

基于核函数特征提取的室内定位算法研究

李华亮¹, 钱志鸿¹, 田洪亮^{1,2}

(1. 吉林大学通信工程学院, 吉林 长春 130012; 2. 东北电力大学信息工程学院, 吉林 吉林 132012)

摘 要: 提出了一种基于核函数特征提取(KPCA, kernel principal component analysis)的室内定位算法。该算法在离线阶段使用核函数特征提取方法训练原始位置指纹(OLF, original location fingerprint), 提取原始位置指纹的非线性特征, 可以有效地利用各个接入节点(AP, access point)的接收信号强度信息; 而在线阶段使用一种改进的加权 k 近邻 (IWKNN, improved weight k -nearest neighbor)算法, 自主选择近邻数进行位置估计。实验结果表明, 提出的算法在平均误差和定位准确率方面优于其他的室内定位算法, 并且该算法需要更少的接收信号强度 (RSS, received signal strength) 采集次数和 AP 个数。

关键词: 无线局域网; 室内定位; 接收信号强度; 核函数特征提取

中图分类号: TN92

文献标识码: A

Research on indoor localization algorithm based on kernel principal component analysis

LI Hua-liang¹, QIAN Zhi-hong¹, TIAN Hong-liang^{1,2}

(1. College of Communication Engineering, Jilin University, Changchun 130012, China;

2. School of Information Engineering, Northeast Electric Power University, Jilin 132012, China)

Abstract: An indoor localization algorithm based on kernel principal component analysis (KPCA) was proposed. It applied KPCA to train the original location fingerprint (OLF) and extract the nonlinear feature of the OLF data at the offline stage, such that the information of all AP was more efficiently utilized. At the online stage, an improved weight k -nearest neighbor algorithm for positioning which could automatically choose neighbors was proposed. The experiments were carried out in a realistic WLAN environment. The results show that the algorithm outperforms the existing methods in terms of the mean error and localization accuracy. Moreover, it requires less times of RSS acquisition and AP number.

Key words: WLAN, indoor localization, RSS, KPCA

1 引言

随着计算机技术和互联网应用的发展, 移动终端定位得到了广泛的关注, 尤其在物联网和设备与设备(D2D, device to device)通信的应用架构中^[1,2], 无论室外还是室内的位置信息都是不可或缺的。传统的基于卫星的定位技术, 如全球定位系统

(GPS, global positioning system), 能够很好地满足室外定位的需求, 但是用于室内定位时, 由于建筑物对卫星信号的影响, 定位能力受到很大的限制^[3], 相反, 基于无线网络技术的定位系统进行室内定位时, 定位效果明显提高。WLAN 技术因其在室内布设简单、部署广泛、价格低廉等特性更加适用于室内定位。近年来, 基于 WLAN 技术的室内定位成

收稿日期: 2016-06-27; 修回日期: 2016-10-14

通信作者: 钱志鸿, droqzh@163.com

基金项目: 国家自然科学基金资助项目 (No.61371092, No.61401175, No.61540022); 吉林省科技发展计划基金资助项目 (No.20140204019GX); 长春市重大科技攻关计划基金资助项目(No.2014026 /14KG021); 吉林大学研究生创新基金资助项目 (No.2019091)

Foundation Items: The National Natural Science Foundation of China (No.61371092, No.61401175, No.61540022), Scientific and Technological Developing Scheme of Jilin Province (No.20140204019GX), Key Science and Technology Program of Changchun (No.2014026/14KG021), Project Supported by Graduate Innovation Fund of Jilin University (No.2019091)

为研究热点^[4-6]，其中，基于位置指纹架构的定位方法成为主流。

一般来说，基于位置指纹架构的定位算法主要分为 2 个阶段^[7]：离线训练阶段和在线定位阶段，如图 1 所示。

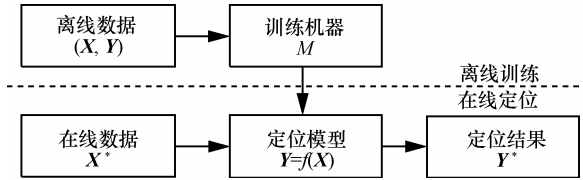


图 1 指纹模型定位框架

离线训练阶段的目的是通过训练机器 M ，得到无线信号向量 X 与物理位置 Y 之间的映射关系 $Y=f(X)$ ，将其提交到在线定位阶段，利用在线采集的无线信号向量 X^* 预测出当前所在位置 Y^* 。

在 WLAN 环境下的基于 RSS 信号的位置指纹定位算法中^[8,9]，离线训练阶段，采集定位区域内各参考节点(RP, reference point)来自各个 AP(access point)的 RSS 信号，将其组成有序向量，与相应 RP 的物理位置，由此构成位置指纹，全部参考节点的位置指纹信息存储在一个数据库中，构成位置指纹空间；在线定位阶段，采集测试地点(TP, testing point)的在线位置指纹数据，利用模式匹配算法与位置指纹空间中的数据相匹配，估计待测用户的位置。

这种基于位置指纹的定位模型，其定位精度很大程度上取决于离线训练的数据和在线采集到的数据是否满足相同的分布模型。在实际的室内 WLAN 环境中，由于室内环境的复杂性、多径效应、阴影效应、人员走动、信道拥塞和节点的通信半径有限等因素的影响^[10]，在 RP 上采集的各个 AP 的 RSS 信号往往表现出复杂多变的特性，严重影响定位算法的精度。

针对于此，Bahl 等^[11]提出了使用 k 近邻 (KNN, k -nearest neighbor) 算法进行定位，加权 k 近邻算法 (WKNN, weight k -nearest neighbor) 是对其提出的后续改进，该方法首先通过移动终端在已知参考位置上接收来自各个 AP 的 RSS 信号建立位置指纹数据库，定位过程中使用最小二乘法找到 k 个最接近的位置，然后对 k 个位置求加权平均，将获得的结果作为目标的位置。这种算法有一缺点，它在定位时，无论位置指纹受多大的干扰都会取前 k 个位置指纹进行位置估计，无疑会增加定位误差。

Chen 等^[12]提出一种基于聚类 and 决策树的方法 (CaDet, clustering and decision-tree-based method)，来

选择定位区域最佳的 AP 的 RSS 值作为位置指纹，这样定位区域未被选择的 AP 及其对应的 RSS 值将被抛弃，因此将影响最后的定位结果，特别是当定位区域可利用的 AP 个数较少时，定位能力受到很大的限制。

