

控制与决策

Control and Decision



无人系统视觉SLAM技术发展现状简析

李云天, 穆荣军, 单永志

引用本文:

李云天, 穆荣军, 单永志. 无人系统视觉SLAM技术发展现状简析[J]. 控制与决策, 2021, 36(3): 513–522.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2019.1149>

您可能感兴趣的其他文章

Articles you may be interested in

基于凸面体圆弧航路的无人机自主避障算法

Autonomous obstacle avoidance algorithm designed for UAV based on convex circular trajectory

控制与决策. 2021, 36(3): 653–660 <https://doi.org/10.13195/j.kzyjc.2019.0768>

机器视觉在轨道交通系统状态检测中的应用综述

A survey of the application of machine vision in rail transit system inspection

控制与决策. 2021, 36(2): 257–282 <https://doi.org/10.13195/j.kzyjc.2020.1199>

数据驱动的综合能源系统运行优化方法研究综述

Review of research of data-driven methods on operational optimization of integrated energy systems

控制与决策. 2021, 36(2): 283–294 <https://doi.org/10.13195/j.kzyjc.2020.0725>

机器人抓取检测技术的研究现状

Recent researches on robot autonomous grasp technology

控制与决策. 2020, 35(12): 2817–2828 <https://doi.org/10.13195/j.kzyjc.2019.1145>

可持续逆向物流网络设计研究进展及趋势

Progress and prospects of sustainable reverse logistics network design

控制与决策. 2020, 35(11): 2561–2577 <https://doi.org/10.13195/j.kzyjc.2019.1175>

无人系统视觉SLAM技术发展现状简析

李云天¹, 穆荣军^{1†}, 单永志^{1,2}

(1. 哈尔滨工业大学 航天学院, 哈尔滨 150001;
2. 中国兵器工业集团有限公司 航空弹药研究院, 哈尔滨 150001)

摘要: 视觉SLAM (visual SLAM, vSLAM) 已成为无人系统在未知环境中实现全自主导航和环境感知不可或缺的重要组成部分。鉴于此, 简要介绍vSLAM的发展历程和典型构成, 总结两种前端位姿估计方法(特征法与直接法)和两种后端优化方法(非线性滤波与非线性优化)的典型代表及优缺点。在此基础上, 依次按照耦合方式和后端优化方法的不同对视觉/惯性SLAM的相关研究成果进行分类。进一步分析代表性开源vSLAM框架的异同, 给出公开数据集下的性能对比, 并分别从泛化能力、高级感知能力、动态适应性和多传感器集成化等方面对当前vSLAM所面临的挑战进行阐述, 探讨vSLAM未来的发展趋势和方向, 为研究者提供有益参考。

关键词: 无人系统; 自主导航; 环境感知; 状态估计

中图分类号: TP242; TP212

文献标志码: A

DOI: 10.13195/j.kzyjc.2019.1149

开放科学(资源服务)标识码(OSID):



引用格式: 李云天, 穆荣军, 单永志. 无人系统视觉SLAM技术发展现状简析[J]. 控制与决策, 2021, 36(3): 513-522.

A survey of visual SLAM in unmanned systems

LI Yun-tian¹, MU Rong-jun^{1†}, SHAN Yong-zhi^{1,2}

(1. School of Astronautics, Harbin Institute of Technology, Harbin 150001, China; 2. Aviation Ammunition Institute, NORINCO Group, Harbin 150001, China)

Abstract: Visual simultaneous localization and mapping (vSLAM) plays a vital role in fully autonomous navigation and perception of unmanned systems in unknown environment. Following a brief introduction to the history and typical structure of vSLAM, a summary about state-of-the-art representatives, advantages and disadvantages of two front-end pose estimation methods (feature based vs. photometric based) and two back-end optimization methods (nonlinear filter vs. nonlinear optimization) is given. On this basis, relevant achievements of visual inertial SLAM (VI-SLAM) are classified into different categories according to their coupled types and back-end optimization methods. Furthermore, the similarities and differences of representative open-source vSLAM frameworks are compared and analyzed, alongside with the comparison of their performance under public datasets. Finally, challenges faced by current vSLAM are elaborated from the aspects of scenario generalization, advanced perception, dynamic adaptability and multi-sensor integration. Future development trends and directions of vSLAM are also discussed to serve as a useful guide for researchers.

Keywords: unmanned systems; autonomous navigation; environment perception; state estimation

0 引言

21世纪以来,以微型航空器(micro aerial vehicle, MAV)和自动导航机器人(automated guided vehicle, AGV)为代表的无人系统呈现出自主化、小型化和智能化的发展趋势,应用场景也逐渐由物资投放、战场侦察、协同作战等军事领域向仓库管理、灾害救援、城市安保、资源勘探、电力巡线等民用领域扩展。这些新的场景普遍具有较高的动态性、未知性和封闭性,要求无人系统在缺乏环境先验信息和可靠外界辅助信息源(例如GPS、遥测系统等)的前提下,具有仅依靠自身传感器实现全自主导航定位和环境感知的

能力,为后续工作提供必要的信息支撑。

同步定位与建图技术(simultaneous localization and mapping, SLAM)正是解决上述问题的首选方案。SLAM技术发展至今,历经了经典阶段(1986~2004年)、算法分析阶段(2004~2015年)和鲁棒感知阶段(2015年~)^{[1]3}个重要阶段。经典阶段的研究重心在于解决SLAM中的状态估计问题,涌现出大批基于扩展卡尔曼滤波(extended Kalman filter, EKF)、无迹卡尔曼滤波(unscented Kalman filter, UKF)和粒子滤波(particle filter, PF)的位姿估计算法。算法分析阶段的研究主要集中于SLAM中的可观测性、收敛性、一致

收稿日期: 2019-08-11; 修回日期: 2019-12-15.

责任编辑: 薛建儒.

[†]通讯作者. E-mail: murjun@163.com.

性、稀疏性等问题. 鲁棒感知阶段的研究目标主要集中于以下3个方面:

1) 通过加入故障感知和参数自适应模块优化SLAM系统性能,提高SLAM系统适应性;

2) 由基本的几何建图向更高级的语义建图扩展,发展智能感知能力;

3) 发展能够根据任务目标自动过滤传感器信息、调配传感器资源的任务驱动能力.

