

DOI: 10.19650/j.cnki.cjsi.J2311407

# 空间指纹测量特征双精简下的室内定位算法\*

郑安琪, 秦宁宁

(江南大学物联网技术应用教育部工程研究中心 无锡 214122)

**摘要:**针对空间结构阻隔、信号弱穿透力等因素导致的参考点和接入点定位匹配冗余问题,提出一种“水平精简参考点,垂直精简接入点”的空间双精简定位算法。首先以最强接收信号的高阶统计信息替代传统均值表征各参考点,并结合小区域融合和边界参考点共享的处理方式,实现目标空间模糊聚类,以此弱化边缘绝对判别的不良影响;其次基于该降维子空间,综合衡量各接入点的空间区分度和覆盖可靠性,为各子空间筛选出高识别价值、高稳定性的精简接入点集合;最后通过判断最强接收信号信源执行一级区域判别,并利用 WKNN 算法实现二级位置估计。经实际路演测试,所提水平精简策略聚类规整也更符合场景结构约束,垂直精简策略较传统接入点选配算法平均定位精度至少提升 17%,并在参考点  $1\text{ m}\times 1\text{ m}$  的分布密度条件下,滤除了约 4.5 m 以上的大定位误差。

**关键词:** 室内定位;指纹定位;模糊聚类;接入点筛选

**中图分类号:** TN92 TH89 **文献标识码:** A **国家标准学科分类代码:** 510.40

## Indoor localization algorithm with dual refinement of spatial fingerprint measurement features

Zheng Anqi, Qin Ningning

(Engineering Research Center of Internet of Things Technology Applications, Ministry of Education, Jiangnan University, WuXi 214122, China)

**Abstract:** Considering the redundancy of localization matching between reference point and access point due to the spatial structure obstruction and weak signal penetration, a spatial dual-refinement localization algorithm named “horizontal refinement of reference points and vertical refinement of access points” is proposed. Firstly, the traditional mean value is replaced with the high-order statistical information of the strongest received signal to characterize each reference point, and the processing methods of small-area fusion and boundary reference point sharing are combined to achieve the fuzzy clustering in the target space, so as to weaken the adverse effect of absolute discrimination at the edges; secondly, based on the dimensionality-reduced subspace, the spatial differentiation and coverage reliability of each access point are measured comprehensively to screen out the subspace with high recognition value and high stability. Finally, the first-level area discrimination is performed by judging the strongest received signal source, and the second-level location estimation is achieved with the WKNN algorithm. The proposed horizontal streamlining strategy possesses the regular clustering and complies with the structural constraints of the scene more precisely according to the realistic roadshow test. It's found that the vertical streamlining strategy improves the average positioning accuracy by at least 17% compared to the traditional access point selection algorithm, and filters out the large positioning error of more than 4.5 m under the condition of distribution density of  $1\text{ m}\times 1\text{ m}$  reference points.

**Keywords:** indoor positioning; fingerprint positioning; fuzzy clustering; access point filtering

## 0 引言

随着智慧物联网和移动互联网的普及,基于位置服务(location based services, LBS)的应用需求持续不断。

以 GPS、BDS 为代表的卫星导航系统在室外定位中发展成熟,但受卫星信号的弱穿透性的影响,卫星导航技术无法有效应用于复杂多变的室内环境中。因此将高效精准的定位核心技术由室外延伸到室内,成为了位置服务领域的一个研究热点。主流的室内定位技术有无线网络

收稿日期:2023-05-09 Received Date: 2023-05-09

\* 基金项目:国家自然科学基金(61702228)、江苏省自然科学基金(BK20170198)、中国高校产学研创新基金(2021ITA10003)项目资助

(WiFi)<sup>[1]</sup>、蓝牙 (bluetooth, BT)<sup>[2]</sup>、射频识别 (radio frequency identification, FRID)<sup>[3]</sup>、超宽带 (ultra wide band, UWB)<sup>[4]</sup>等。鉴于大多数移动智能终端对 WiFi 模块的支持以及室内无线接入点 (access point, AP) 的广泛部署,基于 WiFi 信号的指纹定位技术以其极大的硬件优势被广大研究者所青睐<sup>[5-6]</sup>。其基本思想是借助无线接收信号强度 (received signal strength, RSS) 在不同空间参考点 (reference point, RP) 处的差异性,构建位置-指纹的对应关系,从而以“匹配”的思想实现目标位置的判别。

K 近邻算法 (k-nearest neighbor, KNN)<sup>[7]</sup>、加权 K 近邻算法 (weighted k-nearest neighbor, WKNN)<sup>[8]</sup>作为经典指纹定位算法,默认待定位点 (又称测试点 test point, TP) 匹配全局离线指纹数据库。但当目标空间较大或遮挡严重时,全局匹配方案不但伴随着较大的缓存压力而且会干扰目标位置的确定。因此,可将大型指纹库进行参考点聚类预处理,以保证“匹配 RP 精简,共享 AP 充分”的局部区域优势。基于该思想,文献[9]提出了一种基于信号强度的 K-means 聚类算法,有效改善了聚类中心选择的复杂度,但随机的初始质心和预先设置的簇值,容易导致聚类陷入局部最优,影响定位精度;He 等<sup>[10]</sup>采用谱聚类算法优化定位性能,对数据具备较强的适应性,但很可能因降维幅度不足而拉低运行速度;最强 AP 聚类算法<sup>[11]</sup>,以最强信号信源作为唯一分类判断标准,将不再受限于初始簇值,也不依赖任何距离度量,但均值弱化离线信号波动影响的同时,却忽视了短暂的在线采集信号强度的不确定性带来的区域判别误差。

此外,AP 作为指纹定位的另一重要影响因素,其高效的筛选也十分重要。正如文献[12]所述,为保证通信稳定,运营商所铺设 AP 数量远超本地定位所需;另一方面,在每个区域中具有决定性影响的 AP 通常也是不完全相同的<sup>[13]</sup>。故此,现有研究主要从离线和在线两个角度分析各 AP 性能以实现目标空间内 AP 的有效精简。离线时,文献[14]结合费舍尔 (Fisher) 判别准则,反映类间和类内的样本离散程度,以此构建 AP 的选择框架;Chen 等<sup>[15]</sup>基于信息增益 (information gain, IG) 准则评估各接入点特征的区分价值,并通过筛选信息增益最佳的接入点集提升鉴别纯度;在线时,最大平均值 (max mean, MM) 算法<sup>[16]</sup>通过对目标位置平均接收信号排序,实现指纹特征的相似度动态对比;在线互信息 (online mutual information, OMI)<sup>[17]</sup>选配策略则是以接入点之间的依赖关系为出发点,滤除冗余指纹特征,有效降低计算负载并提高定位精度。以上探究均已证明 AP 精简的必要性,但都需预先确定所选 AP 数量,仍存在过度筛选或筛选不足而导致的粗劣匹配问题。

论文针对聚类算法初始化干扰和边界参考点的绝对判别影响,提出了基于模糊聚类的水平精简 RP 算法,以

单一信源 AP 的信号强度分布特征为导向,结合参考点位置分布信息,以增强空间分类的学习能力和泛化能力,并减小判别子空间的传递误差;针对接入点匹配冗余问题,提出了面向双尺度分析的垂直精简 AP 算法,综合考虑各信源 AP 之间的区分度和可靠性,以此垂直筛选出贡献度充足的子空间 AP 集合。共同构建“水平精简 RP,垂直精简 AP”的空间双精简策略下的指纹定位算法 (fingerprint localization algorithm with spatial dual-refinement strategy, SDRS),降低参考点匹配复杂度的同时,提高位置估计精度。