Fang 等^[13]使用主特征提取 (PCA, principal component analysis) 方法训练离线位置指纹，提取位置指纹样数据间主要的成分，以克服环境变化带来的 RSS 不确定性波动的影响，相比于传统的定位方法，该方法极大地提高了定位精度，且证明了 Fang 等提出的算法优于 CaDet 方法。PCA 方法^[14]能提取数据间的线性特征，但忽略了数据间高于 2 阶的关系，限制了非线性特征的提取，而在实际的 WLAN 环境中，由于 RSS 信号存在的时变特性，使真实的位置指纹数据间通常呈现高于 2 阶的非线性关系^[10]，影响了 PCA 方法的效果。

本文提出一种基于核函数特征提取 (KPCA, kernel principal component analysis) 的室内定位方法。离线阶段时使用 KPCA 方法训练位置指纹数据，KPCA 方法可以有效提取数据间的非线性特征，而且与 PCA 方法不同的是，KPCA 可以将位置指纹空间扩展到更高维 (大于位置指纹空间的维度) 的样本空间进行分析，更大程度地利用了位置指纹数据；在线阶段提出了一种改进的加权 k 近邻 (IWKNN, improved weight k -nearest neighbor) 算法，按照指纹数据间的累计贡献度自主决定 WKNN 算法的近邻数进行位置预测。本文算法使用的 RSS 数据来自真实 WLAN 环境，RSS 数据采集实验表明本文提出的算法优于其他几种 WLAN 环境下的室内定位算法^[11,13]。

2 KPCA 简介

PCA 方法是识别领域中经典的特征提取和降维方法，PCA 是确定一个坐标系统的直交变换，在新的坐标系统下，变换数据点的方差沿新的坐标轴得到了最大化，新的坐标系统的这些坐标轴即是主成分，而 KPCA 是 PCA 的核函数扩展方法^[15]。KPCA 方法是通过非线性映射函数 Φ ，这里采用核方法，将原始空间映射到特征空间 F 上，在特征空间 F 上进行 PCA，其过程如图 2 所示。

核方法是一系列数据处理方法的总称，其共同特征都是采用了核函数进行映射。核方法的思想如式(1)所示， x_i 和 x_j 是数据空间的样本，数据空间到特征空间的映射函数为 Φ ，核函数为 K ，核方法即将特征空间元素的内积运算变换为数据空间元素

的核函数运算。核方法从本质上实现了数据空间、特征空间和类别空间之间的非线性变换。

$$(\mathbf{x}_i, \mathbf{x}_j) \rightarrow K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)\Phi(\mathbf{x}_j) \quad (1)$$

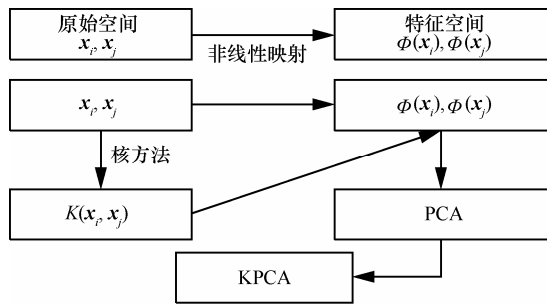


图 2 KPCA 框架

3 基于核函数特征提取的室内定位算法

本文提出了一种基于核函数特征提取的室内定位算法，命名为 KPCA-IWKNN 定位算法，定位流程如图 3 所示。在离线阶段，在选定的定位区域布设 N 个参考节点，每个参考节点的物理位置为 $l_i(x_i, y_i)$ ， N 个参考节点的物理位置信息构成一个位置空间 $L = (l_1, l_2, \dots, l_N)^T$ ，在各个参考节点上可以采集到 n 个非视距 AP 的 RSS 信号及物理地址(MAC, media access control)信息 (MAC 作为 AP 的标识)，在每个参考节点上都要进行 p 次采集，将 RSS 均值作为这个参考节点 $l_i(x_i, y_i)$ 的原始位置指纹信息，这个原始位置指纹信息是一个 n 维向量 $F_i = (rss_1, rss_2, \dots, rss_n)^T, i \in (1, N)$ ，其中， $rss_i = rss_{i,p}$ 为 p 次

采样之后来自第 i 个 AP 的 RSS 均值。将全部参考节点的原始位置指纹信息存储在相关数据库中，构成一个 $N \times n$ 维的原始位置指纹空间 F ，即 $F = (F_1, F_2, \dots, F_N)^T$ ，矩阵 F 中的每个行向量代表一个参考节点的原始位置指纹向量。通过 KPCA 方法，提取原始位置指纹数据的定位特征，这些定位特征构成特征位置指纹空间 $F' = (F'_1, F'_2, \dots, F'_N)^T$ ， F' 与 L 相对应，即 F'_i 是 $l_i(x_i, y_i)$ 的特征位置指纹。在线阶段，在测试节点采集各个 AP 的实时的 RSS 信号，构成在线指纹向量 S ，对 S 进行 KPCA 处理，得到在线特征指纹向量 S' ，使用 IWKNN 方法估计出位置 Position。

3.1 原始位置指纹的 KPCA 变换

设非线性变换 Φ 实现了原始位置指纹空间 F 到特征空间的映射，即 F_1, F_2, \dots, F_N 变换为特征空间的样本点 $\Phi(F_1), \Phi(F_2), \dots, \Phi(F_N)$ (来自特征空间为 $\Phi(F)$)，并且假设特征空间的数据满足中心化条件，如式(2)所示。

