

文章编号: 2095-2163(2022)07-0040-06

中图分类号: TP202

文献标志码: A

移动机器人视觉 SLAM 研究综述

李延真, 石立国, 徐志根, 程超, 夏清泉

(国网山东省电力公司青岛供电公司, 山东 青岛 266001)

摘要: 基于视觉传感器的同步定位与建图技术是移动机器人导航的关键技术之一。视觉 SLAM 系统具有成本低、安装简便等优点, 相较于激光 SLAM 能够获取环境中的大量纹理以及色彩信息, 提取更多的特征信息, 具有更好的场景辨识能力。本文基于当前移动机器人背景下视觉 SLAM 的研究状况, 概述了视觉 SLAM 的原理以及经典方法, 并对研究热点展开综述, 探讨了视觉 SLAM 存在的问题和发展趋势。

关键词: 移动机器人; 同时定位与建图; 视觉; 动态环境

A review of vision SLAM methods on mobile robot

LI Yanzhen, SHI Liguang, XU Zhigen, CHENG Chao, XIA Qingquan

(State Grid QingDao Power Supply Company, QingDao Shandong 266001, China)

【Abstract】 Simultaneous Localization And Mapping (SLAM) base on vision sensors is one of the key technologies of mobile robot navigation. The visual SLAM system has the advantages of low cost and easy installation. Compared with laser SLAM, visual SLAM can obtain a large amount of texture and color information in the environment, extract more feature information, and have better scene recognition capabilities. Based on the current research status of visual SLAM for mobile robots, this paper reviews the principle, classic methods and research hotspots of visual SLAM, and discusses the problems and development trends of visual SLAM.

【Key words】 mobile robot; SLAM; vision; dynamic environment

0 引言

同时定位与建图(Simultaneous Localization And Mapping, SLAM)是移动机器人领域的一个具有挑战性的难题。目前 SLAM 技术研究主要集中于两种方法, 一是基于便携式激光测距仪的方法, 即激光 SLAM, 另一种是基于计算机视觉的方法, 即视觉 SLAM。采用激光雷达构建的点云地图, 在某种程度上无法展现较好的环境细节信息, 以至于智能机器人在环境复杂的场景中, 不能有效的进行环境信息感知和决策。随着计算机技术以及视觉传感器技术的发展, 视觉 SLAM 采用相机来替代激光雷达重构周围环境的 3D 地图, 已取得了巨大且快速的发展。而且, 图像中包含更加丰富的环境特征信息, 以使得机器人能够在更大范围内完成任务。此外, 视觉传感器的发展以及深度相机、立体相机的出现和改进, 吸引了许多学者对视觉 SLAM 技术进行了大量的研究^[1]。视觉 SLAM 主要由特征提取、特征跟踪、运动

跟踪、闭环检测、地图构建、位姿估计等部分组成。

1 视觉 SLAM 概述

经典视觉 SLAM 框架如图 1 所示, 通常包含传感器数据输入、前端、后端、地图构建以及闭环检测。前端获取传感器原始数据, 并对数据进行预处理。例如特征提取、短期和长期数据关联等操作, 以便将几何信息转换为数学模型并将其发送到后端。后端对前端的输入模型进行优化、最小化相机姿态的累计误差, 以及地图信息的优化调整。闭环检测将相机图像的检测结果送到后端处理, 通过计算图像相似性, 对机器人经过的场景进行识别比对, 从而实现累计误差的消除^[2]。

根据视觉传感器的不同, 视觉 SLAM 主要分为单目、RGB-D、立体视觉 SLAM 等方法。其中, 采用单相机解决 SLAM 问题的方案称为单目 SLAM; 而 RGB-D SLAM 方法不仅需要单目相机, 并且需要用到红外传感器; 立体视觉 SLAM 则需要不同方位

基金项目: 国网山东省电力公司科技项目资助(520602200002)。

作者简介: 李延真(1986-) 男, 学士, 高级工程师, 主要研究方向: 智能配电网、智能机器人; 石立国(1984-) 男, 博士, 工程师, 主要研究方向: 智能电网; 徐志根(1983-) 男, 硕士, 高级工程师, 主要研究方向: 主动配电网; 程超(1987-) 女, 硕士, 高级工程师, 主要研究方向: 智能电网信息采集; 夏清泉(1971-) 男, 硕士, 工程师, 主要研究方向: 软件工程。

