



(12) 发明专利申请

(10) 申请公布号 CN 104507097 A

(43) 申请公布日 2015. 04. 08

(21) 申请号 201410815088. 2

(22) 申请日 2014. 12. 19

(71) 申请人 上海交通大学

地址 200240 上海市闵行区东川路 800 号

(72) 发明人 徐昌庆 裴凌 原野 刘乾辰

简洪浩

(74) 专利代理机构 上海科盛知识产权代理有限

公司 31225

代理人 赵志远

(51) Int. Cl.

H04W 16/22(2009. 01)

H04W 64/00(2009. 01)

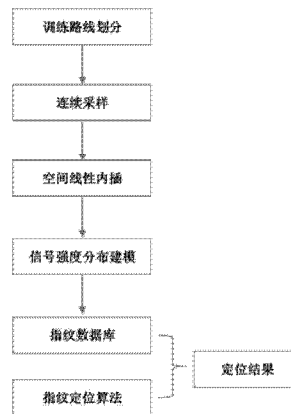
权利要求书2页 说明书4页 附图2页

(54) 发明名称

一种基于 WiFi 位置指纹的半监督训练方法

(57) 摘要

本发明涉及一种基于 WiFi 位置指纹的半监督训练方法,包括:确定目标区域内的离线训练路线;对目标区域内进行样本点的连续采样;根据离线训练路线起止点位置坐标及每一采样点时间戳对以上采集所得样本点进行位置信息的线性内插;根据上述步骤连续采样及线性内插后获得的样本点,进行目标区域信号强度分布建模;将原始样本点数据与得到的目标样本点数据建立数据库。与现有技术相比,本发明采用离线训练时间与定位精度两个评价指标,利用定点测量法对离线训练过程进行评价,得到各自的离线训练过程评价价值,根据原始采样数据与目标区域拓扑结构建立信号强度分布模型,实现了快速、准确的数据库建立。



1. 一种基于 WiFi 位置指纹的半监督训练方法,其特征在于,包括以下步骤:

步骤一,根据目标区域拓扑结构,为目标区域进行训练路线划分,并确定目标区域内的离线训练路线;

步骤二,以确定的离线训练路线为单位,对目标区域内进行样本点的连续采样;

步骤三,根据离线训练路线起止点位置坐标及每一采样点时间戳对以上采集所得样本点进行位置信息的线性内插;

步骤四,根据上述步骤连续采样及线性内插后获得的样本点,进行目标区域信号强度分布建模;

步骤五,将原始样本点数据与步骤四得到的目标样本点数据建立数据库;

步骤六,根据数据库及当前信号强度数据,选定定位算法进行定位。

2. 根据权利要求 1 所述的基于 WiFi 位置指纹的半监督训练方法,其特征在于,包步骤一中将目标区域以多覆盖、少重叠为原则对目标区域内的离线训练路线进行确定。

3. 根据权利要求 1 所述的基于 WiFi 位置指纹的半监督训练方法,其特征在于,步骤二中所述的连续采样具体为:

以确定的离线训练路线为单位,移动终端在训练路线上匀速移动,同时对接收到的各接入点信号强度以及每次采样的时间戳进行连续采样,并且记录当前路线的起止点位置坐标。

4. 根据权利要求 1 所述的基于 WiFi 位置指纹的半监督训练方法,其特征在于,步骤二的连续采样中,训练路线的起止点样本包含位置坐标信息、时间戳信息以及信号强度信息,其他点的样本只包含时间戳信息及信号强度信息。

5. 根据权利要求 4 所述的基于 WiFi 位置指纹的半监督训练方法,其特征在于,步骤三中所所述的空间线性内插具体为:

在路线确定并且移动速度相对稳定条件下,根据起止点位置坐标信息、起止点采样时间戳信息及路线内任意样本点的采样时间戳信息对该样本点位置坐标进行计算,对于样本点具体是:

$$L_i = L_{start} + \frac{t_i - t_{start}}{t_{end} - t_{start}} \times (L_{end} - L_{start}), t_{start} \leq t_i \leq t_{end}$$

