



导航定位学报

Journal of Navigation and Positioning

ISSN 2095-4999, CN 10-1096/P

《导航定位学报》网络首发论文

题目: SLAM 中视觉和激光信息的融合应用
作者: 曾瑞琪, 纪新春, 魏东岩, 巨柳荫, 赵航
网络首发日期: 2025-04-25
引用格式: 曾瑞琪, 纪新春, 魏东岩, 巨柳荫, 赵航. SLAM 中视觉和激光信息的融合应用[J/OL]. 导航定位学报.
<https://link.cnki.net/urlid/10.1096.P.20250425.1351.002>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

引文格式:曾瑞琪, 纪新春, 魏东岩, 等. SLAM 中视觉和激光信息的融合应用[J]. 导航定位学报, 2025, 13(3) (ZENG Ruiqi, JI Xinchun, WEI Dongyan, et al. Fusion application of visual and LiDAR information in SLAM[J]. Journal of Navigation and Positioning, 2025, 13(3))

SLAM 中视觉和激光信息的融合应用

曾瑞琪^{1,2}, 纪新春^{1,3}, 魏东岩¹, 巨柳荫^{1,2}, 赵航^{1,2}

(1. 中国科学院 空天信息创新研究院, 北京 100094; 2. 中国科学院大学 电子电气与通信工程学院, 北京 100049;

3. 西北工业大学 电子信息学院, 西安 710114)

摘要: 为了进一步提升同步定位与建图 (SLAM) 技术的性能, 研究分析 SLAM 中视觉和激光信息的融合应用: 指出 SLAM 技术广泛应用于自动驾驶、智能机器人等移动平台的自主导航定位过程, 视觉相机和激光雷达 (LiDAR) 作为最常用的感知传感器, 各有优劣, 融合二者数据能增强 SLAM 系统的鲁棒性和准确性; 然后在分析视觉和激光信息融合需求的基础上, 结合 SLAM 工作流程, 重点从里程计、回环检测及地图构建 3 个关键环节探讨视觉和激光信息融合方法; 最后分析 SLAM 中将图像与点云融合所面临的挑战, 并对未来的技术发展方向进行展望。

关键词: 同步定位与建图 (SLAM); 视觉相机; 激光雷达 (LiDAR); 数据融合; 里程计; 回环检测; 建图
中图分类号: P228 **文献标志码:** A **文章编号:** 2095-4999(2025)03-0000-00

Fusion application of visual and LiDAR information in SLAM

ZENG Ruiqi^{1,2}, JI Xinchun^{1,3}, WEI Dongyan¹, JU Liuyin^{1,2}, ZHAO Hang^{1,2}

(1. Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China;

2. School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China;

3. School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710114, China)

Abstract: In order to further improve the performance of simultaneous localization and mapping (SLAM) technology, the paper studied and analyzed the fusion application of visual and light detection and ranging (LiDAR) information in SLAM: it was pointed out that SLAM technology is widely used in the autonomous navigation and positioning process of mobile platforms such as autonomous driving and intelligent robots, and as the two most commonly used perception sensors, camera and LiDAR have their own unique advantages and limitations, thus fusion of the two data could give full play to their respective advantages and improve the robustness and accuracy of SLAM system; then, based on the analysis of visual and LiDAR information fusion requirements and, combined with the SLAM workflow, the information fusion scheme of camera and LiDAR was mainly discussed from three aspects of odometry, loop closure detection and mapping; finally, the challenges faced by images and point clouds fusion in SLAM were analyzed, and the future development direction was projected.

Keywords: simultaneous localization and mapping (SLAM); visual camera; light detection and ranging (LiDAR); data fusion; odometry; loop closure; mapping

0 引言

随着自动驾驶、智能机器人等行业的广泛发展, 安全可靠的高精度位置服务需求激增, 对现有的导航定位手段提出了更高要求。同步定位与建图 (simultaneous localization and mapping, SLAM) 技术能在没有任何基站等先验基础设施且具有潜在动态障碍的场景中构建环境地图^[1], 有效弥补全球导航卫星系统 (global navigation satellite system, GNSS) 在隧道、城市峡谷、室内等卫星信号缺失场景中无法提供可靠定位信息的问题^[2], 进而实现安全稳定的自主导航, 受到了

