

## 无基础设施 Wi-Fi 室内定位算法设计

王小建, 薛政, 曾宇鹏, 吴迪

(中山大学 信息科学与技术学院, 广东 广州 510006)

**摘要:** 设计了一套基于信号强度 RSS (received signal strength) 的无基础设施定位算法。它无需借助任何其他的基础设施, 只利用 Wi-Fi; 减轻了构建数据库的工作量, 通过一种特殊的动态方式进行数据库的完善; 通过热点的筛选, 有效地减轻了周围无线信号的干扰; 改进了基于 RSS 定位时的匹配算法, 提高了定位准确度; 定位操作简单, 无需终端用户的复杂参与。本算法适合服务商在特定地点提供定位服务, 通过部署专用的热点, 有效地减轻周围环境的影响, 提高定位服务的质量。

**关键词:** 无基础设施; Wi-Fi 定位; RSS; 匹配定位

中图分类号: TP393.03

文献标识码: A

文章编号: 1000-436X(2012)Z2-0239-05

## Design of infrastructure-free Wi-Fi indoor localization

WANG Xiao-jian, XUE Zheng, ZENG Yu-peng, WU Di

(School of Information Science & Technology, Sun Yat-Sen University, Guangzhou 510006, China)

**Abstract:** A RSS-based infrastructure-free localization algorithm was proposed, which is only based on Wi-Fi signals and does not require any additional infrastructure. It reduces the database construction cost by a special dynamic method. By picking the hot spot, it can effectively alleviate the interference from other wireless signals in the nearby region. It also enhances the RSS-based matching algorithm and hence improves the localization accuracy. The algorithm is easy to operate and doesn't need any complex participation from end users. The algorithm is suitable for those who can provide positioning service in some special locations. Through the deployment of dedicated hot spots, it will reduce the environment interference and improve the quality of location-based service.

**Key words:** infrastructure-free; Wi-Fi localization; RSS; mapping localization

### 1 引言

伴随着现代社会的进步, 各个领域对定位的要求也越来越高。室外定位方面, GPS(global positioning system)能提供高质量的服务。室内定位方面, GPS 受地理环境的影响, 定位效果不佳, Wi-Fi 定位成为解决室内定位问题的一大选择。

根据定位原理的不同, 大体分为 3 类: 基于基础设施、基于信号传输模型、基于信号强度 RSS。

基于基础设施的定位, 需要在定位区域铺设特殊的硬件装置, 然后根据这些硬件设施提供的地理信息来计算定位请求者的位置<sup>[1-5]</sup>。显然, 这种定位方式的缺点就是需要特定硬件设施的支持。

基于信号传输模型的定位, 通过建立信号传输

收稿日期: 2012-10-23

基金项目: 教育部新世纪优秀人才支持计划基金资助项目(110542); 广东省自然科学基金资助项目(10451027501005630); 广州市珠江科技新星专项计划基金资助项目(2011J2200086); 教育部博士点基金资助项目(20100171120047)

**Foundation Items:** New Century Excellent Talents in University (110542); The Natural Science Foundation of Guangdong Province (10451027501005630); Zhujiang Nova Special Programs of Guangzhou (2011J2200086); Doctoral Fund of Ministry of Education of China (20100171120047)

距离与信号强度之间的关系模型,计算信号接收点与信号发射源之间的距离,进而根据发射源的地理信息计算出接收点的地理位置<sup>[6-10]</sup>。由于信号在空间中的传输过程非常复杂,不能单一地认为传输过程是一条直线,因为会造成很多定位的错误。

基于信号强度 RSS 的定位,是通过信号强度的匹配来实现的,也就是根据定位请求者周围的信号强度信息,到预先构建好的数据库中去匹配,得到一个最佳的匹配结果,将相应的地理信息返回给定位请求者。这种方式需要提前构建数据库,涉及大量的时间和精力,而且受环境的影响很大<sup>[11-15]</sup>。

本文设计了一套基于信号强度 RSS 的定位算法,可以不受基础设施的束缚,显著提高了定位准确度,同时通过轻量的预构建和相应的动态完善措施减轻了构建数据库的工作量,增强了对环境的适合性。

## 2 算法框架

算法大体分为 2 部分:预构建数据库、定位。预构建数据库包括物理划分、热点筛选、建立存储结构、信号收集、聚类处理、关联地理信息。定位包括匹配定位、更新数据库。定位流程如图 1 所示。

