

# VSLAM 的研究与发展

吴家伟, 潘凤琳, 胡合斌

(中南民族大学 计算机科学学院, 武汉 430074)

**摘要:** 基于视觉的同步定位与地图构建(Vision based Simultaneous Localization and Mapping, VSLAM)是移动机器人研究非常热门的研究话题,为了进一步讨论 VSLAM 领域未来的发展趋势,本文回顾了 VSLAM 的发展历程,对 VSLAM 的整体框架进行描述,分别介绍了前端、后端的实现方法与目前成果,列举了当前主要的开源 VSLAM 算法,并从多个角度对 VSLAM 技术的研究与发展方向进行了探讨。

**关键词:** VSLAM; 移动机器人; 深度学习

**中图分类号:** TP242

**文献标识码:** A

## Research and Development of VSLAM

Wu Jiawei, Pan Fenglin, Hu Hebin

(School of Computer Science, South-central Minzu University, Wuhan 430074, China)

**Abstract:** In recent years, VSLAM has been a hugely popular research topic in the field of mobile robot. In order to further discuss the trend of VSLAM, we review its development and describe the overall framework of VSLAM. The implementation methods and current achievements of front-end, back-end and loop closure detection are introduced. Many open source VSLAM algorithms are listed in this paper, and we also discuss the research methods and development direction of VSLAM technology from many aspects.

**Key words:** VSLAM; mobile robot; deep learning

## 引言

同步定位与地图构建(Simultaneous Localization and Mapping, SLAM)已经成为这三十年来移动机器人领域的热门研究问题,它要求机器人在放置于未知环境时,能够在确定自身位置的同时构建出与环境相一致的地图<sup>[1-3]</sup>。且在无人驾驶、AR、无人机以及各类移动机器人领域占据着重要的地位,并被认为是实现机器人真正自主的关键<sup>[2]</sup>。

SLAM 最早出现于 1986 年,由 Smith、Self 和 Cheeseman 提出<sup>[4]</sup>,这是 SLAM 问题被引入机器人和 AI(Artificial Intelligence)领域的开始,也是很多研究者开始研究求解定位和建图问题的理论方法。由于相机相对于激光具有价格便宜、体型小巧和低功耗的特点,以及能够获取环境的外观、颜色、纹理等信息的优势,使得在近十年来将相机作为解决 SLAM 问题的主要传感器的趋势越来越明显<sup>[5]</sup>。

VSLAM 系统的架构与经典 SLAM 框架相同,主要由前端(视觉里程计)、后端(滤波、图优化)、回环检测和建图 4 个模块构成(见图 1)。前端和后端是基本部分,前端将相机数据抽象为便于估计的模型,它也被称为视觉里程计,后端对前端得到的模型进行优化,得到全局一致的轨

迹和地图。回环检测主要解决视觉里程计随着时间产生的累积误差问题,它主要使用的方式是词袋模型。而建图则是根据估计的轨迹来构建地图的过程,构建的地图大体上可分成度量地图和拓扑地图,其中,度量地图又可分成稀疏地图和稠密地图。

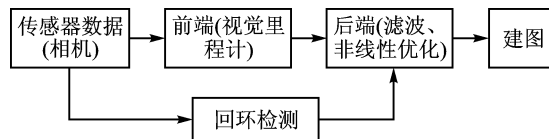


图 1 VSLAM 系统架构

根据传感器来划分,VSLAM 研究主要分为使用一个相机的单目 VSLAM、使用多个相机(主要为两个相机)的多目 VSLAM 和使用立体视觉相机的 RGB-DSLAM。本文对 VSLAM 框架中的主要算法进行了总结,见表 1 和图 2。

## 1 前端

前端也称为视觉里程计(Visual Odometry, VO),视觉里程计关心的是相邻图像之间的传感器运动,它根据传感器运动得到的图像信息估计得到对应传感器的运动方

