

DOI:10.16356/j.1005-2615.2022.06.002

视觉及其融合惯性的 SLAM 技术发展综述

曾庆化^{1,2,3}, 罗怡雪¹, 孙克诚¹, 李一能¹, 刘建业^{1,2,3}

(1. 南京航空航天大学导航研究中心, 南京 211106; 2. 先进飞行器导航、控制与健康管理和工业和信息化部重点实验室, 南京 211106; 3. 卫星通信与导航江苏高校协同创新中心, 南京 211106)

摘要: 当前视觉同时定位与建图 (Simultaneous localization and mapping, SLAM) 技术已经成为导航定位领域重要的导航方式之一。为提升视觉 SLAM 全天候多场景的综合性能, 融合多种信息的视觉 SLAM 技术得到了快速发展。本文系统综述了近年国内外视觉 SLAM 的研究成果。首先归纳整理了视觉及视觉/惯性 SLAM 中的前端里程计、后端优化、回环检测和地图构建关键环节。在纯视觉 SLAM 技术分析方面, 围绕传统基于几何变换和现代基于深度学习的两大类视觉 SLAM 方法开展讨论, 并梳理了相关具有代表性的视觉 SLAM 技术。在详细阐述视觉/惯性组合导航的多信息融合 SLAM 方法基础上, 分析了近年基于异源图像的视觉及其与惯性组合的 SLAM 导航技术进展。最后对视觉 SLAM 的未来发展方向进行了针对性展望。

关键词: 同时定位与建图; 异源图像; 视觉里程计; 惯性导航; 组合导航

中图分类号: TP242

文献标志码: A

文章编号: 1005-2615(2022)06-1007-14

Review on SLAM Technology Development for Vision and its Fusion of Inertial Information

ZENG Qinghua^{1,2,3}, LUO Yixue¹, SUN Kecheng¹, LI Yineng¹, LIU Jianye^{1,2,3}

(1. Navigation Research Center, Nanjing University of Aeronautics & Astronautics, Nanjing 211106, China;
2. Key Laboratory of the MIIT Advanced Aircraft Navigation, Control and Health Management, Nanjing 211106, China;
3. Jiangsu University Collaborative Innovation Center for Satellite Communication and Navigation, Nanjing 211106, China)

Abstract: Visual simultaneous localization and mapping (SLAM) technology has become one of the important navigation methods in the field of navigation and positioning. In order to improve the comprehensive performance of visual SLAM in all-weather and multi-scene environments, the visual SLAM technology integrating various information has been developed rapidly. This paper systematically reviews the research results of visual SLAM in recent years. First, the key links of front-end odometry, back-end optimization, loop closure detection and map construction in visual and visual/inertial SLAM are summarized. Second, in the analysis of pure visual SLAM technology, the two main types of visual SLAM methods based on traditional geometric transformation and modern deep learning are discussed, and related representative visual SLAM technologies are sorted out. Third, based on the detailed description of the multi-information fusion

基金项目: 国家自然科学基金(61533008, 61603181, 61673208, 61873125); 中央高校基本科研业务费专项资金(NJ20170005, NJ20170010); 江苏省高校优势学科建设工程项目。

收稿日期: 2022-01-16; **修订日期:** 2022-06-14

作者简介: 曾庆化, 男, 教授, 博士生导师, 担任《航空学报》《全球定位系统》《导航与控制》编委及某些 SCI 期刊客座编委, 主要研究惯性、视觉、卫星等多种导航方式的智能多信息融合导航理论、技术与应用。曾获国防科学技术进步一等奖 1 项、发明三等奖 1 项、进步三等奖 2 项; 获国家基金留学奖学金项目 2 项, 主持国家自然科学基金、国防预先研究基金等多种重要项目。发表 SCI、EI 等各种论文 170 余篇, 著作 2 部, 申请国家发明专利 60 余项, 获授权 40 余项。

通信作者: 曾庆化, E-mail: zengqh@nuaa.edu.cn。

引用格式: 曾庆化, 罗怡雪, 孙克诚, 等. 视觉及其融合惯性的 SLAM 技术发展综述[J]. 南京航空航天大学学报, 2022, 54(6): 1007-1020. ZENG Qinghua, LUO Yixue, SUN Kecheng, et al. Review on SLAM technology development for vision and its fusion of inertial information[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2022, 54(6): 1007-1020.

SLAM method of visual/inertial combined navigation, the progress of SLAM navigation technology based on heterogeneous image vision and its combination with inertial sensor in recent years is analyzed. Finally, the future development direction of visual SLAM is prospected.

Key words: simultaneous localization and mapping (SLAM); heterogeneous image; visual odometry; inertial navigation; integrated navigation

实时精确的导航与定位信息对于运动载体的高效工作和任务实现具有不可替代的重要作用。为了实现运动载体在未知环境中的高精度动态定位和导航,同时定位与建图(Simultaneous localization and mapping, SLAM)技术获得了飞速的发展。该技术实现了搭载传感器的运动载体在未知环境中运动感知并建立环境地图,同时估计自身在所建地图中位置的功能^[1],最早由Randall C. Smith和Peter Cheeseman于1986年提出^[2],后经Leonard等扩充^[3]。随着传感器类型的扩充以及运动恢复结构(Structure from motion, SfM)技术^[4]的出现,基于不同传感器和优化理论的SLAM技术得到迅速发展^[5-6],并在增强现实(Augmented reality, AR)/虚拟现实(Virtual reality, VR)、机器人、无人驾驶和无人机^[7]等领域得到广泛应用。

SLAM技术所依据的传感器形式多样,如激光、超声波和视觉等多种形式,目前激光SLAM和视觉SLAM是两种主流方式^[8-9]。激光SLAM主要通过激光雷达直接获得载体相对于环境的距离和方位信息^[10],实现环境的建图和载体自身相对位置的确定,该技术理论和工程均比较成熟,但其存在成本较高、探测范围有限等缺陷。视觉SLAM主要依靠视觉传感器,其价格适中,使用方便^[11],因此日益受到重视。随着应用场景和任务的日益复杂化,纯视觉SLAM也呈现了易受光照变化和运动速度干扰与影响等缺点。因此,多信息融合的视觉SLAM技术逐渐获得重视。结合多传感器信息和多层次互补滤波,可大大提升载体的SLAM整体性能。视觉SLAM分类及里程碑技术发展框图如图1所示。

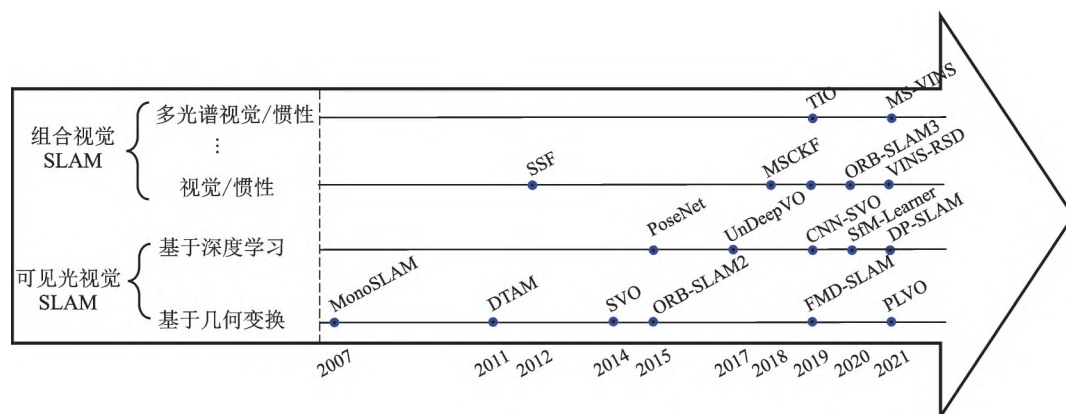


图1 视觉SLAM分类及里程碑技术发展

Fig.1 Visual SLAM classification and milestone technology development

本文集中针对视觉及视觉/惯性SLAM技术发展进行综述研究,分别从视觉及视觉/惯性SLAM中的关键环节、纯视觉SLAM关键技术发展分析、视觉/惯性信息融合SLAM关键技术发展分析等方面展开。

1 视觉及视觉/惯性SLAM关键环节

视觉/惯性SLAM系统框架结构如图2所示,获取视觉及惯性传感器的数据后,通过里程计算法估计载体运动,结合回环检测结果,在后端优化框架中优化运动估计参数,同时确定自身轨迹并实现周围环境地图的构建。本文以前端里程计技术、后端优化技术、回环检测技术和地图构建技术为代表,对视

