

## 基于主成分分析的室内指纹定位模型\*

陈祠<sup>1,3</sup>, 牟楠<sup>1</sup>, 张晨<sup>2</sup>, 陈永乐<sup>3</sup>, 朱红松<sup>3</sup>, 刘燕<sup>1</sup>

<sup>1</sup>(北京大学 软件与微电子学院, 北京 102600)

<sup>2</sup>(中央民族大学 信息工程学院, 北京 100081)

<sup>3</sup>(中国科学院 信息工程研究所 信息安全国家重点实验室, 北京 100093)

通讯作者: 陈祠, E-mail: chenci@pku.edu.cn

**摘要:** 指纹定位是目前最有前途的室内定位方法之一, 基于无线信号强度的指纹模型因其无需额外硬件成本、易于推广等特点被广泛采用. 指纹模型的选择是影响指纹定位精度的关键因素. 传统的通过选择指纹采集点的指纹方法尽管可以减少计算量, 但对定位精度贡献不大. 提出一种基于主成分分析的指纹模型, 通过选择对精度影响最大的一组“成分”作为指导定位的指纹, 在减少指纹计算量的同时, 提高定位精度. 实验结果表明, 与基于欧式距离指纹算法和最近邻指纹算法相比, 基于主成分分析的指纹算法可以将平均定位精度由 5.3m 和 3.9m 降低到 2.7m.

**关键词:** 室内定位技术; 指纹定位算法; 主成分分析法; 误差分析; 指纹模型

中文引用格式: 陈祠, 牟楠, 张晨, 陈永乐, 朱红松, 刘燕. 基于主成分分析的室内指纹定位模型. 软件学报, 2013, 24(Suppl. (1)): 98-107. <http://www.jos.org.cn/1000-9825/13011.htm>

英文引用格式: Chen C, Mu N, Zhang C, Chen YL, Zhu HS, Liu Y. Indoor fingerprint positioning model based on principal component analysis. Ruan Jian Xue Bao/Journal of Software, 2013, 24(Suppl. (1)): 98-107 (in Chinese). <http://www.jos.org.cn/1000-9825/13011.htm>

## Indoor Fingerprint Positioning Model Based on Principal Component Analysis

CHEN Ci<sup>1,3</sup>, MU Nan<sup>1</sup>, ZHANG Chen<sup>2</sup>, CHEN Yong-Le<sup>3</sup>, ZHU Hong-Song<sup>3</sup>, LIU Yan<sup>1</sup>

<sup>1</sup>(School of Software and Microelectronics, Peking University, Beijing 102600, China)

<sup>2</sup>(School of Information Engineering, Minzu University of China, Beijing 100081, China)

<sup>3</sup>(State Key Laboratory of Information Security, Institute of Information Engineering, The Chinese Academy of Sciences, Beijing 100093, China)

Corresponding author: CHEN Ci, E-mail: chenci@pku.edu.cn

**Abstract:** Fingerprint localization is one of the most promising indoor positioning methods, and the fingerprint model based on the wireless signal strength is widely used due to its no-additional hardware cost and easy-to-spread characteristics. The selection of the fingerprint model is the key factor to the fingerprint positioning accuracy. Although the traditional fingerprint method by selecting the fingerprint collection points can reduce the computation, it contributes little to the accuracy of the positioning. In this paper, a fingerprint model based on principal component analysis is proposed. The new model accomplishes improvement in positioning accuracy as well as reduction in fingerprint calculation by selecting a set of “ingredients” with the largest impact on the accuracy to guide the positioning of fingerprint. Experimental results show that compared with the fingerprint algorithms based on Euclidean distance and nearest neighbor, the fingerprint algorithm based on principal component analysis improves average positioning accuracy to 2.7 m from 5.3 m and 3.9 m.

**Key words:** indoor positioning technology; fingerprint positioning algorithm; principal component analysis; error analysis; fingerprint model

\* 基金项目: 国家自然科学基金(61073180); 国家高技术研究发展计划(863)(2012AA013104); 国家科技重大专项(2011ZX 03005-002-02); 中国科学院信息工程研究所前瞻部署项目(Y3Z0071E02)

收稿时间: 2013-05-02; 定稿时间: 2013-08-22

室内定位技术是基于 Wi-Fi、Zigbee、Bluetooth、红外和超声波等无线信号技术,通过测距和相关算法实现室内目标的精确定位,在商业应用、公共安全和军事场景等许多领域有着广泛的需求与应用.室内定位的快速发展将填补室外定位的应用盲区,近年来,诺基亚、三星、苹果等智能手机开始推出针对机场及大型商业区的室内定位导航服务,更是将室内定位技术推广到日常生活中.