研究学者的广泛关注。

SLAM 中包含了定位和建图 2 大基本任务，无人车辆、智能机器人等移动平台搭载特定的传感器，通过对环境的感知确定其运动轨迹并构建地图，可将未知场景转换为合作场景。其中，在智能驾驶领域，特斯拉以基于纯视觉的全自动驾驶（full self-driving, FSD）系统而闻名，但结合激光雷达和相机的方案采用全覆盖的感知方式，在实际应用过程通常具有更高的精度与可靠性能上限。相机图像以稠密二维（two dimension, 2D）数据格式提供丰富的场景语义，构建的环境地图更符合智能感知应用需求，但缺失精确的深度信息。激光雷达扫描则能提供厘米级的精确三维（three dimensions, 3D）点云测量，但同时其测量信息相对稀疏、对环境理解能力差。因此，对于环境感知应用，激光雷达和相机 2 类传感器具有天然的互补性，将二者融合能取长补短，充分发挥各自的观测优势，显著增强对场景的感知能力，进而提高 SLAM 系统性能。

近年来，许多关于激光/视觉 SLAM 的研究已被广泛提出。文献[3-5]梳理了 SLAM 技术的发展脉络，并对视觉 SLAM 和激光 SLAM 分别展开综述，详述了二者的技术处理流程和环境适应性。文献[6-7]分析了激光和视觉传感器各自的性能优势，认为融合处理能够取长补短，从而有效提升 SLAM 系统的综合性能。对于视觉和激光的融合 SLAM，文献[8-9]探讨了视觉相机和激光雷达的时空标定方法，指出高精度的时空同步是保证视觉图像和激光点云信息充分融合的前提。文献[10-11]从特征融合层面总结了图像和点云的融合策略，综述分析了卡尔曼滤波、图优化及深度学习等技术手段在融合 SLAM 系统中的性能特点。然而，以 SLAM 技术环节的视角详述视觉和激光信息融合方案的相关文献尚有缺失。基于此，首先对 SLAM 技术进行概述，简要分析视觉相机和激光雷达的工作原理及二者融合的应用需求；从 SLAM 的工作流程出发，探讨各类视觉和激光信息融合方案，指出 SLAM 中进行图像和点云数据融合所面临的技术挑战，并对未来视觉和激光融合 SLAM 的主要发展趋势及研究热点进行展望。

1 SLAM 技术及不同系统传感器原理

1.1 SLAM 概述

SLAM 系统的工作流程如图 1 所示。大致分为 5 个部分，即环境感知、里程计、后端优化、回环检测和地图构建。环境感知过程中使用激光雷达、相机或其他传感器采集环境数据，交由里程计处理分析，快速估算出相邻数据帧之间的位姿变换。后端优化采用图优化或滤波算法等对里程计估计所得的轨迹进行优化得出精确位姿。回环检测持续执行，识别机器人之前是否访问过该场景，消除里程计的累积漂移。地图构建过程采用运动估计的结果创建与任务要求对应的环境地图。

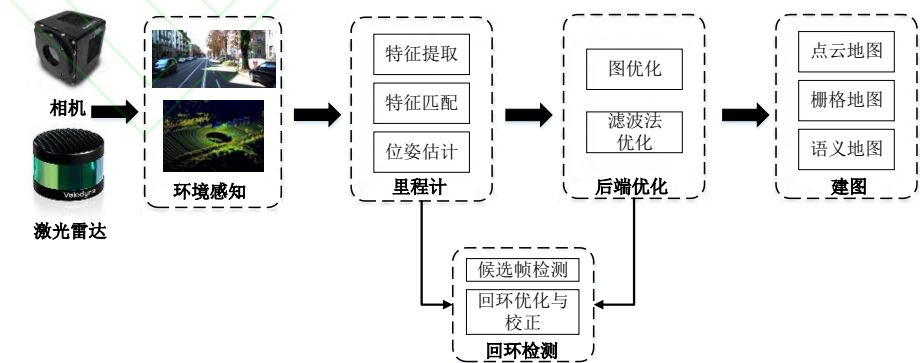


图 1 SLAM 系统基本工作流程

1.2 视觉相机和激光雷达的工作原理

视觉相机和激光雷达是 SLAM 系统主要的 2 类感知传感器。视觉相机通过获取连续的图像以供 SLAM 系统估计自身位姿并构建环境地图，能获得深层次场景语义特征，其成本低，体积小，应用广泛。依据工作方式的不同可分为单目相机、双目相机、深度相机等。

单目相机无法获取绝对深度信息，通常需要初始化过程来设定尺度基准。双目相机通过比较 2 个视角的图像差异来计算深度信息。使用单目或双目相机的视觉 SLAM 所构建的关键帧质量对环境中的光照条件要求严格。深度相机利用了结构光、飞行时间（time of flight, ToF）、衍射光栅或快门速度编码等原理通过深度图获得每一帧图像中的像素深度，弥补了单目相机和双目相机尺度估计的不确定性，但深度相机在大区域环境中的使用受到其精度与量程的限制，同时存在一定的环境敏感性。

激光雷达利用激光脉冲信号作为辐射参考，实时扫描周围环境的三维模型提供易于解释的点云来执行 SLAM 任务。按照运转方式可将激光雷达分为机械式激光雷达和固态激光雷达。