## 1 多变环境下系统模型构建

基于 WiFi 信号的指纹定位算法,利用同一接入点辐射的接收信号强度刻画距离差异,从而递进在线信号和离线指纹的相似度匹配。

根据目标空间实际定位需求、物理结构分布等,设置空间物理坐标为  $l_n = (x_n, y_n)$  的  $N$  个参考点 RP,  $n \in \{1, 2, \dots, N\}$ , 并表示为  $\mathbf{RP} = \{P_{R_1}, P_{R_2}, \dots, P_{R_N}\}$ ;此外,记录目标空间内  $M$  个可探测 WiFi 接入点 AP, 以  $\mathbf{AP} = \{A_{P_1}, A_{P_2}, \dots, A_{P_M}\}$  表示,且  $m \in \{1, 2, \dots, M\}$ 。

离线阶段,采集来自各 AP 辐射于各 RP 处的接收信号强度值 RSS,并录入离线指纹库中,其中将第  $m$  个接入点  $A_{P_m}$  在第  $n$  个参考点  $P_{R_n}$  处共被采集  $Q$  次的接收信号强度样本值 RSS 表征为向量  $\mathbf{RSS}_{nm} = [rss_{nm}^1, rss_{nm}^2, \dots, rss_{nm}^q, \dots, rss_{nm}^Q]$ ,  $rss_{nm}^q$  表示  $P_{R_n}$  处第  $q$  次采集来自  $A_{P_m}$  的 RSS 值,且  $q \in \{1, 2, \dots, Q\}$ 。

在线阶段,在目标位置处实时接收来自  $M$  个 AP 的 RSS 值,共同组成接收信号强度向量  $\mathbf{RSS}_{TP} = [rss_{TP_1}, rss_{TP_2}, \dots, rss_{TP_m}, \dots, rss_{TP_M}]$ ,其中,  $rss_{TP_m}$  表示空间物理坐标为  $l_{TP} = (x_{TP}, y_{TP})$  的待定位点 TP 处接收来自接入点  $A_{P_m}$  的 RSS 样本值,将其与离线指纹库进行相似度比较,再通过匹配算法实现位置确定。

综上,目标空间 RP 部署、离线信息采集、在线信息对比,共同构建了如图 1 所示的目标空间下的指纹定位系统模型。

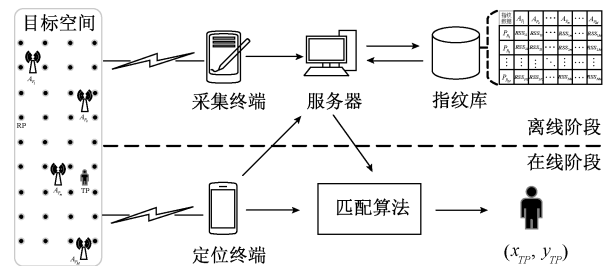


图 1 指纹定位系统模型

Fig. 1 Fingerprint positioning system model

## 2 空间双精简策略

### 2.1 空间双精简思想

受物理空间遮蔽、信号弱穿透力等客观因素的约束,具备决定性影响的定位参量往往并不相同。故在线定位共用全局参考点和相同接入点往往并非最佳方案。为减少不必要的匹配负担和提高定位精度,论文采用“水平精简 RP,垂直精简 AP”的空间双精简策略,如图 2 所示,将全局参量的整体映射转化为各子空间内筛选样本的精简映射,以此构建参考点位置和接入点信号的高质量子空间映射关系库。

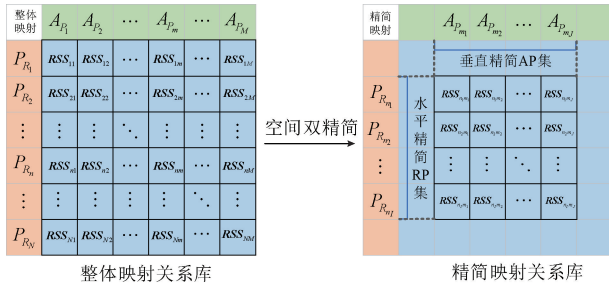


图 2 空间双精简示意图

Fig. 2 Schematic of double streamlined space

### 2.2 水平精简 RP

由于邻近 RP 最强接收信号极有可能来自同一 AP,经典的最强 AP 聚类算法以最大信号均值作为聚类参考标准,将属于同一 AP 最强覆盖范围内的 RP 聚为一类<sup>[18]</sup>,即:

$$\max(\mathbf{RSS}_n) = r_{SS_{nm}} \Rightarrow P_{R_n} \in \Omega_m \quad (1)$$

式中:  $\mathbf{RSS}_n = [r_{SS_{n1}}, r_{SS_{n2}}, \dots, r_{SS_{nm}}, \dots, r_{SS_{nM}}]$  表示  $P_{R_n}$  处接收来自  $M$  个接入点 AP 的信号均值向量,且  $r_{SS_{nm}} = (1/Q) \sum_{q=1}^Q r_{SS_{nm}}^q$ ;  $\Omega_m$  表示最强接收信号来自  $A_{p_m}$  的子空间。该方法简单易行,且无需预先定义集群数量,避免了人为干预的影响。

然而,非理想定位环境下,仅通过探索信号均值表征各参考点很可能丢失原始数据中的高阶统计信息;此外,区域绝对判别容易导致在线阶段差异较小的 RSSI 指纹误匹配的问题,且无法在后续定位过程中被纠正。为此,论文提出了一种基于最强 AP 接收信号的模糊聚类方法。以最强 AP 接收信号频率映射空间 RP 分布特征,对强干扰区具备空间划分决定性质的接入点模糊处理,并兼顾小区域误判和边界信息空白等问题,对目标空间进行聚类优化,以此完成 RP 的水平精简。

### 1) 高频接入点模糊划分

假设在参考点  $P_{R_n}$  处,采集到的最强 RSS 值,共计  $num_{nm}$  次来自接入点  $A_{p_m}$ ,不失一般性,该最强 AP 接收信号频率可定义为  $num_{nm}/Q$ 。此外,为均衡子空间判别精度和在线匹配速度,论文通过设置场景最佳模糊参数  $\delta_{th}$ ,灵活地调整不同场景的模糊范围,提高了算法的普适性。基于此,建立子空间判别模型:

$$num_{nm}/Q \geq \delta_{th} \Rightarrow P_{R_n} \in \Omega_m \quad (2)$$

式中:  $\sum_{m=1}^M num_{nm}/Q = 1, 0 \leq num_{nm} \leq Q$ ,当接收到多个最强 AP 信号频率大于  $\delta_{th}$  时,则判断  $P_{R_n}$  归属多个集群。其中, $\delta_{th}$  值越小,集群归属门槛越低,进而形成越大范围的模糊区域,TP 与  $\Omega_m$  匹配准确率越高,与此同时,也将伴随更多的匹配计算量。不难发现,当  $\delta_{th} > 0.5$  时,模糊区域的设置将变得没有意义,甚至因  $\delta_{th}$  值过大而存在部分 RP 分类漏洞,故此,  $\delta_{th}$  的最佳探索区间为  $(0, 0.5]$ 。

