathbf{z})$ 和协方差 $E_c^k(\mathbf{z}^2)$ 。

利用新增数据的统计参数对分区 Ω_k 所对应的MVGMM模型参数进行自适应优化, 即

$$\hat{w}_c = \left(\frac{n_c^k}{M_k} \alpha_c^w + (1 - \alpha_c^w) w_c \right) \lambda \quad (17)$$

$$\hat{\boldsymbol{\mu}}_c^k = \alpha_c^m E_c^k(\mathbf{z}) + (1 - \alpha_c^m) \boldsymbol{\mu}_c^k \quad (18)$$

$$\hat{\mathbf{P}}_c^k = \alpha_c^v E_c^k(\mathbf{z}^2) + (1 - \alpha_c^v) (\mathbf{P}_c^k + (\hat{\boldsymbol{\mu}}_c^k)^2) - (\hat{\boldsymbol{\mu}}_c^k)^2 \quad (19)$$

其中, $\alpha_i^\rho, \rho \in \{w, m, v\}$ 用于平衡新旧数据对模型参数的影响程度, λ 为归一化因子, ρ 为参数的固有相关因子。新增数据更能体现当前环境内信号的分布状况, 在自适应更新过程中, 如表2所示, 平衡因子 $\alpha_i^\rho, \rho \in \{w, m, v\}$ 更依赖于新增观测数据。

表2 分区MVGMM模型的自适应更新算法

输入: 原始高斯组成元素数量 C_k , 原始高斯组成元素的权值 \hat{w}_c^k , 均值 $\hat{\boldsymbol{\mu}}_c^k$, 协方差 $\hat{\mathbf{P}}_c^k$, 新增分区 Ω_k 内采样数据 $\mathbf{z}^k = [\mathbf{z}_1^k, \mathbf{z}_2^k, \dots, \mathbf{z}_{N_k}^k]$, $c \in \{1, 2, \dots, C_k\}$
输出: 更新后高斯组成元素权值 w_c^k , 均值 $\boldsymbol{\mu}_c^k$, 协方差 \mathbf{P}_c^k
do
$w_c^k \leftarrow \hat{w}_c^k, \boldsymbol{\mu}_c^k \leftarrow \hat{\boldsymbol{\mu}}_c^k, \mathbf{P}_c^k \leftarrow \hat{\mathbf{P}}_c^k$
E步:
for $n \in \{1, 2, \dots, N_k\}, c \in \{1, 2, \dots, C_k\}$
计算 $\gamma_{n,c}^k$ //基于式(5)
end
for $c \in \{1, 2, \dots, C_k\}$
计算 $n_c^k, E_c^k(\mathbf{z})$ 和 $E_c^k(\mathbf{z}^2)$ //基于式(6)–式(8)
计算 $\hat{w}_c, \hat{\boldsymbol{\mu}}_c^k$ 和 $\hat{\mathbf{P}}_c^k$ //基于式(17)–式(19)
end
while($(L(\mathbf{z}^k, \hat{\mathbf{w}}^k, \hat{\boldsymbol{\mu}}^k, \hat{\mathbf{P}}^k) - L(\mathbf{z}^k, \mathbf{w}^k, \boldsymbol{\mu}^k, \mathbf{P}^k)) > 0$)
$w_c^k \leftarrow \hat{w}_c^k, \boldsymbol{\mu}_c^k \leftarrow \hat{\boldsymbol{\mu}}_c^k, \mathbf{P}_c^k \leftarrow \hat{\mathbf{P}}_c^k$

5 仿真分析

5.1 数据处理

试验场景为江南大学物联网工程学院C区某层环形走廊环境，选取移动运营商在学院内均匀铺设的WiFi路由器作为AP信号源。由于AP信号源主要铺设于走廊中，且单侧走廊区域相对开阔，信号受墙壁影响的差异性较小，故将物理结构相对连通且接收AP信号差异较小的单侧走廊区域划分为对应分区，则试验区域可划分为 $K=4$ 个分区。RP位置采用网格拓扑，并以走廊宽幅居中形式排列，相邻RP间隔1 m，共计368个RP点，AP信号源与RP点排布平面图如图1所示。根据各分区内AP信号源的稳定性，选取区域1-4内AP信号源数量分别为{4,5,4,4}。为降低设备差异性对定位算法的影响，实验使用统一型号智能手机进行信号收集。