表 1 VSLAM 主要算法

名称	年份	传感器	特点
MonoSLAM	2007	单目	首个实时单目视觉 SLAM 系统
PTAM	2007	单目	首次使用非线性优化, 引入关键帧机制
DTAM	2011	单目	单目稠密地图, 直接法
SVO	2014	单目	稀疏直接法
DSO	2016	单目	目前最好的单目直接法
LSD-SLAM	2014	单目	单目直接法成功应用, 半稠密地图
ORB-SLAM	2015	单目为主	ORB 特征, 创新使用三个线程完成 SLAM
ORB-SLAM2	2016	单目为主	ORB 特征, 对 ORB-SLAM 的改进
VINS	2017	单目	基于优化和滑动窗口的 VIO
DVO	2013	RGBD	RGBD 直接法, 稠密地图
RTAB-MAP	2013	双目/ RGB-D	完整、集成度高
RGBD-SLAM-V2	2014	RGBD	完整的 RGB-D 稠密建图
Elastic Fusion	2015	RGBD	RGBD 在线重建
ROVIO	2015	单目+IMU	基于 EKF 和 IMU 融合
OKVIS	2015	多目+IMU	紧耦合、基于非线性优化

位和轨迹。将图像的信息转化为矩阵, 通过求解矩阵来得到传感器的运动状态。目前, 视觉里程计的解决方案主要是基于特征的间接法和直接对像素进行处理的直接法。

### 1.1 基于特征的间接法

如上所述, 视觉里程计的主要问题是通过图像信息来估计传感器的运动。需要从图像中选取比较有代表性的点, 通过这些点在图像中的位置变化来估计传感器的运动。图像在计算中以矩阵的形式而存在。我们希望提取的特征能在传感器运动后依然保持稳定, 而单个像素值受外界条件影响易改变, 所以通常采用人工设计的特征提取器来获取具有鲁棒性的图像特征。

研究者们为此已经设计了许多能够提取具有足够鲁棒性特征提取算法。Davod Lowe 等人提出了 SIFT (Scale-invariant feature transform) 算法并于 2004 年将其完善, 它是一种用来检测和扫描图像中的局部性特征图像处理方式。MonoSLAM 是 Davison 等人<sup>[6]</sup>提出的第一个成功用于单目相机的纯视觉的实时定位与地图构建系统。MonoSLAM 的核心是在概率框架内在线创建稀疏但持久的自然地标(地图上选择有意义的部分)地图。Mur-Artal

等人<sup>[7]</sup>开源了 ORB-SLAM 算法, 该算法是目前基于稀疏特征较好的 VSLAM 算法, ORB-SLAM 的提出也对 VSLAM 发展起到了推动作用。在 ORB-SLAM 算法中, 所有 SLAM 任务使用相同的特性: 跟踪、映射、重新定位和闭环。该算法选择重建的点和关键帧, 具有良好的鲁棒性, 可生成紧凑的可跟踪地图, 只有在场景内容发生变化时才会增长, 可重复使用。

### 1.2 直接法

直接法相比于特征法有很大的不同, 直接法根据图像的像素信息来计算相机运动, 不需要花费很多时间在计算特征点上, 也不会因为只使用特征点丢弃大量可能有用的图像信息。直接法分为稀疏、稠密和半稠密三种, 都具有恢复稠密结果的能力。

目前, 已经提出了许多以普通相机为传感器的直接法 VSLAM 系统。Christian 等人提出一种 SVO (Semi-direct Visual Odometry) 算法, 该算法是一种稀疏直接法。它具有精确、鲁棒性强和速度快等特点, 并且不需要昂贵的特征提取和鲁棒匹配技术进行运动估计。Engel 等人<sup>[10]</sup>在论文中提出 DSO (Direct Sparse Odometry) 算法, 该算法是一种新颖、高精度的且基于稀疏直接结构和运动公式的视觉里程计测量方法。Kerl 等人<sup>[11]</sup>提出 DVO 算法, 该算法可以快速、准确地估计 RGB-D 图像的运动, 它首先通过图像流来估算深度相机的运动轨迹。然后, 用一种基于“关键帧”的位姿 SLAM 方法来消除稠密视觉里程计建模过程中产生的累计误差。这个方法在单个 CPU 内核上实时运行时, 具有较小、恒定的内存占用, 因此对于计算资源有限的机器人十分友好。

## 2 后端优化

VSLAM 后端的算法主要分为基于滤波(filter)的算法和基于优化(optimization-based)的算法。基于滤波的算法是通过(k-1)时刻的状态来预测 k 时刻的状态, 而基