根据所使用传感器的不同,SLAM主要分为激光SLAM和视觉SLAM (visual SLAM, vSLAM) 两大类. 激光SLAM起步最早,相关理论、技术和产品均已完善,目前与激光SLAM相关的研究内容主要集中于算法鲁棒性改进、激光/视觉/触觉融合SLAM等方面. 激光发生器在成本、功耗、体积、质量等方面的劣势大幅限制了激光SLAM在小型化无人系统上的应用,性价比更高的vSLAM方案逐渐成为小型无人系统自主导航与感知的主流.

vSLAM前端又称视觉里程计 (visual odometry, VO), 主要分为基于图像特征变化的特征法和基于图像灰度变化的光度法 (又称直接法) 两类. 前者基于特征不变性,通过最小化两帧图像之间的重投影误差实现位姿估计,对环境变化的鲁棒性更强;后者基于光度不变假设,通过最小化两帧图像中同一像素坐标处的灰度误差实现位姿估计,计算效率更高. vSLAM后端亦可分成非线性滤波和非线性优化两大流派. 非线性滤波方法基于Markov假设,在计算资源受限或较为简单的场景下具有良好的应用效果. 非线性优化则通过构建和优化位姿与路标点的联合误差函数获取全局最优估计,在同等计算规模下具有比非线性滤波方法更高的精度.

虽然vSLAM对计算资源的要求正在逐渐下降,精度也在逐渐提高,但数据更新慢(秒级)、高速运动下测量精度下降显著、难以处理动态障碍物等固有缺陷仍无法满足无人系统在动态环境下的强机动和快响应需求. 惯性测量单元 (inertial measurement unit, IMU) 具有数据更新率高、动态响应好、可分离角运动和线运动等优点;由于具有良好的优势互补性,将IMU与视觉传感器相结合构建惯性视觉SLAM系统 (visual inertial SLAM, VI-SLAM) 可有效提高vSLAM系统的测量带宽和抗干扰能力,增强动态环境下的鲁棒性,是当前vSLAM研究的重点方向^[2].

本文从前端位姿估计、后端全局优化、视觉惯性融合等方面出发,对vSLAM关键技术的发展现状予以梳理和总结. 在此基础上,重点分析代表性开

源vSLAM框架的异同,给出公开数据集下的性能对比. 同时,分别从泛化能力、高级感知能力、动态适应性和多传感器集成化等方面对vSLAM所面临的挑战进行阐述,探讨vSLAM未来的发展热点趋势和方向,为研究者提供有益参考.

1 vSLAM前端相关技术研究现状

1.1 特征法vSLAM

视觉特征由特征检测子 (detector) 和特征描述子 (descriptor) 两部分组成,前者主要负责检测图像中的关键点 (keypoint), 后者根据一定的法则生成关键点处的描述信息,用于后续的特征匹配. 最早的特征检测子多为角点、边缘、气泡等天然特征,这类特征对透视等图像变化的鲁棒性较差. 因此,近年来以加速分割测试特征 (features from accelerated segment test, FAST)^[3]、尺度不变特征变换 (scale invariant feature transform, SIFT)^[4]、加速鲁棒特征 (speed-up robust feature, SURF)^[5] 为代表的人工特征逐渐成为主流. FAST是一种基于SUSAN模型的改进角点检测器,具有较快的特征检测速度但对环境噪声较为敏感. 为解决这一问题,提出了一种基于图像高斯差 (difference of gaussian, DoG) 的特征检测子——SIFT, SIFT对光照、旋转、尺度、视角等变化均不敏感,具有较强的鲁棒性,但随之而来的是较大的计算负担,文献[5]在此基础上使用盒式滤波器和图像积分实现对海森矩阵的近似,提出了可显著降低计算量但在鲁棒性等方面稍逊一筹的SURF算法.

早期的特征描述子多基于梯度方向直方图 (histogram of oriented gradient, HOG) 实现,需要耗费较多的计算资源,难以实时应用. Calonder等^[6]提出的二进制鲁棒独立元素特征描述子 (binary robust independent elementary features, BRIEF) 通过在特征点附近随机选取若干点对,基于点对灰度值构建二进制字符串作为描述子,极大地加速了特征匹配过程,但同时使得描述子对环境噪声的鲁棒性有所下降. 针对这一问题,Leutenegger等^[7]提出了另一种旋转和尺度不变性更好、噪声鲁棒性更强的BRIEF改进型——二进制鲁棒不变尺度关键点 (binary robust invariant scale keypoint, BRISK)^[7]; Alahi等^[8]则在BRISK的基础上,参考人类视网膜的拓扑结构,提出了进一步优化的快速视网膜关键点 (fastretina keypoint, FREAK) 描述子及相应的检测匹配算法.

Rublee等^[9]提出的ORB特征标志着特征法的发展迎来了一个巅峰时期. ORB特征继承了FAST特征快速检测的优点,同时采用图像金字塔算法保

证尺度不变性,采用灰度质心提供方向描述保证旋转不变性,这使得ORB特征无论是在平移、旋转还是缩放等情况下均具有良好的检测精度与匹配效率.基于ORB特征设计的ORB-SLAM^[10]已成为特征法SLAM设计标准参考框架,第1代ORB-SLAM只适用于单目相机,2016年面世的ORB-SLAM2^[11]则进一步支持双目和RGB-D.

1.2 直接法vSLAM

虽然特征法目前占据着vSLAM前端设计的主流地位,但仍具有耗时长、图像信息丢失严重、对环境

特征要求高等缺陷.由于许多无人系统应用场景的光度是近似不变的,近年来采用图像光度代替图像特征的直接法日益受到学界重视.由于完全不需要计算描述子,直接法既避免了较长时间的特征识别与匹配过程,又保留了大量图像信息.此外,直接法对场景的特征信息没有要求,只需要场景中存在明暗变化即可.相比于特征法只能基于特征点构建稀疏地图,直接法则可根据需求,自由选择构建稀疏、半稠密或稠密地图,更加灵活.表1给出了特征法与直接法之间的优缺点对比.

表1 特征法与直接法优缺点比较

| | 特征法 | 直接法 |
|----|--|---|
| 优势 | 1) 不需要额外进行特征匹配以实现回环检测和重定位; 2) 对环境光照与相机曝光变化的适应性强; 3) 可实现较大尺度的帧间运动估计 | 1) 不需计算描述子和匹配特征,稀疏工况下位姿估计效率更高 2) 对特征不敏感,可在纹理不显著区域工作 3) 可自由构建稀疏、半稠密或稠密地图 |
| 缺陷 | 1) 特征匹配耗时较多,位姿估计更新率低; 2) 需要对误匹配特征进行筛选; 3) 特征点数目有限,只能构建稀疏地图 | 1) 单个像素缺乏区分度,难以实现有效的回环检测; 2) 对场景变化敏感,只适用于小尺度帧间运动估计 3) 对图像光度变化敏感,需要标校相机曝光参数且在室外场景下应用受限 |

Newcombe等^[12]于2011年提出了首个完整的直接法SLAM框架——稠密追踪与建图(dense tracking and mapping, DTAM).虽然DTAM基于深度相机实现,避免了在深度估计上耗费大量的计算资源,但DTAM本身是一种稠密直接法,依旧需要高性能的图形处理单元(graphic processing unit, GPU)辅助位姿估计与地图重建,目前只在手机AR等实时性要求不高的应用中取得了一定成果.