### 2) 小区域融合处理

最强 AP 聚类算法十分依赖单一信源所辐射的信息特征,而移动信源或者突发噪声很可能引发少数甚至个别 RP 独立成区的现象,使得在线子空间判别困难。为避免最强 AP 聚类带来的极小区域集成误判问题,论文将指纹数量少于所有子区域平均指纹数量的小区域  $\Omega_{m_1}$ ,与其具备高度相似性的大区域  $\Omega_{m_2}$  融合处理,即  $\Omega_{m_2} = \Omega_{m_1} \cup \Omega_{m_2}$ 。基于各子区域信号均值向量所表征的聚类中心  $C_{\Omega_m}$ ,采用式(3)所示的皮尔逊相关性分析:

$$\rho(\Omega_{m_1}, \Omega_{m_2}) = \frac{\text{cov}(C_{\Omega_{m_1}}, C_{\Omega_{m_2}})}{\sigma C_{\Omega_{m_1}} \sigma C_{\Omega_{m_2}}} \quad (3)$$

式中:  $\text{cov}(C_{\Omega_{m_1}}, C_{\Omega_{m_2}})$  表示  $C_{\Omega_{m_1}}$  和  $C_{\Omega_{m_2}}$  的协方差,  $\sigma C_{\Omega_{m_1}} \sigma C_{\Omega_{m_2}}$  则表示  $C_{\Omega_{m_1}}$  和  $C_{\Omega_{m_2}}$  的标准差之积。当  $\Omega_{m_1}$  和  $\Omega_{m_2}$  的相关性  $\rho(\Omega_{m_1}, \Omega_{m_2}) > 0.8$  时,默认二者高度相关<sup>[19]</sup>,并将两区域融合处理,以此实现大区域兼并小区域,有效克服不良分类伴随的区域误判问题。

### 3) 边界参考点共享

在区域划分过程中需要注意的一个关键点是两个相邻区域应共享边界处的参考点,以便它们可以无缝接壤。若边界上的 RP 不共享,如图 3(a) 所示,两个边界之间将存在一个空间不属于两个区域中的任何一个。而在这个盲区,由于没有 RP,用户将无法调用可靠信息。因此,论文默认边界处的 RP 同时归属于近邻区域,如图 3(b) 所示,以此避免边缘信息空白问题。

综上,兼顾信号聚类特征和实际物理分布,通过水平精简 RP 完成目标空间划分,以实现低成本约束在线定位误差范围。并以  $\{P_{R1}^\phi, P_{R2}^\phi, \dots, P_{Rn}^\phi, \dots, P_{Rn}^\phi\}$  标记子空间  $\Omega_\phi$  内的 RP 精简集,其中,  $N_\phi$  表示  $\Omega_\phi$  所包含的 RP

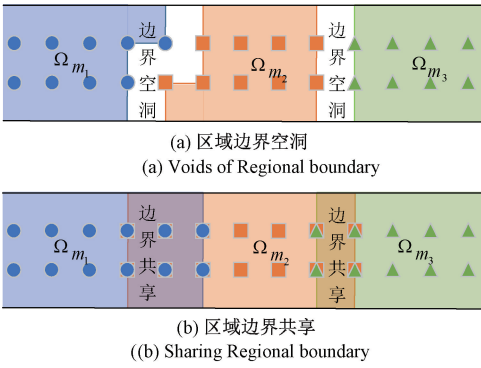


图 3 相邻区域边界划分方式

Fig. 3 Adjacent area boundary demarcation methods

数量,利用  $\phi$  标记子空间以便区分 AP 标签  $m$ 。该方案所确定的各子空间 RP 精简集及对应指纹信息,将进一步用于后续研究。

### 2.3 垂直精简 AP

离线采集的某些 AP 指纹信息并非适用于空间所有位置的在线定位,故有必要在水平精简 RP 后的每个子空间内对用于最终定位的 AP 进行精简。为筛选出贡献度最高的精简 AP 集合,论文从 AP 作用于子空间内的辐射区分度和覆盖可靠性两个角度综合分析,提出了一种新型的接入点 AP 筛选策略。

#### 1) 辐射区分度分析

考虑到 AP 对目标空间的辐射区分度决定其空间识别价值,论文基于各 AP 在不同 RP 处的 RSS 动态范围交叉程度衡量该 AP 的空间辐射区分度<sup>[20]</sup>,以此选取子空间识别力较强的 AP 集作为目标位置筛选近邻 RP 的信源。

以  $\min(\mathbf{RSS}_{nm})$  与  $\max(\mathbf{RSS}_{nm})$  分别表示在  $P_{R_n}$  处接收来自  $A_p$  信号的最小值与最大值,则  $A_p$  在  $P_{R_n}$  处的信号波动范围为  $[\min(\mathbf{RSS}_{nm}), \max(\mathbf{RSS}_{nm})]$ 。基于式(4)和(5),分别计算在参考点  $P_{R_{n_1}}$  与  $P_{R_{n_2}}$  处,接收来自  $A_p$  信号强度的交叉区间长度、并列区间长度:

$$\text{length}_{n_1 \cap n_2}^m = |[\min(\mathbf{RSS}_{n_1 m}), \max(\mathbf{RSS}_{n_1 m})] \cap [\min(\mathbf{RSS}_{n_2 m}), \max(\mathbf{RSS}_{n_2 m})]| \quad (4)$$

$$\text{length}_{n_1 \cup n_2}^m = |[\min(\mathbf{RSS}_{n_1 m}), \max(\mathbf{RSS}_{n_1 m})] \cup [\min(\mathbf{RSS}_{n_2 m}), \max(\mathbf{RSS}_{n_2 m})]| \quad (5)$$

其中,以  $|\cdot|$  表示区间长度的选取。综合考量两参考点组合对的交叉区间与并列区间差异,同时设定区分度因子  $f$  以稀释信号自身波动、临界值计算偏差等情况对 AP 筛选的不利影响。以此构造 RP 组合对的信号差异系数:

$$d_{n_1, n_2}^m = \frac{1}{2} \frac{\text{length}_{n_1 \cup n_2}^m - \text{length}_{n_1 \cap n_2}^m}{\text{length}_{n_1 \cup n_2}^m} + \frac{1}{2} \frac{|\text{length}_{n_1 \cup n_2}^m - f|}{\text{length}_{n_1 \cup n_2}^m} \quad (6)$$

显然,  $d_{n_1, n_2}^m \in [0, 1)$ , 且  $P_{R_{n_1}}$  与  $P_{R_{n_2}}$  组合对的差异越大,  $d_{n_1, n_2}^m$  则越接近于 1。为整体性衡量  $A_p$  在各子空间的识别能力,以  $A_p$  对子空间  $\Omega_\phi$  内所有 RP 组合对的信号差异系数的均值作为其空间辐射区分度的判别标准:

$$D_\phi^m = \frac{N_\phi(N_\phi - 1)}{2} \sum_{n_1=1}^{N_\phi-1} \sum_{n_2=n_1+1}^{N_\phi} d_{n_1, n_2}^m \quad (7)$$