机械式激光雷达通过旋转高频激光设备获取周围环境的详细三维信息，按照单一扫描周期内能够同时发射和接收激光线的数量，可将机械式激光雷达分为单线激光雷达和多线激光雷达。固态激光雷达利用光学相干断层扫描或光学相干测距技术，以及微机电系统（micro-electro mechanical systems, MEMS）或其他固态技术来控制光束的发射和接收，从而实现快速扫描，在短时间内采集到稠密的点云。激光雷达虽然能提供高精度的三维空间数据，但其在解释环境语义信息方面存在局限性，对环境的深层次理解能力不足。

综上所述，激光雷达的显著优点是测距相对准确，对光照变化不敏感，采集的点云之中包含空间信息，不存在尺度漂移的问题。视觉传感器不依赖环境几何结构，能较好地弥补激光雷达在长直走廊、隧道等几何结构单一环境中的精度退化问题，同时能提供语义信息。依赖单一传感器感知不能保证无人车辆、智能机器人在现实环境中的安全部署，二者融合的 SLAM 能弥补单一传感器的缺点，已成为当前研究的热点方向。依据 SLAM 的工作流程，各个环节均可利用视觉和激光信息以达到不同目的的融合优化效果。如表 1 所示总结了经典的视觉和激光信息融合 SLAM 算法及其各自主要特点。

表 1 视觉和激光信息融合 SLAM 算法分析

年份	名称	融合环节	主要特点
2015	激光视觉惯性里程计 (visual LiDAR odometry and mapping, VLOAM) [12]	里程计	<ul style="list-style-type: none"> 视觉里程计输出辅助点云匹配 激光点云和视觉特征进行配准形成立体视觉
2021	具有深度先验的物体级 SLAM (object oriented SLAM with deep shape priors, DSP-SLAM) [13]	建图	<ul style="list-style-type: none"> 通过 3D 语义边界框检测获得姿态的初始估计 利用激光点云对建图效果进行优化
2021	紧耦合的视觉激光 SLAM (tightly-coupled visual-LiDAR SLAM, TVL-SLAM) [14]	里程计/ 回环检测	<ul style="list-style-type: none"> 视觉信息和激光测量值都用于运动估计和回环 依赖预先构建的视觉地图进行匹配和定位
2021	激光视觉惯性融合 SLAM (tightly-coupled LiDAR-visual-inertial odometry via smoothing and mapping, LVI-SAM) [15]	里程计/ 建图	<ul style="list-style-type: none"> 包含激光里程计和视觉里程计 2 个子模块 将 2 个里程计构建的运动因子进行融合后输出建图结果
2022	鲁棒实时彩色激光雷达-惯性-视觉融合 SLAM 系统 (robust, real-time, radiance reconstruction package with a tightly-coupled LiDAR-inertial-visual state Estimator, R ³ LIVE) [16]	里程计/ 建图	<ul style="list-style-type: none"> 包含激光里程计和视觉里程计,2 个子模块 利用激光和视觉观测构建含颜色信息的稠密三维地图

1.3 视觉相机和激光雷达时空同步

激光点云和图像信息表示完全不同，对相机和激光雷达进行时空统一是信息融合的基础步骤，即时间同步和空间标定，如图 2 所示。时间同步旨在同步相机和激光的时间戳，确保二者在相同的参考时间系统下采集数据，消除二者数据的时间表达差异。空间标定即确定相机坐标系与激光雷达坐标系之间的相对变换关系，消除二者相对位姿，实现空间融合。