在所有参考点处采集所有分区所选用共计12个AP的信号强度值，采样间隔为1.2 s，共采集4.8 s (避免因频率原因所导致的数据缓存)，根据第2.1节所述过程构建离线指纹库。测试阶段，实验员手持同款智能手机沿试验区域行走一圈，行进至测试点处通过操作获取实时观测数据，并标记当前位置，测试过程中共获得184个测试点，间隔1 m。

5.2 分区效果和RSSI地图构建效果

目标区域划分完毕后，可对采集参考点做分区标识，通过采样数据与分区标识构建分区模型，构

建过程如第2.2节所述。基于分区模型，可对测试数据进行分区操作，算法将需要启动2级判别准则的测试点划分为区域5以表示信号复杂区域，各分区模型的判别精度如表3所示，各分区的分区精度都在95%以上，已达到后期相应分区内定位操作的精度要求。

基于第2.3节所述MVGMM模型，为验证该算法对RSS指纹地图的构建效果，图2给出目标区域内AP3信号的真实数值图与MVGMM模型和GP模型对其的估计效果图。利用MVGMM模型对目标区域内参考点位置与AP信号的联合分布进行拟合后，区域内AP3信号的RSS分布可使用式(7)和式(8)对相应位置进行调整获得。试验在参考点的观测数据中提取AP3信号的RSS测量值，选取50%测试值作为训练数据，对模型进行训练，利用剩余50%测量值验证RSS分布效果。由于不同分区选用不同AP设施且AP3主要作用于分区1内，故图3给出了分区1中部分参考点处两种模型对AP3信号的拟合效果。通过比较实际测量值与模型估计值，可看出MVGMM模型通过多个高斯组成元素较好地拟合了AP信号在室内环境下的分布状况，尤其体现在分区1内。

5.3 在线定位精度分析

基于已获取MVGMM模型与测试数据，将本文算法(SMVGMM)分别与传统WKNN算法，GP算法做对比，分析算法的定位精度。用户在目标区

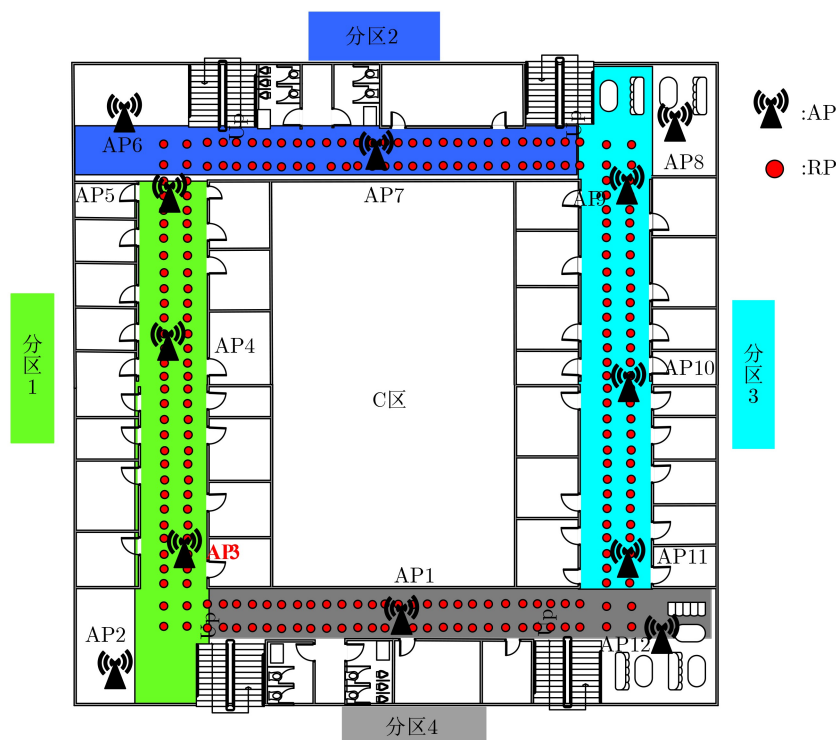


图1 实验场景图

表3 分区间精度(%)