指纹(fingerprint)定位算法利用定位环境中多个无线路由器的信号强度值,通过采集训练建立离线指纹库,然后在定位过程中根据实时采集指纹与指纹库进行匹配,从而获得最佳匹配位置.随着国内外公共环境中 Wi-Fi 热点的普及,指纹定位系统可以利用热点网络中已有的无线路由器提供基本指纹信息,从而使指纹定位算法更容易在室内环境中广泛使用.然而室内环境的复杂多变和人身遮挡等实际干扰因素导致无线信号的传播产生多径和阴影现象,这对室内定位的精度提出了严峻的挑战.目前,指纹定位算法精度提高较多的集中于在线匹配算法的改进,由确定性的 KNN 算法<sup>[1,2]</sup>发展到概略性的贝叶斯估计算法<sup>[3-5]</sup>和顺序蒙特卡洛算法<sup>[6-8]</sup>均被用于指纹定位算法中的指纹匹配过程.但是这些指纹定位算法很少考虑室内多径和阴影造成的影响.

与基于距离的定位方法比较,指纹定位算法的优势是对复杂环境能够进行较为精确的定位.根据对复杂环境的假定,指纹算法分为针对静态环境和动态环境两大类.RADAR 算法是最早利用指纹思想在 Wi-Fi 网络进行定位的系统<sup>[9]</sup>.该算法在理想环境下定位误差达到了 2.94m.尽管该算法揭示了人身遮挡影响的存在,但是仍然没有提出有效的解决办法.对于动态环境下的指纹定位主要研究变化的气候条件、移动的人或物对定位精度的影响.在室内定位系统中,携带标签在运动物体身上的佩戴位置和朝向是影响指纹定位的关键因素之一.文献[9]研究了不同朝向信号的衰减模型以改善指纹定位算法,提出了一种能够感知朝向的指纹定位算法.文献[10]也提出了基于朝向的指纹定位算法,对采集的各个朝向的指纹进行去噪预处理后添加到指纹数据库,但是在线采集的数据并不预知朝向信息,致使各个朝向的指纹并没有很好的区分粒度,很难给出准确定位.COMPASS<sup>[11]</sup>算法利用带有数字指南针的节点,在采集信号指纹的同时也采集了人身体的朝向,定位的精度较上述算法有一定的提高.

上述针对动态环境的指纹定位算法的思路是通过对直接学习位置与各种变化状态间的关系,来实现对各种变化条件的细致分类,从而一定程度上解决了动态环境下可能出现的瞬时定位异常问题.我们在指纹定位研究中发现,影响指纹定位精度和发生异常变化往往与数据源有很大的关系.不同无线路由器的信号强度值对移动节点的位置估计贡献不同,甚至有些信号强度值对位置判断起到相反的作用,如果我们能够通过分析找出真正对位置判断有效的数据分量,就可以明显提高定位系统的精度.为此,本文提出了一种基于主成分分析的指纹定位模型,通过主成分分析技术,找到不同环境条件下影响定位精度的主要分量,提高指纹的可区分度,进而提高指纹定位的精度.同时,主成分分析也能够对指纹向量进行降维处理,提升指纹定位算法的处理效率.主成分分析方法是从事数据源优化的角度来提高定位精度,因此可以应用于大部分的指纹算法以提高精度.实验结果表明,与基于欧式距离指纹算法和最近邻指纹算法相比,基于主成分分析的指纹算法可以将平均定位精度由 5.3m 和 3.9m 降低到 2.7m.

本文第 1 节介绍指纹定位算法的架构.第 2 节介绍主成分分析的指纹生成方法的主要思路.第 3 节通过实例的方式详细讨论主成分分析方法的实施步骤和计算方法.第 4 节针对实例分析主成分方法的有效性原理.第 5 节通过具体的定位实验验证主成分分析方法的有效性.最后一节总结全文.

## 1 指纹定位算法技术架构

指纹定位算法分为离线训练和在线定位两个阶段,如图 1 所示.离线训练阶段是指在定位区域中的每一个位置由人携带的可移动节点接收周围固定位置无线路由器广播的数据包,并将数据包通过无线路由器转发到定位服务器,定位服务器根据移动节点测得的信号强度值,生成对应的信号指纹,以此建立与所有位置对应的指纹数据库.在线定位阶段是指待定位节点在定位区域中移动时,收集到无线路由器的信号强度值,并转发给定位服务器,后台在已有指纹数据库的基础上,通过概率统计和数据挖掘等方法实现实时指纹与数据库指纹的匹配,从而确定待定位节点的位置<sup>[12,13]</sup>.

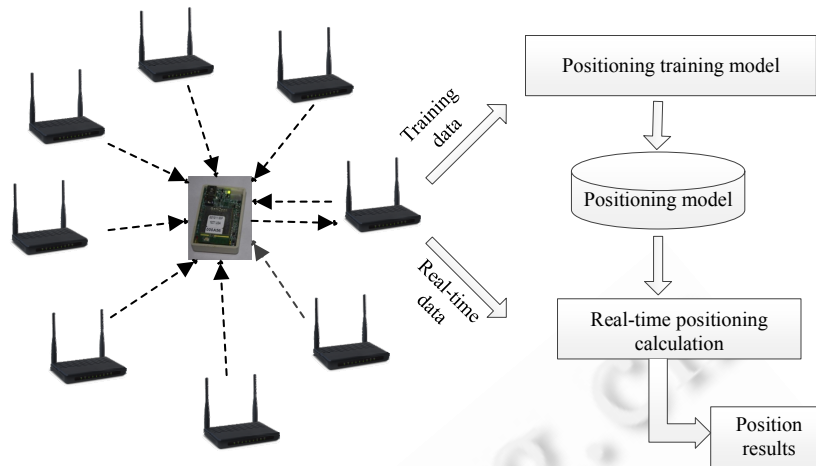


Fig.1 Positioning model of fingerprint library matching algorithm

图1 指纹库匹配算法定位模型

## 2 基于主成分分析的指纹生成方法

主成分分析(principal component analysis,简称 PCA),又称为主成分回归分析法,由英国数学家 Pearson 在 1902 年提出,用于分析数据及建立数理模型.该方法的核心思想是通过随机向量的协方差矩阵或相关性矩阵的特征向量和特征值来计算影响变量的主要成分和它们的权重.