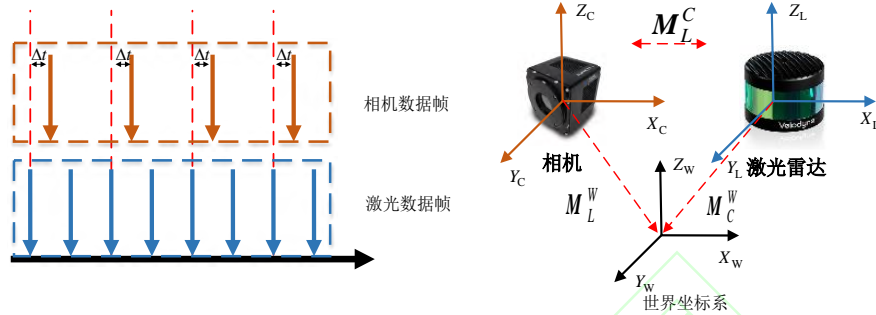


图 2 激光雷达和相机的时空同步

时间同步通常通过硬件^[17]或软件^[18]的方法解决：硬件同步需要使用专门的同步设备或接口来确保相机和激光雷达在同一时刻开始采集数据，精度可达百纳秒级，但额外的硬件设备增加了应用成本；软件同步成本低，利用软件算法分析传感器的时间戳或特定的同步信号即可计算和调整时间差异，精度范围通常在毫秒级，通常能满足中低动态的无人配送车辆等场景应用需求。空间标定常用的方案有基于标定板的方法、基于运动的方法和基于深度学习的方法等。文献[19]利用特定的标定板从激光雷达和相机数据中提取自定义的同步目标参考点，再通过目标点配准找到最优刚性变换。文献[20]通过基于运动的方法计算标定参数，考虑了单目相机运动估计带来的尺度模糊问题，将激光雷达和相机的外参标定结合在同一个优化框架内。为使标定模型具有良好的泛化能力，文献[21]利用神经网络提取点云和图像特征，学习标定过程中的潜在几何关系，并利用帧间约束得到的几何损失和光度损失来优化标定参数，提高同步精度。空间标定包含旋转和平移 2 个重要参数，其中，平移参数标定精度可达厘米级，旋转参数精度不同的方法差别较大，通常在优于 0.3° 范围内，具体取决于所使用的标定方法、标定物的精度、传感器的性能及环境条件。在实际应用中，搭载激光雷达和相机的无人车辆、机器人会在出厂时根据具体应用需求和可用资源选择合适的标定方法，完成时间同步和空间外参标定以满足信息融合需求。然而，当载体平台长期遭受外力作用或传感器配置更新时，现有标定方法通常无法在线完成时空参数的重新精确标定，进而影响融合系统的可靠性和可扩展性，亟须进一步开发适应性更强的在线时空标定方法^[9]。

2 使用视觉和激光信息的里程计

里程计通过对传感器感知数据进行特征提取与匹配，进行前后帧间位姿变化估计，提供载体短期内的移动信息。基于相机的视觉里程计可能会由于特征点的缺失或运动模糊而产生误差，在估计尺度时存在固有的不确定性。激光雷达测量则是直接的深度信息，可精确地确定场景尺度，但同时存在由于反射率变化或遮挡导致数据丢失，进而产生点云失真的问题。因此，二者的互补特性能提高里程计状态估计的准确性和可靠性。此外，图像的语义信息提供了丰富的场景理解能力，将语义信息与三维点云数据相结合可在准确地识别和定位场景中物体的同时降低点云数据量，提取稳健特征进行轻量、精确的运动估计。如图 3 所示简要总结了里程计的数据预处理、特征提取和位姿优化 3 个基本流程中视觉和激光信息融合应用的策略。

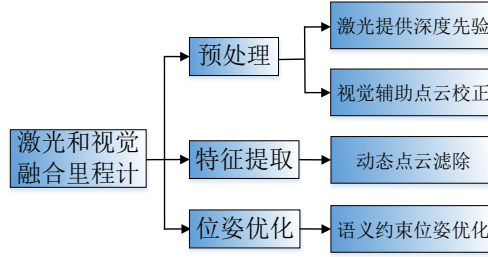


图 3 使用视觉和激光信息的里程计

2.1 激光提供深度先验及视觉辅助点云校正

针对视觉里程计难以估计绝对尺度问题，一类研究利用激光雷达提供的深度信息作为真实值或先验值，为视觉特征点提供准确的深度，帮助优化视觉特征的尺度估计。

文献[22]将激光点云投影到对应图像平面，以此利用激光的深度信息辅助视觉特征点完成尺度估计，将三维点云投影到相机坐标系，有

$$\mathbf{Z}_c \mathbf{Y} = \mathbf{P} \mathbf{M}_L^C \mathbf{X} \quad (1)$$

式中： \mathbf{Z}_c 为点 \mathbf{x} 在相机坐标系下的坐标值； $\mathbf{Y} = (u, v, 1)^T$ 为该点在二维图像中的像素坐标； \mathbf{P} 为激光雷达和相机的内参数矩阵； \mathbf{M}_L^C 为给定从激光雷达到相机的初始外部变换矩阵； $\mathbf{X} = (x, y, z, 1)^T$ 为该点在激光雷达坐标系下的坐标值。该方法通过投影建立点云与视觉特征点的联系，使特征点获得精确深度。在此基础上，文献[23]采用滑动窗口选择关键帧以保证位姿估计实时性，但相机分辨率比激光分辨率高，存在部分像素缺失深度信息问题，且视觉里程计和激光雷达之间的同步精度对结果有显著影响，若同步不准确，运动估计的误差会传递到点云校正中。为弥补此类问题，文献[24]采用了非合作场景下的图像和点云自动同步方法，并利用激光雷达辅助的视觉关键帧完成深度估计。