$$\sum_{i=1}^N \Phi(F_i) = 0 \quad (2)$$

根据式(3)计算特征空间的协方差矩阵 C 。

$$C = \frac{1}{N} \sum_{i=1}^N \Phi(F_i)\Phi(F_i)^T \quad (3)$$

求协方差矩阵 C 的特征值 λ 和特征向量 V ，如式(4)所示。

$$\lambda \Phi(F_i)V = \Phi(F_i)CV, \quad i=1, 2, \dots, N \quad (4)$$

特征向量 V 属于 $\Phi(F_1), \Phi(F_2), \dots, \Phi(F_N)$ 生成的

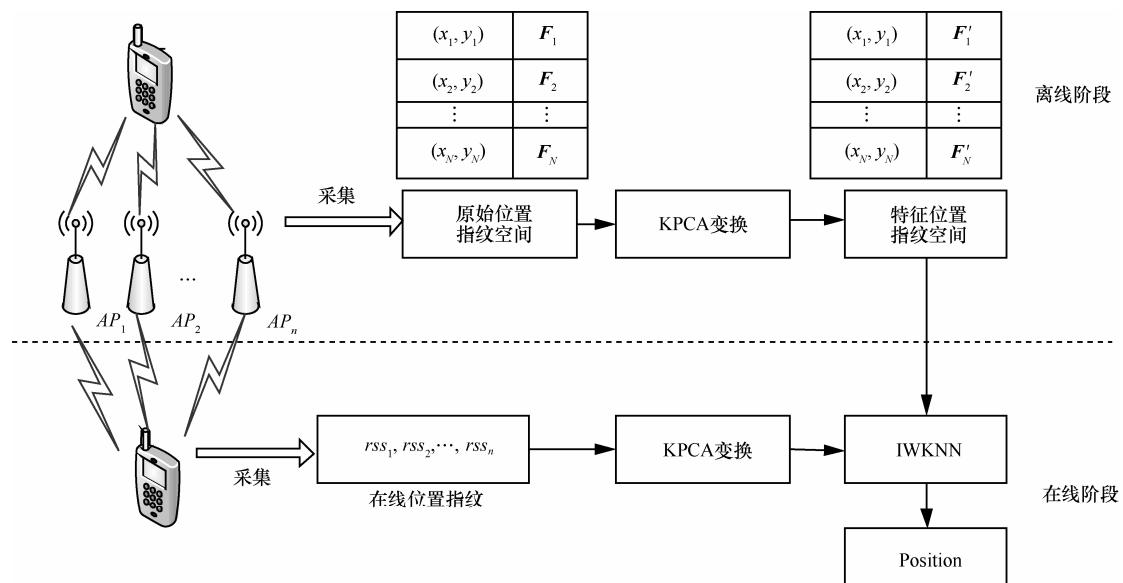


图 3 KPCA-IWKNN 的无线室内定位算法流程

空间，所有的特征向量都可以表示为 $\Phi(F_1), \Phi(F_2), \dots, \Phi(F_N)$ 的线性张成，如式(5)所示。

$$V_i = \sum_{t=1}^N \alpha_t \Phi(F_t) \quad (5)$$

其中， α_t 是 $\Phi(F_t)$ 同阶的系数向量，将式(3)、式(5)代入式(4)中，得到

$$\frac{1}{N} \sum_{t=1}^N \alpha_t \left(\sum_{i=1}^N (\Phi(F_i) \Phi(F_i) \Phi(F_t) \Phi(F_t)) \right) = \lambda \sum_{i=1}^N (\Phi(F) \Phi(F_i)) \quad (6)$$

定义 $N \times N$ 阶核矩阵 K ，它的各个元素表达如式(7)所示

$$K_{ii} = (\Phi(F_i) \Phi(F_i)) \quad (7)$$

因此，式(6)可以化简为

$$N\lambda K\alpha = K^2\alpha \quad (8)$$

对式(8)左右同乘 K^{-1} 可以得到式(9)

$$N\lambda\alpha = K\alpha \quad (9)$$

根据式(9)可知，求解矩阵 K 的特征值和特征向量就能得到 C 的特征值及相对应的特征向量。

设 K 的前 $k(k \leq N)$ 个最大的特征值 $\lambda_1 \geq \lambda_2 \dots \lambda_{k-1} \geq \lambda_k$ ，及对应的 k 个单位正交化特征向量 $\alpha = (\alpha_1, \alpha_2 \dots \alpha_k)^T$ 。对特征空间 $\Phi(F)$ 进行特征提取，就是求 $\Phi(F)$ 到特征向量空间的投影，如式(10)所示。

$$(V\Phi(F)) = \sum_{t=1}^N \alpha_t (\Phi(F_t), \Phi(F)) \quad (10)$$

将式(7)代入式(10)得到

$$(V\Phi(F)) = \sum_{t=1}^N \alpha_t K(F_t, F) \quad (11)$$

在一般情况下特征空间的数据不满足中心化条件，即式(2)不成立，因此，需要对特征空间的数据进行调整，调整方式如式(12)所示。

$$\tilde{\Phi}(F_i) = \Phi(F_i) - \frac{1}{N} \sum_{j=1}^N \Phi(F_j) \quad (12)$$

而修正后的核矩阵元素如式(13)所示。

$$\begin{aligned} \tilde{K}_{ii} &= (\tilde{\Phi}(F_i) \tilde{\Phi}(F_i)) \\ &= \Phi(F_i) \Phi(F_i) - \frac{1}{N} \sum_{n=1}^N \Phi(F_i) \Phi(F_n) - \\ &\quad \frac{1}{N} \sum_{j=1}^N \Phi(F_j) \Phi(F_i) + \frac{1}{N^2} \sum_{j=1}^N \sum_{n=1}^N \Phi(F_j) \Phi(F_n) \end{aligned}$$

$$= K_{ii} - \frac{1}{N} \sum_{n=1}^N K_{in} - \frac{1}{N} \sum_{j=1}^N K_{ji} + \frac{1}{N^2} \sum_{j=1}^N \sum_{n=1}^N K_{jn} \quad (13)$$

进一步简化式(13)，得到式(14)，其中， C 为 $N \times N$ 阶矩阵，每一个元素都是 $\frac{1}{N}$ 。

$$\tilde{K} = K - CK - KC + CKC \quad (14)$$

综上所述，原始位置指纹数据的 KPCA 变换如式(15)所示。

$$(V\Phi(F)) = \sum_{t=1}^N \alpha_t \tilde{K}(F_t, F) \quad (15)$$

KPCA 实现的优良程度取决于核函数的选取，高斯核函数是模式识别中最常用的核函数，高斯核函数有良好的局部特征提取能力和平滑特性^[16]，选用高斯核函数的 KPCA 可以有效挖掘位置指纹间的非线性关系，高斯核函数如式(16)所示。

$$K(x_i, x_j) = \exp\left(\frac{|x_i - x_j|^2}{-\varepsilon^2}\right) \quad (16)$$

基于以上描述，原始位置指纹空间的 KPCA 的处理过程如下所示。

1) 将 F 作为输入空间，根据式(16)计算核矩阵 K ， $K_{ij} = \exp\left(\frac{|F_i - F_j|^2}{-\varepsilon^2}\right)$ ， ε 为高斯核宽度，其中，