觉/惯性SLAM中的关键环节进行归纳整理。去除惯性信息后,即可得视觉SLAM关键环节。

(1) 关键环节1:前端里程计技术

该关键环节在SLAM过程中处于前端位置,狭义的视觉里程计技术仅指代依据传感器数据进行载体位姿估计的过程,对SLAM的结果具有重要影响。在对视觉里程计运动估计时,可采用对极几何、迭代最近邻(Iterative closest point, ICP)或n点透视(Perspective-n-point, PnP)等原理进行解算。

视觉里程计根据工作原理可分为特征点法和直接法:特征点法通过提取和跟踪解算图像特征获得载体与环境之间的相对位姿,理论较为成熟,鲁棒性较强,且对环境的纹理特征要求较高,计算量

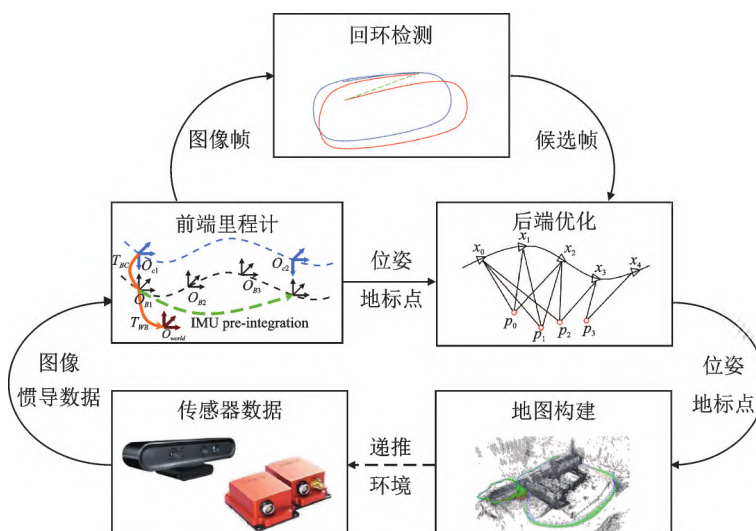


图2 视觉/惯性SLAM系统框架结构

Fig.2 Frame structure of visual/inertial SLAM system

较大;直接法利用图像像素信息计算并累积相机的相对运动,具有速度快、实时性好的优点,但需要满足时间连续、灰度不变等假设模型。

特征点法和直接法均需通过图像匹配实现信息跟踪,存在图像难匹配、误匹配问题。针对特征难匹配问题,国内外学者提出结合多源传感器信息的方法,采用惯导、全球导航卫星系统(Global navigation satellite system, GNSS)、激光雷达等设备信息与视觉信息耦合,提升视觉导航鲁棒性^[12-13];针对特征误匹配问题,大多数采用改进或挖掘特征匹配结构的方法丰富特征信息,如构建基于双向光流法的特征点环形匹配^[14]、提取人造环境结构边缘线特征的运动估计^[15]、结合平面与直线关联特征估计位姿^[16]、结合深度学习中的语义信息剔除动态特征点^[17-18]等算法。

通常采用惯性传感器辅助提高视觉里程计的动态适应性,视觉/惯性里程计的前端关键技术主要有惯性预积分和信息联合初始化。具有短期高动态适应能力的惯性器件信息采集频率通常高于视觉相机的图像采集频率,因此使用图像采集频率对惯性观测数据进行预积分^[19],实现视觉/惯性信息的频率配准和坐标系转化效率的提升。视觉/惯性联合初始化策略目的在于提升视觉里程计的初始信息的快速性和准确性,如采用观测样本情形评估函数评估状态误差信息矩阵,循环判断是否满足初始化所需精度,从而提升初始化效率^[20]。

(2)关键环节2:后端优化技术

后端优化可分为滤波器和非线性优化两类,多源信息松耦合架构常采用滤波器融合方式,紧耦合架构中两种方式均有应用,且基于图优化理论的非线性优化策略更为广泛。

滤波器优化方法通常改进卡尔曼滤波(Kalman filter, KF)模型以适用于非线性系统。如采用联邦滤波将惯性测量单元(Inertial measurement unit, IMU)信息分别与里程计和视觉数据组合,避免视觉失效时组合导航系统无法定位的问题^[21];采用不变卡尔曼滤波融合多层ICP估计的相机位姿与惯性解算位姿,提升相机追踪与重定位效果^[22];基于容积卡尔曼滤波(Cubature Kalman filter, CKF)框架,融合视觉与惯性信息,避免雅可比矩阵计算过程,提升系统位姿估计精度^[23]等。

非线性优化方法通常改进非线性最小二乘问题的求解过程。经典的最小二乘问题求解算法,如高斯牛顿法和列文伯格-马夸尔特(Levenberg-Marquardt, LM)算法运算时间较长,因而可采用改进dogleg算法减少耗时^[24]。此外,部分研究针对于模型的精度优化,如增加模型到模型配准的约束实现大回环局部优化^[25];采用卡尔曼滤波器融合全局优化后的高精度视觉估计与IMU信息,兼顾滤波器的计算效率与优化方法的精度^[26]。

(3)关键环节3:回环检测技术

SLAM中的回环检测技术通过建立当前帧与之前某一帧的位姿约束关系消除视觉里程计所产生的累积误差,进而实现地图优化,主要依赖基于词袋模型的改进算法或结合深度学习的回环检测技术^[27]。

词袋(Bag of words, BoW)模型聚类特征描述形成单词,将图像帧转换为单词集合,对比图像帧相似性判断回环,可以加快特征匹配速度,但存在环境适应性弱、难以应对视角变换等问题。因此,部分研究通过优化BoW模型进行改进,如结合图像均衡化处理实现光照变换场景下的视觉BoW在线增删,提升算法鲁棒性^[28];计算匹配图像的共视

区域协方差矩阵距离确定回环^[29],实现大视角下的回环检测等^[30]。

BoW模型中构建的单词向量近似语义信息,因而可结合语义标注剔除运动物体^[31],提高BoW模型鲁棒性;也可以采用神经网络代替BoW模型获取图像帧的特征向量^[32-33],提高算法召回率。

(4) 关键环节4:地图构建技术

由于SLAM前端理论差异,其构建地图主要分稀疏点云地图、半稠密点云地图和稠密点云地图3类,点云稠密程度与检测特征点的密集程度相关。随着语义信息的引入,部分框架构建语义地图^[34]以提供更好的视觉感知效果。

地图构建通常依据传感器获取信息及数据处理算法的差异而不同。文献[35]提出构建融合类脑感知信息的三维认知地图,其三维位置误差为总行程的1.56%;文献[36]将图像中的语义标注结果与相机估计轨迹相融合,构建增量式稠密语义地图。

SLAM建图过程中通常需要包含载体的相对

位姿信息以及周围环境景物的三维位置信息,评判算法精度的指标采用绝对轨迹误差(Absolute trajectory error, ATE)和相对位姿误差(Relative pose error, RPE)。ATE对比数据序列时间戳中每个相近时刻下算法解算的载体位姿估计值与真实位姿之间的差值,是评估算法定位精度的主要指标。RPE计算数据序列时间戳中每个固定时间间隔内算法估计位姿变化量与真实位姿变化量的差值,是评估系统漂移的主要指标。由于并行处理技术的应用,算法的实时性评估通常较为独立,可依据研究需要对不同环节分别评估其运行时间,也可以针对前端特征跟踪耗时的问题,评估跟踪线程单帧图像的处理时间。鉴于SLAM算法影响因素较多,包括前端理论、环境特征点数目、光照变换、硬件配置等,当前的水平为:室内定位精度可达分米级;室外定位精度多为米级;随着动态性能的增大,其定位精度有所下降。图3汇集了几种常用的视觉SLAM地图结果形式,其中图3(c,d,e)均属于点云地图。