高精度点云匹配通常需要依赖精确的运动畸变补偿及良好的匹配初值，限制了激光里程计在动态及复杂退化场景下的适用性。针对此问题，另一类研究侧重于利用视觉里程计为点云变换提供先验信息以此来进行激光雷达的数据校正，减少激光点云累积误差，提高里程计的长期稳定性。

文献[25]先使用视觉里程计来估计连续 2 帧图像之间的运动，再使用该结果校正激光点云的运动失真，但校正结果依赖于视觉里程计的准确性，仍可能存在尺度不确定性。文献[26]通过视觉模块直接提取 3D 激光点云投影的高梯度像素进行跟踪，将候选帧的点云传播到更接近当前帧的中间关键帧，减小了尺度差异。文献[27]利用双目相机可以提供尺度信息的特点，将其作为里程计提供相对位姿，利用激光点云生成局部子图，采用基于独立成分分析（independent component analysis, ICA）的图像匹配方法将局部子图与预先构建的全局先验地图进行匹配，使得 2 个地图相应的误差平方和最小，以消除里程计的累积误差，提高长时定位的精度。

利用相机对激光点云初值校正的里程计方法提高了状态估计的可靠性，但对相机深层次的场景感知能力利用不充分，识别不同对象实例的效率低下，相机可被惯性测量单元（inertial measurement unit, IMU）或其他高频里程计所替代。随着深度学习技术的发展，视觉语义信息提供了对场景的高级理解，融合语义信息的稳健里程计已成为重要研究趋势。

2.2 借助视觉信息进行动态点云滤除

使用视觉和激光信息的里程计研究均基于静态环境假设，动态环境中的应用存在观测不稳定的缺陷^[28]，且激光雷达对环境的点云特征提取采用的是无差别扫描，动态对象的移动可导致特征误匹配，从而影响位姿估计的精度。通常的解决思路为利用视觉提供的语义信息对点云特征进一步筛选保证特征提取的可靠性，减少动态区域对特征提取和匹配造成的影响，以确保里程计位姿估计的准确性，同时保障后续仍能较好地复原实际的环境地图。在里程计中，借助视觉语义信息进行动态点云滤除的流程如图 4 表示。

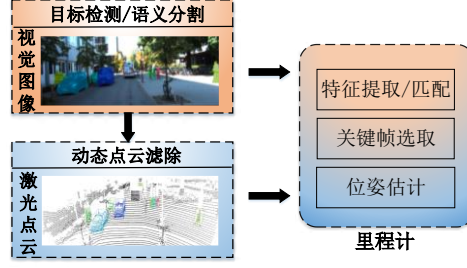


图 4 视觉信息辅助点云滤除

文献[29]将图像进行 2D 目标检测的结果投影到 3D 截锥体从中获得相应的激光点云,但存在庞大的计算量,实时性差。为简化数据处理,文献[30]基于投影和插值将低密度的激光点云与语义分割后的图像进行融合,提取出语义图像对应的激光点云。文献[31]对图像语义信息及其在点云中检测到的 3D 边界框进行滤波跟踪,减轻点云计算量的同时增强对场景的完整理解。借助语义信息进行点云滤除还需要考虑遮挡环境对滤除效果的影响。文献[32]利用鸟瞰视图具有绝对尺度和无遮挡描述环境的优势,将图像信息和激光点云共同投影到鸟瞰视图(bird's eye view, BEV)中提取特征,为点云提供目标物体的语义信息。文献[33]在点云特征提取和匹配过程中为消除动态物体和遮挡物体干扰,考虑了语义概率并以此作为点云滤除过程中的约束,提高动态点云滤除的准确性,过滤结果如图 5 所示。

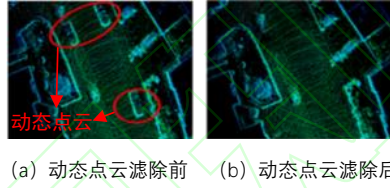


图 5 动态点云滤除结果

在语义信息滤除动态激光点云方案中,对于固态激光雷达稠密的点云信息而言,可直接利用点云语义分割网络进行处理,识别并过滤动态对象的点云,但在大规模室外环境中,稠密点云给系统带来了较大的计算和处理负担。利用视觉提供的语义信息辅助点云过滤能帮助识别具有特定语义信息的关键特征,提升位姿估计精度和计算效率。滤除效果主要取决于语义分割的精度,用于语义分割的模型可能存在泛化能力弱或遮挡环境中性能差的问题。

2.3 语义约束点云位姿优化

语义信息能识别和分类环境中的物体,提取静态特征以提升帧间匹配跟踪的稳定性。同时通过对不同语义类别特征的误差项附加相应的权重作为额外的约束,能有效增强里程计处理非结构化和退化环境的能力。激光里程计不依赖特征点的匹配来估计深度,减少了误差来源,因此通常利用激光里程计作为基础框架,借助视觉语义信息约束点云位姿优化,进一步提高里程计位姿估计的精度与可靠性。