在多变量(室内定位位置信息)分析中,为了更好地确定被定位节点的位置,需要收集室内多个路由器的信号强度值,每一组信号强度值作为辨识定位节点位置信息的依据,即为每个定位节点形成指纹标识.而这些信号强度值具有很大程度的相关性,并且具有信息重叠的现象,这样的数据结果会影响到定位的精确度.因此,需要利用几个不相关的综合指标来提供节点所需的大部分信息,也就是说,将信息的主成分提取出来.这也就是主成分分析的核心思想,即降维的思想.例如,对于  $n$  条定位数据来说,分别由  $k$  种信号强度值作为一条指纹来描述,那么将  $k$  种信号强度值用  $k$  个指标  $x_1, x_2, \dots, x_k$  表示,这样统计出来的原始数据矩阵为

$$X = \begin{pmatrix} x_{11} & x_{12} \cdots & x_{1k} \\ x_{21} & x_{22} \cdots & x_{2k} \\ \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} \cdots & x_{nk} \end{pmatrix} = (X_1 \quad X_2 \cdots X_k) \quad (1)$$

在一些特定的研究问题中,这些指标会具有不同的量纲,并且有的指标在数量级上会存在很大的差异,所以为了消除这些差异所带来的不合理的影响,进行主成分分析之前首先要对这些指标进行标准化处理,消除量纲.对于信号强度值来说,由于量纲相同,故标准化这一步骤可以省略.如公式(1)所示,得到的数据矩阵为  $X = (X_1 \quad X_2 \cdots X_k)$ ,形成的综合指标用  $Z$  表示,进行主成分分析后得到的综合指标满足公式(2).

$$\begin{cases} Z_1 = a_{11}X_1 + a_{12}X_2 + \dots + a_{1k}X_k \\ Z_2 = a_{21}X_1 + a_{22}X_2 + \dots + a_{2k}X_k \\ \dots \\ Z_k = a_{k1}X_1 + a_{k2}X_2 + \dots + a_{kk}X_k \end{cases} \quad (2)$$

由  $k$  个指标  $X_1, X_2, \dots, X_k$  做线性组合会得到不同的综合指标  $Z$ ,这些综合指标会包含各自特有的信息特性.为了达到很好的效果,  $Z_i$  的方差尽量大并且  $Z_i$  之间需要相互独立,对于综合指标的系数还需要满足以下条件:

- (1)  $a_{i1}^2 + a_{i2}^2 + \dots + a_{ik}^2 = 1$ , 其中  $i = 1, 2, \dots, k$ .
- (2)  $Z_i$  与  $Z_j$  ( $i \neq j, i, j = 1, 2, \dots, k$ ) 之间不具有相关性.

(3)  $Z_1$  是  $X_1, X_2, \dots, X_k$  所有线性组合中方差最大的,  $Z_2$  是与  $Z_1$  互不相关的  $X_1, X_2, \dots, X_k$  所有线性组合中方差最大的, 以此类推,  $Z_k$  是与  $Z_1, Z_2, \dots, Z_{k-1}$  都不相关的  $X_1^*, X_2^*, \dots, X_k^*$  所有线性组合中方差最大的.

通过这样的处理后得到的  $Z_1, Z_2, \dots, Z_k$  分别是原指标  $X_1, X_2, \dots, X_k$  的第 1 个, 第 2 个,  $\dots$ , 第  $k$  个主成分.  $Z_1$  是在总方差中占的比例最大者,  $Z_2, Z_3, \dots, Z_k$  依次减小<sup>[14]</sup>.

在研究室内定位的过程中, 会选取在总方差中所占比例较大的前几个指标作为研究定位的主成分, 这样使得定位指标的数目减少, 达到了降维的目的, 并且这些主成分相比原始数据更具有代表性, 是影响问题的主要因素. 抓住这些主要因素既能提高系统的处理效率, 又可以得到更理想的效果<sup>[15]</sup>.

根据上述分析, 主成分定位指纹的提取可分为原始 RSSI 信号强度收集、原始信号强度统计特征分析、主成分选择、指纹生成这 4 个步骤.

### 3 定位指纹主成分生成实例与分析

为了更好地理解主成分提取的具体过程, 我们以中国科学院信息工程研究所 3 号楼 1 层大厅部署的实验系统为例对主成分计算流程和结果进行分析.