式中:辐射区分度  $D_\phi^m$  值越大,指纹信息差异越明显,  $A_p$  在子空间  $\Omega_\phi$  内的辐射区分度则越突出。值得一提的是,该方案的处理过程,剔除了  $\Omega_\phi$  中大多数 RP 处接收不到的不可见 AP。

#### 2) 覆盖可靠性分析

稳定辐射下的高效覆盖是衡量 AP 筛选有效性的另一重要指标。论文以 RP 节点信号可接收映射 AP 信号可覆盖,基于单一节点可接收 AP 集,分析各 AP 信号的相对稳定性和丢失率,以此作为单一节点覆盖的可替代性指标,衡量该 AP 在此节点处的覆盖可靠性;弱化低可靠性 AP 筛选权重的同时,兼顾空间覆盖度,以达到子空间覆盖稳健性与整体覆盖完整度间的制约平衡,筛选出时间域稳定、空间域完整的子空间精简 AP 集。

假设在参考点  $P_{R_{n_i}}$  处,共接收到  $J$  个信源 AP:  $\{A_{p_{m_1}}, A_{p_{m_2}}, \dots, A_{p_{m_j}}, \dots, A_{p_{m_J}}\}$ , 并计算其对应接收信号强度方差和丢失率  $\{\text{var}_{n_i m_j}, LR_{n_i m_j} \mid j = 1, 2, \dots, J\}$ , 其中  $LR_{n_i m_j} = 1 - Q_{n_i m_j} / Q$ ,  $Q_{n_i m_j}$  表示  $P_{R_{n_i}}$  处接收到的  $A_p$  信号次数。以各单一节点可接收 AP 集为初始分析单元,归一化处理  $A_p$  信号稳健性以量化其在节点  $P_{R_{n_i}}$  处的覆盖可靠性:

$$R_{n_i m_j} = \frac{1}{\text{var}_{n_i m_j} \times LR_{n_i m_j} + \varepsilon} \quad (8)$$

其中,为规避分母项为 0, 附加  $\varepsilon \rightarrow 0_+$ 。依据式(8)推断,  $R_{n_i m_j} \in (0, 1]$ , 其值越大,  $A_p$  在  $P_{R_{n_i}}$  处越不可替代;当  $J = 1$  时,表明该节点处仅可被唯一接入点覆盖,此时接收信号稳健性将不再作为该 AP 可靠性的判断条件,而是直接设置  $R_{n_i m_j} = 1$ , 尽可能避免覆盖空洞现象。

考虑到论文以子空间 AP 精简为导向,故将  $A_p$  在子空间  $\Omega_\phi$  内各单一节点处的可靠性进行整体化考量:

$$R_\phi^{m_j} = \sum_{i=1}^{N_\phi} R_{n_i m_j} \quad (9)$$

作为表征空间覆盖可靠性的  $R_\phi^{m_j}$  值越大,  $A_p$  在子空间  $\Omega_\phi$  内被过滤的可能性则越低。该方案综合考量单一接入点信号的稳定性和丢失率,在弱化空间覆盖冗余的同时,最大限度地保障了空间覆盖的可靠性和完整性。

### 3) 最佳接入点筛选

通过对各接入点信号的辐射区分度和覆盖可靠性的双尺度考量,将目标子空间  $\Omega_\phi$  内关于  $A_{P_m}$  的空间贡献度定义如下:

$$H_\phi^m = D_\phi^m \times R_\phi^m \quad (10)$$

通过对各 AP 的空间贡献度进行降序排序,以  $H_\phi$  作为贡献度标准,为各子空间选取前  $M_\phi$  个  $H_\phi^m \geq H_\phi$  的重叠度低、覆盖高效的精简接入点 AP 集  $\{A_{P_1}^\phi, A_{P_2}^\phi, \dots, A_{P_{M_\phi}}^\phi, \dots, A_{P_{M_\phi}}^\phi\}$ ,完成目标空间垂直划分以获取更好的定位精度。在此说明,为保证量纲一致性,  $D_\phi^m$  和  $R_\phi^m$  均为归一化取值。

至此,双精简子空间  $\Omega_{N_\phi \times M_\phi}^\phi$  构建完毕,目标空间内的 RP 划分和 AP 筛选过程,共同组建符合后期定位需求的理想映射关系库,在有效减少本地缓存成本和在线匹配压力的同时,促使每一次的数据采集直接面向强识别、高可靠的 AP 集合。

## 3 面向 $\Omega_{N_\phi \times M_\phi}^\phi$ 的二级定位方法

为保证样本聚类特征与分类特征的统一性,考虑到模糊空间的设置已较大程度降低了一级区域判别难度,故此判别目标位置归属子空间仅需确定 TP 处实时接收信号  $\mathbf{RSS}_{TP} = [r_{SS_{TP_1}}, r_{SS_{TP_2}}, \dots, r_{SS_{TP_m}}, \dots, r_{SS_{TP_M}}]$  的最强接收信号信源,即可以较低误差满足最佳子空间基础判别:

$$\max(\mathbf{RSS}_{TP}) = r_{SS_{TP_m}} \Rightarrow TP \in \Omega_m = \Omega_\phi \quad (11)$$

当 TP 处最强接收信号来自于第  $m$  个接入点  $A_{P_m}$  时,则判断 TP 所属子空间为  $\Omega_m$ ,并将其标记为  $\Omega_\phi$ ,便于区分接入点标签和子空间标签。

在判定当前位置最佳目标子空间  $\Omega_\phi$  后,依据垂直精简结果保留 TP 处与  $\Omega_\phi$  内对应 AP 序列的接收信号强度值,组成新的待定位点接收信号强度向量  $\mathbf{RSS}'_{TP} = [r_{SS_{TP_1}}, r_{SS_{TP_2}}, \dots, r_{SS_{TP_{M_\phi}}}]$ ,并将 TP 到  $\Omega_\phi$  内参考点  $P_{R_{n_\phi}}^\phi$  的接收信号距离  $d_{n_\phi}$  表征如下:

$$d_{n_\phi} = \|\mathbf{RSS}'_{TP} - \overline{\mathbf{RSS}_{n_\phi}}\|_2 \quad (12)$$

式中:  $\overline{\mathbf{RSS}_{n_\phi}} = [r_{SS_{n_\phi}^1}, r_{SS_{n_\phi}^2}, \dots, r_{SS_{n_\phi}^m}, \dots, r_{SS_{n_\phi}^{M_\phi}}]$  表示  $P_{R_{n_\phi}}^\phi$  来自子空间  $\Omega_\phi$  内精简接入点集的接收信号强度均值向量。遍历  $\Omega_\phi$  中包含的  $N_\phi$  个参考点,升序存储其与 TP 的接收信号距离,选取前  $K$  个最小值组成新的距离序列  $\mathbf{D}_{TP} = \{d_{TP}^1, d_{TP}^2, \dots, d_{TP}^k, \dots, d_{TP}^K\}$ ,并记录对应的参考点序列  $\mathbf{P}_{R_{TP}} = \{P_{R_{TP}}^1, P_{R_{TP}}^2, \dots, P_{R_{TP}}^k, \dots, P_{R_{TP}}^K\}$ ,用于二级位置判别。

为保证在线位置确定的时效性和精确性,论文采用依据  $\mathbf{D}_{TP}$  元素值大小衡量 TP 与  $\mathbf{P}_{R_{TP}}$  内相应参考点相似

