

基于张量鲁棒特征分析的室内指纹定位算法

李华亮, 张家伟, 黄平, 程超

(长春工业大学计算机科学与工程学院, 吉林 长春 130102)

摘要: 针对室内CSI指纹定位中受多径效应、尖峰噪声及在线离线时间差影响, 导致指纹失真、接入点状态波动和定位精度下降的问题, 提出一种融合张量鲁棒主成分分析、斜率熵检测与高斯混合加权K近邻的CSI指纹定位方法。首先, 构建由子载波、幅值和时间戳组成的三维CSI张量, 利用张量鲁棒主成分分析分离低秩成分与稀疏噪声, 实现CSI数据去噪。其次, 引入斜率熵检测机制, 动态筛选在线离线阶段中稳定的接入点, 提高算法对环境变化的适应能力。最后, 结合高斯混合模型与加权K近邻算法, 对离线指纹建模并对在线数据分组, 自适应分配K值完成定位估计。结果表明, 该方法能够有效提升复杂室内环境下CSI指纹定位的精度与鲁棒性。

关键词: WIFI室内定位; TRPCA; 斜率熵; AP状态受限检测; 高斯混合模型改进WKNN

Indoor Fingerprint Positioning Algorithm Based on Tensor Robust Feature Analysis

LI Hualiang, ZHANG Jiawei, HUANG Ping, CHENG Chao

School of Computer Science and Engineering, Changchun University of Technology, Changchun, Jilin, 130102, China

Abstract: Aiming at the problems of fingerprint distortion, access point state fluctuation and decreased positioning accuracy caused by multipath effect, spike noise and time difference between online and offline in indoor CSI fingerprint positioning, a CSI fingerprint positioning method integrating tensor robust principal component analysis, slope entropy detection and Gaussian mixture weighted k-nearest neighbor is proposed. Firstly, a three-dimensional CSI tensor composed of subcarriers, amplitudes and timestamps is constructed. Tensor robust principal component analysis is utilized to separate low-rank components from sparse noise, thereby achieving CSI data denoising. Secondly, a slope entropy detection mechanism is introduced to dynamically screen stable access points in both online and offline phases, thereby enhancing the algorithm's adaptability to environmental changes. Finally, by combining the Gaussian mixture model and the weighted K-nearest neighbor algorithm, the offline fingerprint is modeled and the online data is grouped, and the K value is adaptively allocated to complete the location estimation. The results show that this method can effectively improve the accuracy and robustness of CSI fingerprint positioning in complex indoor environments.

Key words: Wi-Fi Indoor Positioning, TRPCA, Slope Entropy, AP State Constraint Detection, GWKNN

0 引言

近些年来, 随着物联网技术的发展, 基于位置信息的服务 (LBS, location based service) 得到广泛的应用, 进而推动了无线技术的发展^[1]。在目前的应用背景下, 室外定位主要以全球定位系统

(Global Positioning System, GPS) 为主^[2], 但由于在室内环境下场景更加复杂并且存在多径效应以及无线信号传输衰减等情况, GPS无法实现精确的定位。为了准确的实现室内定位, 目前的方法主要借助于射频识别^[3] (RFID, radio frequency identification)、无线传感器网络^[4] (WSN, wireless sensor

收稿日期: XXXX-XX-XX; 修回日期: XXXX-XX-XX

通信作者: 程超, chengchao@ccut.edu.cn

基金项目: 吉林省自然科学基金(YDZJ202401313ZYTS)

Foundation Items: Natural Science Foundation of Jilin Province (YDZJ202401313ZYTS)

network)、超宽带^[5] (UWB, ultra wide band) 和无线局域网^[6] (WLAN, wireless local area network) 等技术。其中由于 WiFi 技术无接触、不侵犯隐私、对光照和 LOS 无要求、成本低且普及度高, 所以更加适用于室内定位的场景。成为非常具有潜力的无线定位技术^[7]。

无线定位中使用的无线信号指标包含到达时间^[8] (Time of Arrival, TOA)、到达时间差^[9] (Time Difference of Arrival, TDOA)、到达角度^[10] (Angle of Arrival, AOA)、接收信号强度^[11] (Received Signal Strength, RSS) 和信道状态信息^[12] (Channel State Information, CSI), 其中信道状态信息可以反映 OFDM (Orthogonal Frequency Division Multiplexing, OFDM) 系统中各个子载波的幅度和相位信息, 具有更高的细粒度, 能够更好的刻画信道中的信息^[13]。

文献[14]基于并行因子 (PARAFAC) 分析模型和交替最小二乘法 (ALSs) 迭代算法的张量分解算法降噪, 利用张量小波分解算法提取特征, 再引入 AGA 优化 BPNN 以寻找最优权重和阈值, 从而获得估计位置坐标, 但该方法主要针对随机分布的噪声。文献[15]提出基于扰动理论 AP 选择 (PeAS) 算法, 通过该算法剔除信号干扰强以及无位置区分度的 AP 来减少在线阶段相似性比较的数据量, 但该方法主要针对静态场景, 未针对时间差带来的动态变化。文献[16]提出的指纹定位算法结合 CSO 与 WKNN 指纹定位, 通过分区策略限定 RP 范围, 消除跨区域干扰, 但该方法主要针对距离加权, 在信号波动大的场景下可能带来相应的误差。针对以上问题, 本文提出一种 CSI 指纹定位算法, 具体工作内容如下:

(1) 针对室内环境下由于多径干扰等原因产生尖峰噪声影响 CSI 数据采集精度的问题, 本文从 CSI 张量结构入手, 将采集到的数据根据子载波, 幅值, 时间戳信息构建成三维张量形式。引入张量鲁棒性分析算法 (TRPCA) 将采集到数据信息分解为低秩向量和稀疏噪声, 转为凸优化问题利用增广拉格朗日乘子法 (ADMM) 进行求解, 有效解决室内环境下的问题。该方法能够利用张量的多维结构, 直接在张量空间中分解低秩成分和噪声。利用各维度间的相关性实现更精准的去噪。可同时处理稀疏噪声和低秩张量。