收稿日期: 2021-12-01

哈尔滨工业大学主办 ◆ 学术研究与应用

安装多个相机。视觉传感器一般具有视觉里程测量功能,具有足够的稳定性和鲁棒性,而且易于实现。

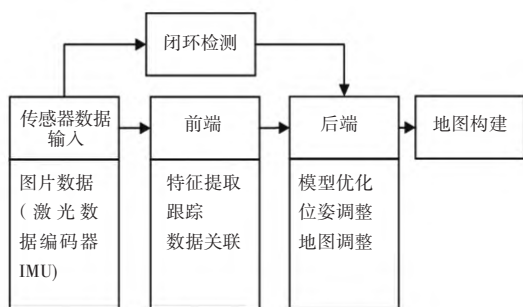


图 1 经典 SLAM 框架

Fig. 1 Classic SLAM framework

近十年来,SLAM 算法有很多,主要分为基于直接法的视觉 SLAM^[3,12-15]和基于特征点法的视觉 SLAM^[4-11]两大类。

1.1 经典 SLAM 方法

1.1.1 基于特征点的视觉 SLAM

Davison 等人^[4]提出了一种单目 SLAM 方法,称为 MonoSLAM。MonoSLAM 采用 EKF 算法建立环境特征点的地图,这种地图虽有一定限制,但在解决单目特征初始化的问题上足够稳定。此外,地图的稀疏性也暴露了机器人在需要更多环境细节的情况下,无法完成定位任务的问题。为此,研究出现了 UKF 方法^[5]以及改进的 UKF 方法^[6],用于解决视觉 SLAM 的线性不确定性。而 Sim 等人^[7]提出的基于 PF 的单目 SLAM 方法可以构建更精确的映射,但 PF 方法的算法具有很高的计算复杂度,以至于无法在大型环境下应用。Klein^[8]等人设计了一种基于关键帧的单目 SLAM 方法,即 PTAM。在该方法中,跟踪和建图分为两个并行化的任务。关键帧提取技术,即通过数个关键图像串联,优化地图和运动轨迹,从而避免了对每一幅图像的细节进行处理^[9]。这种方法采用非线性优化来替代 EKF 方法解决线性化的困难,进而减少了机器人在定位中的不确定性。但是,由于 PTAM 存在全局优化的问题,使得该方法无法在大型环境中应用。

2015 年,Mur-Artal 等人^[10]提出了一种新的实时视觉 SLAM 方法——ORB-SLAM。该方法是一种基于特征法的单目 SLAM,实时估计 3D 特征位置和重建环境地图,其特征计算具有良好的旋转和缩放不变性,具有较高的定位精度。但该方法使得 CPU 运算负担大,生成的地图仅用于定位需求,无法用于

导航和避障需求。为此,Mur-Artal 等于 2017 年又提出了一种改进算法,即基于 ORB-SLAM 的 ORB-SLAM2,其不仅支持 RGB-D 相机以外的单目相机还支持使用立体相机。

然而,ORB-SLAM2 是通过大规模数据生成训练词汇,当词汇量较大时,其过程对于移动机器人来说是非常耗时的。在机器人工作环境固定的情况下,使用大数据集会导致大量无效数据的产生。此外,ORB-SLAM2 还缺乏离线可视化和轨迹建图的能力。

1.1.2 基于直接法的视觉 SLAM

LSD-SLAM^[12]、DTAM^[13]是基于直接法的单目 SLAM 方法,使用 RGB 图像作为输入,通过所有像素强度估计相机的帧轨迹和重建环境的 3D 地图。DTAM 是一种直接稠密的方法,通过在相机视频流中提取多张静态场景图片来提高单个数据信息的准确性,从而实时生成精确的深度地图。该方法计算复杂度比较大,需要 GPU 并行运算,对全局照明处理的鲁棒性较差。

LSD-SLAM 能够构建一个半稠密的全局稳定的环境地图,包含了更全面的环境表示,在 CPU 上实现了半稠密场景的重建。该方法对相机内参敏感和曝光敏感,需要特征点进行回环检测,无法在照明不规律变化的场景中应用。