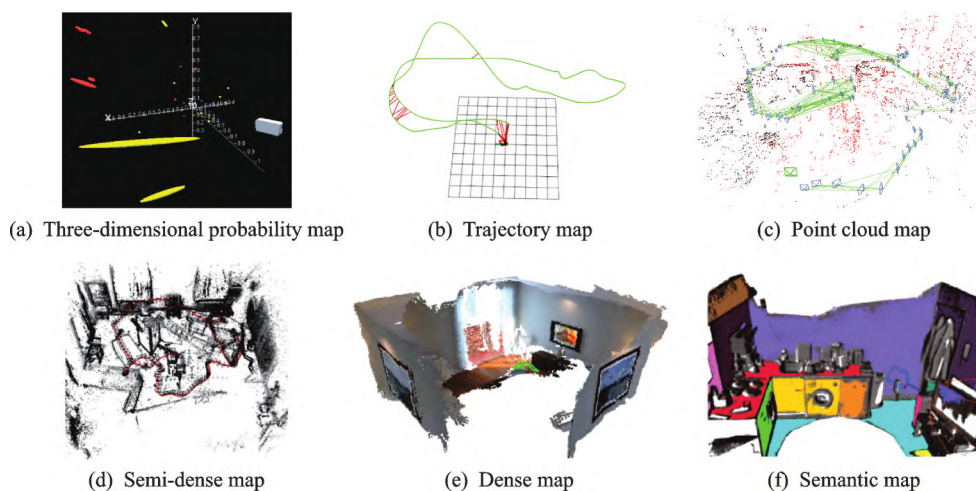


图3 视觉SLAM构建地图类型

Fig.3 Types of maps constructed by visual SLAM

2 纯视觉SLAM关键技术发展分析

自2004年Bergen等提出视觉里程计(Visual odometry, VO)后^[37],基于视觉图像序列估计相机运动的方法获得了科研人员的重视。将建图环节引入VO极大促进了构建环境地图强化数据关联的视觉SLAM技术的发展。视觉SLAM通过多帧图像估计自身的位姿变化,再通过累积位姿变化计算载体在当前环境中的位置,其与激光SLAM技术相比具备低成本、应用方便、信息丰富、隐蔽性强等诸多优势,发展潜力巨大。

近年发展迅速的深度学习技术在视觉SLAM中也得到了重视和应用^[38],基于此可以将纯视觉SLAM分成基于几何变换和基于深度学习这两大

类。其中,基于几何变换的视觉SLAM依据前端原理不同可分为特征点法和直接法两类,基于深度学习的视觉SLAM依据深度学习在视觉SLAM中的学习功能形式不同可分为模块替代SLAM和端到端SLAM两类。

2.1 基于几何变换的视觉SLAM关键技术

基于几何变换的视觉SLAM方法利用二维图像和三维点之间的几何结构映射关系求解导航参数,其主要涉及两类前端理论:特征点法和直接法。

(1) 特征点法SLAM

该方法是VO早期的主流方法,主要地图形式如图3(a,c)所示。第一个视觉SLAM, Mono-SLAM于2007年提出^[39]。首次通过移动端相机获取三维运动轨迹,在个人电脑(Personal computer,

PC)上可以 30 Hz 的频率实时工作,但其单线程结构与实时性要求限制了前端特征跟踪数量。因此 Klein 等^[40]提出具备跟踪和建图双线程结构的并行跟踪与映射(Parallel tracking and mapping, PTAM)方法,首次通过非线性优化,即光束法平差(Bundle adjustment, BA)方式计算相机轨迹和全局一致环境地图^[41]。2013 年,Labbe 等^[42]提出了基于实时外观建图(Real-time appearance-based mapping, RTAB-MAP)的方法,基于 BoW 模型将图像表示为视觉词汇的集合,实现回环检测^[43]。2015 年 Mur-Artal 等^[44-45]进一步提出含有跟踪、局部建图及回环检测多线程的 ORB-SLAM2 算法,视觉 SLAM 的多线程框架基本完善。

近几年基于特征点法的视觉 SLAM 主要针对前端特征结构进行改进,如 2021 年哈尔滨工业大学^[46]提出点线特征融合的方式,结合线特征约束提升线特征估计准确性,降低了 3D 线特征引入的冗余参数对运动估计的影响。南开大学^[16]结合平面和直线之间的几何关系,提出基于平面-直线,即采用直线段检测算子(Line segment detector, LSD)特征混合关联的约束方法 PLVO (Plane-line-based RGB-D visual odometry),解决相机位姿估计时的退化问题。同时,简化跟踪特征,直接依据图像像素信息估计运动的直接法得到广泛研究。

(2)直接法 SLAM

该方法构建的地图形式主要如图 3(c,d,e)所示。2011 年,Newcombe 等^[47]提出了首个基于单像素的直接法 SLAM,稠密跟踪与建图(Dense track-

ing and mapping, DTAM)方法,结合单个 RGB 相机可在商用图像处理器(Graphics processing unit, GPU)中实时定位与跟踪。为了在一定程度上保留关键点信息,随后 Forster 等^[48]提出用稀疏特征点代替像素匹配的半直接法视觉里程计(Semi-direct visual odometry, SVO)方法,Engel 等提出构建大尺度全局一致性半稠密环境地图的大范围直接同步定位与建图(Large-scale direct SLAM, LSD-SLAM)算法^[49],以及通过最小化光度误差计算相机位姿与地图点的位置,将数据关联与位姿估计统一在非线性优化中的直接稀疏里程计(Direct sparse odometry, DSO)方法^[50]。

为兼顾特征点法与直接法的优势,部分学者针对半直接法 SLAM 深入研究。南京邮电大学^[51]提出了基于半直接法的单目视觉 SLAM 算法,结合 KLT(Kanade-Lucas-Tomasi)光流跟踪与 ORB(Oriented fast and rotated brief)特征匹配方法,可处理复杂环境下的低纹理、运动目标和感知混叠等问题。

研究结合几何信息的直接法也获得了进一步的推进,如 2019 年中国科学院大学^[52]提出的融合多视图几何和直接法的同步定位与建图(Fusing MVG and direct SLAM, FMD-SLAM),结合多视图几何和直接法估计位姿信息,并利用最小化重投影误差优化位姿结果。表 1 中梳理了基于几何变换的视觉 SLAM 代表性成果。

从以上基于几何变换的视觉 SLAM 方法的演化趋势可知:特征点法主要对前端关键点进行改进优化,直接法的前端则由基于像素估计趋向结合特

表 1 基于几何变换的视觉 SLAM 研究成果
Table 1 Research results of visual SLAM based on geometric transformation

类型	方案	关键点	优化理论	地图类型	适用模式	特点
特征点法	MonoSLAM	Shi-Tomasi	EKF	三维概率地图	单目	首个单目视觉 SLAM
	PTAM	FAST	BA	点云地图	单目	首次使用双线程框架
	ORB-SLAM2	ORB	Local BA	点云地图	单目、双目、RGB-D	具有完备的多线程框架
	点线单目视觉 SLAM	Shi-Tomasi、LSD	BA	点云地图	单目	利用平面约束去除线特征冗余参数
	PLVO	平面特征、LSD	—	点云地图	RGB-D	基于平面和直线特征关联约束位姿求解
直接法	DTAM	Pixel	位姿图优化	稠密地图	单个 RGB-D	首个直接法 SLAM
	DSO	Pixel	Local BA	半稠密地图	单目、双目	稀疏直接法视觉里程计
	SVO	FAST	Local BA	点云地图	单目、双目	半直接法视觉里程计
	半直接单目视觉 SLAM	KLT、ORB	BA	点云地图	单目	融合视觉词汇间的空间信息进行闭环检测
	FMD-SLAM	FAST	Local BA	点云地图	双目	融合多视图几何与直接法

征点法的半稠密方式,后端优化部分由滤波器过渡到非线性优化为主,地图类型也由点云地图扩展为半稠密/稠密地图,适用模式由单目逐步扩充为双目和RGB-D模式。

2.2 基于深度学习的视觉SLAM关键技术

由于在光照变化、场景更替和物体运动时,基于几何变换的视觉SLAM方法的性能下降^[53],且具有计算量大、计算流程模式固定等问题。因此,随着深度学习技术的蓬勃发展,基于深度学习的视觉SLAM研究也逐渐成为视觉的一个研究热点。基于深度学习的视觉SLAM方法,主要用深度学习代替了SLAM的部分或者全部的模块,其地图形式多如图3(c,e,f)所示。

(1) 利用深度学习替代传统SLAM模块的方法

该方法常用深度学习过程替代特征提取、特征匹配和深度估计等模块^[54-55],而语义分割大多结合语义信息对前端里程计过程进行改进。2008年Roberts等^[56]首次将机器学习应用于VO。2017年佳能公司^[57]提出的CNN-SLAM,将LSD-SLAM中的深度估计和图像匹配模块都替换成基于卷积神经网络(Convolutional neural network, CNN)的方法,提升系统场景适应性。2018年帝国理工学院^[58]提出的CodeSLAM使用U-Net神经网络提取图像特征并估计图像深度的不确定性。2019年博特拉大学^[59]提出一种利用单张图片进行深度预测的方法CNN-SVO,可提升半直接法视觉里程计的建图效果。2020年阿德莱德大学^[60]提出基于CNN实现