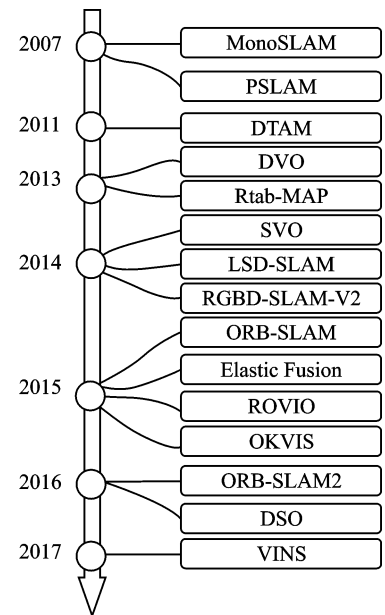


图 2 主要 VSLAM 算法的时间线

于优化的算法则是通过 $(k-1)$ 时刻之前所有的状态来预测机器人完整的轨迹和地图。

## 2.1 基于滤波的方法

基于滤波的方法源于贝叶斯滤波,主要分为预测和更新两个部分进行。预测是根据前一时刻的状态和输入信息来预测当前时刻的状态。而更新则是将当前观测到的传感器数据与地图进行匹配,对之前预测的状态进行校正。这两个步骤迭代并增量地集成传感器数据,以估计机器人的姿态和地图。由于基于滤波的方法只考虑前一时刻的状态以及它的增量特性,也被称为在线 SLAM(online SLAM)。

卡尔曼滤波(KF)是线性最优滤波,使用卡尔曼滤波必须满足两个假设:运动方程和观测方程都可以由线性方程来描述,并且所有的状态都满足高斯分布。所以,卡尔曼滤波只能被用在非线性系统。而由于在现实生活中,系统的运动方程和观测方程大部分都是非线性的,因此使用 KF 来解决很多现实问题会使误差变得很大,SLAM 的后端算法已经很少用到卡尔曼滤波。为此,Moutarlier 等<sup>[12]</sup>对 KF 进行改进使得可以应用于非线性系统中,扩展卡尔曼滤波(EKF)就被用来解决在非线性系统中的概率估计问题。它利用泰勒级数将非线性滤波问题转化为近似的线性滤波问题,然后再通过 KF 问题进行求解,但是扩展卡尔曼滤波 SLAM 对错误数据关联很敏感,因此一个错误的观测会导致滤波产生较大误差。此外,由于扩展卡尔曼滤波的复杂度是地图上地标数量的二倍,因此扩展卡尔曼滤波很难用于大比例尺地图。

粒子滤波是使用序贯蒙特卡罗方法(Sequential Monte Carlo methods, SMC)的滤波方法,它通过一组具有权重的随机样本(粒子)来表示随机时间的后验概率,因此粒子滤波器可以用在任何状态空间的模型上,不受非高斯分布和非线性的影响。Gristetti<sup>[13]</sup>等在粒子滤波的基础上又提出了一种 Rao-Blackwellized 粒子滤波器(RBPF),对于具有许多地标的地图而言,它具有良好的可扩展性。因此, RBPF 可以有效解决数据关联错误。对于粒子滤波算法来说,粒子的数量随着维度的增加而呈指数级增长,计算量会大大增加。所以它仅限于低维问题,对高维问题不适用。

卡尔曼滤波器和粒子滤波器是最常用的算法,这些方法在小空间内都相对成功,由于线性化和更新效率的原因,滤波器算法在大的未知环境中导航时受到限制。

## 2.2 基于图优化的方法

图优化方法是在 2004 年左右兴起的,该方法主要采用 Bundle Adjustment<sup>[14]</sup>来解决位姿变化问题。它将机器人的位姿和地标当作顶点,而用连接它们的边表示它们之

间的约束关系,然后将 Bundle Adjustment 应用于运动恢复结构(Structure from Motion, SfM)<sup>[15]</sup>中,将位姿估计问题转化为图的优化问题,即通过调整图中顶点的位置使得它们更好地满足边的约束关系<sup>[16]</sup>。Golfarelli 等人将图优化问题看作弹簧质量模型<sup>[17]</sup>,则对于后端的求解等价于计算该模型的最小能量。由于机器人运动中的所有观测和控制信息都是用来优化整个轨迹和地图构建的结果,所以这种方法也被称为 Full SLAM。

最近这些年,由于直接线性求解器的发展,图优化算法被广泛使用,常用的优化方法包括松弛法、梯度下降法、流动法等。目前的图优化算法包括 PATM、ORB-SLAM、LSD-SLAM、RTAB-MAP 等。图优化的算法结构如图 3 所示。