其中 L_{start} , L_{end} 为实际测得的训练路线起止点位置坐标, t_{start} , t_{end} 为实际记录的起止点信号强度采样时间戳, t_i 为训练路线内某个有效信号强度采样点的采样时间戳, L_i 为该点位置坐标的计算值。

6. 根据权利要求 1 所述的基于 WiFi 位置指纹的半监督训练方法,其特征在于,步骤四中所所述的信号强度分布建模采用高斯过程建模,具体如下:

1) 将样本信息录入,根据样本信息计算建模核心矩阵 K :

$$K(i, j) = \sigma_f^2 \exp\left(-\frac{1}{2l^2} |L_i - L_j|^2\right)$$

其中, σ_f 代表信号强度标准差, l 代表信号强度空间相关性尺度, l 越大等距离样本点信号强度相关性越小,反之等距离样本点信号强度相关性越大, L_i , L_j 代表原 n 个样本点中两个样本点的空间位置坐标;

- 2) 根据目标区域拓扑图及原输入样本在区域内分布情况选取目标样本点,所述目标样本点为目标区域内任意点,根据分布模型以及位置坐标计算得到信号强度数据的样本点;
- 3) 根据核心矩阵 K 和目标点位置坐标数据计算目标点信号强度数据:

$$rss_* = k_*^T (K + \sigma_s^2 I)^{-1} rss$$

其中, rss_* 代表目标点信号强度, k_* 代表目标点与原 n 个输入点位置坐标的相关性矩阵:

$$k_*(i) = \sigma_f^2 \exp\left(-\frac{1}{2l^2} |L_* - L_i|^2\right)$$

σ_s 代表高斯观测噪声, I 为单位矩阵, rss 代表原 n 个输入点的信号强度数据矩阵, L_* 为目标点位置坐标。

一种基于 WiFi 位置指纹的半监督训练方法

技术领域

[0001] 本发明涉及无线通信技术领域,尤其是涉及一种基于 WiFi 位置指纹的半监督训练方法。

背景技术

[0002] 随着智能手机和无线网络的快速发展,例如公共安全、急救、货物运输等基于位置的服务得到愈发广泛的关注。室外条件下的位置服务可以由全球卫星导航系统提供较好的结果。室内条件下的位置服务根据信号种类有基于 WiFi、蓝牙、终端内置运动传感器等定位方法,WiFi 信号较为稳定且获取方便,是目前较为主流的室内定位信号之一。基于 WiFi 的指纹定位系统一般包括离线训练阶段——用来建立指纹定位数据库,及在线定位阶段——根据当前 RSS 数据及数据库进行当前位置计算两个步骤。其中的离线训练阶段影响着最终定位的精度,是指纹定位系统的关键。

[0003] WiFi 指纹定位系统的训练方法一般以耗时低、可移植性强、定位精度稳定等因素作为标准。传统的离线训练方法分为三个步骤:目标区域网格化、网格点 RSSI 采集及目标区域指纹库建立。传统的离线训练方法通过记录所有参考点的真实位置数据及 RSSI 向量的方式进行指纹库建立,我们视之为全监督训练。全监督训练旨在通过大量的数据采集进行建库,提高数据库对目标区域信息描述的覆盖度及细致度,以期得到高精度定位结果。然而全监督训练中大量耗时与耗工导致的建库效率偏低、系统移植性偏差等问题依然亟需研究解决。

[0004] 经过对现有技术文献的检索发现,B Ferris、D Fox 等人在 2007 年 IJCAI 会议上发表的文章“WiFi-SLAM Using Gaussian Process Latent Variable Models”(使用高斯过程潜变量模型的 WiFi-SLAM 算法)中考虑到目标区域中信号强度根据空间分布的规律性及信号强度分布根据时间的相对稳定性提出了一种通过目标区域少量信号强度数据,建立目标区域拓扑图及信号强度分布模型的算法。但是它存在两个不足之处:1) 数据库中不含真实位置信息导致定位精度较低;2) 只根据 RSS 进行的目标区域拓扑图绘制,不能完全准确的反映地图信息。