(2) 针对在线离线过程中的时间差导致 AP 受限的问题, 本文引入斜率熵检测。该方法用在在线和离线阶段的动态调整代替传统的静态调整, 通过计算 AP 离线与在线阶段时间差的斜率熵, 通过斜率熵的大小来动态调整选择稳定的 AP 点实现在线定位的功能, 有效提升 AP 受限情况下的定位精度。

(3) 针对定位算法方面, 将高斯混合模型与加权 K 近邻算法结合, 利用概率加权代替距离加权, 在离线阶段利用高斯混合模型收集经过 TRPCA 算法去噪后的离线指纹信息, 建立离线指纹数据库。在线阶段, 利用高斯混合模型的聚类能力来对调整后 AP 测得 CSI 数据进行分组。通过将 CSI 数据分类为不同组来分配适当的 K 值, 最后使用改进后的 WKNN 算法进行定位估计。

1 信道状态信息

与 RSSI 不同的是, CSI 信息则相对稳定, 较多的保留了无线信道中的多径信息。^[17]在 802.11n 标准下, OFDM 将信号调制在多个子载波中进行传输, CSI 作为物理层信息可以提取出信号在传输过程中的衰减因子, 从 CSI 中提取出信道的状态属性, 可以用于感知、定位和识别等任务。简化无线通信传播模型, 可以得到 CSI 的一般模型为公式 1 所示:

$$Y = HX + N \quad (1)$$

其中, X 是发射信号向量, Y 是接收端接收信号向量, H 是无线信道系统增益, N 是信道中的高斯白噪声。以信道状态信息 CSI 的形式获取信道频率响应:

$$H = [CSI_1, CSI_2, \dots, CSI_N]^T \quad (2)$$

式中, CSI_i ($i=1, 2, \dots, N$) 表示第 i 个子载波的频率响应:

$$CSI_i = |CSI_i| e^{j\theta_k} \quad (3)$$

同时在 MIMO 技术中, 发射天线和接收天线相互独立, 天线对中形成的信道状态信息矩阵:

$$H = \begin{bmatrix} H_{11}, H_{12}, \dots, H_{1n} \\ H_{21}, H_{22}, \dots, H_{2n} \\ \dots \\ H_{v1}, H_{v2}, \dots, H_{vn} \end{bmatrix} \quad (4)$$

式中, H_{vm} 代表通过发射天线和接收天线之间通信链路的信道状态信息矩阵。矩阵 H 的维度为 $n * v$ 其中 n 代表发射天线数, v 代表接收天线数。考虑

信号在传播过程中的多径效应、幅度衰减、相位偏移和到达时延等因素,则可以得到公式。

$$h(f,t) = \sum_{k=1}^K \alpha_k(t) e^{-j2\pi f \tau_k(t)} \quad (5)$$

其中, K 是传播路径总数量, $\alpha_k(t)$ 是第 k 条路径的复衰减因子, $\tau_k(t)$ 是第 k 条路径的传播时延, f 是载波频率。

相比于 RSS 来说, CSI 综合反映了幅值衰减、相位偏移、以及时间延迟等多种情况,同时还表征了各个子载波的信息,能够更好地描述信号的传播路径。因此, CSI 已成为室内 Wi-Fi 高精度定位中最常用的特征之一。

2 定位模型

CSI 指纹定位通常包含离线阶段和在线阶段。在离线阶段中,需要进行 CSI 数据的采集和预处理,并根据参考点的 CSI 位置指纹和坐标得到标签样本,建立定位的模型,在线阶段的过程中,通过设置测试点进行 CSI 的预处理和采集得到待定位的样本,之后代入到定位模型来进行位置坐标估计。本文的定位系统框架如下图 1 所示:

根据上图 1 所示。在离线阶段,将采集后的数据针对 CSI 幅值信息进行 TRPCA 降噪,将处理后的数据进行特征提取,得到 RP 的位置特征。将 RP 位置特征与位置坐标代入指纹数据库。在线阶段,进行 AP 受限状态检测,根据斜率熵信息选择稳定的 AP 点进行在线定位。在线阶段,通过将选择后的 AP 将采集后的数据进行离线过程中相同的处理,将在线过程中得到的位置特征和离线指纹数据库代入混合高斯模型改进的 WKNN 进行匹配,得到最终的位置坐标。

将离线过程中幅值相位信息处理后,得到每个位置的 (x_i, y_i) 的位置指纹为

$$Finger = \begin{pmatrix} CSI_1^1, CSI_1^2, \dots, CSI_1^n \\ CSI_2^1, CSI_2^2, \dots, CSI_2^n \\ \dots \\ CSI_N^1, CSI_N^2, \dots, CSI_N^n \end{pmatrix}_{N \times n} \quad (6)$$

3 算法分析

3.1 基于张量鲁棒性分析 (TRPCA) 的幅值去噪算法

由于在室内环境中复杂多变,实际定位过程中会存在多径效应等影响,CSI 数据的幅值信息在不同时刻易受到环境的影响,CSI 数据在采集过程中会存在尖峰噪声。为解决该问题,本文引入了一种基于张量鲁棒性分析 (TRPCA) [18] 的去噪算法,通过将 CSI 数据通过幅值,子载波,时间戳信息构建为三维张量形式,三维张量原始图如图 2 所示。

利用该算法将其分解为低秩张量和稀疏张量,低秩张量代表分离出噪声的真实数据,稀疏张量代表噪声矩阵,通过转换为凸优化问题并且引入拉格朗日乘子法 (ADMM) 进行求解,进而去除采集过程中的尖峰噪声,TRPCA 用于去噪的具体推导流程如下:

首先,将原始 CSI 数据构建成三维张量

$$X \in R^{n_1 \times n_2 \times n_3} \quad (7)$$

$$X = L_0 + S_0 \quad (8)$$

其中, L_0 代表分解出的低秩张量, S_0 代表分解出的稀疏张量,通过凸优化恢复出 L_0 和 S_0 。

将张量数据进行 t-SVD 分解,三阶张量 L 的 t-SVD 分解为:

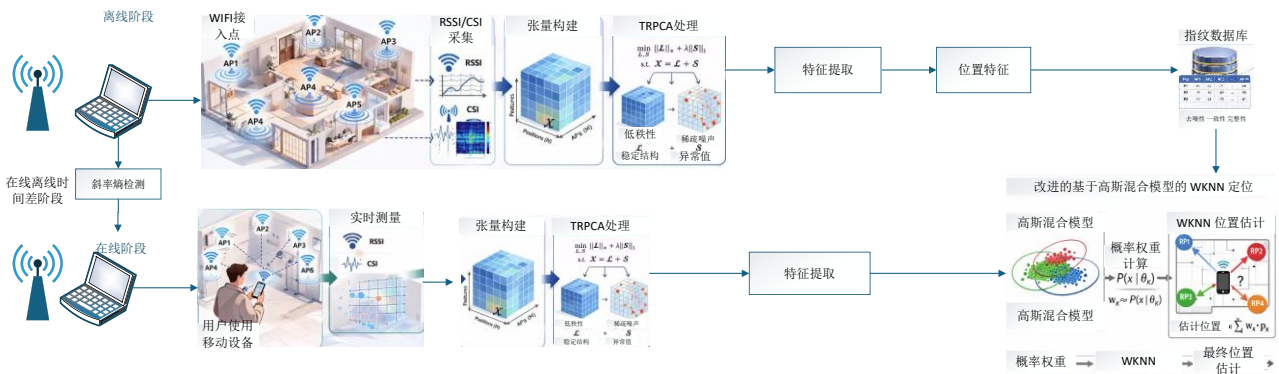


图 1 整体定位流程

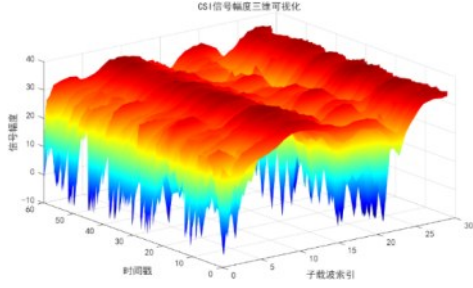


图2 CSI三维张量图

$$L = U * S * V^T \quad (9)$$

其中, U 和 V 正交张量, S 代表 f -diagonal 张量, $*$ 代表张量乘积, 通过沿第三维张量进行 FFT, 可以将张量数据转换为傅里叶域的块对角矩阵, 进而应用矩阵进行 t-SVD 分解。其中定义 L 的张量核范数 $\|L\|_*$ 为傅里叶域中各矩阵核范数的平均, 其中:

$$\|L\|_* = \frac{1}{n_3} \sum_{i=1}^{n_3} \|\bar{L}(i)\|_* \quad (10)$$

其中 $\bar{L}(i)$ 代表 L 经 FFT 变换后的第 i 个前向切片。

去噪算法的凸优化问题可以转换为:

$$\begin{cases} \min \|L\|_* + \mu \|S\|_1 \\ s.t. X = L + S \end{cases} \quad (11)$$

其中, μ 代表平衡参数

$$\|S\|_1 = \sum_{i,j,k} |S_{i,j,k}| \quad (12)$$

接下来引入拉格朗日乘子 y 和惩罚参数 ρ , 构建的增广拉格朗日函数如下所示:

$$\begin{aligned} \varphi(L, S, y) = & \|L\|_* + \mu \|S\|_1 + \langle y, X - L - S \rangle \\ & + \frac{\rho}{2} \|X - L - S\|_F^2 \end{aligned} \quad (13)$$

基于以上条件推导, 具体迭代求解步骤如表 1 所示:

本文采用短时滑动窗口策略构建 WiFi-CSI 三阶时空张量, 限定单窗口张量维度规模固定可控。每一个张量窗口尺寸 $M \times N \times L$ 固定不变, 不随总采集时长增长。TRPCA 算法单次迭代时间复杂度为 $O((MNL)^{\frac{3}{2}})$, 在室内 CSI 指纹定位典型参数下, 实际运算复杂度显著低于传统二维矩阵 PCA 方法, 不存在数量级暴涨问题。

算法运算流程分为离线建模与在线匹配两个阶段: 离线指纹库构建阶段批量完成张量鲁棒主成分分解, 一次性分离低秩有效特征与稀疏突发干扰, 该阶段无实时性约束; 在线定位与动作识别阶段仅

表1 TRPCA 求解流程

算法 1: 原始 CSI 数据的 TRPCA 算法求解过程:

```

输入: 原始 CSI 数据  $X \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ 
    令  $S_0^0 = 0, \rho^0 > 0, a > 1, k=0;$ 
    while 不收敛 do
/*更新  $L^*/L^{k+1} = \arg \min_L \|L\|_* + \frac{\rho}{2} \|X - L - S^k + \frac{y^k}{\rho}\|_F^2$ 
    对  $L = X - S^k + \frac{y^k}{\rho}$  进行 t-SVD 分解
/*更新  $S^*/S^{k+1} = \arg \min_S \mu \|S\|_1 + \frac{\rho}{2} \|X - L^{k+1} - S + \frac{y^k}{\rho}\|_F^2$ 
/*更新乘子  $y^*/$ 
     $y^{k+1} = y^k + \rho(X - L^{k+1} - S^{k+1})$ 
    令  $\rho_{k+1} = a\rho_k$ 
    令  $k \leftarrow k + 1$ 
    结束 while
    令  $L_0 \leftarrow L_0^k, S_0 \leftarrow S_0^k$ 
输出:  $(L_0, S_0)$ 

```

执行低维特征快速匹配运算, 时间复杂度为 $O(MNK)$, 与传统 WKNN 算法持平, 可稳定适配室内实时场景应用。同时分块滑窗张量架构使算法内存占用随时序长度呈线性增长, 规避了长时序数据下张量立方级内存膨胀问题。

3.2 基于斜率熵信息的 AP 受限检测

在线和离线的定位过程中, 会存在一定的时间间隔。所以在线定位时, AP 可能因为位置改变, 断电等问题发生一系列的状态变化。这些变化会使在线测得的 CSI 信号与离线信号的差距较大。进而影响定位的精度。

