

一种基于 C4.5 算法的车位识别方法^{*}

丁 兵 吴允平 李彬雅

(福建师范大学光电与信息工程学院 福州 350007)

摘要: 提出了一种基于 C4.5 算法的车位识别方法。首先通过手机的传感器系统,对车辆行驶轨迹的 GPS 等数据进行采集,然后利用去除飞点、定位补偿、卡尔曼滤波等方法对采集的数据进行一系列的处理,并进一步采用 C4.5 算法识别车辆行驶状态信息,最终通过车位识别算法实现对停车位的自动识别。数据处理结果表明,通过上述数据处理过程和方法能够很好的识别潜在的路边车位信息。

关键词: C4.5 算法;卡尔曼滤波;车位识别;GPS

中图分类号: TP311 **文献标识码:** A **国家标准学科分类代码:** 510.5015

Parking plot recognition based on C4.5 algorithm

Ding Bing Wu Yunping Li Binya

(College of photonic and Electronic Engineering, Fujian Normal University, Fuzhou 350007, China)

Abstract: Proposes a method which is based on C4.5 algorithm to recognize the parking plot. First, we should record the driving trajectory of the car by the sensors of smart phone in order to obtain the data, such as GPS data, velocity, accelerated velocity, etc. Then it is necessary that to have the data processed to improve the precision with several means, such as remove the flying spot, compensation algorithm, Kalman Filter and so on. The next step is to get the data further processed through C4.5 algorithm in order to acquire the state of the car. Finally, the algorithm of recognizing the parking plot is used to make sure whether the spot is a parking plot or not. The results of the processing shows that the method that mentioned above is effective to find the potential roadside parking plot.

Keywords: C4.5 algorithm; Kalman Filter; parking spot recognition; GPS

1 引 言

随着我国国民经济的发展,私家车已成为人们主要的代步工具。而在大城市中,尤其是中心城区,承载着大量的车流量,因为车位紧张和缺乏有效的交通诱导而导致的无效交通量占到整体交通的 40%以上,直接引发了中心城区的道路拥堵^[1-2],而搜寻车位造成的无效巡游更是加剧了这种拥堵。仅在美国,2011年,由于交通拥堵造成的损失就高达 1212 亿美元,耗费 29 亿加仑汽油和 55 亿小时时间^[3]。纵观发达国家中心城区的停车状况发展历程可知,尽管商场和政府在做停车规划的时候已经考虑到未来 5~10 年的需求,但是这些规划在巨大的市场需求面前通常会“早早的落伍”,车位供给能力远远跟不上停车需求的增长步伐^[4]。而不断重新建设大型停车场费时耗力,成本高昂;故而综合各个方面考量,通常做法为在现有停车需求未严重超出停车场停车规划的基础上,在其周围建设一定数量的小型停车场和路边停车位可在一定时间内一定程度上缓解中心城区的交通拥堵问题;但是由于缺乏有效

的停车诱导设施,寻找这些车位并不易。国家也显然已经意识到大城市所面临的行车堵和停车难等窘境,故而在“十二五”重大科技研发专项中将“城市智能交通关键技术”作为五个重点开展的项目之一,同时在《2012 年度国家 863 计划现代交通技术领域支持重点》备选项目中更是有三项关于智能交通的研究,占据重点项目总数的 1/8。但是目前为止,城市的 PGIS (parking guiding information system) 系统仍为基于 VMS (variable message signs) 的交通诱导板来进行引导,而国内外的研究目前也仅限于从思想和理论上提出应用多智能体技术的可行性,具体实施还有待于进一步探索^[5]。随着智能手机和物联网技术的快速发展,这一困境有望得到一定程度上的缓解。通过智能手机 APP 和后端服务器组合成的停车智能引导系统,不仅可以查询实时车位状况和行车路径规划,还可以实时采集车辆行驶数据并定时上传至后端服务器^[6]。但是该系统的弊病在于对于已经废弃、改建、扩建以及新建的停车位不能实现服务器数据及时更新,从而造成下发错误、无效的车位

收稿日期:2015-01

* 基金项目:教育部科技研究重点(212087)、福建省科技厅重大(2011H6009)、福建省高校新世纪优秀人才支持计划(JA11037)资助项目

信息以及行车路径规划,不仅容易让用户对产品失去信任,还会造成行车拥堵和社会资源浪费。整体而言,目前国内基于智能手机和服务器的 ITS 系统研究较少,对于新增车位的自主发现,则尚无此方面的研究。论文即是利用车主共享的行车和停车数据信息,经过一系列数据处理过程,实现新增停车位的自动识别,进而实现新增车位数据及时、有效的添加进车位数据库。