#### 3.1 实验环境

中国科学院信息工程研究所 3 号楼 1 层大厅, 其面积约为  $26.4 \times 38.2 \text{m}^2$ , 内部结构如图 2 所示. 该区域被划分为多个隔断实验室工位, 隔断之间存在全封闭或者半封闭情况, 属于典型的室内办公环境, 具有室内定位环境的代表性. 图 2 中六角星标无线路由器的位置信息, 8 个无线路由器安装在 3m 高的天花板上, 为实验提供无线信号覆盖. 无线信号采集节点采用的是美国 Gainspan 公司的 GS1011WiFi 模块, 可以接收室内环境中无线信号强度值. 实验中为了避免其他无线信号的干扰, 在接收端通过 MAC 地址匹配对 RSSI 值进行过滤筛选. 服务器接收端采用的是联想系列笔记本电脑 Idea Pad Y470, 内置英特尔 Wireless Wi-Fi Link 100 Series 无线网卡, 接收 GS1011WiFi 模块通过无线路由器传回的 RSSI 值.

#### 3.2 信号强度数据收集

在指纹定位的离线训练阶段, 需要在选定的标记位置采集大量 WiFi 信号指纹, 并将指纹数据存入后台数据库, 原始数据经过过滤、平均处理后形成定位指纹. 如图 2 所示, 在该环境中, 我们共部署了 8 个无线路由器, 并对 76 个采样点进行了指纹采集.

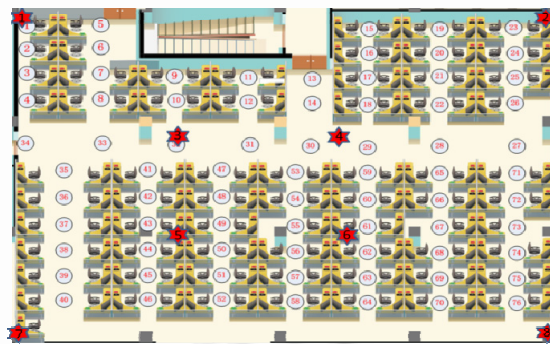


Fig.2 Interior structure in laboratory and fingerprint location maps  
图2 实验室室内结构与指纹采集位置分布图

信号的稳定性是影响定位精度的关键因素, 对于指纹定位模型来说, 信号的强度值越稳定, 由信号强度值生成的指纹就越可靠, 不同测试位置指纹的差异性相对也更大. 在进行离线训练之前, 在大厅随机选取一个测试位置, WiFi 接收节点采集 50 个数据包, 经过后台解析, 得到 8 个无线路由器的信号强度值. 如图 3 所示, 实验环境中无线信号强度值比较稳定, 对 50 个数据进行过滤和平均处理后, 信号强度值基本不变, 为指纹定位模型提供了很好的基础.

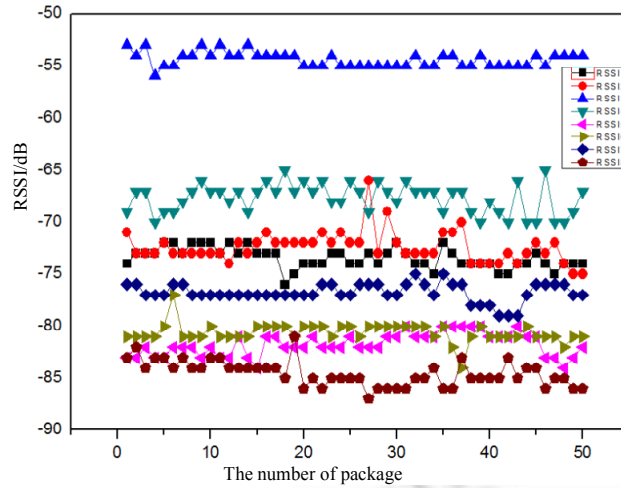


Fig.3 Stability test pattern of signal strength value  
图3 信号强度值稳定性测试图

离线训练时将待定位的接收节点挂于胸前,在每个采集点位置,按照 0°,90°,180°,270°和头顶上方 5 个方向进行采集(图2中向上的方向为0°采集方向,然后逆时针依次旋转90°,180°,270°和头顶方向),并且每个方向采集50个数据包,取信号强度平均值形成一条原始指纹.假设接收到编号为  $n$  的无线路由器信号强度值为  $RSSI_n$ ,则原始指纹可以表示为

$$finger = \{RSSI_1, RSSI_2, RSSI_3, RSSI_4, RSSI_5, RSSI_6, RSSI_7, RSSI_8\} \quad (3)$$

在离线阶段共采集 19 000 条原始指纹数据,由这些数据最终生成 380 条指纹,部分采集点的指纹数据如图4所示.