分区	分区精度(预测正确/测试点)
分区1	99.46(183/184)
分区2	98.91(182/184)
分区3	98.91(182/184)
分区4	99.46(183/184)

域内行进1圈, 3种算法的位置估计对比图如图4所示。由图4可知, 本文算法所得目标行进轨迹预测更为平滑, 且相较于GP算法, 其全程定位精度有所提高。图5则给出了3种算法在各测试点的误差值

箱型图。由图可知, 对于目标轨迹的预测, 相比于GP算法, 本文算法的全程定位精度提高了20%以上, 进一步地印证了本文算法所得目标预测轨迹的平滑性。

由于WKNN算法与其余两种算法在定位精度的巨大差异, 图6仅给出本文算法与GP算法位置估计误差的累计概率对比图。由图可知, 本文算法初始的误差累计速度相比于GP算法较慢, 整体效果优于GP算法, 也从另一方面体现出本文算法通过狭小分区内AP信号间的相关性全面提升了传统算法的定位效果。

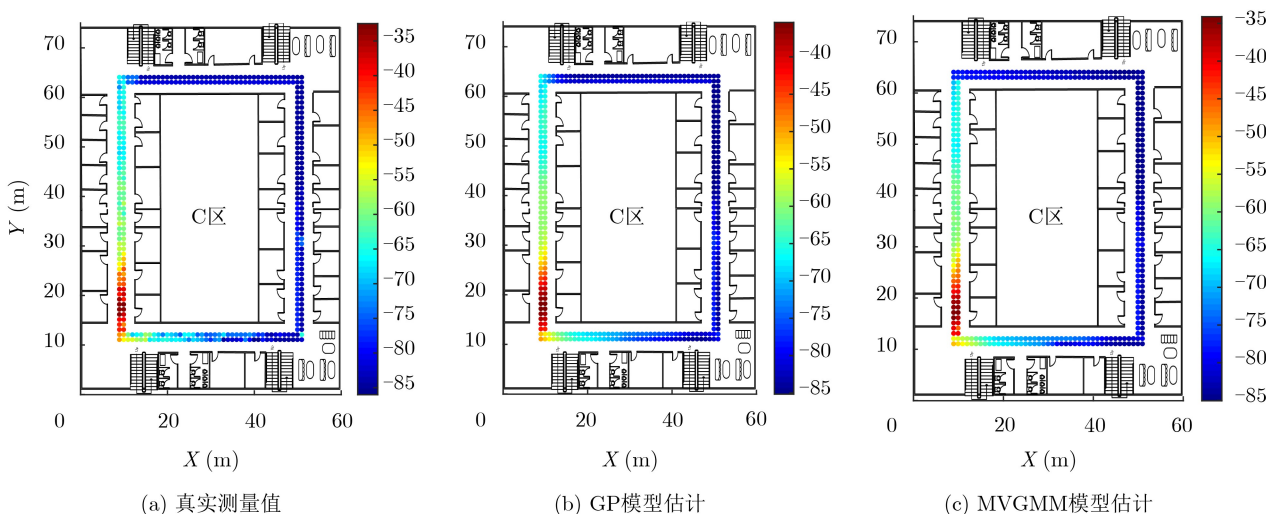


图2 RSS指纹地图构建效果对比图

5.4 指纹库的更新测试

为验证指纹库在线更新的效果与价值, 本实验分两次对目标区域进行数据采集(间隔7 d时间), 利用第2次所采集数据对原始数据所构建MVGMM模型参数进行自适应更新。通过两次采集的测试数据, 比较参数更新前后MVGMM模型对AP3信号的拟合效果。图7给出了参数更新前后模型对AP3信号的拟合状况及其误差。从图中可以看出, 两次采集的测试数据在区域4存在较大差异, 参数更新后模型对最新测试数据的拟合效果优于前次模型的