论文通过采集车辆行驶数据,然后经过滤波处理、决策树算法、车位判别算法等过程以实现停车位的自动识别。

2 数据处理

2.1 数据采集和预处理

通过智能手机的传感器系统和数据采集软件实现对车辆行驶的 GPS、速度、加速度等数据进行采集和存储。由于受到卫星钟差、大气误差、多径时延误差、卫星轨道几何误差以及其他随机误差等因素的影响,采集的 GPS 数据并不精确^[7-8];同时由于设备或其他意外的原因,会使采集的数据中出现一些虚假的数据(一般称为粗大误差),因而在对数据进行处理之前还需预处理以去除粗大误差。

数据预处理的方法是:一般认为 GPS 数据在短时间内变化不会太大,根据车辆行驶的实际状况和位置变化趋势确定 GPS 数据可能的变化值域,超出此值域的数据即为粗大数据,予以剔除。对于剔除后的单点粗大数据,用定位补偿算法^[9]补足。具体方法是利用数据采集软件得到的前一时刻的速度数据,计算出这段时间间隔内车辆的行驶距离。考虑到车辆正常行驶情况下,行驶方向不会突变,加之数据采样时间间隔较小,故可认为车辆在该段时间内匀速行驶,可列方程式如下:

$$\begin{cases} X_n = X_{n-1} + \left[\frac{(X_{n-1} - X_{n-2})}{(t_{n-1} - t_{n-2})} \right] (t_n - t_{n-1}) \\ Y_n = Y_{n-1} + \left[\frac{(Y_{n-1} - Y_{n-2})}{(t_{n-1} - t_{n-2})} \right] (t_n - t_{n-1}) \\ Z_n = Z_{n-1} + \left[\frac{(Z_{n-1} - Z_{n-2})}{(t_{n-1} - t_{n-2})} \right] (t_n - t_{n-1}) \end{cases} \quad (1)$$

2.2 卡尔曼滤波及数据处理

在对 GPS 数据预处理后,仍需进一步处理以提升定位精度。常用的 GPS 数据处理方法有差分定位法以及滤波算法。差分定位由于需要设置定位基准站,并将基站解析的差分数据转发至移动端,成本较高,且受地形地貌的限制,故从成本以及适用性考虑,采用滤波算法较为可行。目前常用滤波算法有卡尔曼滤波算法、最小二乘法滤波、自适应滤波以及粒子滤波等^[10],本实验采用卡尔曼滤波算法,并假定系统噪声和测量噪声为高斯白噪声。

2.2.1 系统建模

采用三维坐标系建立车辆行驶数据模型,为了对车辆行驶数据实现滤波,选取车辆位置、速度、加速度以及等效位移总误差作为状态变量如下:

$$\mathbf{X} =$$

$$\begin{bmatrix} x & v_x & a_x & \varepsilon_x & y & v_y & a_y & \varepsilon_y & z & v_z & a_z & \varepsilon_z \end{bmatrix}^T \quad (2)$$

式中: x 轴、 y 轴、 z 轴为分别对应北、东、天方向的三维坐标系,式中 x 、 y 、 z 分别对应维度、经度、海拔数据, v_x 、 v_y 、 v_z 和 a_x 、 a_y 、 a_z 以及 ε_x 、 ε_y 、 ε_z 为相应的 x 、 y 、 z 轴方向的速度、加速度以及等效位移总误差。其中等效位移总误差包括卫星误差、传播误差、接收误差,并假定其符合一阶马尔科夫过程^[11-12]。参考文献^[11-12]可建立本实验的状态方程为:

$$\bar{\mathbf{X}}(t) = \mathbf{A}\mathbf{X}(t) + \mathbf{U}(t) + \mathbf{W}(t) \quad (3)$$

其中式(3)中系统转移矩阵 \mathbf{A} 为:

$$\mathbf{A} = \begin{bmatrix} A_x & 0_{4 \times 4} & 0_{4 \times 4} \\ 0_{4 \times 4} & A_y & 0_{4 \times 4} \\ 0_{4 \times 4} & 0_{4 \times 4} & A_z \end{bmatrix}$$

$$\text{式中: } \mathbf{A}_i = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{-1}{\tau_{ai}} & 0 \\ 0 & 0 & 0 & \frac{-1}{\tau_i} \end{bmatrix} \quad i = x, y, z$$