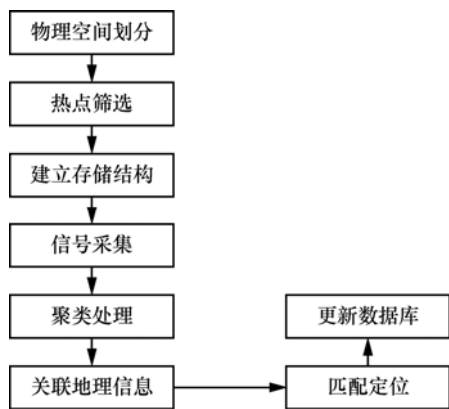


图 1 算法流程

## 3 算法设计

### 3.1 预备知识

无线接入点即无线 WAP (wireless access point) 它是用于无线网络的无线交换机,也是无线网络的核心。当打开终端的 Wi-Fi 扫描周围的无线网络信息时,可以获得周围接入点的相关信息,每个接入点的都有其相应的信号强度 RSS 和 SSID(service set identifier),如图 2 所示。

SSID	接入点的网络标识符
RSS	接入点的信号强度

图 2 接入点信息

### 3.2 物理空间划分

物理空间划分就是将定位区域进行科学的划分,分成具有明显差异的  $k$  个部分。这里的明显差异是指扫描到的 Wi-Fi 信息具有明显的差异,也就是不同的部分之间扫描到的 Wi-Fi 信息差异明显,相同的部分之中,扫描到的 Wi-Fi 信息非常相似。其中的相似与差异的判断本文会在匹配定位一节中进行说明。一般情况下,可以将定位区域以房间为单位进行划分,每个房间就可以作为一部分,  $k$  个房间组成整个定位区域。

### 3.3 热点筛选

热点筛选就是选择一些稳定的常设的热点,这样在一定程度上就减少了其他热点对系统的影响。一般情况下,这些常设热点由定位服务提供商提供,假设筛选出来的热点个数为  $n$ ,笔者记录下这些热点的 SSID,进而建立相应的存储结构来存储与这些 SSID 相对的热点的信号强度,不存储其他热点的信号强度。

### 3.4 建立存储结构

根据上述筛选好的热点,就可以建立存储结构了。假如筛选后的热点数量为  $k$ ,那么就建立一个数组,长度为  $k$ ,用来存储来自  $k$  个热点的信号强度。现在需要做的就是怎样根据热点的 SSID,来映射到相应的数组下标。记录下  $k$  个热点的 SSID 后,就可以随便设置一种映射关系,建立 SSID 与数组下标的一一映射,如图 3 所示。

SSID	数组下标
ID1	1
ID2	2
...	...
ID $k$	$k$

图 3 SSID 与数组下标的映射

综上,对于每一个热点信息,可以用一个 Map 来存储,SSID 为键,RSS 为值。相应的,每一次扫描周围的热点信息时,会得到一组热点的信息,用一个 Array 来存储,Array 里面的元素是上述的 Map,数组下标与 SSID 的映射也就是上述建立的一一映射。

### 3.5 信号收集

现在,当打开 Wi-Fi 扫描周围的无线信号时,

就会得到所有扫描到的热点信号, 对于每一个热点, 先检测它是不是在筛选范围收, 如果不在, 丢弃掉。如果在, 则根据映射表将其存储在相应的数组元素中。这样, 一次扫描完成后, 就可以得到一组信号强度信息, 存储在一个 Array 当中。

因为信号收集不止一次扫描, 本文将每次的扫描结果都保存在同一个 List 中, 用这个 List 来存储所有采集到的信息。

假设, 在每个物理空间中采集  $r$  次, 则最终的采取总量为  $kr$ 。

### 3.6 聚类处理

根据物理空间划分的数量  $s$ , 将收集到的信号也会为  $s$  组, 每组形成一个虚拟空间, 用来代表一个相应的物理空间。具体的分组, 可以根据上面采取到的信号总量  $kr$  和每个空间采取的数量, 每隔  $r$  个元素划分为一组, 这样最终就会形成  $k$  组信号, 分别采集自  $k$  个不同的物理空间。