深度预测和特征匹配的融合深度与光流的视觉里程计(Depth and flow for visual odometry, DF-VO)方法,将深度学习与对极几何和PnP成功结合。2021年中国科学技术大学^[61]提出结合几何约束和语义分割去除动态环境中移动对象的DP-SLAM方法,结合移动概率传播模型进行动态关键点检测,有助于虚拟现实的应用研究。

(2) 利用深度学习实现的端到端SLAM方法

该方法利用端到端的训练方法可直接估计图像序列的帧间运动,在线计算速度快,具有较强的算法迁移能力^[62],其计算速度相比替代传统SLAM模块的方法更快。2015年,剑桥大学^[63]基于图像识别网络GoogleNet^[64]开发基于单张图像信息的绝对位姿估计网络PoseNet。2017年埃塞克斯大学^[65]提出的UnDeepVO采用无监督深度学习训练双目图像,可通过单目图像估计相机位姿并实现绝对尺度的恢复。弗莱堡大学^[66]提出的DeMoN利用连续无约束的图像计算深度及相机运动,还可对图像间的光流和匹配的置信度进行估计。2019年伦敦大学学院^[67]提出的Monodepth2是一种基于CNN的自监督方法,通过在运动图像序列上训练一个建立在自监督损失函数上的架构实现深度及位姿预测。2020年北京大學^[68]提出一种单目深度、位姿估计的无监督学习网络SfM-Learner,可从无标签视频序列中进行深度和相机姿态估计卷积神经网络的训练。表2整理了基于深度学习的视觉SLAM代表方案。

表2 基于深度学习的视觉SLAM研究成果

Table 2 Research results of visual SLAM based on deep learning

类型	方案	训练类型	网络类型	地图类型	适用模式	特点
模块替代	CNN-SLAM	有监督	CNN	稠密地图	单目	融合训练预测和计算深度图
	CodeSLAM	有监督	U-Net	稠密地图	单目	可提取特征、估计深度不确定性
	DP-SLAM	有监督	Mask R-CNN	稠密地图	RGB-D	可检测动态关键点
	CNN-SVO	有监督	CNN	点云地图	单目	利用单张图片预测深度
	DF-VO	无监督	CNN	轨迹地图	双目	可估计单视图深度、双视图光流
端到端训练	PoseNet	有监督	GoogleNet	稠密地图	单目	
	DeMoN	有监督	Bootstrap Net	点云地图	单目	
	Monodepth2	自监督	U-Net/ResNet	—	单目	利用网络估计相机位姿
	UnDeepVO	无监督	DNN	点云地图	单目	
	SfM-Learner	无监督	CNN	—	单目	

基于深度学习的视觉SLAM方法在2015年之后兴起,近年来研究成果越来越丰富,其训练类型以有监督为主,并逐步改进为自监督、无监督方式。采用替代传统视觉SLAM模块的形式,构建地图类型较为多样,适用模式丰富,但仍以单目模式为主;端到端的学习则主要集中在单目视觉模式。

综上,基于几何变换的方式和结合深度学习的方式都存在其优势和不足。几何变换方法理论成熟,特征点法适用于尺度较大的运动,鲁棒性更好,但特征提取耗时;直接法速度快,适用于特征缺失场景,但灰度不变假设不易满足,不适合快速运动。基于深度学习的方法对光线复杂环境的适应性强,对动态场景识别更加有效,可结合语义信息

构建地图,但同样存在训练时间长、计算资源大、可解释性差等问题。因此无论是理论成熟、具有可解释性模型的传统方法还是可移植性强的深度学习方法都亟待进一步深入研究。

3 视觉/惯性 SLAM 关键技术发展分析

由于纯视觉SLAM存在受环境光照影响大、动态适应性差等问题,基于视觉的多信息融合SLAM技术成为新的重要研究方向,如:视觉/惯性、异源视觉/惯性等^[69]组合方式。作为一种常见的自主导航方式^[70-71],传统的可见光视觉与惯性信息融合SLAM技术,在研究成本和隐蔽性方面均具有较大优势。针对可见光图像在弱光照条件下适应性差的问题,可结合不同成像传感器和图像处理技术获得的异源图像进一步丰富视觉导航信息。下文将在阐述传统视觉/惯性SLAM的发展历程及经典算法的基础上,对融合异源图像的视觉、视觉/惯性SLAM方案进行综合分析。

3.1 传统视觉/惯性SLAM关键技术

纯视觉导航对环境特征跟踪较为精确,但在载体快速运动以及低纹理、含有光照变化等场景中应用效果不佳;纯惯性导航则不受运动场景和速度限制,但长时间运行会产生累积误差^[72]。因此,融合视觉与惯性传感器优势,并可规避劣势的视觉/惯性组合SLAM技术成为视觉SLAM领域的主流组合方式,并依据视觉与惯性信息的耦合方式分为松耦合和紧耦合两种组合形式^[73-74],构建地图形式主要如图3(b,c)所示。

(1)视觉/惯性松耦合SLAM

视觉图像与惯性数据的松耦合模式多应用于早期理论探索阶段,如2012年,苏黎世联邦理工学院^[75]提出的一种基于光流(Optical flow,OF)和惯性约束的惯性与单传感器融合(Single sensor fusion,SSF),其具有良好的可扩展性。但由于视觉/惯性松耦合方式通常以惯性数据为主,视觉测

量信息则用于修正惯性传感器的累积误差,因而会引入视觉量测所产生的误差,因此在后续研究中多采用紧耦合方式。

(2)视觉/惯性紧耦合SLAM

相较于松耦合方式,紧耦合SLAM可以更紧密地联系视觉特征与惯性传感器的量测数据,提升视觉/惯性组合SLAM系统的精度,因而形成诸多以图优化理论为主的视觉/惯性紧耦合SLAM成果。

2015年Bloesch等^[76]提出一种单目视觉/惯性里程计鲁棒视觉/惯性里程计(Robust visual inertial odometry,ROVIO),直接利用图像块像素强度误差实现精确跟踪,使用EKF实现优化。Leutenegger等^[77]提出基于滑动窗口的非线性优化紧耦合双目视觉/惯性里程计OKVIS。2018年宾夕法尼亚大学^[78]提出一种基于多状态约束卡尔曼滤波(Multi-state constraint kalman Filter,MSCKF)的双目视觉/惯性里程计MSCKF-VIO。弗莱堡大学^[79]提出将基于ORB-SLAM2的视觉姿态解算与惯性、GPS测量值相结合的姿态测量算法,并在无人机救援任务中得到应用。2019年香港科技大学^[80-81]提出基于优化的多传感器状态估计器VINS-Fusion,支持多种视觉/惯性传感器类型,可实现自主精确定位^[82]。上海交通大学^[83]提出一种在人造环境中利用结构规律的视觉惯性里程计StructVIO,基于结构化线特征完成VIO设计。慕尼黑工业大学^[84]提出了一种利用非线性因子恢复法提取相关信息,进行视觉/惯性建图的里程计Basalt。2020年萨拉戈萨大学^[85]提出适用于单目、双目、RGB-D相机的视觉、视觉/惯性多种模式的ORB-SLAM3,在部分测试场景中定位精度可达分米甚至厘米级,实现了传统视觉SLAM算法在定位精度上的突破。2021年,武汉大学^[86]提出一种适用于卷帘快门RGB-D相机的视觉/惯性SLAM方法VINS-RSD,利用深度相机直接获得深度,解决了初始化特定运动状态的需求。表3整理了视觉/惯性组合SLAM的代表性方案。

表 3 视觉/惯性SLAM研究成果
Table 3 Visual/inertial SLAM research results

耦合类型	方案	前端理论	优化理论	地图类型	适用模式	特点
松耦合	SSF	OF	EKF	点云地图	单目视觉/惯性	自校准即插即用系统
	StructVIO	特征点法	MSCKF	点云地图	单目视觉/惯性	采用结构化线特征
	ORB-SLAM3	特征点法	Local BA	点云地图	单、双目视觉/惯性、RGB-D	基于特征的紧耦合框架
紧耦合	VINS-RSD	特征点法	Local BA	稠密地图	RGB-D单目视觉/惯性	适用卷帘快门RGB-D相机
	MSCKF-VIO	KLT	S-MSCKF	点云地图	双目视觉/惯性	基于滤波器的双目VIO
	VINS-Fusion	KLT	Local BA	轨迹地图	单、双目视觉/惯性	支持静止初始化
	Basalt	KLT	Local BA	点云地图	双目视觉/惯性	非线性因子恢复相关信息