系统状态变量:

$$\mathbf{X}(t) =$$

$$\begin{bmatrix} x & v_x & a_x & \varepsilon_x & y & v_y & a_y & \varepsilon_y & z & v_z & a_z & \varepsilon_z \end{bmatrix}^T$$

系统激励:

$$\mathbf{U}(t) =$$

$$\begin{bmatrix} 0 & 0 & \frac{\bar{a}_x}{\tau_{ax}} & 0 & 0 & 0 & \frac{\bar{a}_y}{\tau_{ay}} & 0 & 0 & 0 & \frac{\bar{a}_z}{\tau_{az}} & 0 \end{bmatrix}^T$$

系统噪声:

$$\mathbf{W}(t) =$$

$$\begin{bmatrix} 0 & 0 & \omega_{ax} & \omega_y & 0 & 0 & \omega_{ay} & \omega_y & 0 & 0 & \omega_{az} & \omega_z \end{bmatrix}^T$$

将手机 GPS 接收的数据中车辆的位置、速度、加速度取为观测变量,可得本实验的观测方程为^[11-12]:

$$\mathbf{L} = \mathbf{H}\mathbf{X} + \mathbf{V} \quad (4)$$

式中: $\mathbf{L} =$

$$\begin{bmatrix} L_x & L_{vx} & L_{ax} & L_y & L_{vy} & L_{ay} & L_z & L_{vz} & L_{az} \end{bmatrix}^T \text{ 为观测矢量矩阵;}$$

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

为观测矩阵;

$$\mathbf{X} =$$

$$\begin{bmatrix} x & v_x & a_x & \varepsilon_x & y & v_y & a_y & \varepsilon_y & z & v_z & a_z & \varepsilon_z \end{bmatrix}^T \text{ 为状态变量矩阵;}$$

$V =$

$[\omega_{Lx} \ \omega_{Lxx} \ \omega_{Lax} \ \omega_{Ly} \ \omega_{Lyy} \ \omega_{Lay} \ \omega_{Lz} \ \omega_{Lxz} \ \omega_{Laz}]^T$ 为观测噪声矢量矩阵。

2.2.2 建立卡尔曼滤波方程

根据卡尔曼滤波原理和以上推得的系统方程和观测方程可建立车辆行驶数据的卡尔曼滤波方程^[11-12]。本实验中离散化后的系统噪声协方差矩阵 $Q(k)$ 为：

$$Q(k) =$$

$$diag [0 \ 0 \ \sigma_{ax} \ \sigma_x \ 0 \ 0 \ \sigma_{ay} \ \sigma_y \ 0 \ 0 \ \sigma_{az} \ \sigma_z]$$

离散化后的观测噪声协方差矩阵 $R(k)$ 为：

$$R(k) =$$

$$diag [R_x^2 \ R_w^2 \ R_{ax}^2 \ R_y^2 \ R_{vy}^2 \ R_{ay}^2 \ R_z^2 \ R_{xz}^2 \ R_{az}^2]$$

2.3 C4.5 算法简介

C4.5 算法是在 ID3 算法基础上改进而来的算法，与 ID3 算法相比，C4.5 算法采用信息增益率作为分类属性，同时在构造决策树的过程中进行剪枝，而且还可以对连续型数据以及不完整数据进行处理^[13-14]。

C4.5 算法构造决策树的基本过程如下：

1: 读取并统计样本数据。

2: 计算信息熵并选取决策树分支：

1) 按照分类标签(假设分为 m 类)对训练数据集 S 的属性集 A (假设有 n 个属性)进行划分，并计算信息熵：

$$I_S(S) = \sum_{i=1}^m -\frac{S_i}{S} \log_2 \frac{S_i}{S}, \text{ 其中 } S_i \text{ 为数据集 } S \text{ 中属于类 } i \text{ 的个数。}$$

i 的个数。

2) 按照属性集 A 中每个属性进行划分，并计算信息熵：

$$I_A(S) = \sum_{j=1}^n \frac{S_j}{S} I_{S_j}(S_j), \text{ 其中 } S_j \text{ 为数据集 } S \text{ 中属于属性 } j \text{ 的样本个数。}$$

j 的样本个数。

$$I_{S_j}(S_j) = \sum_{i=1}^m -\frac{S_{ij}}{S_j} \log_2 \frac{S_{ij}}{S_j}, \text{ 其中 } S_{ij} \text{ 为 } S_j \text{ 中属于类 } i \text{ 的样本个数。}$$