Engel等^[13]于2014年提出的大尺度直接SLAM (large-scale direct SLAM, LSD-SLAM)是半稠密直接法的典型代表,亦是目前最成功的直接法开源SLAM框架.LSD-SLAM利用图像梯度近似光流变化,通过忽略图像中灰度变化不明显或深度信息难以估计区域的像素信息,有效降低了计算负担,使其不再依赖GPU,直接使用CPU即可完成半稠密地图构建,较大程度地降低了对硬件性能的需求.

在LSD-SLAM的基础上,Engel等^[14]于2017年又提出了稀疏版本的直接法SLAM——直接稀疏里程计(direct sparse odometry, DSO).DSO基于图像中灰度强度变化大的像素点进行位姿估计与地图构建,并通过几何与光度透视关系对累积误差进行校正.虽然有效地提升了直接法的精度和计算效率,但在地图构建和利用方面的效果仍然不够理想,因此更多被视为一种VO,而非完整的vSLAM系统.稀疏直接法的另一个代表是Forster等^[15]于2014年提

出的半直接视觉里程计(semidirect visual odometry, SVO).SVO首先基于FAST角点检测器探测图像中的关键点,随后通过跟踪关键点周围4×4像素范围内的像素点进行相机位姿的估计.SVO最大的优势在于轻量化和实时性,SVO 1.0和2.0^[16]的数据更新率分别达到了每秒100帧和每秒400帧.原始的SVO并不具有后端优化和回环检测功能,同样只能作为VO使用.Li等^[17]则为SVO设计了基于关键帧的后端优化和回环检测环节,进一步扩展了SVO的适用范围,使其成为一种完整的vSLAM系统.

2 vSLAM后端相关技术研究现状

2.1 基于非线性滤波的vSLAM

vSLAM的早期研究者们大多将位姿估计视作非线性最大后验估计问题,作为无偏贝叶斯估计的典型代表,以EKF为代表的非线性卡尔曼滤波器(Kalman filter, KF)首先被应用于此方面.

Davison等^[18]于2007年提出了世界上第1个可实时应用的vSLAM系统——Mono-SLAM. Mono-SLAM后端基于EKF实现,但由于vSLAM的系统状态中特征点的数目远大于位姿数目,系统状态维数会随时间显著增加,使得Mono-SLAM只适用于稀疏特征点环境.Howard^[19]和Thrun等^[20]则分别研究了PKF在SLAM中的应用,后者更是提出了著名的Fast-SLAM算法.Fast-SLAM采用Rao-Blackwellized粒子滤波器进行位姿估计,而采用另一个EKF滤波器进

行路标点位置估计,该设计在降低计算复杂度的同时具有较好的鲁棒性. Ali等^[21]在此基础上进一步开发了Fast-SLAM 2.0算法,有效延缓了Fast-SLAM中的粒子退化问题.

2.2 基于非线性优化的vSLAM

Markov假设使得滤波vSLAM算法在计算资源受限或较为简单的场景下具有良好的应用效果,但也为累积误差的消除带来了不利影响. 此外,基于滤波的vSLAM算法需要存储状态量的均值和协方差,而vSLAM的状态量中包含大量路标点信息,这使得滤波算法占用的存储空间急剧增加,计算效率大幅下降,无法满足实时应用的需求. 基于非线性优化的vSLAM算法通过优化所有位姿和路标点的联合误差函数获取所有位姿和路标点的全局最优估计,由于可将当前和更久之前的状态同时进行优化,在同等计算量下具有比滤波算法更高的精度,非线性优化算法逐渐成为vSLAM后端处理的主流. 同时,vSLAM后端的非线性优化问题和机器视觉中的光束平差法(bundle adjustment, BA)十分类似,学界也常用BA代指.

同步追踪与建图算法(parallel tracking and mapping, PTAM)是非线性优化在vSLAM中的首项成功应用. 该算法由Klein等^[22]于2007年提出并开源. PTAM具有跟踪与建图两个并行线程:跟踪线程实时估计相机位姿,建图线程则在后台另行优化全局位姿并构建地图,两者仅在必要时进行同步. PTAM同时引入关键帧机制,减少了需要优化的图像数量,确保了算法的实时性. 尽管PTAM具有不适用于大场景、跟踪容易丢失等缺陷,但其具有的双线程架构思想仍成为当前众多vSLAM开源框架的设计基础.

此外,Di等^[23]研究了可同时利用2D和3D量测信息、适用于RGB-D相机的扩展BA算法;Alismail等^[24]舍弃了采用最小化重投影误差的传统BA框架,取而代之以基于最大化光度连续性的光度BA算法(photometric BA, PBA),精度较传统BA算法取得了显著提升;Liu等^[25]提出的ICE-BA算法(incremental, consistent and efficient BA)采用了增量技术,使该算法在保证同等精度的前提下,仅消耗传统BA 1/10左右的计算资源.

此后,研究者们逐渐注意到BA问题具有天然的稀疏性(即Jacobian矩阵中只有当前待优化变量的对应位置处为非零项,其余位置均为零),进而可基于Schur消元将无关变量边缘化,实现BA中线程方程组的稀疏求解^[26].

虽然稀疏性处理在一定程度上缓解了非线性优化的计算负担,但随着运行时间增加,越来越庞大的路标点数量仍给实时应用带来了困难. 由于路标点的空间位置是近似不变的,有研究者尝试在BA中舍弃对路标点的优化,构建只包含位姿的简化图模型,即位姿图(pose graph, PG). 这一模型将路标点位置作为未知参量而非待优化变量,大幅缩减了优化规模. Deutsch等^[27]研究了利用多局部位姿图进行全局位姿图融合,再基于全局位姿图改善多机器人定位与建图精度,从而实现多机器人协同SLAM的方法;Lee等^[28]设计了一种基于RGB-D传感器和位姿图优化的SLAM系统,以解决动态环境下对场景目标的辨识问题;Mu等^[29]提出了一种新的非参数化位姿图,可同时实现对场景物体的识别和定位,在数据关联与定位两方面均表现出良好的性能.

另一个缓解优化规模随时间快速增长的途径为具有增量特性的因子图优化(factor graph optimization, FGO)算法. 因子图这一概念1998年即已问世^[30],是一种求解最大后验概率的概率图模型算法,但直到2008年才首次被Kaess等应用于vSLAM中BA问题的求解;该团队先后开发出两代因子图优化算法iSAM^[31]和iSAM 2^[32],并于2012年发布了首个开源因子图算法库GTSAM^[33]. 因子图优化的最大优势在于增量特性:在新的变量因子加入图模型前,先筛选出与其关联的因子;随后,只对受新变量因子影响的关联因子进行更新,而保持其余因子不变. 虽然目前因子图由于实现困难较大,开源项目较少等原因尚未普及,但已有越来越多的团队投入到了基于因子图优化的vSLAM后端研究中,相信在不远的将来,会有更多基于因子图优化的成熟vSLAM系统面世.