分组完成后, 形成  $k$  个虚拟空间, 每个虚拟空间有  $r$  个信号信息, 现在构造一个能代表虚拟空间的量  $Center_i (i=1, 2, L, k)$ , 它也是一个数组, 本文将其定义为

$$Center_i[j] = \frac{\sum_{l=(i-1)r}^{ir} List(l)[j]}{r}$$

其中,  $i=1, 2, L, k; j=1, 2, L, n$ 。

现在, 就成功构建了  $k$  个虚拟空间。之后, 对于每次再扫描到的信号信息, 就可以通过一定的匹配算法, 确定与其最相近的中心, 也就是能确定扫描到的信号所在的虚拟空间了。相应的匹配算法将在 3.8 节中进行说明。

### 3.7 关联地理信息

对于每个扫描结果, 可以确定其所在的虚拟空间了, 然而仍然不能提供定位服务, 因为虚拟空间不是物理空间, 没有任何地理信息在里面, 因为我们需要将虚拟空间和物理空间关联起来, 根据虚拟空间就能找到物理空间, 从而实现定位。

现在, 在每个物理空间中, 扫描一次 Wi-Fi 信息, 确定自己所在的虚拟空间, 物理空间是已知的, 则这个虚拟空间和这个物理空间形成了一个对应关系, 将这个关系保存起来。于是, 在  $n$  个虚拟空间中, 都进行一次这样的操作, 就能将虚拟空间和物理空间的对应关系建立起来, 构造一一映射。之后, 就可以根据用户所在的虚拟空间, 给出其所在

的物理空间。

此时, 数据库的预构建已经完成, 系统具有定位功能。

### 3.8 匹配定位

对于每一个扫描结果, 只要判断出它所在的虚拟空间, 就能再根据相应的虚拟空间与物理空间的映射得到最终的地理信息。现在需要解决的问题就是如果判断每个扫描结果所在的虚拟空间, 也就是判断它与哪个虚拟空间的中心最相似。

笔者发现, 这种相似不仅仅是信号强度大小上的相似, 而且各个热点的信号强度的大小顺序也是非常相似的。例如, 在物理空间 A 中进行扫描,

将扫描到的结果根据信号强度 RSS 进行排序, 比如得到的结果如下。

$$RSS_1 > RSS_2 > RSS_3 > L > RSS_k$$

那么任何来自 A 空间的扫描结果, 根据 RSS 进行排序后的结果将与上面的结果最相近。同样, 在 B 空间中的扫描结果进行排序后, 结果可能如下。

$$RSS_2 > RSS_1 > RSS_3 > L > RSS_k$$

来自不同空间的扫描结果进行排序后, 得到的结果是有差异的, 这就是计算扫描结果所在虚拟空间的基本原理。

因此, 在判断扫描结果所在虚拟空间时, 先分别对每个虚拟空间的中心  $Center_i (i=1, 2, L, k)$  按照信号强度进行排序, 然后将扫描结果进行同样的排序, 接着将排序后的扫描结果与排序后的中心一一进行比较, 找出相似度最大的。举例来说, 例如扫描结果与中心 C 进行比较, 排序后, 首先看它们的第一个元素所对应的 SSID 是否相同, 如果相同, 则相似度增加一个, 如果不同, 继续进行第二个元素的比较, 一直比较到第  $k$  个元素为止, 就可以得到扫描结果与中心 C 的相似度了。按照这样的方式, 将扫描结果与其他中心分别进行比较, 最终筛选出相似度最大的那些(可能不止一个)。

最终筛选出的相似度最大的中心可能不止一个, 这时再根据扫描结果与中心之间的距离来找出距离最小的那个, 距离的计算方式采用向量之间的距离计算公式

$$d = \sqrt{\sum_{i=0}^{k-1} (RSS_i - C_i)^2}$$

以上 2 次筛选之后, 就可以更加准确地确定扫

描结果所在的虚拟空间，进而根据相应的物理空间来定位用户。

### 3.9 更新数据库

经过 3.7 节的关联地理信息后，数据库预构建完成，系统能够定位用户。此时，用户可以通过终端发送周围的扫描结果给服务器，服务器将相应的物理空间返回给用户。与此同时，也就用户本次的扫描结果进行存储，根据其所在的虚拟空间，放到相应的存储空间中，同时同步更新相应空间的中心 C 的数据，也就是重新计算中心 C。