样本个数。

3) 计算信息增益：

$$I_G(S) = I_S(S) - I_A(S)。$$

4) 计算信息增益率：

$$I_{GR}(S) = \frac{I_G(S)}{I_{AW}(S)}；$$

$$\text{式中: } I_{AW}(S) = \sum_{j=1}^n -\frac{S_j}{S} \log_2 \frac{S_j}{S}。$$

5) 按照上述过程循环建立决策树。

3: 决策树剪枝：

决策树剪枝主要有两种方法：一种是基于误判的剪枝方法；另一种方法为悲观剪枝法^[15-16]。

4: 输出决策树。

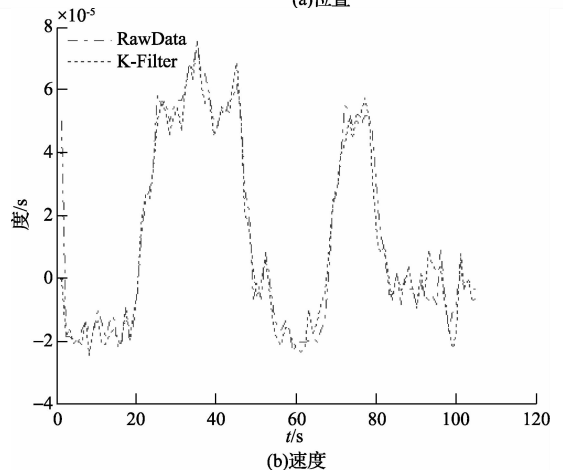
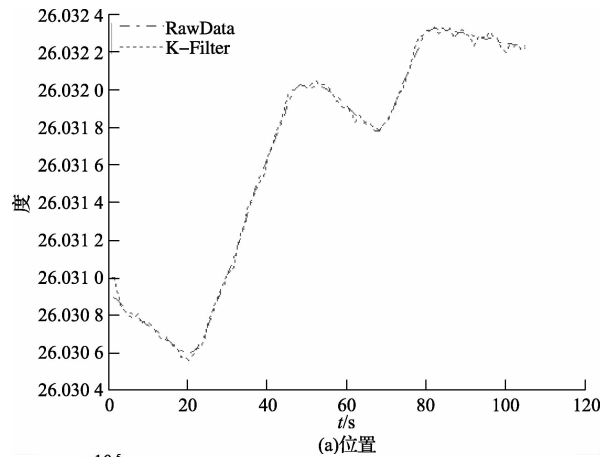
由于实验所用的数据是连续型数据，故在 C4.5 算法中首先需要对数据进行离散化处理；在 C4.5 算法中，一般连续型变量的分支都是 2 条，即通过分支阈值 θ 将连续型变量分

为 2 个分支。计算分支阈值 θ 的方法是：首先将连续型变量按照大小顺序排列(假设最小值为 MIN_{data} ，最大值为 MAX_{data})并将其 N 等分，假设等分点为 D_i ，其中 $i = 1, 2, 3, \dots, N$ 。根据等分点 D_i 分别计算 $[MIN_{data}, D_i]$ ， $[D_i, MAX_{data}]$ 的信息增益，并将计算得的最大信息增益的等分点 D_G ($G = 1, 2, 3, \dots, N$) 所在的值作为分支阈值，即 θ 。

3 实验仿真与结果分析

实验中获取的数据方法是：每次驾车前将智能手机置于车内固定好并打开数据采集软件，然后发动车辆行驶；数据来源是 2014 年 3 月 7 日至 2014 年 5 月 13 日期间车辆日常办公行驶的数据。采集的数据主要有北、东、天 3 个方向的速度和加速度以及经度、纬度、海拔、定位精度、航向角度等数据，采样周期为 1 s，共采集到数据组 196 个，选取其中 50 组精度较好的数据作为本实验数据。