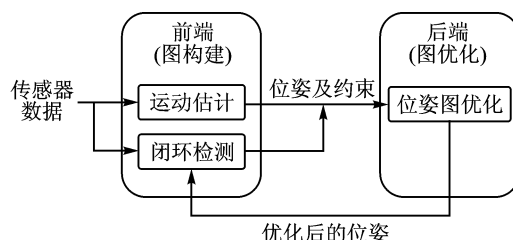


图 3 图优化的算法结构

## 3 趋势

### 3.1 基于深度学习的 VSLAM 算法

深度学习算法作为当前计算机视觉领域主流的一种识别算法,相比于其他已成熟的算法,有着更高的识别准确率<sup>[18]</sup>。通过深度神经网络能够退回到从运动机器人直接获得的两个原始图像的帧间姿态,可以有效代替视觉里程计测量的标准几何形状<sup>[23]</sup>。深度学习算法将图像与语义相结合,还可以构建语义地图<sup>[19]</sup>,提高机器人的交互能力。

目前已经有将深度学习与帧间估计相结合的算法。Konda 和 Memisevic<sup>[20]</sup>提出了基于端到端的深度神经网络架构,该架构用于预测相机速度和方向的改变。该算法能实现对连续帧的帧间估计,有着较快的执行速度,但在精度方面还有一些不足。Costante<sup>[21]</sup>等利用卷积神经网络学习图像数据的最优特征表示进行视觉里程计估计,该算法在应对图像运动模糊、光照变化方面有着较强的鲁棒性。但是该算法对训练数据依赖大,当图像序列帧间的速度过快会产生较大的误差。Handa<sup>[22]</sup>等在空间变换网络的基础上进行了扩展,利用神经网络构建了包含全局变换、像素变换和 M 估计器在内的 gvnv(geometric vision with neural network)软件库。

相比传统的帧间估计算法,基于深度学习的方法代替了繁琐的公式计算,无需人工进行特征提取和匹配,有着



直接和速度快等优点。同时,基于深度学习的 SLAM 算法还处于起步阶段,不同的神经网络结构设计差异较大,算法的适用性还有待深入研究。

### 3.2 多机器人 VSLAM

在机器人领域,多机器人的“协调(coordination)”与“合作(cooperation)”是应用迫切要求的结果也是技术发展的必然趋势。通过多机器人的协作,可以解决一些单机机器人不能完成的作业,或者用来简化单机机器人的作业量和复杂度,从而提高效率<sup>[24]</sup>。

多机器人 VSLAM 主要有两种主要的方向:一种是集中式,每个机器人建立自己当前位置的地图并将本地信息传送到执行推断的中心站;另一种是分布式,没有中央数据中心进行信息融合,它利用本地通信就基本达成一致。多机器人 VSLAM 算法目前还不能像单机机器人 VSLAM 算法一样成熟,面临着各个机器人的任务规划、通信带宽和距离的限制<sup>[25]</sup>、地图融合等问题。


### 3.3 主动 VSLAM

传统的 VSLAM 问题多依赖于相机所得到的数据来获得当前位置环境信息,并完成地图构建和对机器人自身位置的定位,一般都在有人为干预的情况下完成。机器人在未知环境中没有人有干预的情况下,能根据已经获得的信息自主推算控制规律及规划运动路径,并完成机器人自身的定位和地图的构建任务,即为主动式 VSLAM。

在主动式 VSLAM 的研究中,移动机器人需要根据探索到的环境特征与障碍物的情况,在环境作用下自主决定运动规律。Frintrop<sup>[26]</sup>等人将主动视觉引入到 VSLAM 中,并取得了较好的成果。Bosse<sup>[27]</sup>等人针对有障碍物的环境提出了基于势场法路径规划的 SLAM 方法,但是需要在环境中设置若干导引点,避免机器人陷入局部最小点陷阱。

## 结 语

基于视觉的 SLAM 技术在过去的三十多年里取得了巨大的进步,并且可以广泛应用于无人机、无人驾驶等平台。随着新型且功能更加丰富的传感器出现,SLAM 技术也将不断朝着高精度、高鲁棒性、多传感器融合的方向发展。

本文对 SLAM 的 4 个环节已有的方法进行了总结,对其视觉里程计的直接法和特征点法、后端优化的滤波法和图优化方法的主要算法的优缺点和特点进行了客观分析,同时对 VSLAM 技术的发展进行了展望。目前研究者已经在语义地图和主动 VSLAM 领域取得了一定的成果,这对于提高 VSLAM 的性能有着非常大的帮助。但是想要很好地将深度学习与 SLAM 相结合并达到理想的状态,还面临着极大的挑战。