文献[34]根据语义信息自适应地调整特征点对位姿优化的影响程度,保证高动态环境中的位姿估计精度。文献[35]利用语义信息去除动态物体的同时为点云扫描配准添加语义约束,利用基于语义的迭代最近点(iterative closest point, ICP)方法实现了点云数据关联与匹配,通过最小化误差函数寻找 2 帧点云之间的最佳变换,即

$$E(\mathbf{V}_D, \mathbf{V}_M, \mathbf{N}_M) = \sum_{u \in \mathbf{V}_D} w_u n_u^T (\mathbf{T}(u) - \mathbf{v}_u)^2 \quad (2)$$

式中: $u \in \mathbf{V}_D$ 为点云的集合表示; w_u 为语义标签附加的权重信息; $n_u \in \mathbf{N}_M$ 为点云 $u \in \mathbf{V}_D$ 的法向量; $\mathbf{T}(u)$ 为对点集中的每一个点云 u 应用变换矩阵 \mathbf{T} ; $\mathbf{v}_u \in \mathbf{V}_M$ 为给定点云 $u \in \mathbf{V}_D$ 对应的匹配点云表示。显然,语义信息提供了额外的约束,可减少错误匹配,提高位姿估计和后续地图构建的

精度。为更细致地捕捉标签分布的相似性，文献[36]将语义相似度定义为标签分布对应的核函数，即有

$$w_i = \exp\left(-\frac{1}{2}(P_{(k+1,i)} - P_{(k,j)})^T \Sigma^{-1}(P_{(k+1,i)} - P_{(k,j)})\right) \quad (3)$$

式中： $P_{(k+1,i)}$ 为激光扫描特征点对应的预测语义标签分布； $P_{(k,j)}$ 为距离该点最近点的语义标签分布； Σ 为卷积神经网络带来的预测不确定性。该方法增强了数据的非线性映射能力，减少了位姿估计对初始值的敏感性。除上述语义信息以权值参与位姿优化之外，相应的对象语义特征同样能够应用于位姿估计环节。文献[37]利用目标语义边界的平面特征和点云几何分布的最小二乘估计进行目标状态更新，在特征匹配过程中加入与语义置信度相关的评价标准，以更好地优化获得精确位姿。文献[38]通过目标检测获得目标的边界框对点云信息的位姿估计进行约束，通过目标的高层语义信息和低级几何特征相关联实现相对运动增量的鲁棒估计。

仅依赖于几何特征的位姿估计容易产生不同物体特征点间的误匹配，视觉提供物体级语义标签可作为特征点的先验知识，在位姿优化过程中增加额外约束，提高匹配的准确性和位姿估计的鲁棒性。然而，这种方法的有效性高度依赖于准确的语义标签，语义标签错误可能会影响位姿优化的结果，且由于计算复杂度的增加，语义约束的点云位姿优化方法可能难以满足实时应用的需求。

3 使用视觉和激光信息的回环检测

在长时间的运行过程中，里程计估计往往会累积误差，导致定位和建图的准确性下降，回环检测通过识别和利用环境中的重复性特征，帮助 SLAM 系统校正累积误差，实现更可靠的定位和全局一致性的建图。虽然激光雷达能够获得准确的位姿估计，但其分辨率、范围和视场角等限制了它检测回环的能力，而相机捕获的图像信息可作为独特的环境标记，具有更好的环境适应性和区分能力，能帮助识别和验证回环，从而提高回环检测的效率与准确性。按照对视觉和激光信息的关联方式不同可将回环检测方法分为 2 类，即联合使用二者特征，在回环检测过程中同时利用视觉和激光信息，或通过融合图像和点云特征构建新的特征描述子并用于回环检测。

3.1 联合使用视觉和激光特征

采用视觉检测回环的方法初步确定回环关键帧，进而利用相应的点云特征进行回环的位姿校正是联合使用视觉和激光信息最直观的方案，对应的框架流程如图 6 所示。此过程中通常保留相机提供的视觉关键帧，不仅能解决激光分辨率不足带来的点云区分度不高的问题，还能节省计算点云之间相似度的处理时间。