随着数据量的增大，每个虚拟空间的中心也将越来越准确。不但减轻了预构建数据库的负担，同时增强了定位的准确性。

### 3.10 算法小结

本文算法主要分为预构建数据库和定位两大部分。预构建数据库是用来建立提供定位服务的数据系统，由服务提供者预先完成，不涉及定位用户的参与。定位是提供定位服务并不断更新数据库的阶段，定位用户在此阶段享受系统提供的定位服务，同时会将自己的扫描结果贡献给数据库，使其更加准确和完善。

## 4 实验

### 4.1 实验说明

以往的基于 RSS 的定位算法涉及构建数据库的大量工作，本文通过轻量的预构建和相应的动态完善措施，有效的减轻了数据库的构建负担，之后还要进一步对其定位的准确度进行测试。

传统的基于信号强度 RSS 的定位算法，在确定最佳匹配结果时，往往根据向量之间的距离来计算，得到距离小的那个，然后再根据其相应的地理信息进行定位。本文发现了一个新的规律，就是各个热点在同一物理空间中信号强度的大小顺序具有相似性。根据这一规律进行初步筛选，然后再根据向量的距离计算方法再次筛选，同时不断的更新和完善数据库，最终提高了定位的准确度。以下将对这 2 种算法的定位准确率进行比较。

### 4.2 实验过程

对于传统算法和本文算法，本文使用相同的物理空间划分，常设热点以及相同的虚拟空间中心计算方法。对所选的定位区域，划分为 20 个部分，也就是有 20 个房间，常设热点筛选出 30 个。

传统定位算法不会动态的更新数据库，需要预

先进行大量的数据采集来构建数据库。对于每个物理空间中，均匀地进行 50 次采集，包括 4 个角落和房间正中心。以后进行定位服务时，将不再更新数据库。

本文定位算法包括预构建和定位 2 部分，随着定位的进行，会不断的更新和完善数据库。对于每个物理空间，均匀地采集 25 次，包括 4 个角落和房间正中心。以后进行定位服务时，数据库也会不断的更新。

对于每种定位算法，都发送 100 次定位请求，如果系统返回的定位结果显示为当前的房间，则把它定义为一次正确的定位，并统计正确定位的次数。重复进行 3 次以上实验，结果如图 4 所示。

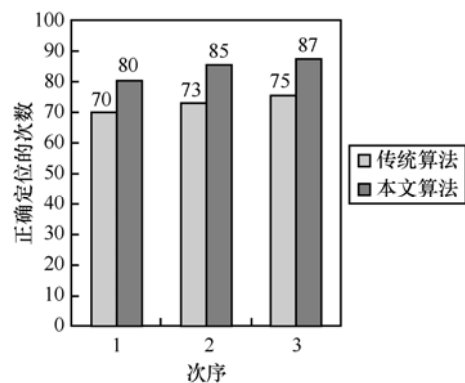


图 4 算法对比

结果表明，本文算法提高了定位的准确度。同时还发现，随着定位次数的增加，定位准确度有增加的趋势。

针对本文算法，同样按照上面的方法进行数据库的预构建，通过定位次数的增加，来观察定位准确度的变化情况，如图 5 所示。

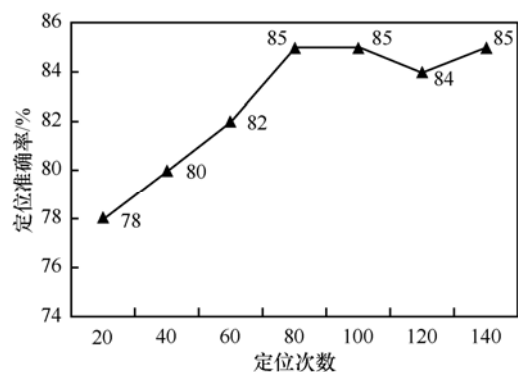


图 5 定位准确率随着定位次数的变化

实验表明，随着定位次数的增加，定位准确度也会有一定的增强。

## 5 结束语

### 5.1 本文工作

本文定位算法适合于特定服务商根据自己提供的定位服务来部署相应的定位环境。

算法分为 2 个部分: 数据库预构建和定位。算法简单, 只利用 Wi-Fi, 摆脱了特殊硬件设施的束缚, 减轻了数据库的构建工作, 在一定程度上提高了定位的准确度, 不涉及用户复杂的参与, 同时增强了对环境的适应性。