## 参考文献

- [1] Bresson G, Alsayed Z, Yu L, et al. Simultaneous Localization and Mapping: A Survey of Current Trends in Autonomous Driving[J]. IEEE Transactions on Intelligent Vehicles, 2017, 2(3):194-220.
- [2] Durrant-Whyte H, Bailey T. Simultaneous localization and mapping: part I[J]. IEEE Robotics & Automation Magazine, 2006, 13(2):99-110.
- [3] Cadena C, Carlone L, Carrillo H, et al. Past, Present, and Future of Simultaneous Localization and Mapping: Towards the Robust-Perception Age[J]. IEEE Transactions on Robotics, 2016, 32(6):1309-1332.
- [4] Smith R, Self M, Cheeseman P. Estimating uncertain spatial relationships in robotics [M]. New York: Springer, 1990: 167-193.
- [5] Fuentes-Pacheco J, Ruiz-Ascencio J, Rendón-Mancha J M. Visual simultaneous localization and mapping: a survey[J]. Artificial Intelligence Review, 2015, 43(1):55-81.
- [6] Davison A J, Reid I D, Molton N D, et al. MonoSLAM: Real-time single camera SLAM[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2007(6):1052-1067.
- [7] Mur-Artal R, Montiel J M M, Tardós J D. ORB-SLAM: a versatile and accurate monocular SLAM system[J]. IEEE Trans on Robotics, 2015, 31(5):1147-1163.
- [8] Lowe D. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2):91-110.
- [9] Forster C, Pizzoli M, Scaramuzza D. SVO: fast semi-direct monocular visual odometry[C]//Proc of IEEE International Conference on Robotics and Automation, 2014:15-22.
- [10] Engel J, Koltun V, Cremers D. Direct sparse odometry[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016(99):1-14.
- [11] Kerl C, Sturm J, Cremers D. Robust odometry estimation for RGB-D cameras [C]//IEEE International Conference on Robotics & Automation, 2013.
- [12] Moutarlier P, Chatila R. An experimental system for incremental environment modelling by an autonomous mobile robot[C]//Experimental Robotics I, 1990.
- [13] Grisetti G, Stachniss C, Burgard W. Improved techniques for grid mapping with rao-blackwellized particle filters [J]. IEEE transactions on Robotics, 2007, 23(1):34.
- [14] Triggs B, McLauchlan P F, Hartley R I, et al. Bundle Adjustment-A Modern Synthesis[C]//International Workshop on Vision Algorithms: Theory & Practice, 1999.
- [15] Koenderink J J, Van Doorn A J. Affine structure from motion[J]. J. opt. soc. america, 1991, 8(2):377-385.
- [16] 梁明杰, 闵华清, 罗荣华. 基于图优化的同时定位

及升级程序量的影响,一个升级过程通常会持续几分钟至几十分钟,因此有效反映当前的升级境况显得尤其重要。本方案中的平台软件严格按照升级流程实时反映当前升级状态,当发生升级错误时,能正确展现错误信息,对升级任务的完成起到保障作用。图6展示了在平台软件上实现的升级步骤。

0123000003	杭州	大表NB-IoT-UDP 协议	查询软件版本 应答	2019-04-09 15:40:22	详情
0123000003	杭州	大表NB-IoT-UDP 协议	下发新版本通 知	2019-04-09 15:40:22	详情
0123000003	杭州	大表NB-IoT-UDP 协议	查询软件版本 应答	2019-04-09 15:40:25	详情
0123000003	杭州	大表NB-IoT-UDP 协议	下发新版本通 知	2019-04-09 15:40:25	详情
0123000003	杭州	大表NB-IoT-UDP 协议	新版本通知应 答	2019-04-09 15:40:31	详情
0123000003	杭州	大表NB-IoT-UDP 协议	请求分片0	2019-04-09 15:40:36	详情
0123000003	杭州	大表NB-IoT-UDP 协议	请求分片0应答	2019-04-09 15:40:36	详情
0123000003	杭州	大表NB-IoT-UDP 协议	请求分片1	2019-04-09 15:40:41	详情