Forster 等人^[14]提出的 SVO (Semi-direct Visual Odometry) 是一种半直接法的视觉里程计,其是特征点和直接法的混合使用,该方法的时间复杂度较低,但是,该方法舍弃了后端优化和回环检测,而且位姿估计会产生累积误差,因此在移动机器人丢失位置后重定位比较困难。

DSO (Direct Sparse Odometry)^[15]也是一种半直接法的视觉里程计,基于高度精确的稀疏直接结构和运动公式。该方法能够直接优化光度误差,考虑了光度标定模型,该方法不仅完善了直接法位姿估计的误差模型,还加入了仿射亮度变换、光度标定、深度优化等方法,在无特征的区域中也可以使其具有鲁棒性。但是,该方法舍弃了回环检测。

2 视觉 SLAM 研究热点

2.1 视觉与惯性传感器融合的 SLAM

VI-SLAM (Visual-Inertial SLAM) 将视觉传感器和 IMU 优势结合^[16],从而为移动机器人提供更加丰富的运动信息和环境信息。其主要方式将视觉前段信息与 IMU 信息结合,即视觉惯性里程计

(VIO)采用滤波技术以及优化方法,对采集的物理量信息进行处理,进而实现对自身的运动和环境信息估计^[17]。当视觉传感器在短时间内快速运动失效时,融合IMU数据能够为视觉提供短时的精准定位,同时利用视觉定位信息来估计IMU的零偏,减少IMU由零偏导致的发散和累积误差^[18]。通过二者的融合,可以解决视觉位姿估计输出频率低的问题,同时位姿估计精度有一定的提高,整个系统也更加鲁棒。目前VI-SLAM已在机器人、无人机、无人驾驶、AR和VR等多个领域有所应用^[19]。

MSCKF^[20]算法将视觉与惯性信息在EKF框架下融合,相较于单纯的VO算法,该算法能够应用在运动剧烈、纹理短时间缺失等环境中,而且鲁棒性更好;相较于基于优化的VIO算法(VINS,OKVIS),MSCKF精度相当、速度更快,适合在计算资源有限的嵌入式平台运行。ROVIO^[21]是基于单目相机开发的紧耦合VIO系统,通过对图像块的滤波实现VIO,利用扩展卡尔曼滤波进行状态估计,使用速度更快的FAST来提取角点,其三维坐标用向量和距离表示。其次,所有角点是通过图像块进行描述,并通过视频流获取了多层次表达;最后利用IMU估计的位姿来计算特征投影后的光度误差,并将其用于后续优化。虽然该算法计算量小,但对应不同的设备需要调参数(参数对精度很重要),并且没有闭环,经常存在误差,会残留到下一时刻。

OKVIS^[22]是一种使用非线性优化基于关键帧的视觉惯性SLAM技术。OKVIS的算法流程是通过IMU测量值对当前状态做预测,根据预测进行特征提取和特征匹配,3D点特征和二维图像特征构成优化中重投影,同时预测IMU状态量和优化的参数之间构成IMU测量误差,两项误差放在一起做优化。OKVIS不支持重定位,也没有闭环检测或校正方案。VINS-Mono^[23]是香港科技大学开源的一个VIO算法,用于估计器初始化和故障恢复,采用基于紧耦合、非线性优化的方法。通过融合预积分的IMU测量数据和特征观测数据获得高精度的视觉惯性里程计。VINS与OKVIS相比,具有更加完善和鲁棒的初始化以及闭环检测过程。

2.2 视觉与激光雷达融合的SLAM

视觉与激光雷达融合的SLAM,是将激光雷达在建图和距离测量时准确度较好的优势,与视觉方法构建环境信息较准确的优势相结合,在一定程度上能够避免单相机在使用过程中的单目尺度漂移、双目深度估计精度不高、户外RGB-D稠密重建困