目前具有代表性的传统视觉/惯性 SLAM 方案多采用紧耦合方式实现多信息融合,前端常采用特征点法和 KLT 光流法,优化理论以非线性优化方式为主。构建地图的类型较为丰富,主要为点云地图和轨迹地图。由于视觉/惯性组合的方式有效解决了单目视觉无法恢复尺度的问题,因此单目及双目视觉/惯性组合模式均得到了广泛研究。

在近几年的多信息融合导航领域,出于全天候全时段导航任务的需求,研究人员深入挖掘了视觉图像信息,通过融合不同模态的视觉图像信息提升视觉解算性能,基于异源图像的视觉导航技术得到发展。

3.2 融合异源图像的视觉及视觉/惯性 SLAM 关键技术

传统可见光视觉导航方式存在光谱范围受限、低光照环境不可用等问题,异源图像则既保留了视觉导航自主抗干扰的优势,又对单色图像信息进行有效补充,在地形勘探、目标提取和精确导航方面具有实用前景,融合异源图像的视觉导航技术也在低光照环境及复杂地形场景中得到应用。早期异源图像多作为信息辅助测量手段应用于导航领域,用于提高视觉感知^[87-88]、实现双目测距^[89]等功能。近年来则涌现了更多融合异源图像信息的视觉导航方案,基于其融合的传感器数据分为异源视觉 SLAM 和异源视觉/惯性 SLAM 方案,地图形式主要为图 3(b)的轨迹地图。

(1) 异源视觉 SLAM

2018 年新加坡国立大学^[90]提出了一种多平面红外视觉里程计,采用激光设备辅助消除尺度恢复中的测量误差,提升了无人机在多平面复杂地形中昼夜飞行的鲁棒性。2021 年大连理工大学^[91]提出一种偏振光/双目视觉仿生组合导航方法,利用仿生偏振定向传感器输出的角度信息修正航向角。克兰菲尔德大学^[92]提出多模态纯视觉里程计 GMS-VO,分别对姿态和尺度进行估计,提升了系统整体定位精度。

(2) 异源视觉/惯性 SLAM

2017 年克兰菲尔德大学^[93]提出了一种融合惯性和多光谱信息的导航系统,适用于载体快速运动及寒冷环境场景。2018 年新加坡国立大学^[94]提出了适用于无人机昼夜飞行的红外视觉惯性里程计算法,采用扩展卡尔曼滤波(Extended Kalman filter, EKF)框架实现视觉及惯性测量信息融合,消除尺度恢复和运动估计误差。2019 年加州理工学院^[95]提出了一种热像惯性里程计,分别将惯性数据与夜间的热图像或白天的视觉图像紧耦合以适应不同光照条件,并集成在运动规划和控制的自主框架中实时运行。2021 年克兰菲尔德大学^[92]提出视觉/惯性里程计 MS-VINS,采用误差状态卡尔曼滤波器融合视觉姿态与惯性数据。表 4 整理了融合异源图像的视觉 SLAM 代表性方案。

表 4 融合异源图像的视觉及视觉/惯性 SLAM 研究成果

Table 4 Visual and visual/inertial SLAM research results fused with heterogeneous images

方案	适用模式	图像信息	特点
多平面红外视觉里程计	视觉	红外热像图	适用于多平面复杂地形
偏振光/双目视觉仿生组合导航	视觉	可见光/偏振光图像	利用偏振定向传感器修正航向角
GMS-VO	视觉	可见光/红外热像图	分别进行姿态估计与尺度估计
红外视觉/惯性里程计	视觉/惯性	红外热像图	可用于全天时 UAV 导航
多光谱视觉/惯性组合导航	视觉/惯性	可见光/LWIR 图像	用于快速运动和寒冷环境
热像-惯性里程计	视觉/惯性	可见光/红外热像图	不同时间段选用不同视觉图像
MS-VINS	视觉/惯性	可见光/红外热像图	惯性数据可解决多模态图像退化问题

异源图像逐渐由辅助导航手段转为视觉导航的主要信息来源,视觉图像也由红外热像图转变为可见光/红外组合图像,并扩展为视觉/惯性组合导航模式,利用多信息融合优势提升 SLAM 系统精度。虽然在异源传感器数据处理、硬件配置等方面仍需进一步探索,但随着异源图像信息的引入,视觉 SLAM 的环境适应性逐渐增强,呈现出在全天时复杂环境应用的发展倾向。

综上,视觉及视觉/惯性 SLAM 在不同领域得到了不同程度的应用。

基于几何变换的视觉及视觉/惯性 SLAM 整体研究已较为成熟,现已在多个领域得到实际应用,如 iRobot、Dyson、科沃斯、小米等公司推出的具备自主导航功能的扫地机器人,应用视觉 SLAM 技术实现室内定位感知及建图任务;Skydio 公司推出的自主飞行无人机,具备厘米级定位精度,可用于高压线塔、桥墩、盖梁等场景检测;特斯拉公司研发的无人车辆智能驾驶,可实现 L2 级别的部分自动化车辆自动驾驶;Nreal 公司推出的增强现实眼镜,结合 SLAM 技术联通虚拟环境与现

实世界,提升用户感知体验。

基于神经网络或融合异源视觉的定位导航技术虽然为视觉SLAM带来了新的变革,但主要处于理论研究与实验验证阶段,凭借神经网络良好的学习能力及异源视觉的环境适应性,多应用于移动机器人巡检、无人机矿井探测、无人驾驶目标识别及导航等自主任务。

4 结 论

本文以视觉及视觉/惯性SLAM技术为主,归纳整理了前端里程计、后端优化、回环检测和地图构建关键环节,系统综述了纯视觉SLAM的关键技术发展情况,并对视觉/惯性SLAM导航方案进行梳理分析,结合异源图像的研究应用情况,探讨了融合异源图像的视觉导航技术研究进展情况,为融合异源图像的视觉及视觉/惯性SLAM研究提供参考。结合现有研究成果,视觉及视觉/惯性SLAM技术主要呈现以下3点发展趋势。

(1)理论优化改进:由于应用场景需求的多样化,结合惯性、异源图像等多传感器的信息融合模式成为SLAM主流,促进了以紧耦合为主的信息融合理论发展,而随着大场景SLAM应用需求及图优化理论的推进,逐步形成了基于扩展卡尔曼滤波框架的改进滤波器优化架构,和以光束法平差为主的非线性优化架构两种研究趋势。

(2)新型技术引入:随着深度学习技术在计算机视觉中的广泛应用,视觉SLAM呈现出由传统几何变换方式逐步转向结合深度学习的智能融合趋势。一方面视觉图像与语义信息的紧密联系,使得集成语义信息的视觉SLAM得到更多探索;另一方面为减少对传统方式依赖,利用神经网络架构替代SLAM的部分模块或端到端强化学习的模式得以广泛研究。

(3)应用领域推广:视觉SLAM目前在智能家居、自动驾驶、无人机等领域得到了不同层次的应用,随着硬件性能的提升,视觉SLAM的应用范围逐步由室内小空间转向动态目标丰富的复杂大空间场景,应用的广度和深度得到较大提升。

参考文献:

- [1] 高翔,张涛,刘毅,等.视觉SLAM十四讲:从理论到实践[M].北京:电子工业出版社,2019.
GAO Xiang, ZHANG Tao, LIU Yi, et al. Fourteen lectures on visual SLAM: From theory to practice [M]. Beijing: Publishing House of Electronics Industry, 2019.
- [2] SMITH R C, CHEESEMAN P. On the representa-

- tion and estimation of spatial uncertainty[J]. International Journal of Robotics Research, 1986, 5(4): 56-58.
- [3] LEONARD J J, DURRANT-WHYTE H F. Mobile robot localization by tracking geometric beacons[J]. IEEE Transactions on Robotics & Automation, 1991, 7(3): 376-382.
- [4] DHOND U R, AGGARWAL J K. Structure from stereo-a review[J]. IEEE Trans System Man & Cybernetics, 1989, 19(6): 1489-1510.
- [5] BAILEY T, DURRANT-WHYTE H. Simultaneous localization and mapping (SLAM): Part II [J]. IEEE Robotics & Automation Magazine, 2006, 13(3): 108-117.
- [6] 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.
- [7] 陈润泽,郝向阳,陶健,等.旋翼无人机视觉/惯性组合导航技术研究[J].电子测量技术,2021,44(12): 121-127.
CHEN Runze, HAO Xiangyang, TAO Jian, et al. Research on integrated navigation of vision/inertial for rotorcraft[J]. Electronic Measurement Technology, 2021, 44(12): 121-127.
- [8] YUAN C, LAI J, ZHANG J, et al. Research on an autonomously tightly integrated positioning method for UAV in sparse-feature indoor environment[C]//Proceedings of the 15th International Bhurban Conference on Applied Sciences and Technology (IBCAST). Islamabad, Pakistan: IEEE, 2018: 318-324.
- [9] 赖际舟,袁诚,吕品,等.不依赖于卫星的无人系统视觉/激光雷达感知与自主导航技术[J].导航定位与授时,2021,8(3): 1-14.
LAI Jizhou, YUAN Cheng, LYU PIN, et al. Unmanned system visual/LiDAR perception and navigation technology independent of GNSS[J]. Navigation Positioning and Timing, 2021, 8(3): 1-14.
- [10] 刘朋,任工昌,何舟.2D激光SLAM中特征角点的提取方法[J].南京航空航天大学学报,2021,53(3): 366-372.
LIU Peng, REN Gongchang, HE Zhou. Method for extracting corner feature from 2D laser SLAM [J]. Journal of Nanjing University of Aeronautics & Astronautics, 2021, 53(3): 366-372.
- [11] 廖鑫森,王正平,刘莉,等.拒止环境下视觉辅助定位与导航技术综述[J].弹箭与制导学报,2020,40(6): 76-81.
LIAO Xinmiao, WANG Zhengping, LIU Li, et al. Survey on vision-assisted positioning and navigation