为解决 AP 状态变化导致定位精度降低的问题, 本文引入斜率熵信息^[19], 通过计算 AP 离线与在线阶段时间差的斜率熵, 通过斜率熵的大小来动态调整选择稳定的 AP 点实现在线定位的功能。假设在室内定位的区域中, 存在 M 个参考节点, 可以扫描到 m 个 AP。离线的指纹数据库中包括 RP 位置和 RP 的位置指纹。

$$G = \left\{ \begin{array}{l} (x_1, y_1) \\ (x_2, y_2) \\ \vdots \\ (x_M, y_m) \end{array} \right\} \quad (14)$$

$$H = \begin{pmatrix} csi_1^1, csi_1^2, \dots, csi_1^m \\ csi_2^1, csi_2^2, \dots, csi_2^m \\ \vdots \\ csi_m^1, csi_m^2, \dots, csi_m^m \end{pmatrix} \quad (15)$$

其中 G 代表 RP 的位置坐标, H 代表 RP 的位置指纹。

设在线和离线的的时间间隔的时间序列为

$$T = \{t_1, t_2, \dots, t_n\} \quad (16)$$

在室内的区域中选择一个固定的位置, 部署一个终端作为 SP 节点, 该节点实时采集时间序列 T 中的 CSI 信号幅值信息, 在线定位过程中, 以当前测试点为中心, 维护一个固定长度的时间序列滑动窗口, 即可实时评估各 AP 的稳定性。连续采集多帧 CSI 幅值构成短时序列, 通过相邻采样点差值构建差值序列, 完成符号映射与熵值统计。具体框图如图 3 所示。时间序列 T 中每个 AP 的测得环境中的幅值信息用 F 表示

$$F_i = \begin{pmatrix} f_1^1 f_2^1, \dots, f_M^1 \\ f_1^2 f_2^2, \dots, f_M^2 \\ \dots \\ f_1^n f_2^n, \dots, f_M^n \end{pmatrix} \quad (17)$$

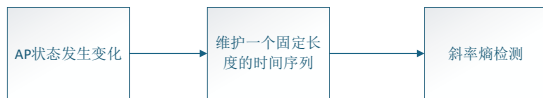


图3 斜率熵检测示意图

在上式中, 每一行代表该 AP 点在某一时刻收集到环境中的 CSI 信号幅值信息。

根据斜率熵的原理, 选取嵌入维度 $u=3$, 提取时间序列内的子序列, 该子序列为每个 AP 在相同参考点的时间序列 T 中测得环境中的连续时间点的 CSI 幅值差值 R_i 。

$$T_i = \{f_i, f_{i+1}, \dots, f_{i+u-1}\}, 1 \leq i \leq n - u \quad (18)$$

$$\begin{aligned} R_i &= \{f_{i+1} - f_i, f_{i+2} - f_{i+1}, \dots, f_{i+u-1} - f_{i+u-2}\} \\ &= \{e_i, e_{i+1}, \dots, e_{i+u-2}\}, 1 \leq i \leq n - u + 1 \end{aligned} \quad (19)$$

其中 $e_i = f_{i+1} - f_i$

之后设置参数 q 和 j, 其中 q 代表主阈值, 决定差值分类的边界。j 代表零附近阈值, 处理接近 0 的差值, 避免微小波动导致分类不稳定。根据设置

的参数 q 和 j, 将每个差值 e 映射为 5 种符号 (-2, -1, 0, +1, +2)。符号配置规则为:

$$\begin{aligned} e > q &\rightarrow +2 \\ j < e \leq q &\rightarrow +1 \\ |e| \leq j &\rightarrow 0 \\ -q \leq e < -j &\rightarrow -1 \\ e < -q &\rightarrow -2 \end{aligned} \quad (20)$$

其中主阈值 q 用于将差值 e 映射为符号, 其取值与信号幅值的标准差相关。本文设 $q = 0.25 \times \text{std}(E)$, 其中 E 为所有 AP 的差值集合, 使分类边界适应环境整体波动水平。零附近阈值 $j = 0.1 \times q$, 用于将极小幅值波动归为“0”类, 避免噪声引起的符号抖动。

根据设置的符号规则, 统计每个符号出现的频率 P_i , 进而计算出每个 AP 在相同参考点的时间序列斜率熵值。

$$se = -P_i \log P_i \quad (21)$$

进而计算出每个 AP 在该环境下的平均斜率熵值

$$SE = \frac{\sum_{i=1}^M se_i}{M} \quad (22)$$

假设环境中需要 k 个 AP, $k < m$ 。根据每个 AP 在该环境下的平均斜率熵值, 选取在线离线时间间隔中前 k 小的熵值对环境中的 AP 进行在线的调整, 来实现更加精准的定位效果。

本文 AP 保留数量 k 结合实验环境 AP 总数、信号稳定性与冗余性综合确定。经斜率熵对各 AP 信号稳定性量化排序后, 选取前 k 个稳定性最优的 AP。若 k 取值过小, 易因单 AP 信号异常导致定位失效; 若 k 过大, 会引入噪声较大、区分度低的冗余 AP, 增加计算量并降低定位鲁棒性。量化排序后可在抗干扰能力与特征维度之间取得最优平衡。

3.3 基于混合高斯模型定位

高斯混合模型是基于概率密度函数的聚类方法^[20]。在高斯混合模型中, 每个数据点由单独几个函数生成。根据每个函数具有不同的均值, 方差和分配的权重。可以将单独的高斯模型按照比例进行线性组合, 能够更加准确的代表数据的真实分布。其概率密度函数如下:

$$P(r) = \sum_{j=1}^M w_j g(r|u_j, \sigma_j^2) \quad (23)$$

$$= \sum_{j=1}^M g(r|u_j, \sigma_j^2) \frac{w_j}{\sqrt{2\pi} \sigma_j} \exp\left[-\frac{(r - u_j)^2}{2\sigma_j^2}\right]$$

其中 r 代表服从混合高斯分布的随机变量，本文中代表接收到 CSI 数据的测量值。M 代表混合模型，也就是高斯分布的数量。 w_j 代表第 j 个单一高斯概率密度函数的权重（加权和为 1 且大于等于 0）。 μ_j 代表均值， σ_j 代表标准差， $g(r|u_j, \sigma_j^2)$ 代表第 j 个单一的高斯概率密度函数。

为了估计上面的参数，我们使用 EM 算法进行计算。EM 算法分为两个基本的步骤： e 步和 m 步， e 步以当前的模型参数作为条件对潜在变量的后验概率分布进行评估， m 步对这些参数计算最大似然估计。