度的方法,分配近邻参考点权重  $\omega_k$ ,以完成待定位点位置  $(x_{TP}, y_{TP})$  的有效估计:

$$(x_{TP}, y_{TP}) = \sum_{k=1}^K \omega_k (x_{TP}^k, y_{TP}^k) \quad (13)$$

式中:  $\omega_k = \frac{1/d_{TP}^k}{\sum_{k=1}^K 1/d_{TP}^k}$ ,且  $(x_{TP}^k, y_{TP}^k)$  表示参考点  $P_{R_{TP}}^k$  实际物理坐标。

## 4 DRFS 的定位流程

论文所提 DRFS 定位算法,首先将离线阶段采集并预处理后的数据进行空间双精简:通过水平精简 RP 算法,完成模糊子空间的构建,其中边界模糊和区域融合的处理过程,为在线子空间判别铺垫了良好条件;基于区域降维结果,结合 AP 信号辐射区分度和覆盖可靠性,共同度量 AP 空间贡献度,以此执行垂直精简 AP 算法;其次,通过空间双精简过程共同凝练高质量指纹映射关系库,为在线阶段提供对比支撑;最后,在线定位依赖于空间进行动态判别,利用 RP 与 TP 间的信号欧氏距离分配近邻权重,完成目标位置的有效估计。相较于整体空间匹配,DRFS 算法以更小的模糊子空间为匹配单元,提高空间判别精度的同时,有效减少了在线匹配时的计算复杂度;此外,离线算法的复杂度并不会影响在线算法的实时性,该算法的核心工作主要体现在离线阶段,故此进一步保证了在线阶段的实时定位效率。

## 5 实验与结果分析

### 5.1 测试环境与数据处理

为评估 DRFS 算法性能,以某大学物联网工程学院 C 区某楼层 60 m×42 m 的环形室内走廊为实验实测场景。沿着走廊以 1 m 间隔均匀部署  $N = 368$  个 RP;并以 2.3 s 的采样间隔在每个 RP 处采集  $Q = 50$  次信号;在离线数据采集阶段,目标区域内共探测到  $M = 112$  个 AP,并按 MAC 地址赋予 1~112 的标签。由于每个 RP 上可采集到的 AP 数量和 MAC 地址是有差异的,为确保指纹向量间的可比性,用 -100 dBm 填充缺失的 AP 信号以保证整体数据维度的一致;同时考虑信号短时干扰可能会造成局部异样值或概率偏差数据,论文采用箱线图法<sup>[21]</sup>,借助同一 RP 下其余非异样接收信号强度均值对其进行平滑处理,并构建全局离线指纹库。此外,测试者手持同一移动终端沿着走廊中间匀速行走一圈,在该矩形轨迹上共记录 185 个 TP 及当前位置的 RSS 值,用于后续定位验证。

### 5.2 参数配置对性能的影响

#### 1) 模糊参数 $\delta_{th}$ 配置

空间水平划分质量直接影响着系统定位性能,为平衡空间 RP 部署伴随的计算压力与判别精度,对 2.2 节中所述模糊参数  $\delta_{th}$  进行配置寻优。实验对比了目标空间内累积 RP 数量和子空间判别正确率随  $\delta_{th}$  值的变化情况。

如图 4(a) 所示,当  $\delta_{th}$  值较小时,部分 RP 处于模糊空间,被不止一次地统计,故目标空间内的 RP 累积量大于原始部署 RP 数量  $N = 368$ ;随着  $\delta_{th}$  值的增大,不断剔除达不到聚类条件的 RP,参考点总量相应减少。当  $\delta_{th} = 0.5$  时,参考点数量仅为  $363 < 368$ ,显然,部分 RP 处采集到的最强 AP 信号频率并未达到采样次数的  $1/2$ ,均值最强未必能够准确表达参考点的实际位置,该结果进一步验证了频率表征比均值表征在执行子空间判别中更加具备合理性,同时验证了  $\delta_{th}$  最佳探索区间的预设。观察图 4(b) 可以发现,当  $\delta_{th} \in (0, 0.46]$  时,区判别正确率

下降略缓,但当  $\delta_{th} > 0.46$  时,该性能急剧下降,根本原因在于,前期  $\delta_{th}$  值的增加仅略微减少了模糊空间的范围,并不会引发过大的错误率;而当  $\delta_{th}$  足够大时,部分 RP 开始被遗漏,自然导致区判别正确率严重不足问题。在保证空间聚类完整的前提下,即确保所有 RP 至少被分配到一个子空间,综合考量匹配速度与判别精度,不失一般性,折中配置本场景模糊参数  $\delta_{th} = 0.23$ 。

#### 2) 子空间最优参数选取

同一子空间内,  $H_\phi$  值越大,表明对接入点贡献度的筛选要求越低,  $M_\phi$  值则越大,反之亦然。但过多或过少的接入点都将削弱定位系统的预估性能,故以  $\Omega_\phi$  平均定位精度最佳为目标设置贡献度标准寻优范围  $H_\phi \in \{0, 0.001, 0.002, \dots, 0.6\}$ ,经实验验证设置空间区分度因子  $f = 10$ 、位置预估近邻数  $K = 5$ ,并分别配置 8 个子空间相关参数:

依据表 1 可知,8 个子空间中,平均定位误差均有效保持在 1 m 左右;同时,受空间结构特征、信源分布不均等影响,各子空间最优精简 AP 集筛选控制存在明显差异。其中,最多精简 AP 数量为  $M_2 = 52$ ,而  $M_{14} = 6$  为最少 AP 筛选量,表明  $\Omega_2$  和  $\Omega_{14}$  分别仅需要 52 个 AP 和 6 个 AP 即可实现子空间信息较为完整的表达。相较各 RP 原始缓存  $M = 112$  个 AP 信息,本算法对于区分度和稳定性的双重筛选,将垂直维度降低 53.6% ~ 94.6%,很大程度上缓解了离线存储和在线匹配的压力。

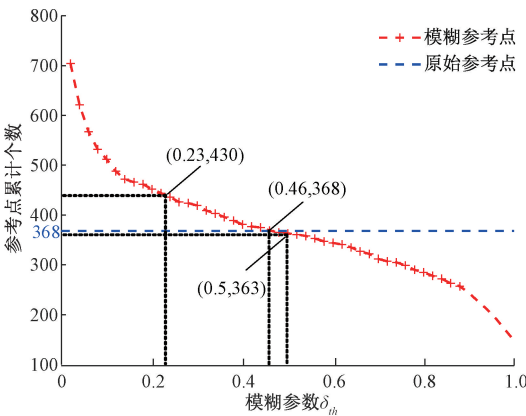
表 1 8 个子空间性能参数一览

Table 1 List of 8 subspace performance parameters

$\phi$	$\Omega_\phi$	$H_\phi$	$M_\phi$	均误差/m	精简 AP 集
2	$\Omega_2$	0.031	52	0.966	$A_{P_2}, A_{P_1}, A_{P_{89}}, \dots, A_{P_{44}}$
4	$\Omega_4$	0.328	29	1.005	$A_{P_4}, A_{P_3}, A_{P_6}, \dots, A_{P_{70}}$
6	$\Omega_6$	0.678	8	1.342	$A_{P_5}, A_{P_6}, A_{P_{100}}, \dots, A_{P_{33}}$
10	$\Omega_{10}$	0.491	28	0.995	$A_{P_9}, A_{P_{10}}, A_{P_{24}}, \dots, A_{P_{26}}$
12	$\Omega_{12}$	0.510	18	0.930	$A_{P_{50}}, A_{P_{11}}, A_{P_{53}}, \dots, A_{P_{55}}$
14	$\Omega_{14}$	0.569	6	0.809	$A_{P_{14}}, A_{P_{63}}, A_{P_{49}}, \dots, A_{P_{65}}$
16	$\Omega_{16}$	0.592	26	0.923	$A_{P_{15}}, A_{P_{16}}, A_{P_{64}}, \dots, A_{P_6}$
32	$\Omega_{32}$	0.548	22	0.910	$A_{P_{32}}, A_{P_{16}}, A_{P_{31}}, \dots, A_{P_{14}}$