在用 MATLAB 对数据进行仿真处理时，为了提高处理速度，对前述的卡尔曼滤波方程进行了降维处理，将其分为 x 轴、 y 轴、 z 轴(分别对应北、东、天方向的数据坐标)分别进行滤波，每个轴向对应的取位置、速度、加速度以及等效位移总误差作为状态变量。经过卡尔曼滤波后的车辆位置、速度、加速度如图 1 所示(图中以 x 轴为例，取一组采集样本的部分数据；所取数据为车辆停驶前 100 s 以及停驶后 5 s 的数据，加速度相关时间常数 $\tau_i = 0.5$)。



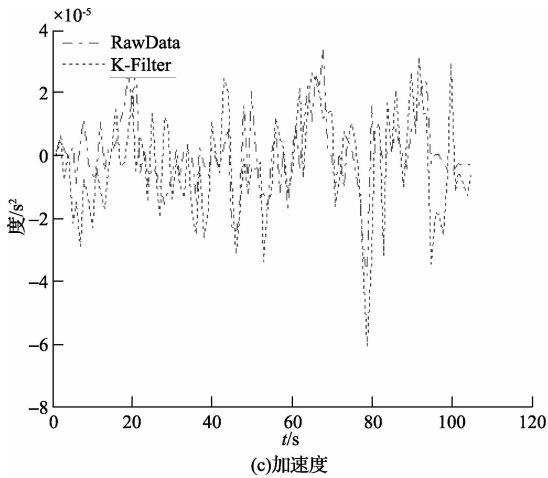


图 1 卡尔曼滤波

将滤波所得数据集前 25 组作为训练集数据，并根据滤波后的数据和车辆行驶记录，判定其分类状态。分别取 x 、 y 、 z 轴的速度、加速度作为 C4.5 算法的属性，并将车辆的行驶状态停驶和行驶中作为分类属性，分别记为离散变量 1 和 2。按照前述 C4.5 算法过程构造决策树如图 2 所示。