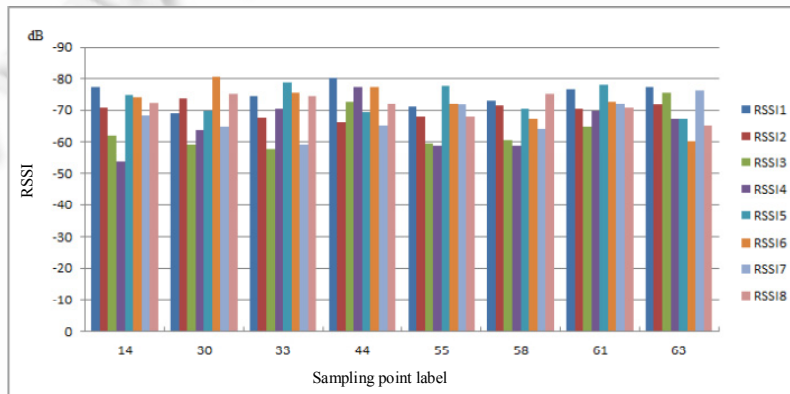


Fig.4 Original fingerprint of sampling points  
图4 采样点原始指纹

### 3.3 原始信号强度的统计特征

将收集好的 380 条数据作为指纹样本,将 8 个路由器的 RSSI 值作为指标,用  $RSSI_1, RSSI_2, \dots, RSSI_8$  表示,这样就构造出了原始数据矩阵:

$$RSSI = \begin{pmatrix} rssi_{11} & rssi_{12} \dots & rssi_{18} \\ rssi_{21} & rssi_{22} \dots & rssi_{28} \\ \vdots & \vdots & \vdots \\ rssi_{3801} & rssi_{3802} \dots & rssi_{3808} \end{pmatrix} \quad (4)$$

统计在 76 个位置点上 8 个路由器的信号强度值 RSSI 向量的统计特征,见表 1.

**Table 1** Statistical characteristics of the signal strength

**表1** 信号强度统计特征

	均值	标准差	分析 <i>N</i>
<i>RSSI</i> <sub>1</sub>	-76.379 132	9.990 887 4	380
<i>RSSI</i> <sub>2</sub>	-70.897 132	8.699 473 3	
<i>RSSI</i> <sub>3</sub>	-69.674 842	7.905 955 2	
<i>RSSI</i> <sub>4</sub>	-69.524 526	8.806 152 8	
<i>RSSI</i> <sub>5</sub>	-78.169 211	7.000 732 7	
<i>RSSI</i> <sub>6</sub>	-78.656 395	7.611 058 0	
<i>RSSI</i> <sub>7</sub>	-69.182 316	8.042 135 6	
<i>RSSI</i> <sub>8</sub>	-71.609 395	9.451 184 2	

其中,第 1 列得到的是相应信号强度值的算术平均值,第 2 列得到的是相应信号强度值的标准差,第 3 列得到的是数据个数.其中,在线定位阶段算术平均值和标准差会用于对实时采集的信号强度值进行归一化处理.

**3.4 选取影响位置区分的主成分**

根据公式(2),  $RSSI = (RSSI_1, RSSI_2, \dots, RSSI_8)$  可以产生 8 个综合成分  $(Z_1, Z_2, \dots, Z_8)$ .

令  $\sigma$  为成分  $(Z_1, Z_2, \dots, Z_8)$  的方差,公式(5)表示前  $m$  个成分的方差对总体方差的贡献率  $T$ .

$$T = \frac{\sum_{i=1}^m \sigma_i}{\sum_{i=1}^k \sigma_i} \tag{5}$$

其中, $k$  表示综合成分的个数,所以  $k=8$ .

根据实验系统测试数据,可以得到 8 个综合成分的方差以及它们相对总体方差贡献的百分比,见表 2.

**Table 2** Variance components and their contribution to the overall variance

**表2** 成分方差及其对总体方差的贡献

成分	成分方差	方差贡献百分比	累积贡献百分比
1	3.435	42.931	42.931
2	1.515	18.932	61.863
3	1.208	15.096	76.959
4	0.574	7.175	84.134
5	0.404	5.045	89.179
6	0.371	4.637	93.816
7	0.291	3.642	97.458
8	0.203	2.542	100.000

我们定义累计方差贡献百分比的门限作为选择主成分的衡量标准,如果累计方差超过门限,则选取相应个数的成分作为主成分.一般来说,通常所取的主成分使得累积的贡献率达到 80%以上为宜<sup>[16]</sup>,所以前 4 个成分被选择为定位模型的主成分.

**3.5 主成分定位的指纹模型**

基于指纹的定位方法要为每个可区分的位置建立指纹,采样点的指纹是区分不同位置的唯一指标.基于主成分分析的定位算法的指纹由主成分对每个位置点的贡献之和来计算.

令  $b_i$  为第  $i$  个主成分的方差  $\sigma$  占全部主成分的方差的比例,即

$$b_i = \frac{\sigma_i}{\sum_{i=1}^k \sigma_i} \tag{6}$$

其中,  $i=1, 2, \dots, m$  ( $m$  为选取的主成分个数),  $k=8$ . 那么指纹  $F$  是各主成分  $Z_1, Z_2, \dots, Z_m$  的线性函数,取前 4 个主成分,则有:

$$F = (b_1Z_1 + b_2Z_2 + b_3Z_3 + b_4Z_4) / T \tag{7}$$

利用公式(6)、公式(7)得到 380 个采样点位置的综合得分.可以看到得分值比较小,为了更好地分析结果,将

综合得分放大<sup>[16]</sup>.

考虑到综合得分同时存在正数和负数,为了保证综合得分的正负性不变,规定最高得分为 50 分,最低得分为-50 分,其他得分按照比例进行放大.得到转换后的综合得分见表 3.