$i, j \leq N$ ， K 为 $N \times N$ 维矩阵。

2) 根据式(14)修正核矩阵 \tilde{K} 。

3) 找到 \tilde{K} 的前 k 个最大特征值 $\lambda_1 \geq \lambda_2 \dots \lambda_{k-1} \geq \lambda_k$ ，及对应的特征向量 v_1, v_2, \dots, v_k 。

4) 通过施密特正交化方法将 v_1, v_2, \dots, v_k 单位正交化，得到 $\alpha_1, \alpha_2, \dots, \alpha_k$ 。

5) 根据式(15)计算特征位置指纹空间 $F' = \tilde{K}\alpha$ ，其中， $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)^T$ 为 $N \times k$ 维矩阵。

通过 KPCA 处理过程可以将 $N \times n$ 阶原始位置指纹空间变换为 $N \times k$ 特征位置指纹空间 F' 。

3.2 在线位置指纹的处理

在测试地点采集来自各个 AP 的 RSS 信号构成在线指纹向量 S ，对 S 进行 KPCA 变换后，得到在线特征指纹向量 S' ，如式(17)和式(18)所示，得到的 S' 方可用于在线定位。

$$S' = (V\Phi(S)) = \sum \alpha_t \tilde{K}(F_t, S) \quad (17)$$

$$\begin{aligned} \tilde{\mathbf{K}}(\mathbf{F}_i, \mathbf{S}) &= (\Phi(\mathbf{F}_i) - \frac{1}{N} \sum_{i=1}^N \Phi(\mathbf{F}_i))(\Phi(\mathbf{S}) - \frac{1}{N} \sum_{j=1}^N \Phi(\mathbf{F}_j)) \\ &= \mathbf{K}(\mathbf{F}_i, \mathbf{S}) - \frac{1}{N} \sum_{j=1}^N K_{ij} - \frac{1}{N} \sum_{i=1}^N K(\mathbf{F}_i, \mathbf{S}) + \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N K_{ij} \end{aligned} \quad (18)$$

3.3 IWKNN 定位

计算 \mathbf{S}' 与特征位置指纹空间中各个特征位置指纹 \mathbf{F}'_i 的欧氏距离, 如式(19)所示。

$$D_i(\mathbf{S}', \mathbf{F}'_i) = \sqrt{\sum_{j=1}^k (s'_j - rss_{ij}')^2}, i \in (1, N) \quad (19)$$

$D_i(\mathbf{S}', \mathbf{F}'_i)$ 可以表征 \mathbf{S}' 与 \mathbf{F}'_i 间的相似程度, 其值越小, 二者越相似。在特征位置指纹空间中, \mathbf{F}'_i 是位置 $l_i(x_i, y_i)$ 的特征位置指纹信息。按 $D_i(\mathbf{S}', \mathbf{F}'_i)$ 的大小排列, 找到前 $m(m < N)$ 个最小的 D_i 及其相对应的 m 个特征位置指纹和位置信息 $l_i(x_i, y_i)$, 使其满足

$$\frac{\sum_{i=1}^m \frac{1}{D_i + d_0}}{\sum_{i=1}^N \frac{1}{D_i + d_0}} \geq \sigma \quad (20)$$

其中, d_0 是一个非常小的数, 防止分母为 0, σ 是一个小于 1 的正数, 为位置指纹的累计贡献度, 根据式(20)可以自主确定 m 值大小, 最后位置估计如式(21)所示。

$$position = \frac{\sum_{i=1}^m \left(\frac{1}{D_i + d_0} l_i \right)}{\sum_{i=1}^m \frac{1}{D_i + d_0}} \quad (21)$$

3.4 算法优势

根据以上对算法的描述, 可以得出算法有如下优势。

1) 算法在离线阶段首次使用 KPCA 方法训练原始位置指纹, 提取原始位置指纹的非线性特征, 将这些非线性特征组成特征位置指纹数据库, 而且根据式(15)可以看出, KPCA 方法可以将位置指纹空间扩展到更高维 (大于原位置指纹空间的维度) 的样本空间进行分析, 增加了指纹对比特征, 更加有效地利用位置指纹信息。

2) 在线定位的阶段, 通过引入式(20), IWKNN 算法可以自适应地选择定位输出所需的位置指纹

数目, 避免了 WKNN 和 KNN 算法中使用固定数目的位置指纹进行定位所带来的误差。

3) 为了分析算法在现实 WLAN 环境下的特性, 本文使用的 RSS 数据来自真实的 WLAN 环境, 而不是通过仿真得到, 数据更加具有代表性。

4 实验结果与仿真分析

4.1 实验设置

本文通过仿真的形式对提出的定位算法进行了详细的性能分析, 仿真使用的数据均来自于真实的环境, 数据采集实验在吉林大学通信工程学院第一教学楼进行。综合考虑诸多因素, 选择第一教学楼 3 楼的一块 20 m × 2 m 的走廊区域作为实验区域, 在这个区域, 人员走动频繁, 干扰复杂多变, 这个环境下采集的 RSS 数据具有很强的代表性, 其地形如图 4 所示, 实验环境以 313 室为起点, 设其坐标为 (0,0), 每隔 2 m 取一个参考节点, 共计 22 个参考节点, 分布如图 4 中 RP; 又在实验区域选择 31 个位置点, 作为在线测试节点, 其分布如图 4 中 TP, TP 的实际位置已知, 在线定位时, 实时采集各个 TP 上来自各个 AP 的 RSS 信号, 构成在线位置指纹, 通过定位算法得到 TP 的估计位置, 将 TP 的估计位置与 TP 的实际位置进行比较, 来评价定位算法的性能。

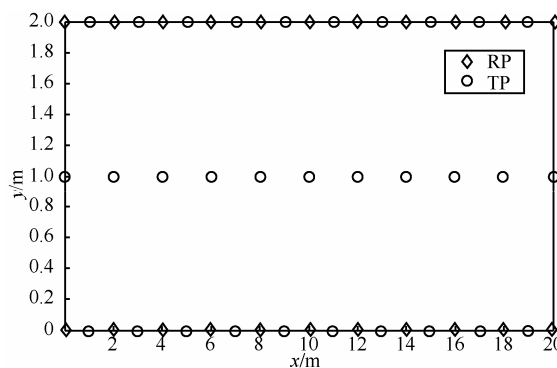


图 4 数据采集实验环境

借助现有的 WLAN 基础设施, 在实验区域可以检测到 8~13 个 AP, 所有 AP 在数据采集区域内都是非视距的, 实验中忽略了 AP 位置因素以及 AP 硬件对算法的影响; RSS 信号采集的终端设备使用 MI2SC 手机, 处理器为四核 1.7 GHz, 操作系统 MIUI-JL50.0, 每个参考节点和测试节点的 RSS 数据都经过多次采集 (最高 100 次) 计算均值, 得到原始位置指纹数据, 部分数据如表 1 所示, 表 1 中