难的缺陷,将有效提升SLAM性能。其确定是标定和融合比较困难。

视觉与激光雷达融合的SLAM主要分为:改进的视觉SLAM、改进的激光SLAM以及并行激光与视觉SLAM。

J.Graeter等人^[24]提出了一种激光-视觉里程计方法(LiDAR-Monocular Visual Odometry, LIMO)。该方法从LiDAR中提取图片中特征点的深度,不仅考虑局部平面假设的外点,并考虑了地面点。Shin等人^[25]利用LiDAR提供的稀疏深度,提出了一种基于单目相机直接法的视觉SLAM框架。采用滑动窗口进行追踪的方法,忽略旧的关键帧,不在大规模场景下集成了深度的帧与帧的匹配方法。这种方法相机分辨率比激光雷达分辨率高,从而导致许多像素缺失深度信息。为处理相机和激光雷达分辨率匹配的问题,De Silva等人^[26]在计算两个传感器之间的几何变换后,采用高斯过程回归,对缺失值进行插值。Scherer等人^[27]采用VIO对机器人的状态进行估计,采用激光雷达进行障碍物及边界检测,但是激光雷达的点云数据可能包含遮挡点,从而对精度有一定的影响。为提高精度,Huang等人^[28]提出一种基于直接法的SLAM方法,采用遮挡点检测器和共面点检测器解决这一问题。

在一些研究中,视觉-激光SLAM,采用激光雷达扫描匹配进行运动估计,相机进行特征检测,从而对纯激光SLAM的缺陷进行相应的弥补。Liang等人^[29]提出了一种解决激光和相机传感器集成的大规模激光碰撞中的闭环问题。其利用ORB特征和词袋特征,实现了环路检测的快速、鲁棒性。通过在不同大规模环境下的实验,验证了该方法的有效性。Zhu等人^[30]将3D激光SLAM和视觉关键帧词袋回环检测相融合,并对最近点迭代(ICP)进行优化。Pandey等人^[31]利用3D点云与可用相机图像的共配准,将高尺度特征描述符(如尺度不变特征变换(SIFT)或加速鲁棒特征(SURF))与3D点相关联,即利用视觉信息对刚性转换做了预测,并且建立通用的ICP框架。

此外,并发视觉激光融合也是一个研究方向。Seo等人^[32]同时采用激光雷达和视觉传感器并行构建了两个地图:激光雷达立体像素地图和具有地图点的视觉地图,并在后端优化中运用残差对里程求解,使其保持全局一致,从而能够更好的进行状态估计,是一种紧耦合的方法。Zhang等人^[33]提出一种视觉雷达里程计的通用框架。该方法使用高频运

行的视觉里程计作为图像帧速率(60 Hz)来估计运动,并使用低频率(1 Hz)激光雷达测距仪细化运动估计,并消除由视觉测距漂移引起的点云失真。

2.3 基于深度学习的视觉SLAM

将深度学习理论引入视觉SLAM中,主要应用包含:使用深度神经网络对单目视觉深度进行估计;将深度学习与视觉SLAM前端结合,从而提高图像特征提取的准确度;融入物体识别、目标检测、语音分割等技术,进而增加对周围环境信息的感知与理解。

Zhou等人^[34]提出了一种单目深度和位姿估计的无监督学习网络,其特点是完全无监督网络,从非结构化视频序列进行单视图深度估计和多视图位姿估计,但得到的深度和位姿缺乏系统尺度。Godard等人^[35]提出了一种卷积神经网络,替代了直接用深度图数据训练,采用容易获得的双目立体视觉的角度,在没有参考深度数据的情况下,估计单个图像的深度,从而可以执行端到端无监督单眼深度估计,并加强左右视差图的一致性,从而提升性能和鲁棒性。Mahjourian等人^[36]提出了一种基于无监督学习的新方法,对单目视频中的深度与自我运动的估计,最终作者在KITTI数据集和手机拍摄的微景观标定的视频数据集上进行了算法验证。Daniel DeTone等人^[37]提出了一种DeepSlam,在存在图像噪声的情况下进行特征点检测,相比传统方案,具有显著的性能差距。此外,DeTone等人^[38]还提出了一个自监督框架SuperPoint,适用于计算机视觉中大量多视图几何问题的兴趣点检测器和描述符的训练。Jiexiong Tang等人^[39]提出了一种基于深度学习的GCNv2网络(GCN-SLAM),用于生成关键点和描述符。

JohnMcCormac等人^[40]提出了Semanticfusion方法,使用CNN进行语义分割,加上条件随机场对分割结果的优化,设计了semantic mapping系统,是一种将CNN和最先进的稠密SLAM融合方案。Thomas Whelan等人^[41]提出了ElasticFusion,用来构建语义三维地图。Bowman等人^[42]提出一种融合尺度信息与语义信息的理论框架,通过对目标进行相应的检测,将尺度信息和语义信息结合,从而实现对运动估计和地图的优化。