拟合状况, 尤其体现于区域4内。

6 结束语

针对室内环境中样本数据与采集位置间映射关系的波动变化, 本文利用区域间的物理连通特性对

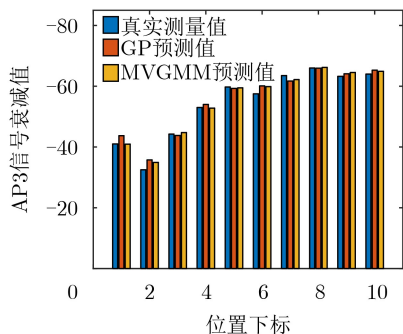


图3 分区1内AP3信号的拟合效果对比图

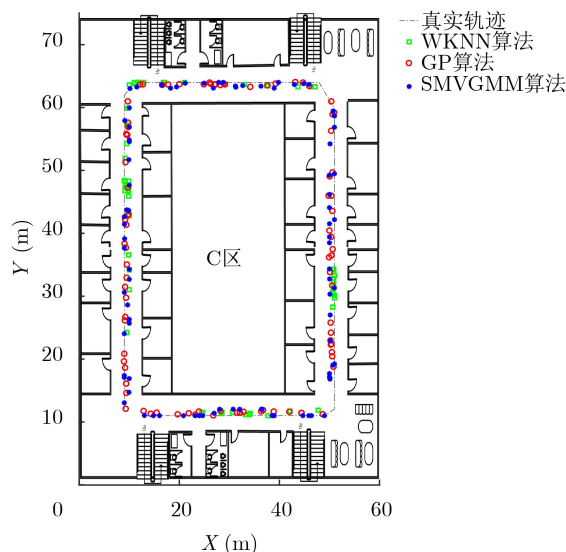


图4 目标运动的轨迹预测对比图

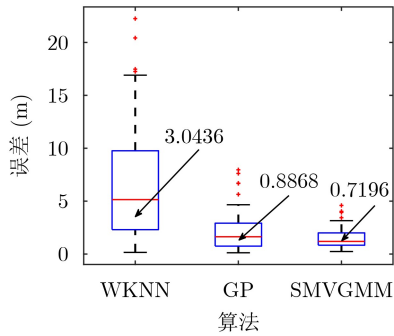


图5 轨迹估计误差箱型图

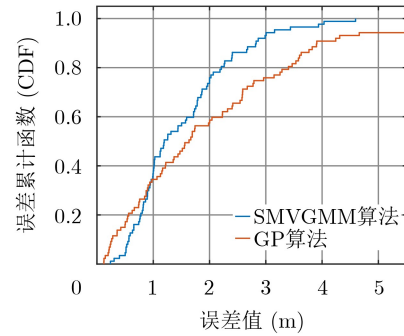
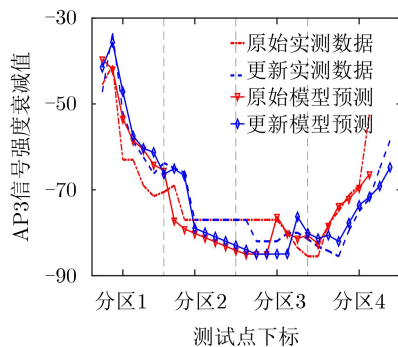
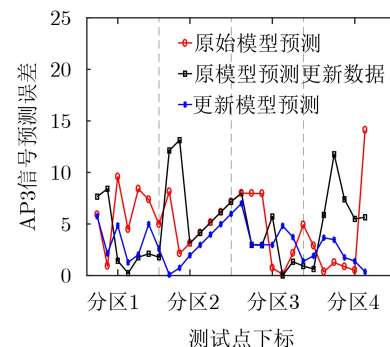


图6 误差累计函数对比图



(a) 参数更新前后模型的拟合状况



(b) 参数更新前后模型的拟合误差

图7 指纹库更新前后AP3数据拟合效果对比图

目标区域进行划分,进而构建基于参考点位置与所采集数据联合概率分布的分区多元高斯混合模型。算法通过分区操作精确室内环境中AP信号的变化区域,强化分区内信号间的耦合程度,以此建立基于分区内信号间相互干扰关系的多元高斯混合模型,并且分区采集也在一定程度上降低了指纹数据库的后期维护成本。实验结果表明,算法可在少量样本数据下拟合信号在室内环境下的分布情况,其定位精度相较于与传统算法也有一定程度提高。