的数据为原始位置指纹，一般情况下可以直接用来定位。

将原始位置指纹经过 PCA 处理后，提取 5 个主要特征后得到数据如表 2 所示。PCA 方法能够提取数据的主要线性特征，达到数据降维的目的，但不能增加原数据的特征。

将原始位置指纹经过 KPCA 处理，在高斯核宽

度为 2 时，得到 5 维和 15 维的数据库，如表 3 和表 4 所示。原始位置指纹数据库中的数据为设备的 RSS 值，KPCA 是一个非线性过程，变换后的数据便不具有实际的物理意义，仅作为特征位置指纹数据使用。

实验采用平均误差(ME, mean error)和定位准确率为标准来比较 4 种方法。

表 1 进行 40 次采集的原始位置指纹数据库(RSS/dBm)

(x, y)	AP ₁	AP ₂	AP ₃	AP ₄	AP ₅	AP ₆	AP ₇	AP ₈	AP ₉	AP ₁₀	AP ₁₁	AP ₁₂	AP ₁₃
(0, 0)	-67	-67	-72	-72	-72	-76	-77	-78	-79	-83	-84	-92	-93
(0, 2)	-66	-68	-71	-65	-72	-76	-72	-78	-77	-72	-73	-80	-88
(2, 2)	-68	-73	-77	-89	-79	-75	-64	-83	-76	-80	-78	-88	-78
(2, 0)	-64	-64	-69	-72	-82	-81	-73	-71	-76	-79	-80	-79	-84
(4, 0)	-72	-73	-61	-73	-70	-76	-67	-70	-82	-83	-81	-76	-83
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

表 2 PCA 提取 5 维的位置指纹数据

(x,y)	AP' ₁	AP' ₂	AP' ₃	AP' ₄	AP' ₅
(0, 0)	-30.76	-1.62	0.00	7.41	0.02
(0, 2)	-18.84	-5.01	-1.47	5.62	2.77
(2, 2)	-26.19	-3.19	-1.23	8.32	10.27
(2, 0)	-18.08	-3.19	-6.33	-2.45	-2.42
(4, 0)	-14.95	-13.26	-0.90	9.62	-1.37
⋮	⋮	⋮	⋮	⋮	⋮

表 3 KPCA 变换 5 维的特征指纹数据

(x,y)	AP' ₁	AP' ₂	AP' ₃	AP' ₄	AP' ₅
(0, 0)	0.05	0.08	-0.11	0.09	0.13
(0, 2)	-0.66	-0.10	0.00	0.01	0.00
(2, 2)	0.05	0.08	-0.11	0.10	0.10
(2, 0)	0.05	0.09	-0.11	0.10	-0.68
(4, 0)	0.05	0.08	-0.10	-0.52	-0.01
⋮	⋮	⋮	⋮	⋮	⋮

表 4 KPCA 变换 15 维的特征位置指纹数据

(x,y)	AP' ₁	AP' ₂	AP' ₃	AP' ₄	AP' ₅	AP' ₆	AP' ₇	AP' ₈	AP' ₁₀	AP' ₁₁	AP' ₁₂	AP' ₁₃	AP' ₁₄	AP' ₁₅
(0, 0)	0.06	0.08	-0.11	0.09	0.13	0.20	0.16	0.08	-0.38	-0.56	-0.02	0.00	-0.02	0.00
(0, 2)	-0.66	-0.10	0.00	0.01	0.00	0.00	0.01	0.16	0.00	0.11	0.00	0.05	0.13	-0.25
(2, 2)	0.05	0.08	-0.11	0.10	0.10	0.18	0.19	0.82	-0.11	0.34	0.00	0.17	-0.02	0.01
(2, 0)	0.05	0.09	-0.11	0.10	-0.68	-0.13	0.01	0.01	-0.02	-0.03	-0.01	0.05	-0.02	0.03
(4, 0)	0.05	0.08	-0.10	-0.52	-0.01	-0.02	0.00	0.00	-0.01	-0.01	-0.01	0.01	-0.61	-0.05
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

假设某待测目标的真实位置为 Y_i ，根据定位系统预测的位置为 Y'_i ，则 ME 的定义如式(22)所示。

$$ME = \frac{1}{N} \sum_{i=1}^N \sqrt{|Y'_i - Y_i|} \quad (22)$$

定位系统在定位时难免会产生定位误差，但是人们对于一定的误差是可以接受的。首先定义误差距离(ED, error distance)，如图 5 所示。待测目标的实际位置 Y_i ，若定位系统预测的位置 Y'_i 与实际位置 Y_i 之间的距离小于 ED，则此次定位可以认定为准确定位。即 $|Y'_i - Y_i| \leq ED$ ，则 Y'_i 为准确定位；反之 $|Y'_i - Y_i| > ED$ ，则 Y'_i 为错误定位。定位准确率是指定位系统多次定位中准确定位次数占全部的定位次数的比率。

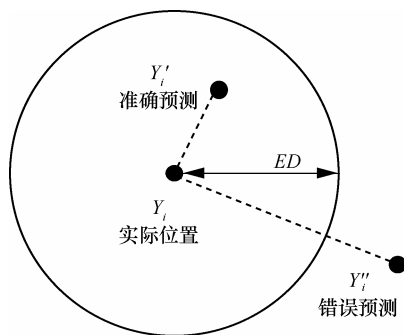


图 5 准确定位概念

在仿真中将本文提出的算法 (KPCA-IWKNN) 与 3 种传统的定位方法^[11,13] (KNN 算法、WKNN 算法和 PCA-WKNN 算法) 进行比较。KPCA 的核函数选择高斯核函数，高斯核宽度 ε 按经验取值 2，WKNN 和 KNN 算法的近邻数目取 4。定位算法的软件环境为环境为 Matlab 8.0, Windows XP, 硬件环境为 Pentium Dual Core 4 GHz CPU, 2 GB 内存的 PC 机。