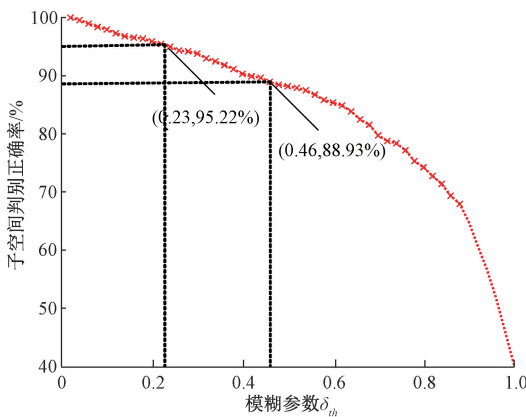
### 5.3 空间聚类结果比较

模糊参数配置完毕实现整体目标空间水平划分,为验证所提精简 RP 算法的有效性,将其与传统经典最强 AP 聚类算法<sup>[11]</sup>进行对比分析。在数据集完全一致的情况下,如图 5(a) 所示的经典最强 AP 聚类结果,因信号波动带来的位置判别不确定性,引发局部空间近邻点划分散乱、个别点独立成区等不良现象,在边界、拐角等干扰高发地表现尤为明显。由于兼顾高阶信息考量和子空



(a)  $\delta_{th}$  对 RP 数量的影响

(a) The effect of  $\delta_{th}$  on the amount of RP



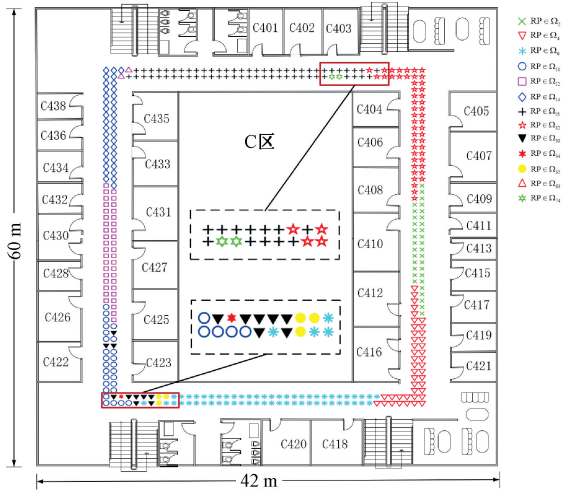
(b)  $\delta_{th}$  对子空间判别正确率的影响

(b) The effect of  $\delta_{th}$  on the accuracy of subspace discrimination

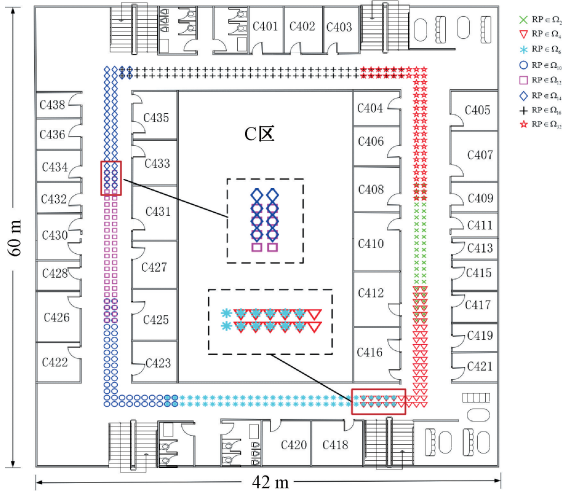
图 4  $\delta_{th}$  对目标空间性能的影响

Fig. 4 The effect of  $\delta_{th}$  on the performance of the target space

间再优化,相较于传统聚类下的 13 个子空间,所提高频最强 AP 模糊聚类仅划分 8 个子空间,如图 5(b)所示,聚类规则且完整,其分布结果也更加符合规律的环状空间结构带来的实际地理约束。



(a) 经典最强AP聚类  
(a) The classical strongest AP clustering



(b) 高频最强AP模糊聚类  
(b) The strongest AP fuzzy clustering at high frequency

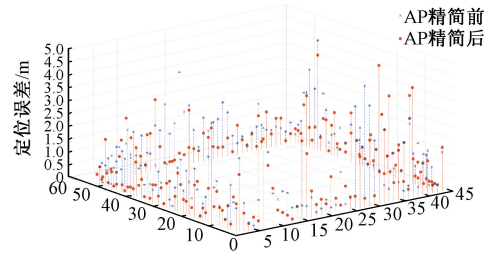
图 5 空间聚类效果对比

Fig. 5 Comparison of spatial clustering effects

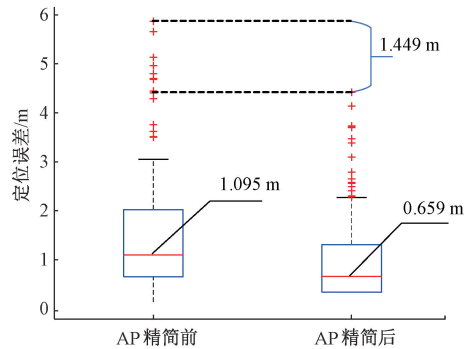
### 5.4 AP 精简前后性能分析

为实现各子空间内 AP 的有效精简,论文提出基于信号空间识别程度和覆盖效益共同约束 AP 筛选的策略。为衡量该方案对位置预估性能的影响,图 6(a)展示了精简前后的全场景各单一测试点 TP 的定位误差。因受限于部分 AP 信息完整性,少量待定位点处呈现精简后定位性能略弱于全局信源定位。但将全局接入点用于定位,将无法避免高重叠、低稳定的 AP 带来的匹配冗余和定位误差,如图 6(b)所示,通过全场景整体

定位效果分析,AP 精简后的中位数误差以 0.436 m 的优势领先于 AP 精简前的定位效果,且垂直精简 AP 算法对于不稳定信源的剔除明显改善了大误差分布,使得精简后的 AP 集定位性能整体优势更为突出。