3 视觉/惯性SLAM技术研究现状

根据是否对相机和IMU状态进行联合建模,VI-SLAM可分为松耦合和紧耦合两种. 松耦合中,相机和IMU首先分别对位姿进行估计,随后融合获得当前最优位姿估计. Späenlehauser等^[34]设计了一种基于Kalman滤波的松耦合单目相机/IMU SLAM系统,利用IMU定位数据辅助确定图像特征点的深度信息,有效改善了单目SLAM的尺度不确定性;Sirtkaya等^[35]研究了利用Kalman滤波融合IMU与立体相机的测量信息,同时对惯性器件多种误差参数进行在线标定的方法;Haachih等^[36]研究了利用误差状态卡尔曼滤波(error state Kalman filter, ESKF)取代传统EKF进行松耦合VI-SLAM设计的方法,基于Odroid XU4

嵌入式开发板进行了原型验证。

与松耦合相比,紧耦合模式具有更高的精度和鲁棒性,已成为VI-SLAM的主流结构。在后端处理方面,与松耦合大多基于滤波算法不同,紧耦合目前处在滤波和优化并存阶段。作为最经典的滤波VI-SLAM系统,Mourikis等^[37]在2007年提出的多状态约束Kalman滤波(multistate constrained Kalman filter, MSCKF)虽已面世十年有余,但仍有诸多研究者专注于其相关改进^[38-39]。除MSCKF外,Huang等^[40-41]研究了SLAM中的滤波一致性问题,提出了两种保证EKF线性化后具有适当维度可观测子空间的算法:可观测性约束EKF(observability constrained EKF, OC-EKF)和首次估计雅克比矩阵EKF(first estimates Jacobian EKF, FEJ-EKF);Bloesch等^[42]通过将图像光度误差作为EKF新息项、将路标点分解为方向向量与逆深度等方法,降低了模型非线性的影响、提高了系统鲁棒性,同时提出了鲁棒VIO算法(Robust VIO, ROVIO),基于迭代EKF滤波器对ROVIO进行了改进,并将适用对象由单目相机扩展到多目相机^[43];Schneider等^[44]在ROVIO的基础上,进一步采用BRISK和FREAK特征替换FAST角点,开发了一种更为成熟的开源滤波紧耦合VI-SLAM框架Maplab。

基于非线性优化的紧耦合VI-SLAM的典型代表是Leutenegger等^[45]提出的OKVIS。OKVIS采用滑动窗口方法约束优化的规模,窗口内包含一定数量的关键帧和路标点,窗口大小则由当前有效关键帧数所决定。每当有新的图像帧进入窗口时,较旧的关键帧或较新的冗余图像帧会被边缘化以保证窗口规模满足实时性要求。OKVIS还首次采用IMU预积分方法解决了IMU和相机量测不同步的问题。

在OKVIS框架下,Qin等^[46]提出了VINS-Mono算法,并先后发布了专用于手机增强现实(augmented reality, AR)的VINS-Mobile及适用于更多视觉传感器的VINS-Fusion。这3种算法在总体架构上基本一致,主要分为以下4部分:

- 1) 图像预处理:利用光流法跟踪序列图像特征点,同时对IMU的量测数据进行预积分;
- 2) 参数初始化:将相机估计位姿与IMU预积分位姿对比,完成速度矢量、重力矢量和尺度因子的初始化以及陀螺零偏的标定;
- 3) 后端BA优化:基于滑动窗口算法对较远的关键帧进行边缘化,保证系统估计实时性;
- 4) 回环检测:基于DBow2词袋库实现回环检测,在位置与航向4个维度上优化位姿图。

面对VI-SLAM这一新浪潮,传统vSLAM的领

军者们自然也不甘示弱,先后推出了各自算法的VI版本。Forster等^[16]于2016年发布的SVO 2.0除进一步提升处理速率外,同样基于预积分技术处理IMU量测,后端则采用iSAM算法实现BA优化;Mur-Artal等^[47]构建的惯性视觉ORB SLAM (visual inertial ORB-SLAM, VI-ORB SLAM)进一步完善了惯性视觉初始化模块,将位置和加速度计零偏也归入估计矢量。虽然受限于MEMS加速度计噪声特性,所需估计时间较长(一般为十几秒),不利于实时应用,但其完善的回环、建图和重定位能力仍不失为当前VI-SLAM研究的典范。

4 开源vSLAM框架对比分析

已面世的vSLAM方案虽多达上百种,但大多数都可归纳为在不同开源vSLAM框架下,针对不同应用需求所进行的算法细节改进。表2给出了目前主流开源vSLAM框架在前端类型、初始化方式、回环检测等方面的异同。可以看出,微小型航空器(micro aerial vehicle, MAV)依旧是目前vSLAM中应用最为广泛的无人平台。在前端算法中,出于平衡计算效率和精度的考虑,特征法多采用角点特征,并在某些特定场景下辅以线、面等特征;直接法和半直接法则统一采用了像素梯度。

特征法大多均基于本质矩阵分解(essential matrix decomposition, EMD)初始化位姿;半直接法的SVO由于应用场景为MAV下视相机,基于平面假设采用了更为简化的单应矩阵分解(homogeneous matrix decomposition, HMD);ORB-SLAM2出于系统鲁棒性考虑,同时采用EMD和HMD获取位姿初值,再选择两者中误差最小的作为初始化位姿;LSD-SLAM、DSO等直接法由于并不需要基于视图几何求解位姿,均采用为第1帧视图中所有空间点赋予随机深度的方式进行初始化。

无论是在纯视觉SLAM还是VI-SLAM中,局部BA(local BA, LBA)和位姿图优化(pose graph optimization, PGO)等非线性优化方法仍是后端优化的主流,非线性滤波则主要应用于特征法VI-SLAM中。此外,纯视觉版本的SVO为了提高实时性,使用了固定位姿,只优化空间点的单结构BA(struct only BA, SoBA),VI-SVO则进一步应用PGO优化,是目前为数不多的开源因子图vSLAM框架。

现阶段大部分vSLAM的目的在于实现无人系统的自主导航与定位,对于地图的需求并不高,因此绝大部分vSLAM都采用了几何度量地图(metric map)和拓扑地图(topological map)。半直接法的SVO