4.2 AP 数量对算法的影响

在指纹定位算法中，位置指纹空间使用的 AP 数量是一个影响定位性能的重要参数，显然，AP 数量越多，定位效果越好。本文忽略 AP 位置因素对各定位算法带来的影响，故依照表 1 中的数据依次选取前 n 个 AP 作为参数。当测试区仅有 1 个 AP 的时候，指纹架构的定位算法不能进行定位，故 n 取 2~13 之间的整数。在这个实验中，KPCA 将原始位置指纹空间训练到 14 维的特征位置指纹空间， σ 取值为 0.3，PCA 方法选取 n 个主元用以定位。图 6 和图 7 是 4 种算法随着 AP 数量增加的性能比

较，进行了 100 次定位实验，无论是平均定位误差还是定位准确率，本文提出的算法都优于其他 3 种算法。

图 6 是平均定位误差的变化情况，随着 AP 数量增加，4 种算法的平均定位误差逐渐变小，在 AP 数量为 10 的时候，KPCA-IWKNN 算法的平均定位误差为 1.10 m，而 PCA-WKNN 算法在 AP 数量为 12 时平均误差与本文算法相当，其他算法需要更高的 AP 数量，当 AP 数量为 13 时，KPCA-IWKNN 算法的平均定位误差为 0.83 m。

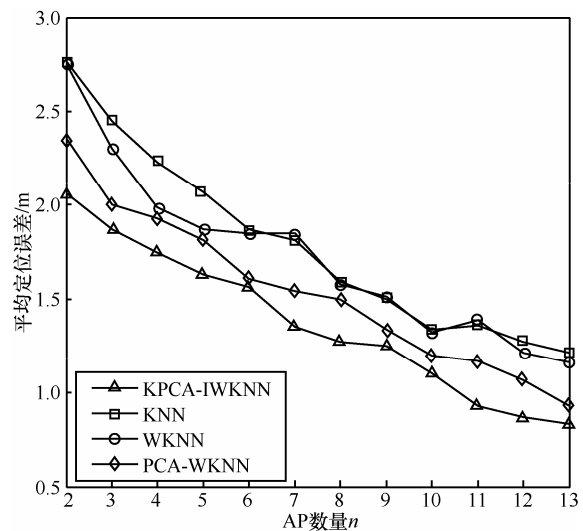


图 6 平均定位误差随 AP 数量增加变化情况

图 7 是误差距离为 1.5 m 时 4 种算法的定位准确率的变化情况，可以看出，随着 AP 数量增加，4 种算法的定位准确率呈现增高的趋势，且本文算法优于其他几种算法。在 AP 的个数为 8 时，本文算法的定位准确率到达 63%，而其他的算法若想达到这个精度，则需要更多的 AP。从图 6 和图 7 还可以看出，本文算法在同样的定位精度下需要的 AP 更加少。随着 AP 数量的增加，定位准确率并不是一味地升高，产生这样结果的原因是随着 AP 数量的增加，引入了更多的噪声，这些噪声会干扰算法的定位效果。

4.3 离线阶段 RSS 样本数量对算法的影响

在指纹架构的定位算法中，影响算法性能的另一个因素是每个参考节点上 RSS 的采集次数。它用来表征指纹架构定位算法在离线阶段的时间消耗，在相同定位效果下，某算法在离线阶段采集 RSS 的次数越少，代表该算法在离线阶段需要的时间越少。

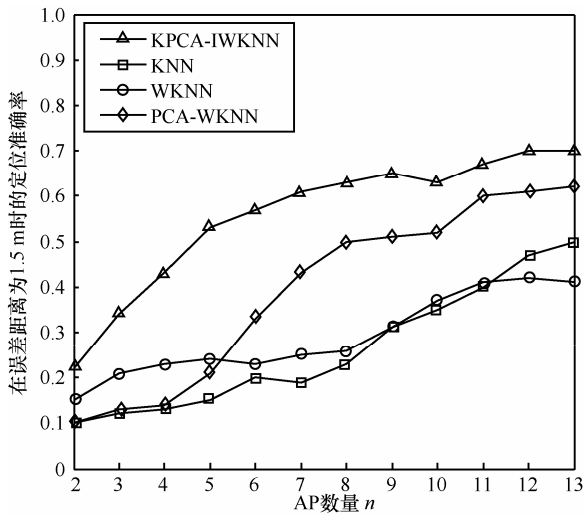


图 7 定位准确率随 AP 数量变化情况

在这个实验中，AP 数量为 10， σ 的值为 0.3，特征位置指纹空间的维度为 10，PCA 方法选取 10 个主元用以定位。

图 8 是在误差距离为 1.5 m 时，定位算法在不同 RSS 样本数量下的定位准确率的比较。从图 8 中可以看出，随着 RSS 采集样本数的增加，几种算法的定位准确率都在增加，在 RSS 采集样本数为 20 时，本文算法的准确率已经远远优于其他几种算法，PCA-WKNN 算法若想达到相同的定位效果，需要更多的样本数量，由此可以判定，本文算法与其他几种算法相比，需要更少的离线 RSS 采集次数。