发明内容

[0005] 本发明的目的就是为了解决上述现有技术存在的缺陷而提供一种基于 WiFi 位置指纹的半监督训练方法,采用建库原始样本数和定位精度两个评价指标对离线训练阶段进行评价,实现了快速、准确的离线训练过程。

[0006] 为准确评价离线训练质量,本发明采用了训练时间和定位精度作为评价指标。前者是反应离线训练效率的参数,后者是反映离线训练质量的重要参数。训练时间指在离线训练阶段中,记录了原始采样过程所用的时间总和;定位精度指在定位阶段,根据当前数据库,以定点测量法进行定位的结果与真实位置以米计的偏差量。

[0007] 定位终端接收到的信号强度在空间的分布是连续的。无线信号传输模型中信号强

度距离接入点 (AP) 的距离是负相关的,当终端向背离 AP 的方向移动时,接收到的信号强度就会逐渐变小,反之,当终端迎着 AP 的方向移动时,接收到的信号强度将逐渐变强;而引入空间遮挡及信道间干扰的影响后,信号强度在目标区域内表现为不规则分布,但仍然为连续变化。

[0008] 本发明的目的可以通过以下技术方案来实现:

[0009] 一种基于 WiFi 位置指纹的半监督训练方法,其特征在于,包括以下步骤:

[0010] 步骤一,根据目标区域拓扑结构,为目标区域进行训练路线划分,并确定目标区域内的离线训练路线;

[0011] 步骤二,以确定的离线训练路线为单位,对目标区域内进行样本点的连续采样;

[0012] 步骤三,根据离线训练路线起止点位置坐标及每一采样点时间戳对以上采集所得样本点进行位置信息的线性内插;

[0013] 步骤四,根据上述步骤连续采样及线性内插后获得的样本点,进行目标区域信号强度分布建模;

[0014] 步骤五,将原始样本点数据与步骤四得到的目标样本点数据建立数据库;

[0015] 步骤六,根据数据库及当前信号强度数据,选定定位算法进行定位。

[0016] 包步骤一中将目标区域以多覆盖、少重叠为原则对目标区域内的离线训练路线进行确定。

[0017] 步骤二中所述的连续采样具体为:

[0018] 以确定的离线训练路线为单位,移动终端在训练路线上匀速移动,同时对接收到的各接入点信号强度以及每次采样的时间戳进行连续采样,并且记录当前路线的起止点位置坐标。

[0019] 步骤二的连续采样中,训练路线的起止点样本包含位置坐标信息、时间戳信息及信号强度信息,其他点的样本只包含时间戳信息及信号强度信息。

[0020] 步骤三中所述的空间线性内插具体为:

[0021] 在路线确定并且移动速度相对稳定条件下,根据起止点位置坐标信息、起止点采样时间戳信息及路线内任意样本点的采样时间戳信息对该样本点位置坐标进行计算,对于样本点具体是:

$$[0022] \quad L_i = L_{start} + \frac{t_i - t_{start}}{t_{end} - t_{start}} \times (L_{end} - L_{start}), t_{start} \leq t_i \leq t_{end}$$

[0023] 其中 L_{start} , L_{end} 为实际测得的训练路线起止点位置坐标, t_{start} , t_{end} 为实际记录的起止点信号强度采样时间戳, t_i 为训练路线内某个有效信号强度采样点的采样时间戳, L_i 为该点位置坐标的计算值。

[0024] 步骤四中所所述的信号强度分布建模采用高斯过程建模,具体如下:

[0025] 1) 将样本信息录入,根据样本信息计算建模核心矩阵 K :

$$[0026] \quad K(i, j) = \sigma_f^2 \exp\left(-\frac{1}{2l^2} |L_i - L_j|^2\right)$$

[0027] 其中, σ_f 代表信号强度标准差, l 代表信号强度空间相关性尺度, l 越大等距离样本点信号强度相关性越小,反之等距离样本点信号强度相关性越大, L_i , L_j 代表原 n 个样本

点中两个样本点的空间位置坐标；

[0028] 2) 根据目标区域拓扑图及原输入样本在区域内分布情况选取目标样本点,所述目标样本点为目标区域内任意点,根据分布模型以及位置坐标计算得到信号强度数据的样本点；