通过上式可以得到对数似然函数 L 的表达式：

$$L(R|\theta) = \log(P(R|\theta)) = \sum_{i=1}^n \log \sum_{j=1}^M \frac{w_j}{\sigma_j \sqrt{2\pi}} \exp\left[-\frac{1}{2} \frac{(r_j - u_j)^2}{\sigma_j^2}\right] \quad (24)$$

其中 R 代表测量到的 CSI 数据， θ 代表 EM 算法待估计的参数，具体步骤如下所示：

在第 E 步中，引入隐含变量 $\alpha_j(r_i)$ ，该隐含变量表示 r_i 是由第 j 个高斯分量产生的概率，得到每个测量到的数据属于每个分布的概率：

$$\alpha_j^t(r_i) = \frac{w_j g(r_i|u_j, \sigma_j^2)}{\sum_{j=1}^M w_j g(r_i|u_j, \sigma_j^2)} \quad (25)$$

在第 M 步中，通过期望步可以最终得到参数的估计：

$$w_j^t = \frac{1}{M} \sum_{j=1}^M \alpha_j^t(r_i) \quad (26)$$

$$u_j^t = \frac{\sum_{j=1}^M \alpha_j^t(r_i) r_i}{\sum_{j=1}^M \alpha_j^t(r_i)} \quad (27)$$

$$\alpha_j^t = \frac{\sum_{j=1}^M \alpha_j^t(r_i) (r_i - u_j^t)^2}{\sum_{j=1}^M \alpha_j^t(r_i)} \quad (28)$$

上面公式是通过对对数似然函数求偏导所得到

的。通过对以上的期望步和最大化步不断地进行迭代，直至似然函数最终收敛，收敛条件为：

$$\|L(R|\theta^t) - L(R|\theta^{t-1})\| \leq \varepsilon \quad (29)$$

其中 ε 表着收敛域值，选值通常选择很小的常量。

设离线阶段共采集 M 个参考点 RP，每个 RP 的 CSI 特征向量为 $r_i \in R^d$ 。采用高斯混合模型对 $\{r_i\}_{i=1}^M$ 进行聚类，得到 C 个高斯分量。高斯混合模型聚类数 C 依据室内场景物理功能分区与 CSI 指纹空间分布特征对应的聚类分量数确定，使每个高斯分量均可对应室内连续空间子区域。并为每个 RP 分配一个簇标签 $c_i \in \{1, 2, \dots, C\}$ 。在线测试点的分组阶段，对于在线阶段采集的测试点特征 r_{test} ，计算其属于各高斯分量的后验概率：

$$p(c_j|r_{test}) = \frac{\omega_j N(r_{test}|\mu_j, \sum_j)}{\sum_{k=1}^C \omega_k N(r_{test}|\mu_k, \sum_k)} \quad (30)$$

将测试点划分至最大后验概率对应的簇 c_{test} 。在同簇参考点匹配阶段，在定位时，仅选择与测试点同簇的参考点参与后续匹配，即：

$$R_{cand} = \{RP_i | c_i = c_{test}\} \quad (31)$$

若同簇参考点数量不足 K_{min} ，则放宽至邻接簇。

在自适应 K 值分配阶段，每个簇 c 预先统计其参考点数量 N_c 与空间分布范围。K 值按如下规则分配：

$$K_c = \min\left(\max\left(\left\lceil \frac{N_c}{2} \right\rceil, 5\right), 15\right) \quad (32)$$

即 K 值随簇内密度调整，密度越高（参考点越多），K 值适当增大，避免跨区域干扰。

在概率加权融合与位置估计阶段，计算测试点与候选参考点之间的欧氏距离 $d_i = \|r_{test} - r_i\|_2$ ，并选择距离最小的 K_c 个参考点。定义概率权重为：

$$W_i = p(r_i|\theta_{c_{test}}) = \omega_{c_{test}} N(r_i|\mu_{c_{test}}, \sum_{c_{test}}) \quad (33)$$

其中 W_i 代表离线指纹库中第 i 个指纹点分配的权重。该权重反映了参考点 i 在其所属高斯分量中的概率密度，值越大表示该参考点的 CSI 特征越典型。将这些权值合并到 WKNN 算法得到的最终定位坐标中，这种方法用概率为权值代替了以距离为权值，对以距离为权值的误差进行了有效的减少，对定位精度加以提升。最后将相应的权值分配给 k 个最近的参考点代入到 WKNN 算法，计算出最终

的位置坐标。WKNN 算法公式为:

$$(x,y) = \frac{\sum_{i=1}^k W_i}{\sum_{i=1}^k W_i} (x_i, y_i) \quad (34)$$

4 实验结果及分析

实验结果分析过程首先对实验的参数环境加以介绍,并给出数据采集的具体流程。接下来,结合降噪的效果图以及性能参数来验证 TRPCA 降噪的可行性。其次,针对斜率熵的对 AP 的动态调整,利用高斯混合模型改进的加权 K 近邻算法进行定位。之后从定位误差的方面来验证整体 CSI 指纹定位算法的有效性,最后对比不同方法的定位误差的累计分布函数来验证该方法的有效性。

4.1 实验场景

实验场景选择 20M 长 15M 宽的实验室,实验场景图如图 4 所示,环境内设置四个 AP 点。实验设备包括 TP-Link 路由器、支持多种 WiFi 标准和模式的 ESP32 开发板;且笔记本电脑具体配置为:系统为 Windows 11,中央处理器(central processing unit, CPU)为 Intel(R)Core i7-7300HQ,随机存储器(random access memory, RAM)为 8 GB。

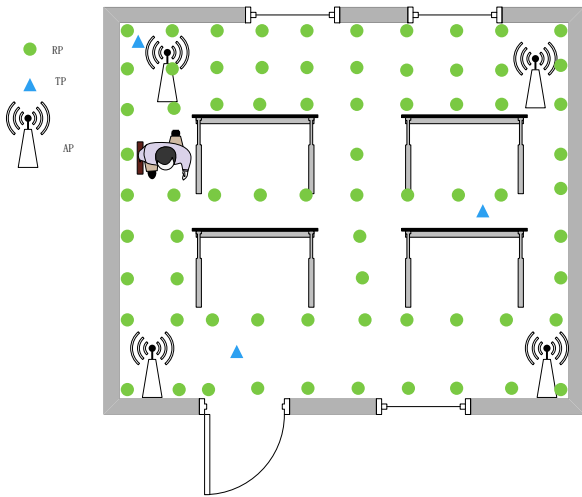


图4 实验场景图

本实验在实验室环境下进行,其中绿色点为参考点(Reference Point, RP),三角形为测试点(Test Point, TP),实验室内有书桌和学生可视作非视距环境(Non Line Of Sight, NLOS)。