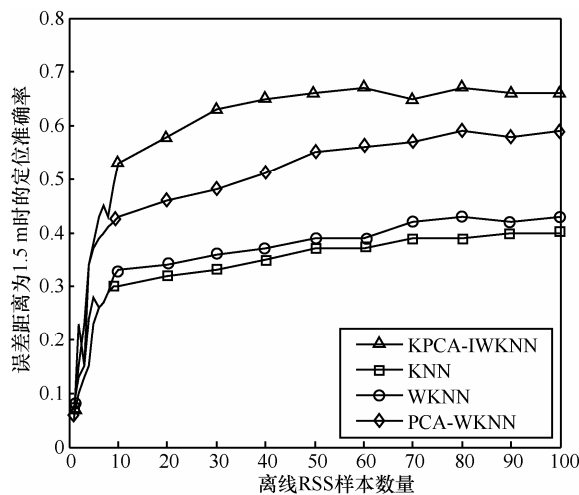


图 8 定位准确率随 RSS 样本数量的变化情况

4.4 特征位置指纹空间的维度对定位算法的影响

本文提出的 KPCA-IWKNN 算法使用 KPCA 训练原始位置指纹空间 F ，由于 KPCA 的特性，既能将 F 映射到更高维度的特征指纹空间，也能映射到

低维度的特征指纹空间。

图 9 和图 10 是在 AP 数量为 12， σ 取值为 0.3 时，本文提出的 KPCA-IWKNN 算法性能随特征位置指纹空间维度 k 变化的情况。

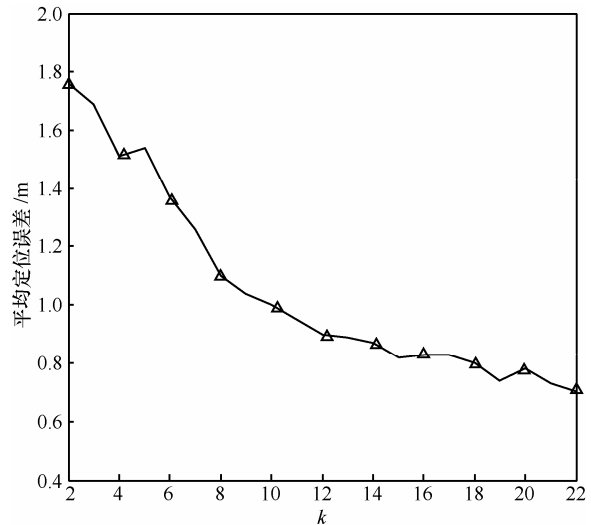


图 9 平均定位误差随 k 值变化情况

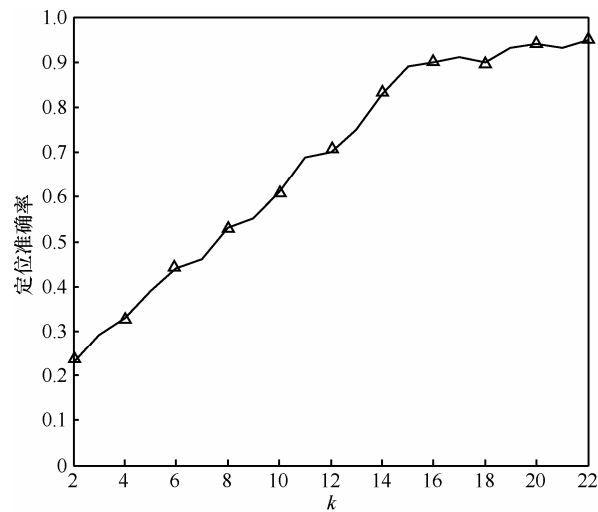


图 10 定位准确率随 k 值变化情况

图 9 是平均定位误差随特征位置指纹空间维度 k 的变化情况，平均定位误差随着 k 值增大而减小，当 k 值达到 22 时，平均定位误差可以达到 0.71 m。图 10 是误差距离为 1.5 m 时误差准确率随 k 值的变化情况，可以看出 k 值越大，定位准确率越大，在 k 值达到 16 时，定位准确率可以达到 90%，当 $k > 16$ 时，定位准确率趋于不变，当 k 为 22 时，定位准确率可达到 95%。

4.5 参数 σ 对定位性能的影响

参数 σ 是一个在 (0,1) 范围内的常数，根据式(20)

可以自适应确定 KPCA-IWKNN 算法定位所使用的近邻数 m 。在参数 σ 值确定后, 根据式(20)可知, 不同位置上的位置估计选用的近邻数不同, 避免了 KNN 和 WKNN 算法中使用固定近邻数带来的问题。在这个实验中 AP 的数量选择 12, 特征位置指纹空间的维度选择 14, σ 在(0,1)范围内变化。

表 5 列出了 $\sigma=0.3$ 时部分测试节点在定位时使用的近邻数, 可以看到不同测试节点在定位时近邻数 m 会发生改变。图 10 所示为算法的平均定位误差随 σ 值变化的情况, 在 $\sigma=0$, 近邻数为 1, 定位的结果就是某一参考节点的位置, 此时的误差最大, 想要达到准确定位非常困难; 随着 σ 的增大, 定位所需的近邻数变多, 平均定位误差变小, 在 $\sigma=0.3$ 时, 平均误差达到最小; $\sigma>0.8$ 时, 定位时选用的近邻数变的更多, 引入了更多无用的位置指纹, 使平均定位误差增大。

表 5 $\sigma=0.3$ 时部分位置的近邻数

测试节点	近邻数
(0,1)	6
(6,1)	5
(12,1)	6
(18,1)	4

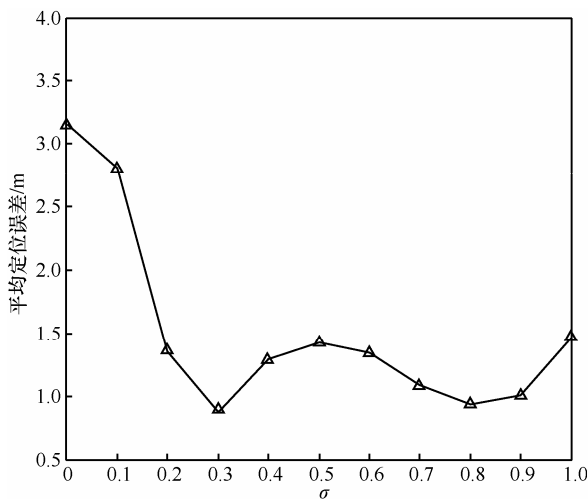


图 11 平均定位误差随参数值 σ 变化情况

4.6 几种算法在其他环境下的性能比较

通过以上仿真可以看出, 本文算法在上述场景下, 定位误差和定位准确率优于其他算法, 为了验证算法在其他环境下的性能, 使用了 King 等^[17]研究的无线定位系统的指纹数据进行分析, 整个系统覆盖面积为 221 m², 整个区域内部署了 25 个无线

接入点, 即 $n=25$, 在 King 的实验数据中, 在每一个参考点进行 100 次 RSS 样本采集。

图 12 和图 13 分别是 4 种算法在 AP 数量和 RSS 样本数量变化下的实验结果。在图 12 中, KPCA 将原始位置指纹空间训练到 14 维的特征位置指纹空间, σ 取值为 0.3, PCA 方法选取 n 个主元用以定位, 使用的是 RSS 经过 40 次采集得到的数据。在图 13 中, AP 的数量为 10, σ 的值为 0.3, 特征位置指纹空间的维度为 10, PCA 方法选取 10 个主元用以定位, 误差距离为 1.5 m。从图 12 和图 13 可以看出, 本文算法在 King 的实验场景下优于其他几种方法。