2.4 动态环境下的视觉SLAM

目前,大部分的研究都是基于静态环境,而且光线良好,且为非单调纹理特征的情形。但实际生活

场景中还会有大量动态的行人或者物体,所以研究动态环境下的SLAM也是极其重要的。其关键技术就是将动态的行人或物品等特征点,能够在地图中过滤移除,避免对定位和闭环检测产生不良影响。

为此,Wei Tan等人^[43]提出了RDSLAM(Robust monocular slam)方法,其是一种基于关键帧的在线表示和更新方法的实时单目SLAM系统。该方法可以处理缓慢变化的动态环境,能够检测变化并及时更新地图。Chao Yu等人^[44]提出了一个面对动态环境的语义视觉SLAM系统DS-SLAM。该方法结合语义信息和运动特征点检测,来滤除每一帧中的动态物体,从而提高位姿估计的准确性,同时建立语义八叉树地图。其是基于优化ORB-SLAM的方法,使其具有更好的鲁棒性。MaskFusion^[45]是由RÜNZ M等人提出的一个实时的、具备对象感知功能的、语义和动态RGB-D SLAM系统。该方法在连续的、自主运动中,能够在跟踪和重建的同时,识别分割场景中不同的物体并分配语义类别标签。DynaSLAM^[46]是一个在动态环境下辅助静态地图的SLAM系统,通过增加运动分割方法使其在动态环境中具有稳健性,并且能够对动态物品遮挡的部分进行修复优化,生成静态场景地图。StaticFusion^[47]是一种面向动态环境基于面元的RGB-D SLAM系统,能够在动态环境中检测运动目标并同时重建背景结构,但该方法的初始若干帧内不能有大量动态物体,否则初始静态场景面元地图的不准确性增加。

3 结束语

在视觉SLAM发展的三十年里,已取得了重大的成果,形成一些常用的框架方法,在机器人、无人机、无人驾驶、AR和VR等多个领域已有所应用。针对视觉SLAM经典研究方法的特点做一总结,见表1。

综上所述,目前已有的SLAM方法仍然在计算力问题、室外动态大规模地图构建、地图复用等方面存在不足,以及在实时性与准确性问题上难以兼顾。目前,SLAM应用场景不断变化,移动环境场景也不断增多,SLAM算法在动态物体及动态环境方面要求也不断升高,同时实时性要求也随之提高,这也成为今后SLAM算法的研究方向,随着新型传感器以及更多算法的研究与发展,视觉SLAM也不断提高精度及鲁棒性,平衡实时性和准确性,向可移植、多传感器融合以及智能语义SLAM的方向发展。

表1 视觉SLAM经典方法总结
Tab. 1 Summary of classical visual SLAM

名称	方法类别	应用特点或局限性	发表年份
MonoSLAM	特征点法、单目	地图的稀疏性缺少环境细节	2007
PTAM	特征点法、单目	无法在大型环境中应用	2007
ORB-SLAM2	特征点法、单目/双目/RGB-D	地图较大时,运算耗时;缺乏离线可视化和轨迹建图的能力	2015
LSD-SLAM	直接法、单目	无法在照明不规则变化的场景中应用	2014
DTAM	直接法、RGB-D	该方法计算复杂度较大,需要GPU性能较高	2011
SVO	半直接法、单目	舍弃了后端优化和回环检测,重定位比较困难	2014
DSO	半直接法、单目	舍弃了回环检测	2016
MSCKF	基于滤波的方法、单目	适合在计算资源有限的嵌入式平台运行	2007
ROVIO	基于滤波的方法、单目	缺少闭环,误差会残留在下一时刻	2015
OKVIS	基于优化的方法、单目/立体	不支持重定位,也没有闭环检测或校正方案	2015
VINS-MONO	基于优化的方法、单目	优于OKVIS	2017
LIMO	改进的视觉SLAM	方法从LiDAR中提取图片中特征点的深度	2018
VLOAM	并行激光与视觉SLAM	能够消除由视觉测距漂移引起的点云失真	2015
DEEPSLAM	双卷积神经网络	能够在存在图像噪声的情况下进行特征点检测	2017
SuperPoint	自监督框架	用于计算机视觉中检测训练	2018
GCN-SLAM	基于深度学习GCNv2网络	用于生成关键点和描述符	2019
RDSLAM	关键帧、单目	可以处理缓慢变化的动态环境	2013
MaskFusion	基于Mask R-CNN	实时的、语义和动态RGB-D SLAM系统	2018
DynaSLAM	单目/立体/RGB-D	在动态环境下辅助静态地图的SLAM系统	2018
StaticFusion	基于面元的RGB-D SLAM	初始若干帧内不能有大量动态物体	2018