- technology in denied environment[J]. Journal of Projectiles, Rockets, Missiles and Guidance, 2020, 40(6): 76-81.
- [12] 孙伟, 苗宏胜, 李海军, 等. 一种基于证据推理的惯性/视觉里程计深组合导航方法[J]. 中国惯性技术学报, 2020, 28(5): 615-623.
- SUN Wei, MIAO Hongsheng, LI Haijun, et al. An INS/visual odometer deeply integrated navigation method based on evidence reasoning[J]. Journal of Chinese Inertial Technology, 2020, 28(5): 615-623.
- [13] 赵耀忠, 咸金龙, 高巍. 基于视觉、LiDAR与IMU的实时无人车里程计研究[J]. 红外与激光工程, 2022, 51(8): 207-215.
- ZHAO Yaozhong, XIAN Jinlong, GAO Wei. Research on a real-time odometry system integrating vision, LiDAR and IMU for autonomous driving[J]. Infrared and Laser Engineering, 2022, 51(8): 207-215.
- [14] 黄平, 曹镇, 王欢. 基于环形特征匹配的双目视觉里程计[J]. 光学学报, 2021, 41(15): 166-176.
- HUANG Ping, CAO Zhen, WANG Huan. Stereo visual odometry based on ring feature matching[J]. Acta Optica Sinica, 2021, 41(15): 166-176.
- [15] 黄平, 曹镇, 黄俊杰. 一种基于线特征的RGB-D视觉里程计算法[J]. 中国惯性技术学报, 2021, 29(3): 340-349.
- HUANG Ping, CAO Zhen, HUANG Junjie. A RGB-D visual odometry method based on line features[J]. Journal of Chinese Inertial Technology, 2021, 29(3): 340-349.
- [16] 孙沁璇, 苑晶, 张雪波, 等. PLVO: 基于平面和直线融合的RGB-D视觉里程计[J]. 自动化学报, 2022. DOI:10.16383/j.aas.c200878.
- SUN Qinxuan, YUAN Jing, ZHANG Xuebo, et al. PLVO: Plane-line-based RGB-D visual odometry[J]. Acta Automatica Sinica, 2022. DOI: 10.16383/j.aas.c200878.
- [17] 卢金, 刘宇红, 张荣芬. 面向动态场景的语义视觉里程计[J]. 激光与光电子学进展, 2021, 58(6): 273-282.
- LU Jin, LIU Yuhong, ZHANG Rongfen. Semantic-based visual odometry towards dynamics scenes[J]. Laser & Optoelectronics Progress, 2021, 58(6): 273-282.
- [18] 李博, 段中兴. 室内动态环境下基于深度学习的视觉里程计[EB/OL]. (2021-10-20). <http://kns.cnki.net/kcms/detail/21.1106.TP.20211019.1809.020.html>.
- LI Bo, DUAN Zhongxing. Visual odometry based on deep learning in dynamic indoor environment[EB/OL]. (2021-10-20). <http://kns.cnki.net/kcms/detail/21.1106.TP.20211019.1809.020.html>.
- [19] 刘刚, 葛洪伟. 视觉惯导SLAM初始化算法研究[J]. 计算机科学与探索, 2021, 15(8): 1546-1554.
- LIU Gang, GE Hongwei. Research on initialization algorithm for visual-inertial SLAM system[J]. Journal of Frontiers of Computer Science and Technology, 2021, 15(8): 1546-1554.
- [20] 凌有铸, 郭俊阳, 陈孟元, 等. 基于双重初始化和分级优化的改进视觉惯性SLAM方法[J]. 中国惯性技术学报, 2021, 29(2): 191-198.
- LING Youzhu, GUO Junyang, CHEN Mengyuan, et al. An approach to improve visual-inertial SLAM method based double initialization and hierarchical optimization[J]. Journal of Chinese Inertial Technology, 2021, 29(2): 191-198.
- [21] 万文辉, 李宇, 胡文敏, 等. 基于联邦滤波进行立体相机/IMU/里程计运动平台组合导航定位[J]. 武汉大学学报(信息科学版), 2018, 43(1): 101-106.
- WAN Wenhui, LI Yu, HU Wenmin, et al. Mobile platform localization by integration of stereo cameras, IMU and wheel odometer based on federated filter[J]. Geomatics and Information Science of Wuhan University, 2018, 43(1): 101-106.
- [22] 黄伟杰, 张国山. 基于快速不变卡尔曼滤波的视觉惯性里程计[J]. 控制与决策, 2019, 34(12): 2585-2593.
- HUANG Weijie, ZHANG Guoshan. Visual-inertial odometry based on fast invariant Kalman filter[J]. Control and Decision, 2019, 34(12): 2585-2593.
- [23] 闫德立, 喻薇, 宋宇, 等. 基于矩阵李群表示及容积卡尔曼滤波的视觉惯导里程计新方法[J]. 控制与决策, 2020, 35(8): 1823-1832.
- YAN Deli, YU Wei, SONG Yu, et al. A new method for visual inertial odometry based on cubature Kalman filter and matrix Lie group representation[J]. Control and Decision, 2020, 35(8): 1823-1832.
- [24] 吴禹彤, 张林. 新型优化策略在视觉惯性导航系统中的应用[J]. 计算机应用, 2020, 40(S2): 215-220.
- WU Yutong, ZHANG Lin. Application of novel optimization strategy in visual inertial navigation system[J]. Journal of Computer Applications, 2020, 40(S2): 215-220.
- [25] 李永锋, 张国良, 王峰, 等. 基于快速视觉里程计和大回环局部优化模型的改进VSLAM算法[J]. 机器人, 2015, 37(5): 557-565.
- LI Yongfeng, ZHANG Guoliang, WANG Feng, et al. Improved VSLAM algorithm based on fast visual odometry and large loop local optimization model[J]. Robot, 2015, 37(5): 557-565.
- [26] 代波, 何玉庆, 谷丰, 等. 结合滤波与优化的无人机多传感器融合方法[J]. 中国科学: 信息科学, 2020, 50(12): 1919-1931.
- DAI Bo, HE Yuqing, GU Feng, et al. Multi-sensor