[0029] 3) 根据核心矩阵 K 和目标点位置坐标数据计算目标点信号强度数据；

$$[0030] \quad r_{SS*} = k_*^T (K + \sigma_s^2 I)^{-1} r_{SS}$$

[0031] 其中, r_{SS*} 代表目标点信号强度, k_* 代表目标点与原 n 个输入点位置坐标的相关性矩阵；

$$[0032] \quad k_*(i) = \sigma_f^2 \exp\left(-\frac{1}{2l^2} |L_* - L_i|^2\right)$$

[0033] σ_s 代表高斯观测噪声, I 为单位矩阵, r_{SS} 代表原 n 个输入点的信号强度数据矩阵, L_* 为目标点位置坐标。

[0034] 与现有技术相比,本发明具有如下有益效果：

[0035] 1) 采用连续采样与线性内插代替传统的离散采样方法,大大降低离线训练阶段所需时间。

[0036] 2) 根据连续采样及线性内插得到的样本数据进行信号强度分布建模,并根据分布模型计算目标区域内任意点信号强度数据,一方面降低离线训练阶段耗工,同时可以保证数据库对目标区域信号强度分布的描述较优,使定位阶段准确性得到保证。

附图说明

[0037] 图 1 为本发明的工作流程图；

[0038] 图 2 为本发明的实施例中的目标区域半监督训练路线图,其中虚线表示采样路线,实线表示区域中的墙壁或隔断,黑点表示采样点；

[0039] 图 3 为本发明的空间线性内插方法示意图；

[0040] 图 4 为本发明的实施例中的目标区域单个接入点信号强度分布图, x , y 轴表示二维空间坐标轴, z 轴表示信号强度,单位 dBm；

[0041] 图 5 为本发明的实施例中的目标区域单个接入点信号强度目标点示意图,虚线表示采样路线,实线表示区域中的墙壁或隔断,黑点表示采样点,三角为目标点分布示意点。

具体实施方式

[0042] 下面结合附图和具体实施例对本发明进行详细说明。

[0043] 下面结合附图对本发明的实施例作详细说明：本实施例在以本发明技术方案为前提下进行实施,给出了详细的实施方式和具体的操作过程,但本发明的保护范围不限于下述的实施例。

[0044] 本实施例中,目标区域结构如图 2 所示,区域内被 11 个接入点覆盖。终端用户在选定路线之间来回匀速运动并实时采集样本点信号强度。分别使用本发明所述方法与现有的传统训练方法采样建库进行定位,最终得到表 1 所示的样本数与定位精度对比图。

[0045] 如图 1 所示,本实施例包括如下步骤：

[0046] 步骤一：根据目标区域拓扑结构，为目标区域进行训练路线划分，在本实例中，将目标区域划分为七条路线。

[0047] 步骤二：以训练路线为单位进行连续采样，在本实例中，进行往复重复连续采样共 10 次，即每条训练路线包括 20 组连续采样数据。

[0048] 步骤三：根据离线训练路线起止点位置坐标及每一采样点时间戳对以上测量所得样本点进行位置信息的线性内插，本实例中共得到有效样本点 862 个。

[0049] 步骤四：根据前述步骤连续采样及线性内插获得的样本点，进行目标区域信号强度分布建模。本实例中对目标区域的 11 个接入点分别进行了信号强度分布建模，每个接入点信号强度的空间分布如表 1 所示。

[0050] 表 1

[0051]

| 训练方法 | 离线训练时间 (小时) | 样本点数 | 平均定位误差 (米) | 最大定位误差 (米) |
|-------|----------------|------|---------------|---------------|
| 半监督训练 | 1 | 148 | 1.85 | 4.04 |
| 传统训练 | 10 | 2100 | 1.58 | 3.13 |

[0052] 步骤五，将原始样本点数据与步骤四得到的目标样本点数据建立数据库，本实例中针对每个接入点在目标区域生成了覆盖目标区域梯形状的 1000 个目标样本点。