**Table 3** Fingerprint of principal component positioning algorithm

**表3** 主成分定位算法的指纹

采样点编号	1	2	3	...	380
综合得分	-42.092	-44.0191	-50	...	34.3666

## 4 主成分特征分析

### 4.1 主成分与RSSI之间的关系分析

前面我们通过最大方差条件选取了 4 个主成分,并没有给出达到最大方差条件时,  $Z_i$  与不同无线路由器信号强度之间的关系矩阵 ( $a_{ij}$ ). 表 4 给出了满足最大方差条件下的载荷矩阵 ( $a_{ij}$ ).

第 1 主成分对 1 号和 7 号路由器的信号强度值  $RSSI_1$  和  $RSSI_7$  有较大的负相关系数,对 2 号和 8 号路由器的信号强度值  $RSSI_2$  和  $RSSI_8$  有较大的正相关系数.从中可以分析出,第 1 主成分主要是对处于边界的路由器信号强度值进行评价,并且以离图 2 右侧边界距离近为正相关,以离左侧边界距离近为负相关.

第 2 主成分对 3 号和 4 号路由器的信号强度值  $RSSI_3$  和  $RSSI_4$  有较大的正相关系数.从中可以得出,第 2 主成分主要是对编号为 3 和 4 的路由器信号强度值进行评价,并且以离路由器距离近为正相关.

第 3 主成分对 5 号和 6 号路由器的信号强度值  $RSSI_5$  和  $RSSI_6$  有较大的正相关系数,主要评价采样点离 5 号 6 号路由器距离的远近程度.

由于第 4 主成分对所有路由器信号强度值的相关系数都很小,所以可以分析出其对所有路由器的信号均不敏感.

**Table 4** Loading matrix

**表4** 载荷矩阵

$a_{ij}$	主成分			
	1	2	3	4
$RSSI_1$	-0.853	-0.038	0.070	-0.342
$RSSI_2$	0.756	-0.192	0.325	0.320
$RSSI_3$	-0.562	0.376	0.609	0.023
$RSSI_4$	0.535	0.196	0.716	-0.061
$RSSI_5$	0.106	0.899	-0.110	0.153
$RSSI_6$	0.559	0.633	-0.227	-0.346
$RSSI_7$	-0.716	0.294	-0.234	0.455
$RSSI_8$	0.834	-0.023	-0.307	0.018

### 4.2 主成分指纹特征分析

为了更好地体现主成分在不同采样点位置的差别,本文选取每个采样点头顶方向的指纹数据作为样本,所得到的 76 个样本的前 4 个主成分得分分别如图 5 所示.

从图 5(a)中可以看出,1~8 号采样点、19~26 号采样点、36~46 号采样点和 65~76 号采样点中大部分数据的绝对值较大,结合指纹采集位置分布图 2 可知,这些采样点正是离 4 个“边界”无线路由器比较近的点.从图 5(b)和图 5(c)同理可以分析得出,图中样本得分绝对值较大的位置均是分布在对主成分影响较大的无线路由器的周围.

主成分综合得分就是主成分指纹,同理,将主成分指纹用柱状图表示,如图 6 所示.从图中可以明显看出,不同区域采样点的指纹有明显的正负区别,并且同一区域不同位置采样点的主成分指纹也存在较大的区分度.