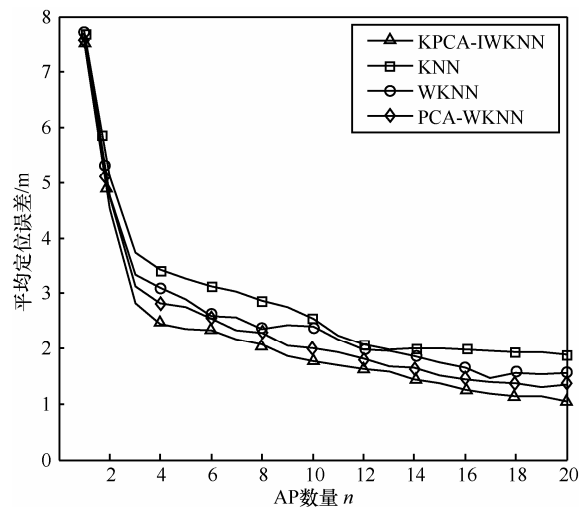


图 12 在 King 场景下的 AP 数量实验

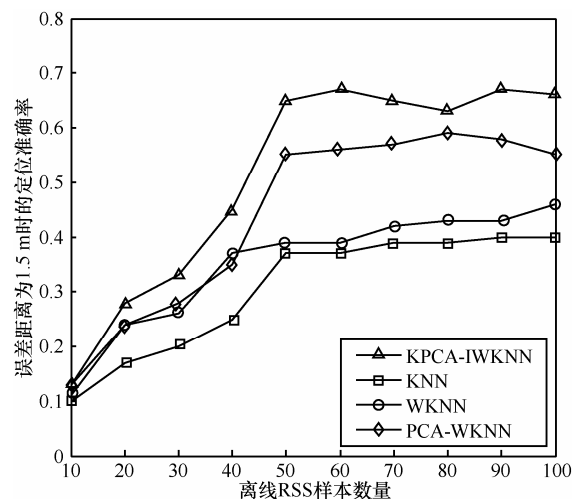


图 13 在 King 场景下的 RSS 样本数量实验

5 结束语

本文针对现实 WLAN 环境中由于 RSS 信号的时变性而引入的冗余和噪声问题, 提出了一种基于核函

数特征提取的室内定位算法。该算法在离线阶段是采用 KPCA 训练原始位置指纹,提取原始位置指纹空间的非线性特征,作为位置指纹数据,在线阶段采用 IWKNN 算法,自主选择位置指纹数进行位置估计。实验表明本文提出的算法优于其他几种算法,提高了定位准确率,降低了平均定位误差。但是在实验中没有考虑到不同 AP 设备、不同接收终端对 RSS 信号的影响,而且在实验时待测目标是静止的,对于 WLAN 环境下移动目标的定位情形有待下一步解决。

参考文献:

- [1] 钱志鸿, 王义君. 面向物联网的无线传感器网络综述[J]. 电子与信息学报, 2013, 35(1): 215-227.
QIAN Z H, WANG Y J. Internet of things-oriented wireless sensor networks review[J]. Journal of Electronics & Information Technology, 2013, 35(1): 215-227.
- [2] 钱志鸿, 王雪. 面向 5G 通信网的 D2D 技术综述[J]. 通信学报, 2016, 37(7): 1-14.
QIAN Z H, WANG X. Reviews of D2D technology for 5G communication networks[J]. Journal on Communications, 2016, 37(7): 1-14.
- [3] KOLODZIEJ K W, HJELM J. Local positioning systems: LBS applications and services[M]//Local Positioning System. LBS Applications and Services. 2006: 101-158.
- [4] MA L, XU Y. Received signal strength recovery in green WLAN indoor positioning system using singular value thresholding[J]. Sensors, 2015, 15(1): 1292-1311.
- [5] JAIN V K, TAPASWI S, SHUKLA A. Performance analysis of received signal strength fingerprinting based distributed location estimation system for indoor WLAN[J]. Wireless Personal Communications, 2013, 70(1): 113-127.
- [6] SCHMIDT E, AKOPIAN D. Indoor positioning system using WLAN channel estimates as fingerprints for mobile devices[C]//IS&T/SPIE Electronic Imaging, International Society for Optics and Photonics, 2015: 94110R-94110R-9.
- [7] DENG Y B, ZHI A X, LIN M. WLAN indoor positioning algorithm based on KDDA and SVR [J]. Journal of Electronics & Information Technology, 2011, 4:23.
- [8] HE S, CHAN S H G. Wi-Fi fingerprint-based indoor positioning: recent advances and comparisons[J]. IEEE Communications Surveys & Tutorials, 2016, 18(1): 466-490.
- [9] MENGUAL L, MARBÁN O, EIBE S. Clustering-based location in wireless networks[J]. Expert Systems with Applications, 2010, 37(9): 6165-6175.
- [10] WEN Y, TIAN X, WANG X, et al. Fundamental limits of RSS fingerprinting based indoor localization[C]//2015 IEEE Conference on Computer Communications (INFOCOM). IEEE, 2015: 2479-2487.
- [11] BAHL P, PADMANABHAN V N. RADAR: An in-building RF-based user location and tracking system[C]//INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. 2000, 2: 775-784.
- [12] 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.
- [13] FANG S H, LIN T N. Principal component localization in indoor WLAN environments[J]. IEEE Transactions on Mobile Computing, 2012, 11(1): 100-110.
- [14] MARTÍNEZ A M, KAK A C. PCA versus LDA[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(2): 228-233.
- [15] DE SA J P M. Pattern recognition: concepts, methods and applications[M]. Springer Science & Business Media, 2012.
- [16] DING M, FAN G. Articulated and generalized Gaussian kernel correlation for human pose estimation[J]. IEEE Transactions on Image Processing, 2016, 25(2): 776-789.
- [17] KING T, HAENSELMANN T, EFFELSBURG W. On-demand fingerprint selection for 802.11-based positioning systems[C]//World of Wireless, Mobile and Multimedia Networks, 2008 (WoWMoM 2008). 2008: 1-8.

作者简介:



李华亮 (1990-), 男, 吉林长春人, 吉林大学硕士生, 主要研究方向为无线定位技术。



钱志鸿 (1957-), 男, 吉林长春人, 吉林大学教授、博士生导师, 主要研究方向为基于物联网、D2D、Wi-Fi、RFID 等无线网络与通信技术。



田洪亮 (1981-), 男, 吉林省吉林市人, 吉林大学博士生, 东北电力大学讲师, 主要研究方向为无线个域网。