在离线阶段,被测试的人员在参考点进行数据

采集,在每个 RP 处进行观察约 2 分钟,期间被测人员随机改变身体的朝向来使数据具有多样性以构建训练样本。采集完成后来重复采集过程,每个点采集 1 分钟作为验证数据。在线阶段,对测试点随机选择进行预测,在选择的测试点上多次采集 CSI 数据,记录坐标使用定位算法进行预测。

4.2 定位性能分析

为直观地表示 TRPCA 的降噪效果,随机选取室内环境中的两个位置。根据图 5 可以看出,通过原始图像可以看出 CSI 数据存在突出的位置,这就是环境中的尖峰噪声。经过 TRPCA 处理后,可以看出该方法对不同位置的尖峰噪声均能进行有效的去除。

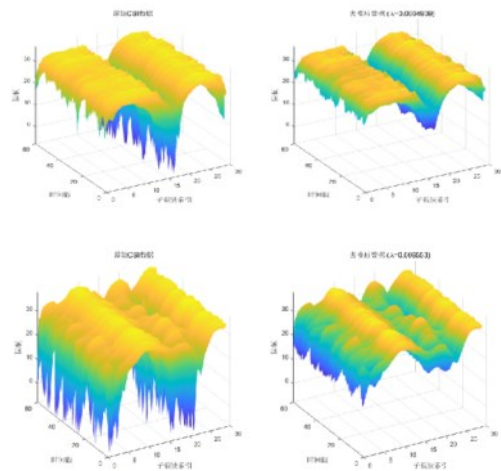


图5 降噪效果图

为进一步验证去噪效果,选取原始 CSI 数据和使用 TRPCA 降噪后的热图进行对比,降噪后热图如图 6 所示,其中的颜色代表 CSI 幅值数值的大小,能够反映出信道衰落,幅值的信号强度等信息。通过热力图可以看出,原始 CSI 热力图颜色的分布比较杂乱,黄色和橙色占比较高,说明原始数据受到环境干扰等影响,信道的特征被噪声所覆盖,数值的波动较大。通过本文引入的 TRPCA 算法去噪后的噪声颜色更加规整,蓝色和绿色的区域更加明显,能够看出噪声得到有效的抑制,信道特征更加清晰,数据的稳定性得到有效的提升。

进而通过不同算法的信噪比(SNR)、均方误差(MSE)、峰值信噪比(PSNR)作为评价指标来验证 TRPCA 对 CSI 数据的降噪性能。由表可以看出,该表定量对比了 TRPCA、PARAFAC、Coif-

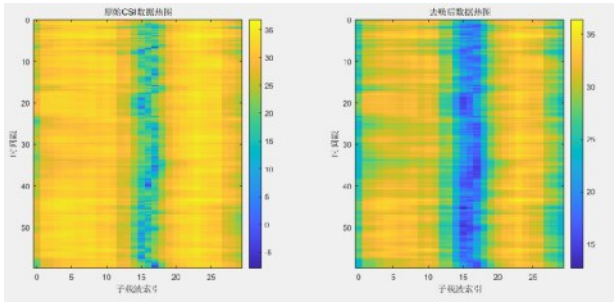


图6 降噪热力图

lets、Symlets 及 Haar 五种方法在 CSI 降噪中的性能，采用 TRPCA 处理降噪性能最好。

表2 降噪性能

	降噪性能				
	TRPCA	PARAFAC	Coiflets	Symlets	Haar
SNR/dB	65.3	63.8	58.5	57.6	57.8
MSE	6.15	6.67	7.86	8.11	7.85
PSNR/dB	40.5	40.1	39.7	39.6	38.5

本文降噪性能评价指标信噪比 SNR、均方误差 MSE、峰值信噪比 PSNR，均以经 TRPCA 算法分离得到的低秩无噪 CSI 幅值数据作为纯净参考真值，将各算法降噪后输出的 CSI 幅值数据作为待测信号，进行指标量化计算。

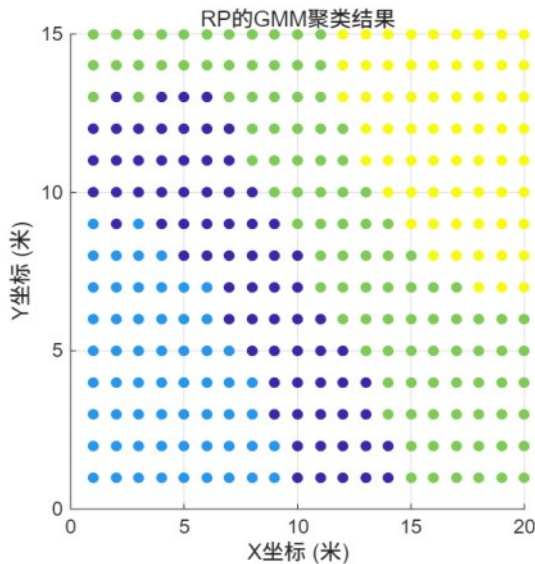


图7 GMM 聚类结果

从图7中可以看出，利用高斯混合模型进行的空间区域划分具有合理性，不同种颜色分别对应的四个簇大致对应在特定的空间区域，根据上图所示的坐标点，蓝色簇对应在左下的区域，黄色簇对应

在右上的区域，绿色簇和紫色簇对应应在中间区域。根据聚类效果可以看出高斯混合模型符合同一区域的信号特征通常更相似，并且所有 RP 都被分配到对应的簇当中，不存在完全异常点的情况。

在实验环境中存在四个 AP 点分布在不同位置，根据下图 8，计算出离线和在线时间差内每个 AP 点的斜率熵值，通过动态调整，选择斜率熵值较小的 1 号和 3 号 AP 点进行在线定位。