表 2 代表性开源 vSLAM 框架特点总结

| 名称 | 目标无人平台 | 前端类型 | 前端特征 | 初始化方式 | 后端策略 | 地图类型 | 路标点估计 | 故障位姿恢复 | 回环检测 | 场景动态 |
|---------------------------|--------|------|------------|---------|----------|-------|--------|---------|--------|------|
| ORB-SLAM2 | 未注明 | 特征法 | FAST | HMD+EMD | LBA&PGO | 几何+拓扑 | 帧间三角化 | 词袋识别 | DBow2 | 静态 |
| SVO | MAV | 半直接法 | 像素梯度 | HMD | LBA&SoBA | 几何 | 深度滤波器 | 对齐最近关键帧 | 无 | 静态 |
| DSO | 未注明 | 直接法 | 像素梯度 | 随机深度 | LBA | 几何 | 深度地图优化 | 对齐最近帧 | 无 | 静态 |
| LSD-SLAM | 未注明 | 直接法 | 像素梯度 | 随机深度 | PGO | 几何 | 深度地图递推 | 对齐随机关键帧 | FabMap | 静态 |
| S-MSCKF | MAV | 特征法 | Harris | EMD | EKF | 几何 | 帧间三角化 | 无 | 无 | 静态 |
| StructVIO ^[48] | MAV | 特征法 | Harris+结构线 | EMD | EKF | 几何 | 帧间三角化 | 无 | 无 | 静态 |
| ROVIO | 未注明 | 特征法 | Harris | EMD | EKF | 几何 | 帧间三角化 | 无 | 无 | 静态 |
| OKVIS | 无人车 | 特征法 | Harris | EMD | PGO | 几何 | 帧间三角化 | 无 | 无 | 静态 |
| VINS | MAV | 特征法 | Harris | EMD | PGO | 几何 | 帧间三角化 | 对齐最近关键帧 | DBow2 | 静态 |
| VI-ORB | MAV | 特征法 | FAST | HMD+EMD | LBA&PGO | 几何+拓扑 | 帧间三角化 | 词袋识别 | DBow2 | 静态 |
| VI-SVO | MAV | 半直接法 | 像素梯度 | EMD | LBA&FGO | 几何 | 深度滤波器 | 对齐最近关键帧 | 无 | 静态 |
| VI-DSO ^[49] | 未注明 | 直接法 | 像素梯度 | 随机深度 | LBA | 几何 | 深度地图优化 | 对齐最近帧 | 无 | 静态 |

首创了深度滤波器用于路标点位置估计. 直接法中, LSD-SLAM 基于当前帧位姿对上一帧的已知路标点位置进行递推; DSO 则在每次关键帧到来时向地图中添加新的路标点, 随后利用关键帧之间的全部图像帧对路标点位置进行优化. 特征法对路标点的估计基本都基于视图几何帧间三角化实现.

ORB-SLAM2 和 VI-ORB 基于开源词袋库 DBow2 设计了故障位姿恢复和回环检测环节; VINS 在这一思想的基础上, 采用更为简单的对齐最近关键帧策略实现故障情况下的位姿恢复; LSD-SLAM 在采用另一种开源词袋库 FabMap 的同时, 直接对齐全局地图中的随机关键帧以进行位姿恢复; SVO、DSO、VI-SVO 和 VI-DSO 由于只是里程计版本的 vSLAM, 并未加入回环检测环节, 只分别采用了对齐最近帧

或对齐最近关键帧策略以实现系统故障时的快速位姿恢复. 同时值得注意的是, 几乎所有当前的开源 vSLAM 都只支持静态环境.

表 3 (单位: m) 和表 4 (单位: m) 进一步给出了上述开源 vSLAM 框架在 EuRoC MAV 公开数据集下的均方根误差 (root mean squared error, RMSE) 对比 (表中粗体表示最小误差, x 表示误差过大算法不可用). 该数据集共有 3 类场景, 11 组数据序列; 数据序列标号越大, 场景的动态性越强, 更具有挑战性. 总体而言, 相比 VI-SLAM, 纯视觉 SLAM 在强动态环境下的鲁棒性较差, 各算法均出现了不同程度的丢失, 其中以 LSD-SLAM 最为明显; 相反, 较为平稳的数据序列 (MH_01_easy 和 MH_02_easy) 中, ORB-SLAM2、DSO 等纯视觉 SLAM 具有更高的位姿估计精度, 而

表 3 EuRoC 下 VI-SLAM 开源框架 RMSE 对比

| 数据集编号 | 特征法 | | | | | | 直接法 | |
|--------|-------|-----------|-------|-------|-------|--------------|--------------|--------|
| | 非线性滤波 | | | 非线性优化 | | | VI-DSO | VI-SVO |
| | MSCKF | StructVIO | ROVIO | OKVIS | VINS | VI-ORB | | |
| MH_01 | 0.227 | 0.079 | 0.236 | 0.321 | 0.113 | 0.075 | 0.062 | 0.100 |
| MH_02 | 0.231 | 0.145 | 0.247 | 0.318 | 0.126 | 0.084 | 0.054 | 0.120 |
| MH_03 | 0.201 | 0.103 | 0.427 | 0.272 | 0.124 | 0.087 | 0.117 | 0.410 |
| MH_04 | 0.351 | 0.130 | 1.170 | 0.406 | 0.220 | 0.217 | 0.132 | 0.430 |
| MH_05 | 0.213 | 0.182 | 0.863 | 0.447 | 0.200 | 0.082 | 0.121 | 0.300 |
| VR1_01 | 0.062 | 0.060 | 0.216 | 0.140 | 0.068 | 0.027 | 0.059 | 0.070 |
| VR1_02 | 0.161 | 0.130 | 0.210 | 0.159 | 0.076 | 0.028 | 0.067 | 0.210 |
| VR1_03 | 0.281 | 0.090 | 0.381 | 0.224 | 0.154 | 0.074 | 0.096 | 0.200 |
| VR2_01 | 0.074 | 0.045 | 0.298 | 0.143 | 0.073 | 0.032 | 0.040 | 0.110 |
| VR2_02 | 0.152 | 0.066 | 0.232 | 0.196 | 0.141 | 0.041 | 0.062 | 0.110 |
| VR2_03 | 0.366 | 0.110 | 0.263 | 0.280 | 0.109 | 0.074 | 0.174 | 1.080 |

表4 EuRoC下纯视觉SLAM开源框架RMSE对比

| 数据集编号 | 特征法 | | 直接法 | | |
|------------------|--------------|--------------|-------|-------|-------|
| | ORB-SLAM2 | | SVO | DSO | LSD |
| MH_01_easy | 0.036 | 0.054 | 0.179 | 0.050 | 0.180 |
| MH_02_easy | 0.019 | 0.056 | 0.270 | 0.050 | 0.560 |
| MH_03_medium | 0.024 | 0.033 | 0.420 | 0.260 | 2.690 |
| MH_04_difficult | 0.184 | 0.072 | 1.000 | 0.240 | 2.130 |
| MH_05_difficult | 0.176 | 0.053 | 0.600 | 0.150 | 0.850 |
| VR1_01_easy | 0.037 | 0.047 | 0.220 | 0.470 | 1.240 |
| VR1_02_medium | 0.028 | 0.043 | 0.350 | 0.100 | 1.110 |
| VR1_03_difficult | x | 0.187 | x | x | x |
| VR2_01_easy | 0.032 | 0.038 | 0.260 | 0.050 | x |
| VR2_02_medium | 0.034 | x | 0.400 | 0.190 | x |
| VR2_03_difficult | x | x | x | x | x |