[0053] 步骤六，根据数据库及当前信号强度数据，选定定位算法进行定位。本实例中选择 K-Nearest-Neighbor (KNN) 算法作为定位算法。

[0054] 所述的 KNN 算法在本实例中指得到当前信号强度向量后，将其与数据库中的信号强度向量进行对比，选出信号强度向量空间内与当前向量欧氏距离最小的 K 个样本点，滤波后进行加权相加，得到当前定位结果。

[0055] 计算信号强度向量空间内欧氏距离的具体公式为：

$$[0056] \quad d_{ij} = \sqrt{(rss_{i1} - rss_{j1})^2 + (rss_{i2} - rss_{j2})^2 + \dots + (rss_{im} - rss_{jn})^2}$$

[0057] 其中 d_{ij} 指样本点 i 与样本点 j 的信号强度向量空间欧氏距离， rss_{ij} 指样本点 i 接收到接入点 j 的信号强度值。

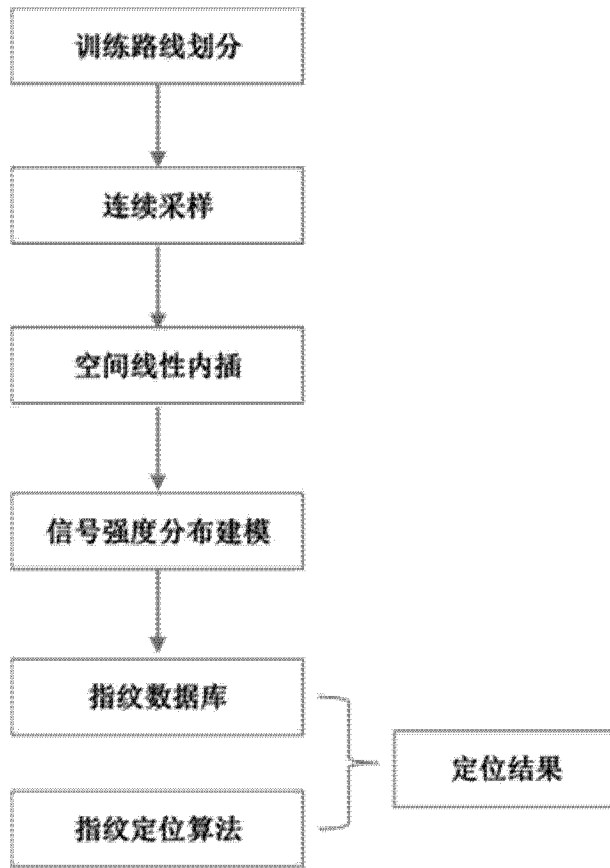


图 1

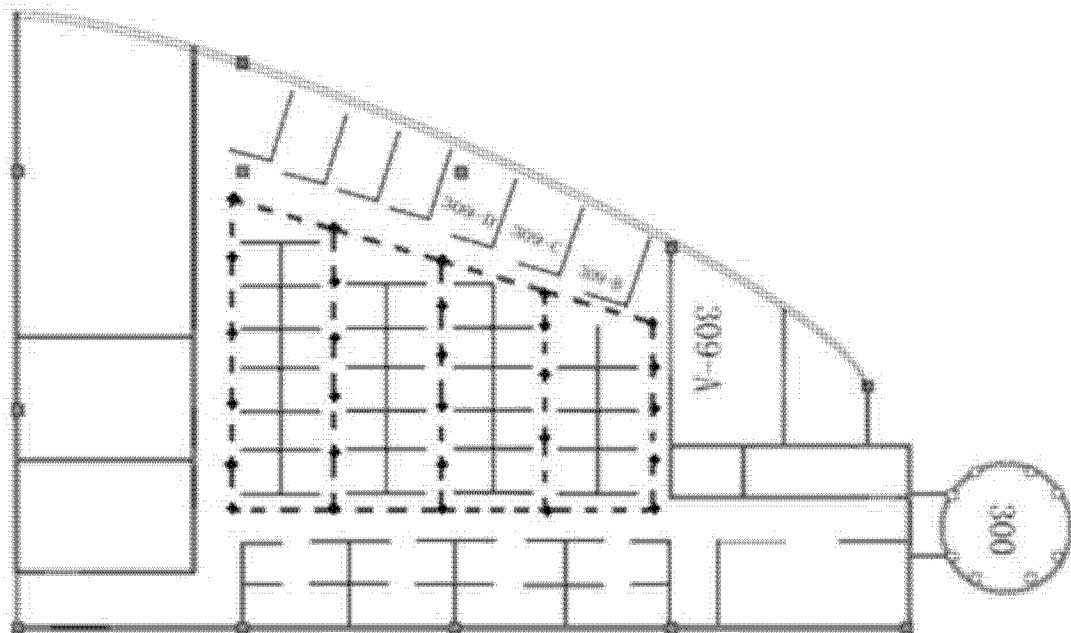


图 2



图 3

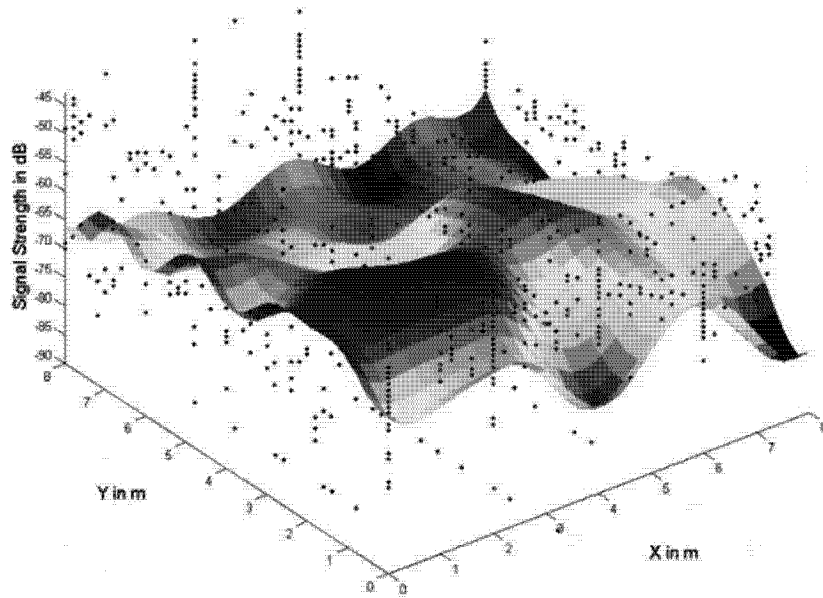


图 4

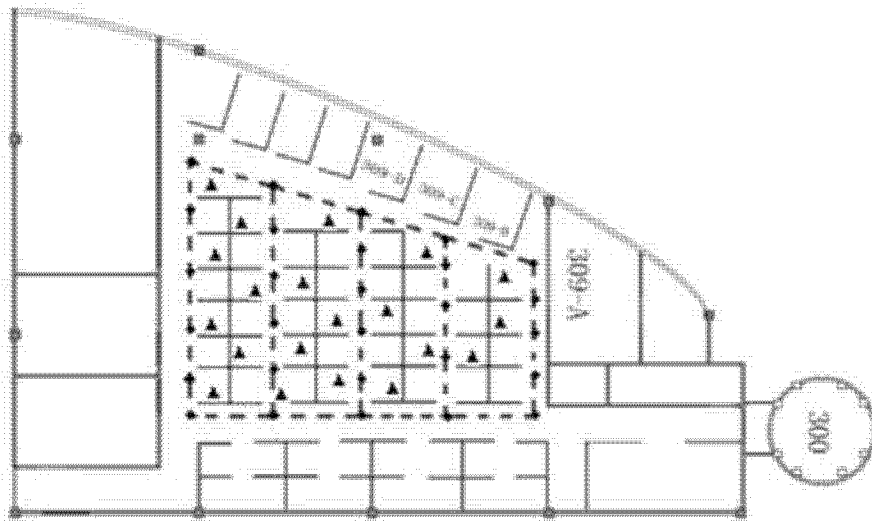


图 5