### 5.2 未来工作

定位的精度不够高, 很大程度上与物理空间的划分有关系, 今后的研究工作可以进一步将定位的基本单位由房间缩小到房间的 4 个角落, 进一步提高定位的精度。

定位的准确度有待进一步提升, 可以在特定空间部署特定的热点, 进一步提高定位的准确度。

### 参考文献:

- [1] WARD A, JONES A, HOPPER A. A new location technique for the active office[J]. *Personal Communications*, 1997, 4(5):42-47.
- [2] YOUSEF N R. Robust multipath resolving infading conditions for mobile-positioning systems[A]. *Radio Science Conference[C]*. Minufiya, Egypt, 2000. C19/1-C19/8.
- [3] YOUSSEF M, YOUSSEF A, RIEGER C, *et al.* Pinpoint: an asynchronous time-based location determination system[A]. *Proceedings of the ACM MobiSys[C]*. 2006. 165-176.
- [4] HWANG D G, HWANG J H, JANG S J, *et al.* A fast ToA position estimation technique based on MHP pulse[A]. *Proceedings of Communications and Information Technology[C]*. Icheon, Korea, 2009. 1472-1476.
- [5] TIAN H, WANG S, XIE H Y. Localization using cooperative AOA approach[A]. *Proceedings of Wireless Communications, Networking and Mobile Computing[C]*. Shanghai, China, 2007. 21-25.
- [6] JI Y, BIAZ S, PANDEY S, AGRAWAL P. ARIADNE: a dynamic indoor signal map construction and localization system[A]. *Proceedings of MobiSys[C]*. Sweden, 2006. 151-164.
- [7] CANDY J V, SULLIVAN E J. Model-based localization for a shallow ocean experiment[A]. *Proceedings of OCEANS '95, MTS/IEEE, Challenges of Our Changing Global Environment[C]*. San Diego, USA, 1995. 969-974.
- [8] GWON Y, JAIN R. Error characteristics and calibration-free techniques for wireless LAN-based location estimation[A]. *Philadelphia, USA, 2004*. 2-9.
- [9] LIM H, KUNG L C, HOU J C, *et al.* Zero configuration indoor localization over IEEE 802.11 wireless infrastructure[J]. *Wireless Networks*, 2010, 16(2): 405-420.
- [10] CHINTALAPUDI K, PADMANABHAIYER A, PADMANABHAN V N. Indoor localization without the pain[A]. *Proceedings of the ACM MobiCom*, 2010. 173-184.
- [11] YOUSSEF M, AGRAWALA A. The horus WLAN location determination system[A]. *Proceedings of MobiSys[C]*. Seattle, USA, 2005. 205-218.
- [12] AZIZYAN M, CONSTANDACHE I, CHOUDHURY R R. Surroundsense: mobile phone localization via ambience fingerprinting[A]. *Proceedings of MobiCom[C]*. Beijing, China, 2009. 261-272.
- [13] WANG H Y, ZHAO J H, YANG Q. Indoor localization in multi-floor environments with reduced effort[A]. *Proceedings of Pervasive Computing and Communications (PerCom)[C]*. Mannheim, Germany, 2010. 244-252.
- [14] GRISWOLD W G, SHANAHAN P, BROWN S W, *et al.* Active campus: experiments in community-oriented ubiquitous computing[J]. *Proceedings of Computer*, 2004, 37(10): 73-81.
- [15] PARK J, CHARROW B, CURTIS D, *et al.* Growing an organic indoor location system[A]. *Proceedings of the ACM MobiSys[C]*. San Francisco, USA, 2010. 271-284.

### 作者简介:



王小建 (1986-), 男, 河北廊坊人, 中山大学硕士生, 主要研究方向为多媒体无线网络。



薛政 (1989-), 男, 湖南益阳人, 中山大学硕士生, 主要研究方向为网络与分布式计算。

曾宇鹏 (1986-), 男, 广东河源人, 中山大学硕士生, 主要研究方向为网络与分布式计算。

吴迪 (1978-), 男, 安徽肥西人, 博士, 中山大学副教授、硕士生导师, 主要研究方向为网络与分布式计算。