VI-SLAM受IMU量测噪声的影响稍逊一筹。同样,以VI-DSO为代表的直接法亦在较为平稳的数据序列中具有更高的估计精度,其他数据序列中则不如以ORB-SLAM2和VI-ORB为代表的特征法vSLAM鲁棒性强。同时,由于EuRoC数据集普遍具有较多的回环运动,ORB-SLAM、VI-ORB、VINS等具有回环检测的完整vSLAM系统位姿估计精度亦比OKVIS、StructVIO等里程计版本vSLAM更高。

5 vSLAM面临的挑战与机遇

5.1 动态适应性

vSLAM中的动态适应性包括两方面:一是在自身高速运动或大姿态机动下依旧能实现准确的位姿估计,即自身动态适应性;二是能够在对环境中动态物体进行识别、跟踪和重建的同时实现自身的位姿估计,即环境动态适应性。现有vSLAM多基于固定帧率曝光的传统相机设计,这类传感器对环境光照度较为敏感且大运动下极易出现成像模糊,自身动态适应性较差。同时,现有vSLAM多基于静态世界假设,而真实世界往往是动态多变的,这在很大程度上降低了vSLAM的动态环境适应性。

以动态主动像素视觉传感器(dynamic active-pixel vision image sensor, DAVIS)和异步时序图像传感器(asynchronous time-based image sensor, ATIS)为代表的事件相机相比传统相机具有更小的时间延迟(毫秒级)、更快的更新速率(最大1 MHz)、更强的动态范围(140 dB)、更低的功耗(20 mW)和更小的存储空间要求^[1],是解决这一问题的有效途径,陆续有基于事件相机的特征检测^[50]、惯性/视觉里程计^[51]和光度建图^[52]等方面的研究成果面世。此外,事件相机完全依赖于场景动态,低动态环境下性能反而不如传统相机,因此亦有尝试融合传统相机和事件相机以实现全动态vSLAM的相关研究出现^[53]。

5.2 泛化能力

一个完善的vSLAM系统应具有良好的场景和平台泛化能力,而现有vSLAM每次更换应用场景和应用平台后,小到各项参数,大到系统结构,都需要进行大量的调整,这无疑会增加大量的工作量,阻碍了vSLAM的快速产品化。随着深度学习在计算机视觉等领域的广泛应用,作为一种可从大量数据中自主学习场景有效表示方式的主流人工智能方法,使用其替换vSLAM中的若干模块(如特征提取、深度估计、帧间运动估计、回环检测等)可有效增强vSLAM在不同场景和平台下的泛化能力。

Xiang等^[54]基于堆积去噪自动编码器(stacked denoising auto-encoder, SDA)设计了一种无监督学习的回环检测方法,检测精度与传统方法持平;Li等^[55]设计了一种可同时估计六自由度位姿和空间深度的无监督学习网络UnDeep,在KITTI数据集下取得了令人满意的位姿估计精度;Bloesch等^[56]提出了一种采用自动编码器训练的空间点深度估计方法,并以此为基础开发了著名的CodeSLAM。

5.3 高级感知能力

现有vSLAM系统大多基于简单的环境几何属性构建度量地图或拓扑地图,这一类地图在浪费存储空间、缺失环境细节的同时,也丢失了环境中的大量有效信息,使得无人系统在导航定位以外,难以实现类似人类的、更高级的环境感知。如何在提高vSLAM自主定位精度和鲁棒性的同时,完成路径规划向任务规划的升级,实现更高级的人机交互和环境感知是vSLAM未来研究的另一个热点,语义VSLAM应运而生。

语义vSLAM的相关研究成果大致分为vSLAM辅助语义和语义辅助vSLAM两大类,前者又称为语义建图,大多基于卷积神经网络(convolutional neural networks, CNNs)实现。Tateno等^[57]研究了基于CNNs的单目相机深度估计方法,并在此基础上采用随机梯度下降法对场景目标进行语义分割;McCormac等^[58]基于CNNs开发了能融合像素级稠密度量地图和2D语义地图,进而构建3D语义地图的SemanticFusion,有效增强了动态环境下的目标识别能力;Ma等^[59]研究了基于FuseNet-CNN的RGB-D语义建图方法,通过应用多视角连续检验优化方法,大幅缓解了单视角下的检测不确定性。

相较于如火如荼的语义建图,语义辅助vSLAM的相关研究仍处于起步阶段。Civera等^[60]首先尝试基于目标识别优化的单目语义SLAM方法;Bowman

等^[61]基于最大期望算法(expectation maximization, EM)研究了语义信息和特征点位置之间的数据关联方法,进而将语义信息归入重投影误差函数,是目前语义辅助vSLAM方面最具有代表性的成果;在此基础上,Lianos等^[62]提出了一种引入语义信息作为中期约束的视觉里程计,进一步减少了vSLAM长期定位中的累积平移误差;Yu等^[63]将场景语义分割结果与运动一致性检测相结合,提高了vSLAM在动态环境下的鲁棒性。

5.4 多传感器集成化

纯视觉vSLAM在精度、鲁棒性等方面已逐渐难以满足无人系统高动态、强鲁棒的需求,以VI-SLAM为代表的多传感器融合vSLAM成为vSLAM研究的一个重要分支。目前多传感器vSLAM的硬件配置通常是松散的,缺乏有效的集成。这在阻碍vSLAM系统小型化的同时,使得各传感器的量测信息之间缺乏有效的同步和空间同步,进而引起较大的位姿估计偏差,降低系统的可靠性。因此,研究高效可靠的多传感器集成方法,开发成熟的硬件产品,对于vSLAM的大规模普及具有重要的意义。

6 结论

作为无人系统自主导航与感知的重要组成部分,vSLAM技术目前正处于蓬勃发展时期。基于特征的前端帧间位姿估计依旧是当前vSLAM的主流,但DSO、LSD-SLAM等算法的面世亦标志着直接法开始由理论走向实际,以SVO为代表的、融合二者优势的半直接法受到学界的广泛关注。尽管因子图优化算法目前尚无法撼动位姿图在后端全局优化中的统治地位,但随着相关算法和开源代码的普及,未来会有更多基于因子图优化的成熟开源SLAM系统面世。由于视觉传感器自身特性受限,紧耦合VI-SLAM已成为vSLAM中不可或缺的重要组成部分。可以预见未来的开源SLAM框架都将在发布最初的单目纯视觉版本后,进一步开发多目和VI版本。与纯视觉vSLAM不同,VI-SLAM中非线性滤波和非线性优化方法无优劣之分,不同应用场景下有不同的最优选择。

尽管经过几十年的研究,vSLAM在位姿估计、地图构建等方面日趋成熟,精度和实时性方面亦取得了不错的成果,但仍在动态适应性、泛化能力、高级感知能力和多传感器集成化等方面存在许多挑战,催生出了事件相机、深度学习、语义vSLAM等未来研究的热点方向。目前事件相机的相关理论和数据集初具规模,有望成为继RGB-D相机后又一应用于vSLAM