图6 平台软件上传流程展现

当升级任务完成时,平台会提示用户升级完成,否则用户可以随时中止升级任务的进行或者直接清除任务。

## 结 语

本文从嵌入式终端设备维护困难的角度出发,描述了一种用于嵌入式终端设备的远程升级方案。在不同的应用场景下,当存在中心节点时,终端设备以TTL、RS485、RS232等有线方式,或者以LoRa、ZigBee等无线方式与中心节点进行数据传输。当进行终端节点的升级时,为了减

少传输环节所引起的升级中断,可以采用中心节点亦主亦从方式,即中心节点先作为从机,按照升级流程负责从远程服务器中接收终端的升级程序,并把它存储于中心节点特意开辟的存储区中,当正确接收完成后,中心节点作为主机,把升级程序按流程下发到指定的终端设备中。这个方案对中心节点的CPU存储空间提出了更高的要求。本设计方案包括远程服务器、终端或包含中心节点,主要从中心节点的程序架构阐述其工作流程。该方案具有较好的兼容性能,适用于使用NB-IoT、GPRS等方式组网的终端节点,具有很好的实用性能。

## 参考文献

- [1] 杭万里,张荣. 嵌入式设备远程升级方案设计[J]. 电子设计工程,2018,26(14):167-171.
- [2] 张博强,卜祥蕊. 嵌入式设备远程在线升级技术研究[J]. 中国新通信,2018,20(12):149.
- [3] 方兵兵. 基于STM32的嵌入式软件远程升级研究[D]. 宁波:宁波大学,2017.
- [4] 夏令伦. 基于电力集中器的远程升级技术与管理系统的研究[D]. 青岛:青岛大学,2016.
- [5] 周捷. 基于无线网络的LED显示屏远程升级[D]. 南京:南京大学,2015.

汪芳君(工程师),主要研究方向为无线远程通信与控制;童孝波(工程师),主要研究方向为嵌入式系统设计;周荷玲、毛德兴、封小刚(工程师),主要研究方向为NB-IoT无线远程通信。

(责任编辑:薛士然 收稿日期:2019-04-23)

## 7 与地图创建综述[J]. 机器人,2013,35(4):500-512.

- [17] Golfarelli M, Maio D, Rizzi S. Elastic correction of dead-reckoning errors in map building [C]//Proceedings. 1998 IEEE/RSJ International Conference on Intelligent Robots and Systems. Innovations in Theory, Practice and Applications (Cat. No. 98CH36190). IEEE,1998,2:905-911.
- [18] Gao X, Zhang T. Loop closure detection for visual SLAM systems using deep neural networks [C]//2015 34th Chinese Control Conference (CCC). IEEE,2015.
- [19] Fischer P, Dosovitskiy A, Brox T. Descriptor Matching with Convolutional Neural Networks; a Comparison to SIFT [J]. Computer Science,2014.
- [20] Konda K R, Memisevic R. Learning Visual Odometry with a Convolutional Network [C]//VISAPP(1),2015:486-490.
- [21] Costante G, Mancini M, Valigi P, et al. Exploring representation learning with cnns for frame-to-frame ego-motion estimation [J]. IEEE robotics and automation letters,2016,1(1):18-25.
- [22] Handa A, Bloesch M, Pă trăucean V, et al. gvn: Neural network library for geometric computer vision [C]//European Conference on Computer Vision. Springer, Cham,2016:67-82.

- [23] Costante G, Mancini M, Valigi P, et al. Exploring representation learning with cnns for frame-to-frame ego-motion estimation [J]. IEEE robotics and automation letters,2016,1(1):18-25.
- [24] Dahl TS, Matari c M, Sukhatme G S. Multi-robot task allocation through vacancy chain scheduling [J]. Robotics and Autonomous Systems,2009,57(6-7):674-687.
- [25] Ferranti E, Trigon N, Levene M. Brick and Mortar: An On-Line Multi-Agent Exploration Algorithm [J]. 2015.
- [26] Frintrop S, Jensfelt P, Christensen H. Detecting Useful Landmarks for Visual SLAM [J]. Robotic 3D Environment Cognition,2006:35.
- [27] Bosse M, Newman P, Leonard J, et al. Simultaneous localization and map building in large-scale cyclic environments using the Atlas framework [J]. The International Journal of Robotics Research,2004,23(12):1113-1139.

吴家伟,主要研究方向为机器人、人工智能;潘凤琳(本科),主要研究方向为人工智能;胡合斌(本科),主要研究方向为数据挖掘。

(责任编辑:薛士然 收稿日期:2019-03-11)