参考文献

- [1] FUENTESPACHECO J, JOSÉ RUIZASCENCIO, JUAN MANUEL RENDÓNMANCHA. Visual simultaneous localization and mapping: a survey [M]. Kluwer Academic Publishers, 2015: 43-55.
- [2] 高翔. 视觉SLAM十四讲: 从理论到实践 [M]. 北京: 电子工业出版社, 2017: 18-23.
- [3] 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.
- [4] 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.
- [5] MARTINEZ - CANTIN R, CASTELLANOS J A. Unscented SLAM for large-scale outdoor environments [C]//2005 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2005: 3427-3432.
- [6] HOLMES S, KLEIN G, MURRAY D W. A square root unscented Kalman filter for visual monoSLAM [C]//2008 IEEE International Conference on Robotics and Automation. IEEE, 2008: 3710-3716.
- [7] SIM R, ELINAS P, GRIFFIN M, et al. Design and analysis of a framework for real-time vision-based SLAM using Rao-Blackwellised particle filters [C]//The 3rd Canadian Conference on Computer and Robot Vision (CRV'06). IEEE, 2006: 21-21.
- [8] KLEIN G, MURRAY D. Parallel tracking and mapping for small AR workspaces [C]//2007 6th IEEE and ACM international symposium on mixed and augmented reality. IEEE, 2007: 225-234.
- [9] ROSTEN E, DRUMMOND T. Machine learning for high-speed corner detection [C]//European conference on computer vision. Springer, Berlin, Heidelberg, 2006: 430-443.
- [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] ENGEL J, SCHÖPS T, CREMERS D. LSD-SLAM: Large-scale direct monocular SLAM [C]//European conference on computer vision. Springer, Cham, 2014: 834-849.
- [13] NEWCOMBE R A, LOVEGROVE S J, DAVISON A J. DTAM: Dense tracking and mapping in real-time [C]//2011 international conference on computer vision. IEEE, 2011: 2320-2327.
- [14] FORSTER C, PIZZOLI M, SCARAMUZZA D. SVO: Fast semi-direct monocular visual odometry [C]//2014 IEEE international conference on robotics and automation (ICRA). IEEE, 2014: 15-22.
- [15] ENGEL J, KOLTUN V, CREMERS D. Direct sparse odometry