中的新型视觉传感器。作为人工智能领域最具吸引力的研究方向之一,深度学习良好的自组织、自学习能力使其在特征提取、回环检测等方面大放异彩。语义vSLAM中,vSLAM辅助语义建图的相关工作由于本质上与机器视觉中的语义建图差别不大,可应用的成熟方法较多,代表性研究成果较为丰富,而语义辅助vSLAM的研究成果则稍逊一筹。尽管目前语义vSLAM的相关研究尚不完善,但作为最具潜力的研究方向之一,未来将会有更多的成熟研究成果和开源框架面世。

参考文献(References)

- [1] Cadena C, Carlone L, Carrillo H, et al. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age[J]. IEEE Transactions on Robotics, 2016, 32(6): 1309-1332.
- [2] Gui J, Gu D, Wang S, et al. A review of visual inertial odometry from filtering and optimisation perspectives[J]. Advanced Robotics, 2015, 29(20): 1289-1301.
- [3] Rosten E, Drummond T. Machine learning for high-speed corner detection[C]. European Conference on Computer Vision. Graz: Austria: Springer, 2006: 430-443.
- [4] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [5] Bay H, Tuytelaars T, Van Gool L. Surf: Speeded up robust features[C]. European Conference on Computer Vision. Graz: Springer, 2006: 404-417.
- [6] Calonder M, Lepetit V, Strecha C, et al. Brief: Binary robust independent elementary features[C]. European Conference on Computer Vision. Crete: Springer, 2010: 778-792.
- [7] Leutenegger S, Chli M, Siegwart R. Brisk: Binary robustinvariant scalable keypoints[C]. 2011 IEEE International Conference on Computer Vision (ICCV). Barcelona: IEEE, 2011: 2548-2555.
- [8] Alahi A, Ortiz R, Vanderghenst P. Freak: Fast retina keypoint[C]. 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence: IEEE, 2012: 510-517.
- [9] Rublee E, Rabaud V, Konolige K, et al. Orb: An efficient alternative to sift or surf[C]. 2011 IEEE International Conference on Computer Vision (ICCV). Barcelona: Citeseer, 2011: 2564-2571.
- [10] Mur-Artal R, Montiel J M M, Tardos J D. Orb-slam: A versatile and accurate monocular slam system[J]. IEEE Transactions on Robotics, 2015, 31(5): 1147-1163.
- [11] Mur-Artal R, Tardós J D. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras[J]. IEEE Transactions on Robotics, 2017, 33(5): 1255-1262.
- [12] Newcombe R A, Lovegrove S J, Davison A J. Dtm: Dense tracking and mapping in real-time[C]. 2011 International Conference on Computer Vision. Zurich:

- IEEE, 2011: 2320-2327.
- [13] Engel J, Schöps T, Cremers D. Lsd-slam: Large-scale direct monocular slam[C]. European Conference on Computer Vision. Zurich: Springer, 2014: 834-849.
- [14] Engel J, Koltun V, Cremers D. Direct sparse odometry[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(3): 611-625.
- [15] Forster C, Pizzoli M, Scaramuzza D. Svo: Fast semi-direct monocular visual odometry[C]. 2014 IEEE International Conference on Robotics and Automation (ICRA). Miami: IEEE, 2014: 15-22.
- [16] Forster C, Zhang Z, Gassner M, et al. Svo: Semidirect visual odometry for monocular and multicamera systems[J]. IEEE Transactions on Robotics, 2016, 33(2): 249-265.
- [17] Li S P, Zhang T, Gao X, et al. Semi-direct monocular visual and visual-inertial slam with loop closure detection[J]. Robotics and Autonomous Systems, 2019, 112: 201-210.
- [18] Davison A J, Reid I D, Molton N D, et al. Monoslam: Real-time single camera slam[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 1052-1067.
- [19] Howard A. Multi-robot simultaneous localization and mapping using particle filters[J]. The International Journal of Robotics Research, 2006, 25(12): 1243-1256.
- [20] Thrun S, Montemerlo M, Koller D, et al. Fastslam: An efficient solution to the simultaneous localization and mapping problem with unknown data association[J]. Journal of Machine Learning Research, 2004, 4(3): 380-407.
- [21] Ali S S, Hammad A, Eldien A S T. Fastslam 2.0 tracking and mapping as a cloud robotics service[J]. Computers & Electrical Engineering, 2018, 69: 412-421.
- [22] Klein G, Murray D. Parallel tracking and mapping for small ar workspaces[C]. Proceedings of the 6th IEEE and ACM International Symposium on Mixed and Augmented Reality. Nara: IEEE, 2007: 1-10.
- [23] Di K, Zhao Q, Wan W, et al. Rgb-d slam based on extended bundle adjustment with 2d and 3d information[J]. Sensors, 2016, 16(8): 1285-1300.
- [24] Alismail H, Browning B, Lucey S. Photometric bundle adjustment for vision-based slam[C]. Asian Conference on Computer Vision. Taipei: Springer, 2016: 324-341.
- [25] Liu H, Chen M, Zhang G, et al. Ice-ba: Incremental, consistent and efficient bundle adjustment for visual-inertial slam[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 1974-1982.
- [26] Sibley G, Mantthies L, Sukhatme G. Sliding window filter with application to planetary landing[J]. Journal of Field Robotics, 2010, 27(5): 587-608.
- [27] Deutsch I, Liu M, Siegwart R. A framework for multi-robot pose graph slam[C]. 2016 IEEE International Conference on Real-time Computing and Robotics (RCAR). Angkor Wat: IEEE, 2016: 567-572.
- [28] Lee D, Myung H. Solution to the slam problem in low dynamic environments using a pose graph and an rgb-d sensor[J]. Sensors, 2014, 14(7): 12467-12496.
- [29] Mu B, Liu S Y, Paull L, et al. Slam with objects using a nonparametric pose graph[C]. 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Daejeon: IEEE, 2016: 4602-4609.
- [30] Frey B J, Kschischang F R, Loeliger H A, et al. Factor graphs and algorithms[C]. Proceedings of the Annual Allerton Conference on Communication Control and Computing. Illinois: Citeseer, 1997: 666-680.
- [31] Kaess M, Ranganathan A, Dellaert F. Isam: Incremental smoothing and mapping[J]. IEEE Transactions on Robotics, 2008, 24(6): 1365-1378.
- [32] Kaess M, Johannsson H, Roberts R, et al. Isam 2: Incremental smoothing and mapping with fluid relinearization and incremental variable reordering[C]. 2011 IEEE International Conference on Robotics and Automation. Shanghai: IEEE, 2011: 3281-3288.
- [33] Dellaert F. Factor graphs and gtsam: A hands-on introduction[Z]. Atalantic: Georgia Institute of Technology, 2012.
- [34] Spaenlehauer A, Frémont V, ekeriolu Y A, et al. A loosely-coupled approach for metric scale estimation in monocular vision-inertial systems[C]. 2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI). Daegu: IEEE, 2017: 137-143.
- [35] Sirtkaya S, Seymen B, Alatan A A. Loosely coupled kalman filtering for fusion of visual odometry and inertial navigation[C]. Proceedings of the 16th International Conference on Information Fusion. Istanbul: IEEE, 2013: 219-226.
- [36] Haochih Lin F. Loosely coupled stereo inertial odometry on low-cost system[C]. International Micro Air Vehicle Conference and Flight Competition (IMAV). Toulouse: IEEE, 2017: 143-148.
- [37] Mourikis A I, Roumeliotis S I. A multi-state constraintkalman filter for vision-aided inertial navigation[C]. Proceedings 2007 IEEE International Conference on Robotics and Automation. Roma: IEEE, 2007: 3565-3572.
- [38] Sun K, Mohta K, Pfrommer B, et al. Robust stereo visual inertial odometry for fast autonomous flight[J]. IEEE Robotics and Automation Letters, 2018, 3(2): 965-972.
- [39] Ramezani M, Khoshelham K. Vehicle positioning in gnss-deprived urban areas by stereo visual-inertial odometry[J]. IEEE Transactions on Intelligent Vehicles, 2018, 3(2): 208-217.
- [40] Huang G P, Mourikis A I, Roumeliotis S I. Observability-based rules for designing consistent ekf slam estimators[J]. The International Journal of Robotics Research, 2010, 29(5): 502-528.
- [41] Huang G P, Mourikis A I, Roumeliotis S I.