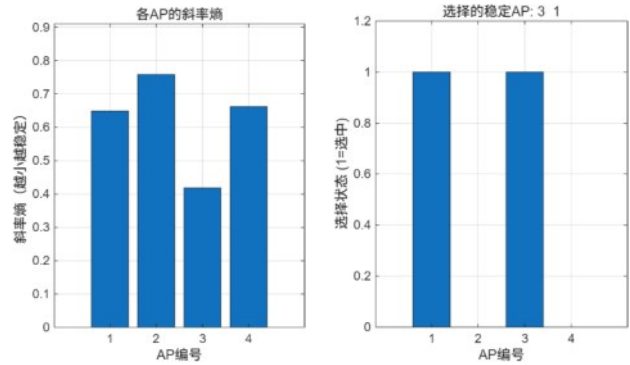


图8 AP 斜率熵选择结果

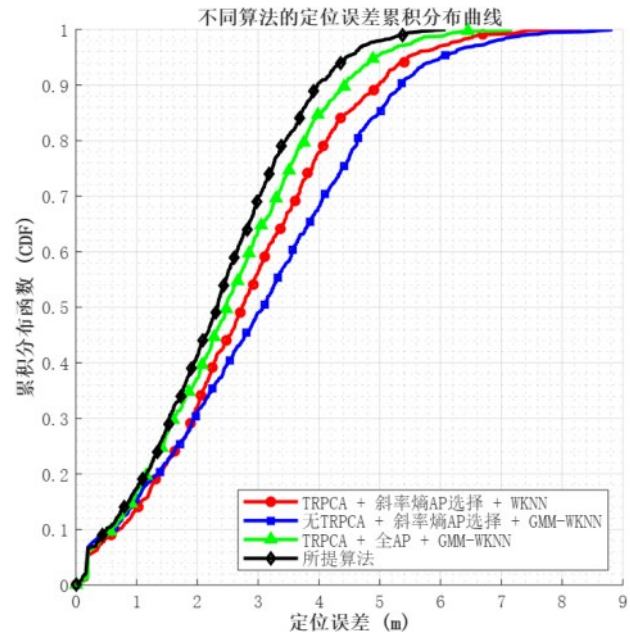


图9 消融实验误差 CDF 曲线

图9为4种不同定位方案的定位误差累积分布(CDF)对比曲线，绿色曲线采用 TRPCA 去噪与 GMM-WKNN 定位，但未进行斜率熵 AP 状态选择、直接使用全部 AP 数据，定位性能次于所提算法，验证了斜率熵 AP 选择策略能够剔除不稳定异常接入点，有效抑制环境动态干扰、优化指纹空间

质量。红色曲线保留 TRPCA 降噪与斜率熵 AP 选择模块,但采用传统距离加权 WKNN 匹配算法,精度进一步下降,证明 GMM 概率-欧氏距离融合权重策略,相较传统单一距离加权方式具备更优异的指纹匹配精度。蓝色曲线未引入 TRPCA 去噪处理,原始 CSI 信号噪声干扰严重,充分验证 TRPCA 降噪能够有效剔除室内多径噪声。综上消融实验结果表明,TRPCA 信号去噪、斜率熵 AP 选择、GMM-WKNN 融合匹配三大模块具备协同增益效应,三者联合优化可显著降低室内复杂场景定位误差,提升 WiFi-CSI 指纹定位的稳态精度与环境适应性。

为验证本文的算法精度,本文选取 WKNN, GraphLoc, CNN 来与本文算法进行对比,各种算法的定位误差累计分布函数如图 10 所示,对比结果表明,本文融合 TRPCA 去噪、斜率熵动态 AP 选择、GMM-WKNN 概率距离加权匹配的联合定位框架,所提算法在 4m 内定位精度为 0.9,能够有效抑制室内复杂环境干扰,精准刻画 CSI 指纹空间分布特征,在误差区间均具备更优异的定位稳定性与定位精度优势,相较于传统机器学习、深度学习主流定位方案具备显著性能增益。

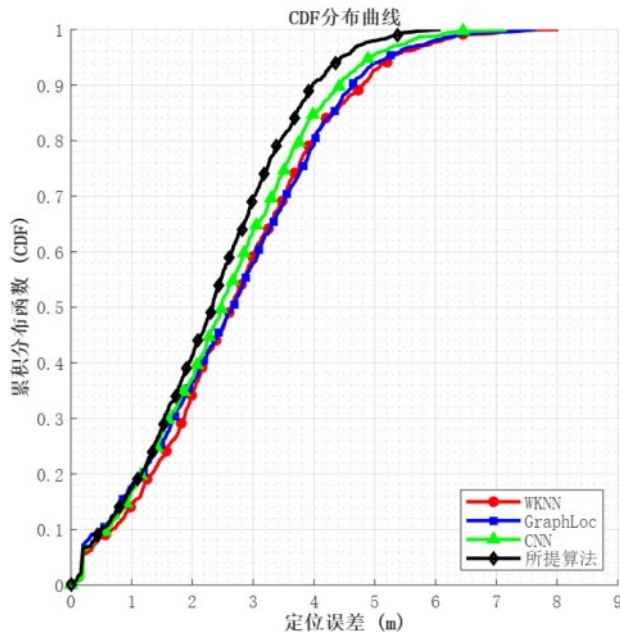


图 10 各定位算法误差 CDF 曲线

为使验证更加准确性,对误差数值进行对比,根据表 3 误差数值所示,本文算法误差中位数,均值,标准差均低于对比算法,证明本文算法优于以上算法。

	中位数	均值	标准差
本文算法	2.24m	2.29m	2.32m
WKNN	2.71m	2.68m	2.66m
GraphLoc	2.54m	2.59m	2.62m
CNN	2.49m	2.48m	2.53m

	WKNN	GraphLoc	CNN	本文算法
1m	0.14	0.15	0.15	0.17
2m	0.35	0.37	0.36	0.41
3m	0.60	0.59	0.64	0.69
4m	0.81	0.79	0.85	0.90