(a) 单一TP误差对比  
(a) Comparison of single TP error



(b) 整体误差对比  
(b) Comparison of overall error

图 6 AP 精简前后定位误差对比

Fig. 6 Comparison of positioning errors before and after AP streamlining

### 5.5 AP 筛选方法性能比较

为验证论文所提垂直精简 AP 算法的筛选表现,实验选取了 Fisher<sup>[14]</sup>、IG<sup>[15]</sup>、MM<sup>[16]</sup>、OMI<sup>[17]</sup> 共 4 种 AP 筛选算法进行仿真对比。为保证算法性能对比的公平性,论文所提接入点筛选方法和上述 4 种对比算法均基于所提水平精简参考点划分子空间结果,并保证子空间内各算法参数最优。限于篇幅,随机选取  $\Omega_{10}$ 、 $\Omega_{12}$ 、 $\Omega_{14}$  共 3 个代表性子空间进行匹配计算量和定位准确率的分析。如图 7(a)所示,相较于于各子空间成本优势最高的接入点筛选算法,所提接入点精简策略匹配量略高 1~8 个;但就子空间平均定位误差分析,如图 7(b)所示,在 3 个代表性子空间内,论文所提 SDRS 与其他 4 种算法相比较均具备最小平均定位误差,这得益于垂直精简 AP 算法对难以识别、稳定性差的接入点的滤除作用,该算法以微小的接入点筛选增量有效保留了各子空间的高价值指纹信息。

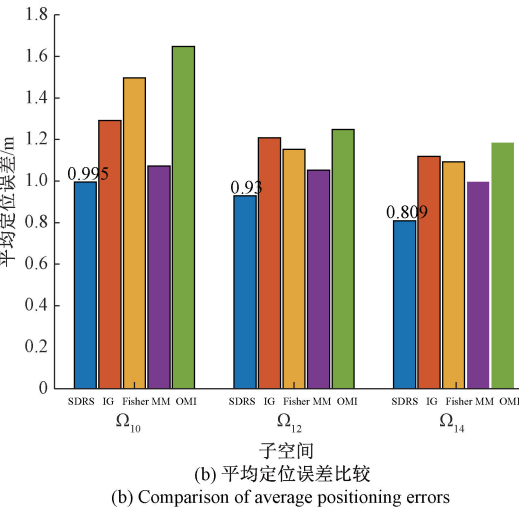
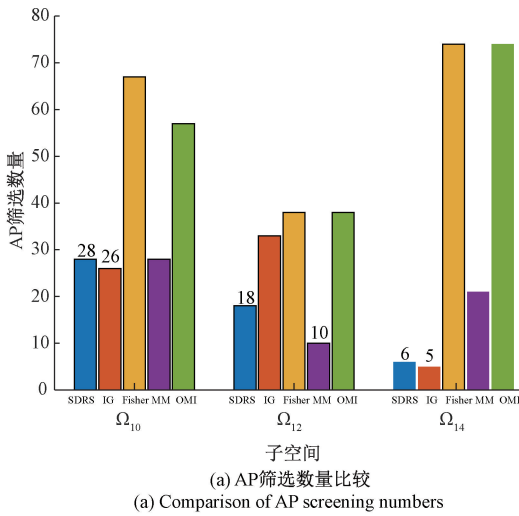


图 7 代表性子空间 AP 筛选方法性能比较

Fig. 7 Performance comparison of representative subspace AP selecting methods

### 5.6 SDRS 整体性能评估

为整体性评估 SDRS 的定位性能,将 SDRS 与 Fisher、IG、MM、OMI 及全局接入点 All AP 共计 6 种算法进行仿真对比。全场景测试点定位结果以误差累积概率形式表征,如图 8 所示。并将定位误差关键统计值总结为表 2。

实验结果表明,6 种算法的误差累积概率均随定位误差允许界值的扩大而不断升高,在同等误差范围内,SDRS 定位性能超越其他 5 种算法,且 2 m 内误差累积概率达到 87.30%,最佳优化 21.35%。原因在于,IG 和 OMI 忽视了筛选 AP 的目标空间可靠性;Fisher 则缺少对空间部署 RP 识别区分能力的考虑;MM 单一考虑最大接收信号值,却过滤了信号强度低但稳定度高的有效 AP。以上 AP 集筛选的局限性导致定位精度提升受阻。而论文所述垂直精简 AP 方法,在保证优选 AP 辐射区分度的

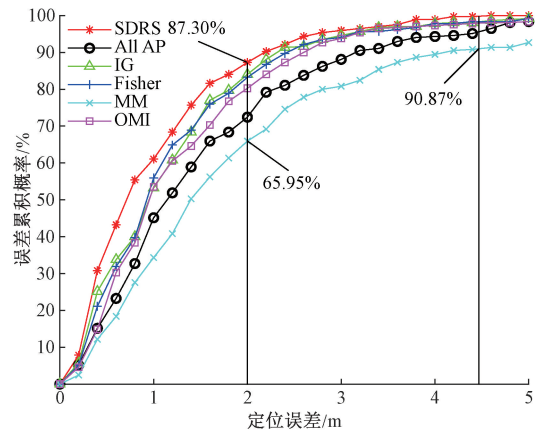


图 8 6 种 AP 筛选误差累积概率对比

Fig. 8 Comparison of the cumulative probability of error for 6 AP selecting methods

表 2 6 种 AP 筛选方法定位误差

Table 2 Positioning errors for 6 AP selecting methods

AP 筛选方法	平均值 /m	最小值 /m	最大值 /m	方差值 /m	2 m 内误差比/%
SDRS	<b>1.003</b>	0.068	<b>4.420</b>	<b>0.738</b>	<b>87.30</b>
All AP	1.505	0.120	5.869	1.438	72.43
IG	1.179	0.081	5.855	0.929	84.05
Fisher	1.205	<b>0.054</b>	9.330	1.105	83.24
MM	1.964	0.067	12.580	3.574	65.95
OMI	1.296	0.059	10.097	1.297	80.27

同时过滤了子空间不可监测的冗余 AP;此外,精选高可靠性 AP 兼顾定位区域覆盖完整性避免空间覆盖效益不足,以此筛选高效 AP 集提升定位质量。

由表 2 可知,相对论文所提 SDRS 算法, Fisher 算法以大量的匹配代价保留了某些 TP 处的信号完整性,故导致 SDRS 计算结果与最优最小值存在 0.014 m 的差距,但从平均值、最大值、方差值等 3 个方面考量 SDRS 都是最优的,较 All AP、IG、Fisher、MM、OMI 这 5 种算法的全场景平均误差分别降低了 0.502、0.176、0.202、0.961、0.293 m。综合实验结果证明,SDRS 算法在多变的室内环境下以较低的匹配成本,维持了良好的定位性能。

### 5.7 实验系统流程

论文仿真数据均为实际场景下的实测数据,并通过实际设备采集,如图 9 所示,将采集到的数据通过 MATLAB 进行处理,并在该平台完成算法性能的验证。基于参数配置,在评估论文所提高频最强 AP 聚类算法效益的同时,主要从单一坐标、子空间、全局空间等不同角度分析验证了 DRFS 算法性能的优越性。