- A first-estimates jacobian ekf for improving slam consistency[M]. Berlin: Springer, 2009: 373-382.
- [42] Bloesch M, Omari S, Hutter M, et al. Robust visual inertial odometry using a direct ekf-based approach[C]. 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Hamburg: IEEE, 2015: 298-304.
- [43] Bloesch M, Burri M, Omari S, et al. Iterated extended kalman filter based visual-inertial odometry using direct photometric feedback[J]. The International Journal of Robotics Research, 2017, 36(10): 1053-1072.
- [44] Schneider T, Dymczyk M, Fehr M, et al. Maplab: An open framework for research in visual-inertial mapping and localization[J]. IEEE Robotics and Automation Letters, 2018, 3(3): 1418-1425.
- [45] Leutenegger S, Lynen S, Bosse M, et al. Keyframe-based visual-inertial odometry using nonlinear optimization[J]. The International Journal of Robotics Research, 2015, 34(3): 314-334.
- [46] Qin T, Li P, Shen S. Vins-mono: A robust and versatile monocular visual-inertial state estimator[J]. IEEE Transactions on Robotics, 2018, 34(4): 1004-1020.
- [47] Mur-Artal R, Tardós J D. Visual-inertial monocular slam with map reuse[J]. IEEE Robotics and Automation Letters, 2017, 2(2): 796-803.
- [48] Zou D, Wu Y, Pei L, et al. Structvio: Visual-inertial odometry with structural regularity of man-made environments[J]. IEEE Transactions on Robotics, 2019, 35(4): 999-1013.
- [49] Von Stumberg L, Usenko V, Cremers D. Direct sparse visual-inertial odometry using dynamic marginalization[C]. 2018 IEEE International Conference on Robotics and Automation (ICRA). Brisbane: IEEE, 2018: 2510-2517.
- [50] Vasco V, Glover A, Bartolozzi C. Fast event-based harris corner detection exploiting the advantages of event-driven cameras[C]. IEEE/RSJ International Conference on Intelligent Robots & Systems. Stockholm: IEEE, 2016: 9-14.
- [51] Elias M, Guillermo G, Henri R, et al. Continuous-time visual-inertial odometry for event cameras[J]. IEEE Transactions on Robotics, 2017, 99: 1-16.
- [52] Gallego G, Lund J E, Mueggler E, et al. Event-based, 6-dof camera tracking from photometric depth maps[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(10): 2402-2412.
- [53] Vidal A R, Rebecq H, Horstschaefter T, et al. Ultimate slam combining events, images, and imu for robust visual slam in hdr and high-speed scenarios[J]. IEEE Robotics & Automation Letters, 2018, 3(2): 994-1001.
- [54] Xiang G, Tao Z. Unsupervised learning to detect loops using deep neural networks for visual slam system[J]. Autonomous Robots, 2017, 41(1): 1-18.
- [55] Li R, Wang S, Long Z, et al. Undeepvo: Monocular visual odometry through unsupervised deep learning[C]. 2018 IEEE International Conference on Robotics and Automation (ICRA). Brisbane: IEEE, 2018: 7286-7291.
- [56] Bloesch M, Czarnowski J, Clark R, et al. Codeslam—Learning a compact, optimisable representation for dense visual slam[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 2560-2568.
- [57] Tateno K, Tombari F, Laina I, et al. Cnn-slam: Real-time dense monocular slam with learned depth prediction[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 6243-6252.
- [58] McCormac J, Handa A, Davison A, et al. Semanticfusion: Dense 3d semantic mapping with convolutional neural networks[C]. 2017 IEEE International Conference on Robotics and Automation (ICRA). Singapore: IEEE, 2017: 4628-4635.
- [59] Ma L, Stückler J, Kerl C, et al. Multi-view deep learning for consistent semantic mapping with rgb-d cameras[C]. 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Vancouver: IEEE, 2017: 598-605.
- [60] Civera J, cálviz-López D, Riazuelo L, et al. Towards semantic slam using a monocular[C]. 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems. San Francisco: IEEE, 2011: 1277-1284.
- [61] Bowman S L, Atanasov N, Daniilidis K, et al. Probabilistic data association for semantic slam[C]. 2017 IEEE International Conference on Robotics and Automation (ICRA). Singapore: IEEE, 2017: 1722-1729.
- [62] Lianos K N, Schönberger J L, Pollefeys M, et al. Vso: Visual semantic odometry[C]. European Conference on Computer Vision (ECCV). Munich: Springer, 2018: 234-250.
- [63] Yu C, Liu Z, Liu X J, et al. Ds-slam: A semantic visual slam towards dynamic environments[C]. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Madrid: IEEE, 2018: 1168-1174.

作者简介

李云天(1991—),男,博士生,从事惯性导航、视觉导航的研究, E-mail: yuntianlee91@hotmail.com;

穆荣军(1967—),男,教授,博士生导师,从事飞行器组合导航与自主导航、惯导系统初始对准、滤波理论与应用等研究, E-mail: murjun@163.com;

单永志(1970—),男,高级工程师,博士生导师,从事武器系统设计等研究, E-mail: zhiyong0451@sina.com.

(责任编辑: 郑晓蕾)