- fusion for unmanned aerial vehicles based on the combination of filtering and optimization[J]. *Scientia Sinica (Informationis)*, 2020, 50(12): 1919-1931.
- [27] 杨雪梅, 李帅永. 移动机器人视觉SLAM回环检测原理、现状及趋势[J]. *电子测量与仪器学报*, 2022, 36(8): 1-14.
- YANG Xuemei, LI Shuaiyong. The principle, current situation and trend of visual SLAM loop closure detection for mobile robot[J]. *Journal of Electronic Measurement and Instrumentation*, 2022, 36(8): 1-14.
- [28] 刘凌云, 蔡成林, 吴芊. 室外光照易变场景下的回环检测方法[J]. *传感器与微系统*, 2022, 41(4): 121-124.
- LIU Lingyun, CAI Chenglin, WU Qian. Loop closure detection method in outdoor scene with variable illumination[J]. *Transducer and Microsystem Technologies*, 2022, 41(4): 121-124.
- [29] 应桃, 严怀成, 吴滢跃, 等. 基于最短路径的大视角变换下SLAM回环检测[J]. *控制工程*, 2021, 28(11): 2158-2164.
- YING Tao, YAN Huaicheng, WU Yingyue, et al. SLAM loop closure detection based on the shortest path under perspective distortions[J]. *Control Engineering of China*, 2021, 28(11): 2158-2164.
- [30] 张光耀, 倪益华, 吕艳, 等. 一种基于改进词袋模型的视觉SLAM算法[J]. *河南理工大学学报(自然科学版)*, 2021, 40(4): 154-161.
- ZHANG Guangyao, NI Yihua, LYU YAN, et al. A visual SLAM algorithm based on improved word bag model[J]. *Journal of Henan Polytechnic University (Natural Science)*, 2021, 40(4): 154-161.
- [31] 郑冰清, 刘启汉, 赵凡, 等. 一种融合语义地图与回环检测的视觉SLAM方法[J]. *中国惯性技术学报*, 2020, 28(5): 629-637.
- ZHENG Bingqing, LIU Qihan, ZHAO Fan, et al. Loop detection and semantic mapping algorithm fused with semantic information[J]. *Journal of Chinese Inertial Technology*, 2020, 28(5): 629-637.
- [32] 郭烈, 葛平淑, 王肖, 等. 基于卷积神经网络优化回环检测的视觉SLAM算法[J]. *西南交通大学学报*, 2021, 56(4): 706-712, 768.
- GUO Lie, GE Pingshu, WANG Xiao, et al. Visual simultaneous localization and mapping algorithm based on convolutional neural network to optimize loop detection[J]. *Journal of Southwest Jiaotong University*, 2021, 56(4): 706-712, 768.
- [33] 杨慧, 张婷, 金晟, 等. 基于二进制生成对抗网络的视觉回环检测研究[J]. *智能系统学报*, 2021, 16(4): 673-682.
- YANG Hui, ZHANG Ting, JIN Sheng, et al. Visual loop closure detection based on binary generative adversarial network[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(4): 673-682.
- [34] SALASMORENO R F, GLOCKEN B, KELLY P, et al. Dense planar SLAM[C]//*Proceedings of 2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. Munich, Germany: IEEE, 2014: 157-164.
- [35] 王雅婷, 刘建业, 熊智, 等. 基于双目视觉的类脑三维认知地图构建方法[J]. *导航定位与授时*, 2021, 8(5): 9-17.
- WANG Yating, LIU Jianye, XIONG Zhi, et al. A method of constructing brain-inspired 3D cognitive maps based on binocular vision[J]. *Navigation Positioning and Timing*, 2021, 8(5): 9-17.
- [36] 江文婷, 龚小谨, 刘济林. 基于增量计算的大规模场景致密语义地图构建[J]. *浙江大学学报(工学版)*, 2016, 50(2): 385-391.
- JIANG Wenting, GONG Xiaojin, LIU Jilin. Incremental large scale dense semantic mapping[J]. *Journal of Zhejiang University (Engineering Science)*, 2016, 50(2): 385-391.
- [37] BERGEN J, NISTER D, NARODITSKY O. Visual odometry[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Washington DC, USA: IEEE, 2004: 652-659.
- [38] 马科伟, 张镔石, 康宇航, 等. 移动机器人中视觉里程计技术综述[J]. *计算机工程*, 2021, 47(11): 1-10.
- MA Kewei, ZHANG Qieshi, KANG Yuhang, et al. Overview of visual odometry technology in mobile robots[J]. *Computer Engineering*, 2021, 47(11): 1-10.
- [39] DAVISON A J, REID I D, MOLTON N D. MonoSLAM: Real-time single camera SLAM[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2007, 29(6): 1052-1067.
- [40] KLEIN G, MURRAY D. Parallel tracking and mapping for small AR workspaces[C]//*Proceedings of the 6th IEEE & ACM International Symposium on Mixed & Augmented Reality*. Nara, Japan: IEEE, 2007: 225-234.
- [41] 柴松. 基于视觉SLAM的室内建图和路径规划研究[D]. 天津: 天津理工大学, 2021.
- CHAI Song. Research on indoor mapping and path planning method based on visual SLAM Technology [D]. Tianjin: Tianjin University of Technology, 2021.
- [42] LABBE M, MICHAUD F. Online global loop closure detection for large-scale multi-session graph-based SLAM [C]//*Proceedings of 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Chicago, USA: IEEE, 2014: 2661-2666.
- [43] CUMMINS M, NEWMAN P. Appearance-only

- SLAM at large scale with FAB-MAP 2.0[J]. The International Journal of Robotics Research, 2011, 30(9): 1100-1123.
- [44] 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.
- [45] MUR-ARTAL R, TARDOS J D. ORB-SLAM2: An open-source SLAM system for monocular, stereo and RGB-D cameras[J]. IEEE Transactions on Robotics, 2017, 33(5): 1-8.
- [46] QUAN M, PIAO S, HE Y, et al. Monocular visual SLAM with points and lines for ground robots in particular scenes: Parameterization for lines on ground[J]. Journal of Intelligent & Robotic Systems, 2021, 101(4): 72-90.
- [47] NEWCOMBE R A, LOVEGROVE S J, DAVISON A J. DTAM: Dense tracking and mapping in real-time[C]//Proceedings of 2021 International Conference on Computer Vision (ICCV). Barcelona, Spain: IEEE, 2011: 2320-2327.
- [48] FORSTER C, PIZZOLI M, SCARAMUZZA D. SVO: Fast semi-direct monocular visual odometry [C]//Proceedings of 2014 IEEE International Conference on Robotics & Automation (ICRA). Hong Kong, China: IEEE, 2014: 15-22.
- [49] ENGEL J, SCHOPS T, CREMERS D. LSD-SLAM: Large-scale direct monocular SLAM[C]//Proceedings of European Conference on Computer Vision. Heidelberg: Springer, 2014: 834-849.
- [50] ENGEL J, KOLTUN V, CREMERS D. Direct sparse odometry[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2018, 40(3): 611-625.
- [51] LIANG Z, WANG C. A semi-direct monocular visual SLAM algorithm in complex environments[J]. Journal of Intelligent & Robotic Systems, 2020, 101(1): 25-43.
- [52] TANG F, LI H, WU Y. FMD stereo SLAM: Fusing MVG and direct formulation towards accurate and fast stereo SLAM[C]//Proceedings of 2019 International Conference on Robotics and Automation (ICRA). Montreal, Canada: IEEE, 2019: 133-139.
- [53] FUENTES-PACHECO J, ASCENCIO J R, RENDON-MANCHA J M. Visual simultaneous localization and mapping: A survey[J]. Artificial Intelligence Review: An International Science and Engineering Journal, 2015, 43(1): 55-81.
- [54] 陈俊芬, 赵佳成, 翟俊海, 等. 基于无监督学习视觉特征的深度聚类方法[J]. 南京航空航天大学学报, 2021, 53(5): 718-725.
- CHEN Junfen, ZHAO Jiacheng, ZHAI Junhai, et al. Deep clustering method based on unsupervised visual features learning[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2021, 53(5): 718-725.
- [55] 张大伟, 苏帅. 自主移动机器人视觉 SLAM 技术研究[J]. 郑州大学学报(理学版), 2021, 53(1): 1-8.
- ZHANG Dawei, SU Shuai. Research on visual SLAM technology of autonomous mobile robot[J]. Journal of Zhengzhou University (Natural Science Edition), 2021, 53(1): 1-8.
- [56] ROBERTS R, NGUYEN H, KRISHNAMURTHI N, et al. Memory-based learning for visual odometry[C]//Proceedings of 2008 IEEE International Conference on Robotics and Automation (ICRA). Pasadena, USA: IEEE, 2008: 47-52.
- [57] TATENO K, TOMBARI F, LAINA I, et al. CNN-SLAM: Real-time dense monocular SLAM with learned depth prediction[C]//Proceedings of 2017 IEEE Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017: 6566-6574.
- [58] BLOESCH M, CZARNOWSKI J, CLARK R, et al. CodeSLAM-learning a compact, optimisable representation for dense visual SLAM[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE Computer Society Press, 2018: 2560-2568.
- [59] LOO S Y, AMIRI A J, MASHOHOR S, et al. CNN-SVO: Improving the mapping in semi-direct visual odometry using single-image depth prediction[C]//Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). Montreal, Canada: IEEE, 2019: 5218-5223.
- [60] ZHAN H Y, WEERASEKERA C S, BIAN J W, et al. Visual odometry revisited: What should be learnt? [C]//Proceedings of 2020 IEEE International Conference on Robotics and Automation (ICRA). Paris, France: IEEE, 2020: 4203-4210.
- [61] LI A, WANG J, XU M, et al. DP-SLAM: A visual SLAM with moving probability towards dynamic environments[J]. Information Sciences, 2021, 556: 128-142.
- [62] 曹剑飞, 余金城, 潘尚杰, 等. 采用双视觉里程计的 SLAM 位姿图优化方法[J]. 计算机辅助设计与图形学学报, 2021, 33(8): 1264-1272.
- CAO Jianfei, YU Jincheng, PAN Shangjie, et al. A SLAM pose graph optimization method using dual visual odometry[J]. Journal of Computer-Aided Design & Computer Graphics, 2021, 33(8): 1264-1272.
- [63] KENDALL A, GRIMES M, CIPOLLA R, et al. PoseNet: A convolutional network for real-time 6-DOF camera relocalization[C]//Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015: 2938-2946.