- [J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 40(3): 611-625.
- [16] HUANG G. Visual-inertial navigation: A concise review [C]// 2019 international conference on robotics and automation (ICRA). IEEE, 2019: 9572-9582.
- [17] MARTINELLI A. Vision and IMU data fusion: closed-form solutions for attitude, speed, absolute scale, and bias determination [J]. IEEE Transactions on Robotics, 2012, 28(1): 44-60.
- [18] CORKE P, LOBO J, DIAS J, et al. An introduction to inertial and visual sensing [J]. The International Journal of Robotics Research, 2007, 26(6): 519-535.
- [19] 孙永全, 田红丽. 视觉惯性 SLAM 综述 [J]. 计算机应用研究, 2019, 36(12): 3530-3552.
- [20] MOURIKIS A, ROUMELIOTIS S. A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigation [C]// Proceedings 2007 IEEE International Conference, 2007: 3565-3572.
- [21] BLOESCH, MICHAEL, et al. Robust visual inertial odometry using a direct EKF-based approach [C]// 2015 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, 2015.
- [22] LEUTENEGGER S, LYNEN S, BOSSE M, et al. Keyframe-based visual-inertial odometry using nonlinear optimization [J]. International Journal of Robotics Research, 2015, 34(3): 314-334.
- [23] TONG Q, PEILIANG L, SHAOJIE S. VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator [J]. IEEE Transactions on Robotics, 2017, 33(1): 1-17.
- [24] GRAETER J, WILCZYNSKI A, LAUER M. Limo: Lidar-monocular visual odometry [C]// Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 2018: 7872-7879.
- [25] SHIN Y S, PARK Y S, KIM A. Direct visual slam using sparse depth for camera-lidar system [C]// 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018: 5144-5151.
- [26] DE SILVA V, ROCHE J, KONDOZ A. Fusion of LiDAR and camera sensor data for environment sensing in driverless vehicles [J]. arXiv Preprint arXiv: 1710.06230, 2018.
- [27] SCHERER S, REHDER J, ACHAR S, et al. River mapping from a flying robot: state estimation, river detection, and obstacle mapping [J]. Autonomous Robots, 2012, 33(1): 189-214.
- [28] HUANG K, XIAO J, STACHNISS C. Accurate direct visual-laser odometry with explicit occlusion handling and plane detection [C]// 2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019: 1295-1301.
- [29] LIANG X, CHEN H, LI Y, et al. Visual laser-SLAM in large-scale indoor environments [C]// 2016 IEEE International Conference on Robotics and Biomimetics (ROBIO). IEEE, 2016: 19-24.
- [30] ZHU Z, YANG S, DAI H, et al. Loop detection and correction of 3d laser-based slam with visual information [C]// Proceedings of the 31st International Conference on Computer Animation and Social Agents, 2018: 53-58.
- [31] PANDEY G, SAVARESE S, MCBRIDE J R, et al. Visually bootstrapped generalized ICP [C]// 2011 IEEE International Conference on Robotics and Automation. IEEE, 2011: 2660-2667.
- [32] SEO Y, CHOU C C. A tight coupling of vision-lidar measurements for an effective odometry [C]// 2019 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2019: 1118-1123.
- [33] ZHANG J, SINGH S. Visual-lidar odometry and mapping: Low-drift, robust, and fast [C]// 2015 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2015: 2174-2181.
- [34] ZHOU T, BROWN M, SNAVELY N, et al. Unsupervised learning of depth and ego-motion from video [C]// Proceedings of the IEEE conference on computer vision and pattern recognition, 2017: 1851-1858.
- [35] GODARD C, MAC AODHA O, BROSTOW G J. Unsupervised monocular depth estimation with left-right consistency [C]// Proceedings of the IEEE conference on computer vision and pattern recognition, 2017: 270-279.
- [36] MAHJOURIAN R, WICKE M, ANGELOVA A. Unsupervised learning of depth and ego-motion from monocular video using 3d geometric constraints [C]// Proceedings of the IEEE conference on computer vision and pattern recognition, 2018: 5667-5675.
- [37] DETONE D, MALISIEWICZ T, RABINOVICH A. Toward geometric deep slam [J]. arXiv preprint arXiv: 1707.07410, 2017.
- [38] DETONE D, MALISIEWICZ T, RABINOVICH A. Superpoint: Self-supervised interest point detection and description [C]// Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2018: 224-236.
- [39] TANG J, ERICSON L, FOLKESSON J, et al. GCNv2: Efficient correspondence prediction for real-time SLAM [J]. IEEE Robotics and Automation Letters, 2019, 4(4): 3505-3512.
- [40] 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). IEEE, 2017: 4628-4635.
- [41] WHELAN T, SALAS-MORENO R F, GLOCKER B, et al. ElasticFusion: Real-time dense SLAM and light source estimation [J]. The International Journal of Robotics Research, 2016, 35(14): 1697-1716.
- [42] 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). IEEE, 2017: 1722-1729.
- [43] TAN W, LIU H, DONG Z, et al. Robust monocular SLAM in dynamic environments [C]// 2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). IEEE, 2013: 209-218.
- [44] 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). IEEE, 2018: 1168-1174.
- [45] RUNZ M, BUFFIER M, AGAPITO L. Maskfusion: Real-time recognition, tracking and reconstruction of multiple moving objects [C]// 2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). IEEE, 2018: 10-20.
- [46] BESCOS B, FÁCIL J M, CIVERA J, et al. DynaSLAM: Tracking, mapping, and inpainting in dynamic scenes [J]. IEEE Robotics and Automation Letters, 2018, 3(4): 4076-4083.
- [47] SCONE R, JAIMEZ M, PETILLOT Y R, et al. Staticfusion: Background reconstruction for dense rgb-d slam in dynamic environments [C]// 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018: 3849-3856.