表 4 统计了 WKNN、GraphLoc、CNN 及本文所提算法在 1 m、2 m、3 m、4 m 定位误差阈值下的累积定位准确率,量化表征不同算法在近距离定位区间内的定位性能。整体来看,所提算法通过 CSI 降噪、AP 优选与 GMM 聚类优化,有效增强了指纹特征区分度,在近距离定位场景中误差约束能力更强、定位可靠性更高,相较传统 WKNN、GraphLoc 及 CNN 方法具备更优的定位精度与鲁棒性。

5 结束语

为解决 WIFI 指纹定位中多径效应导致的 CSI 数据尖峰噪声、在线离线时间差引发的 AP 状态受限及传统算法定位精度低的问题,本文提出基于信道状态信息(CSI)的定位算法:通过将 CSI 数据按子载波、幅值、时间戳构建三维张量,采用张量鲁棒性分析(TRPCA)结合增广拉格朗日乘子法(ADMM)分解为低秩向量与稀疏噪声实现降噪;引入斜率熵动态筛选稳定 AP 点;将高斯混合模型与加权 K 近邻(WKNN)结合优化定位,经实验验证,该算法在定位精度和鲁棒性上均优于传统算法。

参考文献:

[1] JO K, CHU K, SUNWOO M. Interacting multiple model filter-based sensor fusion of GPS with in-vehicle sensors for real-time vehicle positioning[J]. IEEE Transactions on Intelligent Transportation Systems, 2012, 13(1): 329-343.

[2] Thanapon K, Prasert K. The study of total electron content on iono-

- sphere by using single frequency GPS receiver[J]. *Advances in Space Research*, 2025, 75(5): 4245-4259.
- [3] 桂林卿, 蒋祎, 张卿云, 等. RFreeTrack: 基于差分相位全息图的 RFID 非接触式鲁棒定位系统[J]. *通信学报*, 2024, 45(03): 92-103.
- [4] Abdulzahra A M K, Al-Qurabat A K M, Abdulzahra S A. Optimizing energy consumption in WSN-based IoT using unequal clustering and sleep scheduling methods[J]. *Internet of Things*, 2023, 22: 100765.
- [5] Cheraghinia M, et al. A Comprehensive Overview on UWB Radar: Applications, Standards, Signal Processing Techniques, Datasets, Radio Chips, Trends and Future Research Directions[J]. *IEEE Communications Surveys & Tutorials*, 2025, 27(4): 2283-2324.
- [6] Fang F C, Sun J, Han Y L, Zhu H B. Performance analysis of ZigBee under WLAN and multiple bluetooth piconets interferences[C]. *IET International Conference on Information Science and Control Engineering 2012 (ICISCE 2012)*, Shenzhen, 2012: 1-7.
- [7] Peerzade S N, Mudda S. A Compact Asymmetric Coplanar Strip (ACS) Antenna for WLAN and Wi-Fi Applications[C]. *2023 International Conference on Emerging Smart Computing and Informatics (ESCI)*, Pune, India, 2023: 1-4.
- [8] Ha G Y, Seo S B, Oh H S, Jeon W S. LoRa ToA-Based Localization Using Fingerprint Method[C]. *2019 International Conference on Information and Communication Technology Convergence (ICTC)*, Jeju, Korea (South), 2019: 349-353.
- [9] 赵越, 李赞, 李冰, 等. TDoA 定位盲区分析与节点部署策略研究[J]. *通信学报*, 2023, 44(01): 1-13.
- [10] 蒋锐, 李俊, 徐友云, 等. 基于联邦卡尔曼滤波器的容错 GPS-AOA-SINS 组合导航算法[J]. *通信学报*, 2022, 43(08): 78-89.
- [11] 张学军, 李梅, 陈惠, 等. 面向指纹室内定位的高鲁棒性集成对抗训练方法[J]. *通信学报*, 2025, 46(08): 105-118.
- [12] Zhang L, Qiu Y, Wu S, Zhang T, Zhang Q. GraphLoc: Enhancing Fingerprint-Based Localization With Graph Representation Learning [J]. *IEEE Internet of Things Journal*, 2025, 12(12): 21593-21603.
- [13] Luong T V, Ko Y. The BER Analysis of MRC-Aided Greedy Detection for OFDM-IM in Presence of Uncertain CSI[J]. *IEEE Wireless Communications Letters*, 2018, 7(4): 566-569.
- [14] Zhou M, et al. Adaptive Genetic Algorithm-Aided Neural Network With Channel State Information Tensor Decomposition for Indoor Localization[J]. *IEEE Transactions on Evolutionary Computation*, 2021, 25(5): 913-927.
- [15] Ma L, Zhang Y, Qin D. A Novel Indoor Fingerprint Localization System Based on Distance Metric Learning and AP Selection[J]. *IEEE Transactions on Instrumentation and Measurement*, 2022, 71: 1-15.
- [16] Zheng W, Pang S, Liu N, Chai Q, Xu L. A Compact Snake Optimization Algorithm in the Application of WKNN Fingerprint Localization [J]. *Sensors*, 2023, 23(14): 6282.
- [17] SONG Q W, GUO S T, LIU X, et al. CSI amplitude fingerprinting-based NB-IoT indoor localization[J]. *IEEE Internet of Things Journal*, 2018, 5(3): 1494-1504.
- [18] Moreira A R, Ramos L P, Da Silva F G, Alves D I, Machado R. Automatic Classification of Maritime Targets Based on TRPCA Pre-Processing[C]. *2024 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Athens, Greece, 2024: 9748-9752.
- [19] Li Y, Tian G, Cao Y, Yi Y, Zhou D. Optimized Fuzzy Slope Entropy: A Complexity Measure for Nonlinear Time Series[J]. *IEEE Transactions on Instrumentation and Measurement*, 2024, 73: 1-14.
- [20] Gogebakan M. A Novel Approach for Gaussian Mixture Model Clustering Based on Soft Computing Method[J]. *IEEE Access*, 2021, 9: 159987-160003.



李华亮, 男, 长春工业大学计算机科学与工程学院讲师, 博士毕业于吉林大学通信与信息系统专业, 主要从事无线定位技术研究, 吉林省高层次人才 E 类人才。

张家伟, 男, 长春工业大学计算机科学与工程学院硕士, 主要从事无线定位技术研究

黄平, 女, 长春工业大学计算机科学与工程学院讲师, 博士毕业于吉林大学控制科学与工程专业, 主要从事自动驾驶智能决策与控制