参考文献

- [1] WANG Chun, LUO Juan, and ZHENG Yanliu. Optimal target tracking based on dynamic fingerprint in indoor wireless network[J]. *IEEE Access*, 2018, 6: 77226–77239. doi: [10.1109/ACCESS.2018.2880247](https://doi.org/10.1109/ACCESS.2018.2880247).
- [2] 江小平, 王妙羽, 丁昊, 等. 基于信道状态信息幅值-相位的被动式室内指纹定位[J]. *电子与信息学报*, 2020, 42(5): 1165–1171. doi: [10.11999/JEIT180871](https://doi.org/10.11999/JEIT180871).
JIANG Xiaoping, WANG Miaoyu, DING Hao, et al. Passive fingerprint indoor positioning based on CSI amplitude-phase[J]. *Journal of Electronics & Information Technology*, 2020, 42(5): 1165–1171. doi: [10.11999/JEIT180871](https://doi.org/10.11999/JEIT180871).
- [3] 左燕, 陈志猛, 蔡立平. 基于约束总体最小二乘的单站DOA/TDOA联合误差校正与定位算法[J]. *电子与信息学报*, 2019, 41(6): 1317–1323. doi: [10.11999/JEIT180655](https://doi.org/10.11999/JEIT180655).
- [4] ZUO Yan, CHEN Zhimeng, and CAI Liping. Single-observer DOA/TDOA registration and passive localization based on constrained total least squares[J]. *Journal of Electronics & Information Technology*, 2019, 41(6): 1317–1323. doi: [10.11999/JEIT180655](https://doi.org/10.11999/JEIT180655).
- [5] PATWARI N, HERO A O, PERKINS M, et al. Relative location estimation in wireless sensor networks[J]. *IEEE Transactions on Signal Processing*, 2003, 51(8): 2137–2148. doi: [10.1109/TSP.2003.814469](https://doi.org/10.1109/TSP.2003.814469).
- [6] 李世宝, 王升志, 刘建航, 等. 基于接收信号强度非齐性分布特征的半监督学习室内定位指纹库构建[J]. *电子与信息学报*, 2019, 41(10): 2302–2309. doi: [10.11999/JEIT180599](https://doi.org/10.11999/JEIT180599).
LI Shibao, WANG Shengzhi, LIU Jianhang, et al. Semi-supervised indoor fingerprint database construction method based on the nonhomogeneous distribution characteristic of received signal strength[J]. *Journal of Electronics & Information Technology*, 2019, 41(10): 2302–2309. doi: [10.11999/JEIT180599](https://doi.org/10.11999/JEIT180599).
- [7] 刘坤, 吴建新, 甄杰, 等. 基于阵列天线和稀疏贝叶斯学习的室内定位方法[J]. *电子与信息学报*, 2020, 42(5): 1158–1164. doi: [10.11999/JEIT190314](https://doi.org/10.11999/JEIT190314).
LIU Kun, WU Jianxin, ZHEN Jie, et al. Indoor localization algorithm based on array antenna and sparse Bayesian learning[J]. *Journal of Electronics & Information Technology*, 2020, 42(5): 1158–1164. doi: [10.11999/JEIT190314](https://doi.org/10.11999/JEIT190314).