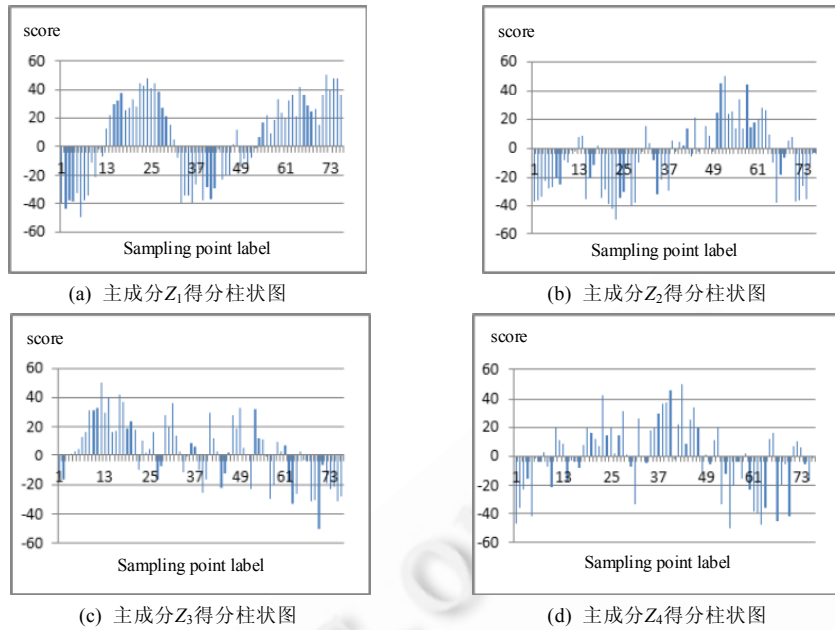


Fig.5 Different locations of the principal component score

图 5 不同位置点主成分的得分图

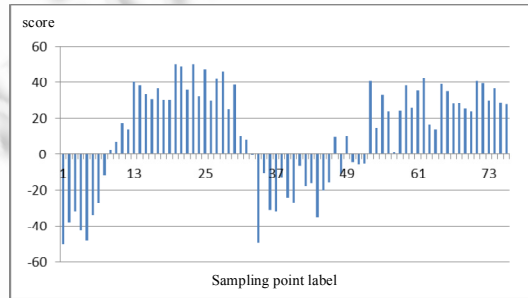


Fig.6 Histogram of principal component fingerprint

图 6 主成分指纹柱状图

### 5 定位实验结果分析

在指纹匹配算法的离线训练阶段,本文选取了 76 个采样点进行原始无线信号强度值采集,选用两种不同的方法生成定位模型的指纹库.第 1 种方法如公式(3)所示,将 8 个路由器的信号强度值组合作为一个向量  $finger$ ,形成一条指纹;第 2 种方法是采用主成分分析法,如公式(7)所示,将每个样本的综合得分  $F$  作为一条指纹,两种方法均形成指纹库.

在指纹库匹配算法的在线定位阶段,在实验大厅中随机选取 31 个位置进行定位测试,本文采用 3 种不同的指纹匹配方法.第 1 种方法是:引入信号在室内空间的欧式距离  $d_j$ ,如公式(8)所示.

$$d_j = \sqrt{\sum_{i=1}^8 (RSSI_i - S_{ij})^2} = \sqrt{(RSSI_1 - S_{1j})^2 + \dots + (RSSI_8 - S_{8j})^2} \quad (8)$$

其中  $j=1,2,\dots,380$ ,  $RSSI_i$  表示在线定位阶段接收到第  $i$  个无线路由器的信号强度值,  $S_{ij}$  表示第 1 种方法形成的指纹库中第  $j$  条第  $i$  个无线路由器的信号强度值.

第 2 种方法是基于最近邻的指纹匹配算法<sup>[17]</sup>,引入欧式距离向量  $D = \{d_1, d_2, \dots, d_M\}$ ,用  $D$  来衡量未知节点距离各个采样参考点远近的标准,对  $D$  向量的元素进行大小排序,选取最小的  $K$  个元素,组成新的向量



$D_K = \{d_1, d_2, \dots, d_k\}$ . 其中,  $K$  个采样点对未知节点位置估计的影响大小不一样, 由于未知节点和采样点间的距离与欧氏距离成反比, 加权计算这  $K$  个点的质心作为节点的位置, 定义权值:

$$\gamma_j = \frac{1}{d_j} \sum_{j=1}^K \frac{1}{d_j}, j=1, 2, \dots, K \quad (9)$$

第 3 种方法是: 引入衡量位置相似性的空间距离  $d_j$ , 如公式(10)所示.

$$d_j = |F - F_j| \quad (10)$$

其中,  $F$  表示在线定位阶段的样本综合得分,  $F_j$  表示第 2 种方法中第  $j$  条指纹. 两种指纹匹配方法得到的定位误差如图 7 所示.

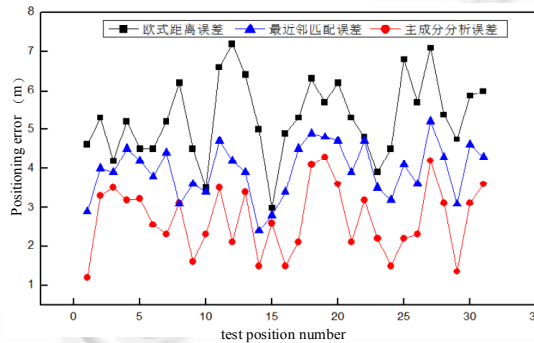


Fig.7 Three kinds of fingerprint algorithm error comparator

图 7 3 种指纹算法误差比较

从图 7 可知, 主成分匹配的指纹定位平均误差为 2.7m, 明显小于欧氏距离匹配的指纹定位平均误差 5.3m 和基于最近邻指纹匹配算法平均误差 3.9m. 这主要是由于主成分分析所得各采样点的综合得分具有较大的区分度, 不同位置对应的综合得分明显不同, 使得主成分匹配具有较高的匹配准确度, 从而保证了较低定位误差.

## 6 结 语

指纹定位由于对硬件布设要求低, 定位匹配过程简单, 受到国内外学者的广泛关注, 已经发展成为一种热门的室内定位方法. 但是, 由于室内特殊的物理环境, 无线信号指纹容易受到不确定性因素的干扰, 导致指纹定位在实际的日常生活中并没有得到快速的普及. 本文将主成分分析法应用于室内定位中, 用样本的综合得分代替由原始 8 个信号强度值构成的指纹向量, 在保留大部分原始信息的前提下, 大大减小了定位指纹信息的冗余程度, 减少了干扰因素的影响, 确保了指纹信息的稳定性和可靠性, 从而, 在减少定位系统计算复杂度的同时提高了定位精度.