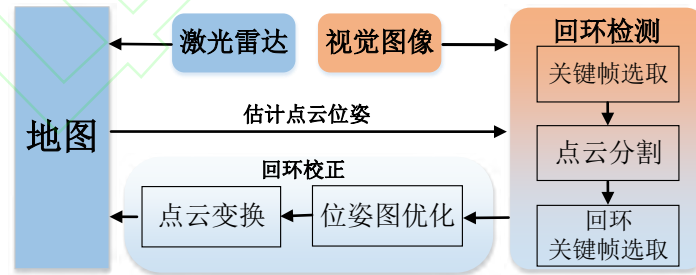


图 6 使用视觉和激光信息进行回环检测流程

文献[39]采用视觉关键帧查询词袋数据库判断是否为回环关键帧，并将确定的回环关键帧与相对应的点云对齐，计算视觉特征点的三维坐标进而进行回环检测和校正。文献[40]利用视觉的鸟瞰图特征变换（bird's eye view feature transform, BVFT）描述子进行回环检测，通过 ICP 算法对点云数据进行优化，将优化结果作为回环校正的真实姿态。文献[41]根据图像关键帧执行最近邻匹配检索对应的激光点云，触发局部地图检测回环并校正地图轨迹，实验结果表明，该方法的平均重定位误差小于 5cm，但实时性较差。针对回环检测实时性差的问题，改进方法主要集中在

提高视觉回环检测帧的搜索时间上。文献[42]使用全景投影将点云强度信息投影到二维图像空间，保持视觉关键帧执行词袋回环检测过程，利用视觉关键帧与激光点云的关联关系节省计算点云之间相似度的时间。文献[43]提出了在视觉系统中使用分布式词袋模型为激光里程计构建回环约束的方法，该方法在提高回环精度的同时，加快了回环检测的搜索速度。

通过视觉信息确定回环关键帧进而利用点云信息进行回环校正的方法对回环关键帧选取的视觉信息有较强的依赖性，在光照、视角变化或纹理稀疏的环境中仍存在回环检测稳定性不高的问题。另一类联合使用二者特征的方法在回环关键帧选取过程中同时考虑了视觉和激光信息，将激光基于几何特征的回环检测和视觉基于图像特征的回环检测方法结合，以增强回环检测过程的稳健性。

回环关键帧的选取可使用各类不同特征提取和描述方法，文献[44]采用视觉词袋方法进行回环检测的同时，并行基于激光“虹膜”（Iris）特征的局部搜索，同时考虑视觉和激光特征寻找回环候选帧。文献[45]使用视觉图像的尺度不变特征（scale invariant feature transform, SIFT）和变维局部形状描述子（variable-dimensional local shape descriptor, VD-LSD）训练视觉和激光传感器的词袋模型，将信息转换为各自的外观向量用于回环检测过程，处理流程如图 7 所示。但是，在回环过程中同时处理视觉和激光数据需要更多的计算资源，可能导致 SLAM 系统的实时性下降。除以上直接采用图像特征和点云信息进行回环检测外，视觉提供的语义信息为回环检测提供了额外的区分度，利用语义信息能够实现快速的场景相似性检测，主要通过构建对象共视图或语义图来实现。文献[46]在映射的三维点云语义目标上构建对象共视图，同时使用高层语义信息和低层几何信息，通过比较对象共视图来进行回环检测。文献[47]通过逐步维护每帧点云的语义图，根据语义图的相似性得分识别相似性场景。但在特征稀疏或混乱的应用场景中，提取和表示语义信息可能较为困难，且高重复性对象，例如路灯、树木等可能导致共视图或语义图无法提供足够的区分度，从而影响回环检测的性能。

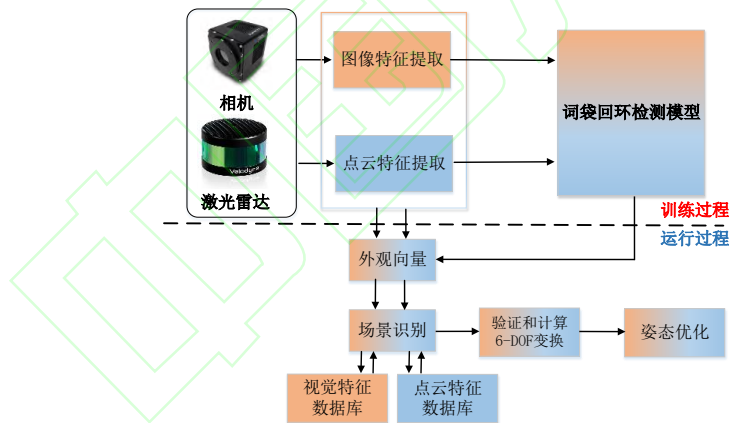


图 7 基于视觉和激光信息的词袋回环检测

综上所述，使用视觉和激光信息进行回环检测的方法通常借助视觉特征对相似场景具有良好的识别能力来判断回环，并在此基础上利用激光点云在回环位姿校正过程中获得更精确的位姿信息。但需要在视觉和激光信息严格对齐的基础上，确保 2 种模态数据在时间和空间上的一致性，才能找到匹配对应的正确信息对场景进一步描述并进行回环检测。