- [64] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE, 2015: 1-9.
- [65] LI R H, WANG S, LONG Z Q, et al. UnDeepVO: Monocular visual odometry through unsupervised deep learning[C]//Proceedings of 2018 IEEE International Conference on Robotics and Automation. Brisbane, Australia: IEEE, 2018: 7286-7291.
- [66] UMMENHOFER B, ZHOU H Z, UHRIG J, et al. DeMoN: Depth and motion network for learning monocular stereo[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017: 5038-5047.
- [67] GODARD C, AODHA O M, FIRMAN M, et al. Digging into self-supervised monocular depth estimation[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea: IEEE, 2019: 3827-3837.
- [68] ZHANG L Q, LI G, LI T H. Temporal-aware SfM-Learner: Unsupervised learning monocular depth and motion from stereo video clips[C]//Proceedings of 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). Shenzhen, China: IEEE, 2020: 253-258.
- [69] AHN S, LEE K, CHUNG W K, et al. SLAM with visual plane: Extracting vertical plane by fusing stereo vision and ultrasonic sensor for indoor environment[C]//Proceedings 2007 IEEE International Conference on Robotics and Automation. Rome, Italy: IEEE, 2007: 4787-4794.
- [70] 张礼廉, 屈豪, 毛军, 等. 视觉/惯性组合导航技术发展综述[J]. 导航定位与授时, 2020, 7(4): 50-63.
ZHANG Lilian, QU Hao, MAO Jun, et al. A survey of intelligence science and technology integrated navigation technology[J]. Navigation Positioning and Timing, 2020, 7(4): 50-63.
- [71] 胡凯, 吴佳胜, 郑翡, 等. 视觉里程计研究综述[J]. 南京信息工程大学学报(自然科学版), 2021, 13(3): 269-280.
HU Kai, WU Jiasheng, ZHANG Fei, et al. A survey of visual odometry[J]. Journal of Nanjing University of Information Science & Technology (Natural Science Edition), 2021, 13(3): 269-280.
- [72] 杨观赐, 王霄远, 蒋亚汶, 等. 视觉与惯性传感器融合的SLAM技术综述[J]. 贵州大学学报(自然科学版), 2020, 37(6): 1-12.
YANG Guanci, WANG Xiaoyuan, JIANG Yawen, et al. Review of SLAM technology based on visual and inertial sensor fusion[J]. Journal of Guizhou University (Natural Sciences), 2020, 37(6): 1-12.
- [73] LUO W, XIONG Z, XING L, et al. An IMU/visual odometry integrated navigation method based on measurement model optimization[C]//Proceedings of 2018 IEEE CSAA Guidance, Navigation and Control Conference (CGNCC). Xiamen, China: IEEE, 2018: 1-5.
- [74] 杨子寒, 赖际舟, 吕品, 等. 一种可参数自标定的鲁棒视觉/惯性定位方法[J]. 仪器仪表学报, 2021, 42(7): 259-267.
YANG Zihan, LAI Jizhou, LYU PIN, et al. A robust visual/inertial positioning method with parameter self-calibration[J]. Chinese Journal of Scientific Instrument, 2021, 42(7): 259-267.
- [75] WEISS S. Vision based navigation for micro helicopters[D]. Switzerland: ETH, 2012.
- [76] BLOESCH M, OMARI S, HUTTER M, et al. Robust visual inertial odometry using a direct EKF-based approach[C]//Proceedings of 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Hamburg, Germany: IEEE, 2015: 298-304.
- [77] LEUTENEGGER S, FURGALE P, RABAU D V, et al. Keyframe-based visual-inertial SLAM using non-linear optimization[C]//Proceedings of Robotics: Science and Systems. Berlin, Germany: The Technische Universitat Berlin Press, 2013.
- [78] 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.
- [79] MITTAL M, VALADA A, BURGARD W. Vision-based autonomous landing in catastrophe-struck environments[EB/OL]. (2018-09-15). <https://arxiv.org/abs/1809.05700>.
- [80] 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.
- [81] QIN T, PAN J, CAP S, et al. A general optimization-based framework for global pose estimation with multiple sensors[EB/OL]. (2019-01-11). <https://arxiv.org/abs/1901.03642>.
- [82] 苗升, 刘小雄, 黄剑雄, 等. 无人机视觉SLAM环境感知发展研究[J]. 计算机测量与控制, 2021, 29(8): 1-6, 41.
MIAO Sheng, LIU Xiaoxiong, HUANG Jianxiong, et al. Research on development of UAV visual SLAM environment perception[J]. Computer Measurement & Control, 2021, 29(8): 1-6, 41.
- [83] 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.
- [84] USENKO V, DEMMEL N, SCHUBERT D, et al. Visual-inertial mapping with non-linear factor recovery[J]. IEEE Robotics and Automation Letters, 2020, 5(2): 422-429.
- [85] CAMPOS C, ELVIRA R, RODRÍGUEZ J J G, et al. ORB-SLAM3: An accurate open-source library for visual, visual-inertial and multi-map SLAM[J]. IEEE Transactions on Robotics, 2020, 37(6): 1-17.
- [86] 曹力科, 肖晓晖. 基于卷帘快门 RGB-D 相机的视觉惯性 SLAM 方法[J]. 机器人, 2021, 43(2): 193-202.
- CAO Like, XIAO Xiaohui. A visual-inertial SLAM method based on rolling shutter RGB-D cameras[J]. Robot, 2021, 43(2): 193-202.
- [87] BHATNAGAR G, WU Q M J. A fractal dimension based framework for night vision fusion[J]. IEEE/CAA Journal of Automatica Sinica, 2019, 6(1): 220-227.
- [88] 马翰飞, 范海震, 李强, 等. 基于多光谱融合图像的飞机导航系统设计[J]. 电子设计工程, 2019, 27(24): 161-166.
- MA Hanfei, FAN Haizhen, LI Qiang, et al. Navigation system design based on multi-spectral fusion image[J]. Electronic Design Engineering, 2019, 27(24): 161-166.
- [89] 赵崑岑. 双目姿态测量与多光谱双目测距技术研究[D]. 西安: 中国科学院大学(中国科学院西安光学精密机械研究所), 2018.
- ZHAO Yincen. Research on binocular attitude measurement and multispectral binocular ranging technology[D]. Xi'an: University of Chinese Academy of Sciences (Xi'an Institute of Optics & Precision Mechanics, Chinese Academy of Sciences), 2018.
- [90] YAN X, BI Y, GAO Z, et al. Multi-plane visual odometry for unmanned aerial vehicle using a thermal camera[C]//Proceedings of 2018 IEEE the 14th International Conference on Control and Automation (ICCA). Anchorage, USA: IEEE, 2018: 588-593.
- [91] 褚金奎, 陈建华, 李金山, 等. 一种偏振光/双目视觉仿生组合导航方法[J]. 光子学报, 2021, 50(5): 192-201.
- CHU Jinkui, CHEN Jianhua, LI Jinshan, et al. Polarized light/binocular vision bionic integrated navigation method[J]. Acta Photonica Sinica, 2021, 50(5): 192-201.
- [92] BEAUVISAGE A, AHISKA K, AOUF N. Robust multispectral visual-inertial navigation with visual odometry failure recovery[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(7): 9089-9101.
- [93] BEAUVISAGE A, AOUF N. Multimodal visual-inertial odometry for navigation in cold and low contrast environment[C]//Proceedings of 2017 European Conference on Mobile Robots (ECMR). Paris, France: IEEE, 2017: 1-6.
- [94] YAN X, BI Y, GAO Z, et al. Laser-aided infrared visual inertial odometry for unmanned aerial vehicle with multi-plane constraint[C]//Proceedings of the 3rd International Conference on Advanced Robotics and Mechatronics (ICARM). Singapore: IEEE, 2018: 323-327.
- [95] DELAUNE J, HEWITT R, LYTLE L, et al. Thermal-inertial odometry for autonomous flight throughout the night[C]//Proceedings of 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Macau, China: IEEE, 2019: 1122-1128.

(编辑: 张蓓)