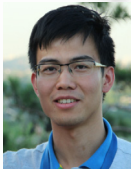
## References:

- [1] Wassi GI, Despins C, Grenier D, Nerguizian C. Indoor location using received signal strength of IEEE 802.11b access point. In: Proc. of the IEEE CCECE 2005. Piscataway: IEEE Press, 2005. 1367-1370.
- [2] Ault A, Zhong X, Coyle EJ. K-Nearest-Neighbor analysis of received signal strength distance estimation across environments. In: Proc. of the WiNMe 2005. Trentino: ICST Transactions, 2005. 1-6.
- [3] Castro P, Chiu P, Kremenek T, Muntz R. A probabilistic room location service for wireless network environments. In: Proc. of the ACM Ubicomp 2001. New York: ACM Press, 2001. 18-34.
- [4] Haeberlen A, Flannery E, Ladd AM, Rudys A, Wallach DS, Kavraki LE. Practical robust localization over large-scale 802.11 wireless networks. In: Proc. of the ACM Mobicom 2004. New York: ACM Press, 2004. 70-84.
- [5] Roos T, Myllymaki P, Tirri H, Misikangas P, Sivänen J. A probabilistic approach to wlan user location estimation. Int'l Journal of Wireless Information Networks, 2002,9(3):155-164.

- [6] Youssef M, Agrawala A. The Horus WLAN location determination system. In: Proc. of the ACM/USENIX Mobisys 2005. New York: ACM Press, 2005. 205–218.
- [7] Seshadri V, Zaruba G V, Huber M. A Bayesian sampling approach to in-door localization of wireless devices using received signal strength indication. In: Proc. of the IEEE Percom 2005. Piscataway: IEEE Press, 2005. 75–84.
- [8] Morelli C, Nicoli M, Rampa V, Spagnolini V, Alippi C. Particle filters for RSS-based localization in wireless sensor networks: An experimental study. In: Proc. of the IEEE ICASSP 2006. Washington: IEEE Press, 2006. 957–960.
- [9] Bahland P, Padmanabhan VN. RADAR: An in-building RF-based user Location and tracking system. In: Proc. of the IEEE INFOCOM 2000. Piscataway: IEEE Press, 2000. 775–784.
- [10] Papapostolou A, Chaouchi H. Orientation-Based radio map extensions for improving positioning system accuracy. In: Proc. of the Int'l Conf. on Wireless and Mobile Computing: Connecting the World Wirelessly. 2009. 947–951.
- [11] Lihan M, Tsuchiya T, Koyanagi K. Orientation-Aware indoor localization path loss prediction model for wireless sensor networks. In: Proc. of the NBIS. 2008. 169–178.
- [12] Yang XY, Kong QR, Dai XJ. An improved weighted centroid location algorithm. Journal of Xi'an Jiaotong University, 2010,44(8):1–4.
- [13] Mhatre V, Rosdenberg C. Design guidelines for wireless sensor networks: Communication, clustering and aggregation. Ad Hoc Networks, 2004,45–63.
- [14] Yu X. Operating performance analysis of commercial banks in china based on principal components analysis [MS. Thesis]. Chengdu: Southwestern University of Finance And Economics, 2011.
- [15] Fang SH, Lin TN. Principal component localization in indoor wlan environments. IEEE Trans. on Mobile Computing, 2012,11(1): 100–110.
- [16] Du M. Studies on comprehension index environmental quality based on principal component analysis [MS. Thesis]. Chengdu: Sichuan University, 2006.
- [17] Song YN. Design and implementation of an indoor positioning node based on wireless sensor network [MS. Thesis]. Beijing: Beijing University of Posts and Telecommunications, 2011.

#### 附中文参考文献:

- [12] 杨新宇,孔庆茹,戴湘军.一种改进的加权质心定位算法.西安交通大学学报,2010,44(8):1–4.
- [14] 虞茜.基于主成分分析法的我国商业银行的经营绩效分析[博士学位论文].成都:西南财经大学,2011.
- [16] 杜敏.基于主成分分析法的环境质量综合指数研究[博士学位论文].成都:四川大学,2006.
- [17] 宋宇宁.基于无线传感器网络的室内定位节点的设计及实现[博士学位论文].北京:北京邮电大学,2011.



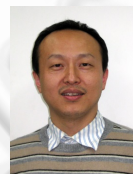
陈祠(1988—),男,湖北咸宁人,硕士生,主要研究领域为无线传感器网络,室内定位.  
E-mail: chenci@pku.edu.cn



陈永乐(1983—),男,博士,主要研究领域为室内定位.  
E-mail: yonglech@gmail.com



牟楠(1989—),女,硕士生,主要研究领域为无线传感器网络,软件工程.  
E-mail: munan@pku.edu.cn



朱红松(1973—),男,博士,副研究员,CCF高级会员,主要研究领域为无线通信,无线传感器网络,虚拟技术,物联网大数据安全分析与智能处理.  
E-mail: zhuhongsong@iie.ac.cn



张晨(1991—),女,学士,主要研究领域为无线传感器网络,计算机网络.  
E-mail: zhangchen9445@163.com



刘燕(1971—),女,博士,教授,CCF高级会员,主要研究领域为计算机网络,软件工程.  
E-mail: ly@ss.pku.edu.cn