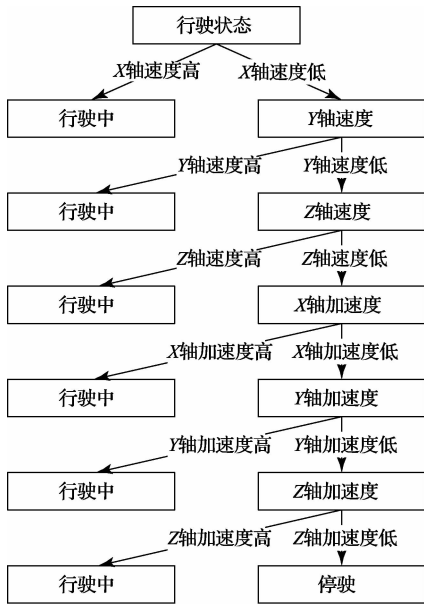


图 2 C4.5 算法构造决策树

然后用构造的决策树对剩余 25 组样本集数据进行测试，并用车位判别算法判定其是否为停车位，车位判别算法流程如图 3 所示。

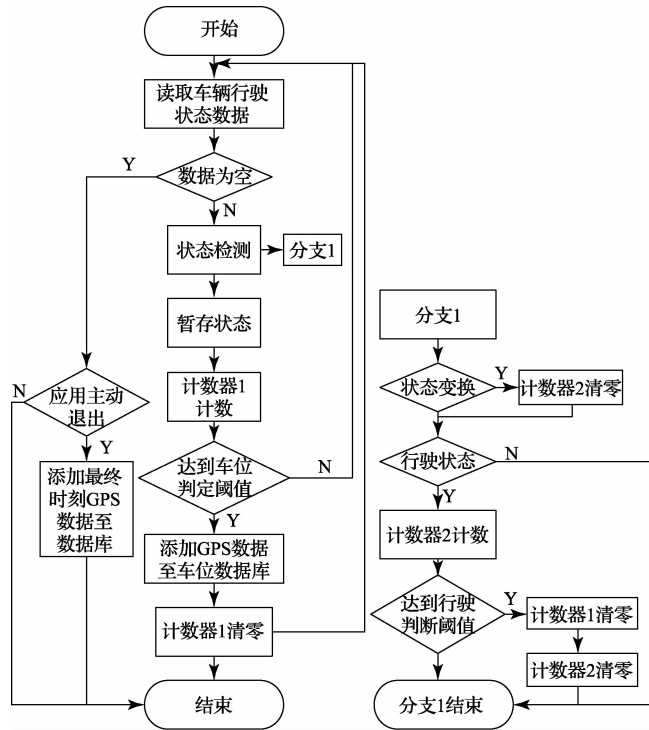


图 3 车位判别算法流程

算法流程中“车位判定阈值”为避免遇红灯及临时停车等情况引起的车位误判，将其设为大于 600 小于 1 800(即时间大于 10 min 小于 30 min)的值，实验取 600；“行驶判断阈值”为避免停车时由于 GPS 定位误差以及传感器误差造成的定位漂移而引起的车辆伪行驶而设置，一般设为 10~15(即时间为 10~15 s)，实验取 10；“应用主动退出”假定为车辆停在固定车位后人为退出数据采集软件，并由软件按协议在采集数据中予以标记；“添加 GPS 数据至车位数据库”中 GPS 数据为车位判别算法识别某点为车位后，该点所在的 GPS 数据。

将算法所得结果与实际停车情况对比后，能准确识别测试集中 22 组数据中的停车位，也即本实验方法的准确率为 0.88。

4 结 论

提出了一种基于 C4.5 算法的车位识别方法，通过卡尔曼滤波提升车辆行驶过程中的 GPS、速度以及加速度数据精度，再经过 C4.5 算法确定车辆的行驶状态，并最终通过车位识别算法实现对室外小型停车场车位和路边车位的自动识别并添加至车位数据库，通过实验发现，基于 C4.5 算法的车位识别方法有良好的精度，而且具有一定的实用价

值。由于手机定位精度及传感器精度一般,故采用更好的数据采集设备会获得更好的车位识别精度。

参考文献

- [1] 张洋,魏连雨.基于模糊理论的停车诱导决策方法研究[J].交通工程,2011(16):144-148.
- [2] 张滢,齐美彬,周云,等.基于特征提取和多示例学习的图像区域标注[J].电子测量与仪器学报,2014,28(8):909-914.
- [3] SCHRANK D, EISELE B, LOMAX T. TTI's 2012 urban mobility report [J]. Texas A&M Transportation Institute, The Texas A&M University System, 2012.
- [4] 安筱,李克平,孙剑.上海 CBD 共享停车方法研究[J].交通标准化,2010(228):58-62.
- [5] 廖瑞辉,陈星光.智能交通系统研究现状与对策分析[J].交通企业管理,2014(1):6-9.
- [6] WATENE G, NDEGWA C, MUSIEGA D. A GIS-based parking management and dissemination system [C]//Proceedings of Sustainable Research and Innovation Conference, 2014: 117-124.
- [7] 喜文飞,方源敏,王宇鸿.附有权值条件的卡尔曼滤波定位方法[J].江西科学,2011,29(1):39-42.
- [8] 孙尚龙,苏寒松.遮挡情况下及与卡尔曼滤波的运动目标跟踪[J].电子测量技术,2011,34(5):8-11.
- [9] SUMMERFELT S T, WILSON D G, WALDRUP T. Advances in fish harvest technologies for circular tanks[J]. Aquacultural Engineering, 2009(40):62-71.
- [10] 刘胜,张青春,张兰勇.聚类算法的 GPS 静态单点定位方法[J].哈尔滨工业大学学报,2012,44(11):71-74.
- [11] 刘延飞,郭锁利,侯荣昌.卡尔曼滤波在 GPS 定位误差处理中的应用[J].电子科技,2011,24(9):140-142.
- [12] 徐宏宇,王浩,王尔申.基于扩展卡尔曼滤波的 GPS 定位数据处理方法研究[J].科学技术与工程,2012,12(31):8131-8147.
- [13] 陈丽萍,武文波.基于决策树 C4.5 算法的面向对象分类方法研究[J].遥感信息,2013,28(2):116-120.
- [14] 苗红星,余建坤.基于决策树的 ID3 算法和 C4.5 算法的比较[J].研究与开发,2014,15:7-10.
- [15] 刘耀南.C4.5 算法的分析及应用[J].东莞理工学院学报,2012,19(5):47-51.
- [16] 郑宝芬,苏宏业,罗林.无监督特征选择在时间序列数据挖掘中的应用[J].仪器仪表学报,2014,35(4):834-840.

作者简介

丁兵,硕士研究生。主要研究方向为嵌入式系统设计。
E-mail: bush1988@126.com

吴允平,教授,研究生导师。主要研究方向为检测技术与自动化装置。
E-mail: wyp@fjnu.edu.cn