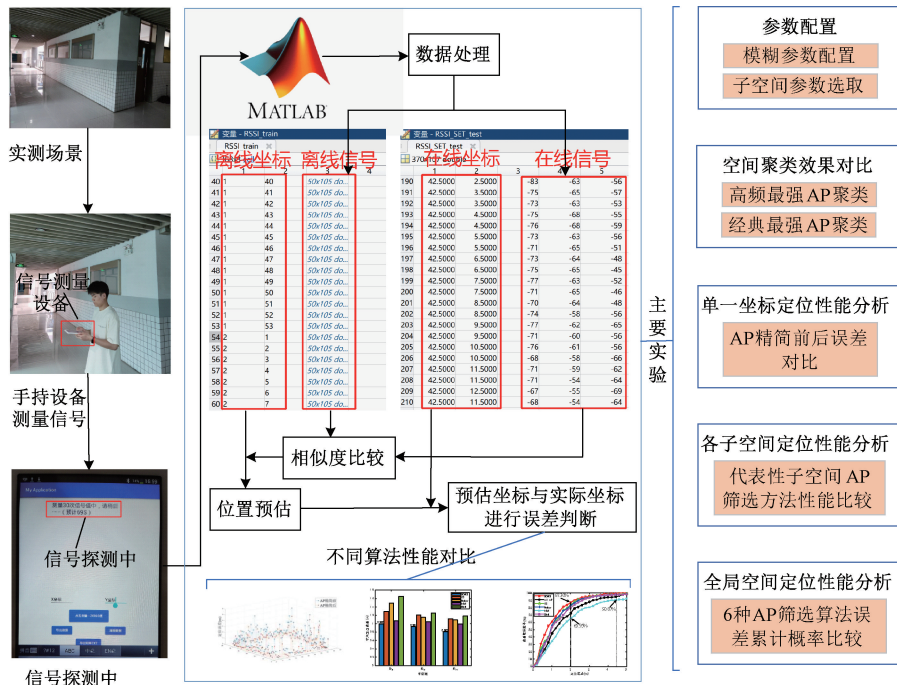


图9 实验系统流程

Fig. 9 The flow of experimental system

## 6 结论

论文针对室内指纹定位需求,提出一种空间双精简策略,以解决全场景匹配冗余问题,构建高价值指纹库,实现目标位置的有效估计。本算法预处理因噪声、传输等不确定因素导致的异常信号,避免信号采集误差;通过分析最强接收信号高阶统计信息,模糊处理低稳定性边界,并综合考虑参考点空间聚类实际物理特征,高效完成目标空间水平精简;基于各降维子空间结果,分析每个AP信号的空间分布特性,以此垂直精简出区分度高、可靠性强的AP集,保证后续定位质量。经实验验证,所提算法采用双精简后的有限维指纹基进行定位,仍能表现出较优异的预估效果。

## 参考文献

- [1] WU Y, CHEN R, LI W, et al. Indoor positioning based on walking-surveyed Wi-Fi fingerprint and corner reference trajectory-geomagnetic database [J]. IEEE Sensors Journal, 2021, 109:18964-18977.
- [2] LUO R C, HSIAO T J. Indoor localization system based on hybrid Wi-Fi/BLE and hierarchical topological fingerprinting approach [J]. IEEE Transactions on Vehicular Technology, 2019, 68(11): 10791-10806.
- [3] LIU M, WANG H, YANG Y, et al. RFID 3-D indoor localization for tag and tag-free target based on

interference [J]. IEEE Transactions on Instrumentation and Measurement, 2018, 68(10): 3718-3732.

- [4] COSTANZO A, DARDARI D, ALEKSANDRAVICIUS J, et al. Energy autonomous UWB localization [J]. IEEE Journal of Radio Frequency Identification, 2017, 1(3): 228-244.
- [5] REMPEL P, BORISOV A, SIEMENS E, et al. Local system of positioning using a WiFi network [J]. Matec Web of Conferences, 2018, 155: 1014-1018.
- [6] GUO X, LI L, ANSARI N, et al. Accurate WiFi localization by fusing a group of fingerprints via global fusion profile [J]. IEEE Transactions on Vehicular Technology, 2018:7314-7325.
- [7] BAHL P, PADMANABHAN V N. RADAR: An in-building RF-based user location and tracking system [C]. Proceedings of the IEEE Infocom 2000. Piscataway: IEEE, 2000:775-784.
- [8] BRUNATO M, BATTISTI R. Statistical learning theory for location fingerprinting in wireless LANs [J]. Computer Networks, 2005, 47(6):825-845.
- [9] LI Z T, ZANG L, TIAN S J, et al. A hybrid compression-aware based data collection method for cluster based networks [J]. Computer Research and Development, 2017, 54(3):493-501.
- [10] HE S, CHAN S H G. Tilejunction: Mitigating signal noise for fingerprint-based indoor localization [J]. IEEE

- Transactions on Mobile Computing, 2016, 15 (6): 1554-1568.
- [11] QUEZADA-GAIBOR D, TORRES-SOSPEDRA J, NURMI J, et al. Lightweight Wi-Fi fingerprinting with a novel RSS clustering algorithm [C]. 2021 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2021:1-8.
- [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] TAO Y, ZHAO L. AIPS: An accurate indoor positioning system with fingerprint map adaptation [J]. IEEE Internet of Things Journal, 2022, 9(4): 3062-3073.
- [14] ZHANG X, MA L, TAN X, et al. A novel algorithm based on clustering and access points selection for indoor fingerprint localization [C]. International Conference on Materials Science and Information Technology, 2013.
- [15] CHEN Y, YANG Q, YIN J, et al. Power-efficient access-point selection for indoor location estimation[J]. IEEE Transactions on Knowledge & Data Engineering, 2006, 18(7):877-888.
- [16] XIA Y, ZHANG Z, MA L. Radio map updated method based on subscriber locations in indoor WLAN localization[J]. Systems Engineering and Electronics, 2015, 26(6):1202-1209.
- [17] ZOU H, LUO Y, LU X, et al. A mutual information based online access point selection strategy for WiFi indoor localization[J]. IEEE International Conference on Automation Science and Engineering (CASE), 2015, 180-185.
- [18] GHAFARIAN H. Reducing search area in indoor localization applications [J]. Wireless Personal Communications, 2021, 117:1243-1258.
- [19] XIAO J, LIAO B, LI C. Improved indoor positioning

based on fingerprint algorithm using reference point subregion and access point optimization[C]. 2019 IEEE 19th International Conference on Communication Technology (ICCT), 2019:461-468.

- [20] WU B, MA Z, POSLAD S, et al. An efficient wireless access point selection algorithm for location determination based on RSSI interval overlap degree determination[C]. 2018 IEEE Wireless Telecommunications Symposium (WTS), 2018:1-8.
- [21] DOVOEDO Y H, CHAKRABORTI S. Boxplot-based outlier detection for the location-scale family [J]. Communications in Statistics-Simulation and Computation, 2015, 44(6): 1492-1513.

### 作者简介



郑安琪, 2018 年于合肥师范学院获得学士学位, 2023 年于江南大学获得硕士学位, 现为无锡太湖学院助教, 主要研究方向为智能物联网、室内定位。

E-mail:1157710222@qq.com

**Zheng Anqi** received her B. Sc. degree from Hefei Normal University in 2018, received her M. Sc. degree from Jiangnan University in 2023. She is currently an assistant in Wuxi Taihu University. Her main research interests include intelligent internet of things and indoor positioning.



秦宁宁(通信作者), 2002 年于江南大学获得学士学位, 2008 年于江南大学获得博士学位, 现为江南大学教授, 主要研究方向为智能网络组建、信道数据的网络化应用。

E-mail:ningning801108@163.com

**Qin Ningning** (Corresponding author) received her B. Sc. degree and Ph. D. degree both from Jiangnan University in 2002 and 2008, respectively. She is currently a professor at Jiangnan University. Her main research interests include intelligent network formation and networked applications of channel data.