- [7] ACHUTEGUI K, MÍGUEZ J, RODAS J, *et al.* A multi-model sequential Monte Carlo methodology for indoor tracking: Algorithms and experimental results[J]. *Signal Processing*, 2012, 92(11): 2594–2613. doi: [10.1016/j.sigpro.2012.03.017](https://doi.org/10.1016/j.sigpro.2012.03.017).
- [8] KHALAJMEHRABADI A, GATSIS N, and AKOPIAN D. Modern WLAN fingerprinting indoor positioning methods and deployment challenges[J]. *IEEE Communications Surveys & Tutorials*, 2017, 19(3): 1974–2002. doi: [10.1109/COMST.2017.2671454](https://doi.org/10.1109/COMST.2017.2671454).
- [9] CHEN Lina, LI Binghao, ZHAO Kai, *et al.* An improved algorithm to generate a Wi-Fi fingerprint database for indoor positioning[J]. *Sensors*, 2013, 13(8): 11085–11096. doi: [10.3390/s130811085](https://doi.org/10.3390/s130811085).
- [10] HONKAVIRTA V, PERALA T, ALI-LOYTTY S, *et al.* A comparative survey of WLAN location fingerprinting methods[C]. The 6th Workshop on Positioning, Navigation and Communication, Hannover, Germany, 2009: 243–251. doi: [10.1109/WPNC.2009.4907834](https://doi.org/10.1109/WPNC.2009.4907834).
- [11] ZHANG Tianyu, ZHAO Qian, SHIN K, *et al.* Bayesian-optimization-based peak searching algorithm for clustering in wireless sensor networks[J]. *Journal of Sensor and Actuator Networks*, 2018, 7(1): 2. doi: [10.3390/jsan7010002](https://doi.org/10.3390/jsan7010002).
- [12] ZHAO Yuxin, FRITSCHÉ C, YIN Feng, *et al.* Sequential Monte Carlo methods and theoretical bounds for proximity report based indoor positioning[J]. *IEEE Transactions on Vehicular Technology*, 2018, 67(6): 5372–5386. doi: [10.1109/TVT.2018.2799174](https://doi.org/10.1109/TVT.2018.2799174).
- [13] ZHAO Yuxin, LIU Chao, MIHAYLOVA L S, *et al.* Gaussian processes for RSS fingerprints construction in indoor localization[C]. The 21st International Conference on Information Fusion, Cambridge, UK, 2018: 1377–1384. doi: [10.23919/ICIF.2018.8455842](https://doi.org/10.23919/ICIF.2018.8455842).
- [14] HE Suining, TAN Jiajie, and CHAN S H G. Towards area classification for large-scale fingerprint-based system[C]. 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Heidelberg, Germany, 2016: 232–243. doi: [10.1145/2971648.2971689](https://doi.org/10.1145/2971648.2971689).
- [15] 傅惠民, 杨海峰, 张勇波, 等. 基于空间特征分区和前点约束的WKNN室内定位方法[J]. *软件学报*, 2019, 30(11): 3427–3439. doi: [10.13328/j.cnki.jos.005569](https://doi.org/10.13328/j.cnki.jos.005569).
FU Huimin, YANG Haifeng, ZHANG Yongbo, *et al.* WKNN indoor positioning algorithm based on spatial characteristics partition and former location restriction[J]. *Journal of Software*, 2019, 30(11): 3427–3439. doi: [10.13328/j.cnki.jos.005569](https://doi.org/10.13328/j.cnki.jos.005569).
- [16] LI Yan, WILLIAMS S, MORAN B, *et al.* High-dimensional probabilistic fingerprinting in wireless sensor networks based on a multivariate Gaussian mixture model[J]. *Sensors*, 2018, 18(8): 2602. doi: [10.3390/s18082602](https://doi.org/10.3390/s18082602).
- [17] RAITOHARJU M, GARCÍA-FERNÁNDEZ Á F, HOSTETTLER R, *et al.* Gaussian mixture models for signal mapping and positioning[J]. *Signal Processing*, 2020, 168: 107330. doi: [10.1016/j.sigpro.2019.107330](https://doi.org/10.1016/j.sigpro.2019.107330).
- [18] REYNOLDS D A, QUATIERI T F, and DUNN R B. Speaker verification using adapted Gaussian mixture models[J]. *Digital Signal Processing*, 2000, 10(1/3): 19–41. doi: [10.1006/dspr.1999.0361](https://doi.org/10.1006/dspr.1999.0361).

秦宁宁: 1980年生, 教授, 博士, 研究方向为无线传感器网络的研究.

王超: 1995年生, 硕士生, 研究方向为室内定位.

杨乐: 1979年生, 副教授, 研究方向为无线信号统计与应用.

孙顺远: 1984年生, 副教授, 研究方向为无线传感网.

责任编辑: 余蓉