3.2 融合图像和点云特征构建描述子

回环检测过程中要求描述子对地点或场景的标识具有强的特异性，以便能稳健区分不同的场景，实现可靠的回环检测。除 3.1 节所述直接使用激光或视觉特征作为回环检测的描述子外，不少研究者为充分利用点云捕获的详细几何信息和图像中固有的丰富视觉信息，通过构建全新的描述子来检测并校正回环。按照对图像和点云数据融合层面的不同，可分为浅层融合和深层融合，浅层融合基于二者数据特征进行叠加，深层融合则通过深度学习等方法实现深度集成。二者

模型如图 8 所示。



图 8 融合构建描述子模型

浅层融合借助空间投影将预处理后的图像数据与点云数据在空间上进行叠加融合，通常涉及到将图像像素映射到对应的点云上或将点云投影到图像平面上，借助提取到的图像语义、边缘、纹理和点云几何、强度、法线等特征组合构建新的回环描述子。文献[48]将点云投影到语义检测框内，利用点云的视点直方图特征和图像的语义特征构建语义点云拓扑描述子，以解决单一特征回环过程的误匹配问题。文献[49]中将语义信息以扇形块的形式组织编码到点云中，通过比较每个扇区场景点云的相似度得分选择回环候选帧，以提高回环检测速度。文献[50]在融合视觉特征和点云特征过程中考虑了离群特征的干扰，使用神经网络为特定位置的特征 (I, P) 生成描述子 X ，即

$$S = \{(I, P)\} \xrightarrow{F(\cdot)} X \xrightarrow{M(\cdot)} X \quad (4)$$

式中： S 为输入图像数据 I 和点云数据 P 的统一空间表示； $F(\cdot)$ 为特征提取模块； X 为提取到的全局融合特征； $M(\cdot)$ 为相似性的度量模块； X 为最终生成的全局描述子。该方法未利用图像的语义信息，相似性场景的回环检测能力不足。文献[51]将视觉特征和激光点云信息描述的三维空间结构的几何属性融合构建视觉扫描上下文（visual scanning context, VSC）描述子，且不依赖于训练和里程计结果，有效地增强了回环准确率。浅层融合描述子的构建需要根据不同场景选择不同的特征组合策略，环境的变化可能会影响特征的稳定性，难以自适应地调整特征组合策略以满足不同场景的需求，从而影响回环的准确性。

与浅层融合侧重于直接将图像信息与点云信息进行组合不同，深层融合更注重挖掘图像特征和点云特征之间的内在联系，可在局部区域或特征上进行匹配并融合，降低融合时对时空一致性的要求。文献[52]将点云和图像特征编码为与位置相关的网格特征，借助随时间的交互关系对点云和图像特征进行融合，减轻回环过程中的计算负荷；但上述融合方法所使用的图像数据视野有限，在安装位置或观测角度改变时引入了固有缺陷，不同模态特征之间信息不平衡时易导致融合时信息丢失。针对此类问题的改进策略主要是采用注意力机制挖掘点云特征和图像特征之间的深层关系。文献[53]引入交叉注意力对图像和点云之间的映射关系进行建模，先将点云转换成体素表示，再融合体素和图像像素特征，即有

$$f_i^{\text{att}} = \text{Att}(Q_i, K, V) = \sum_{j=1}^{hw} o_{i,j} V_j \quad (5)$$

式中： f_i^{att} 为交叉注意力对图像像素和点云体素的融合输出； Q_i 为第 i 个由原始点云生成的查询向量； K 和 V 分别为根据图像特征生成键向量和值向量； hw 为特征维度； $j=1$ 为加权求和从第一个特征开始； $o_{i,j}$ 为像素和体素之间关联关系的归一化权重； V_j 为第 j 个由图像特征生成的值向量。文献[54]通过引入多尺度空间和通道注意力的自适应权重来动态调整视觉和点云特征在描述子中的重要性，增强回环的鲁棒性，该方法的平均回环查全率可达 98.2%。文献[55]利用独立的骨干网络从图像和点云中提取特征，并借助交叉注意力融合点云和图像的深层特征，利用融合的描述子和局部特征分步进行回环检测和校正，平均回环查全率可达 95.4%，回环搜索和校正的平均时间约 4 ms。该方法的实施不仅确保了回环检测的准确性，而且显著提升了回环搜索的效

率。深度融合通常需要设计复杂的神经网络结构以对视觉特征和点云特征进行提取和融合，提取到的特征往往是高度抽象的，导致描述子的可解释性较差，不利于回环检测过程的故障诊断和系统优化。

浅层融合和深层融合旨在利用视觉和激光数据获得覆盖信息更全面的回环描述子，实现对场景的稳健表征，从而提升回环检测的性能。此外，深度学习技术可更深层地利用感知数据，减少对高成本、高精度传感器的依赖。与直接使用视觉和激光信息进行回环检测的方法相比，利用深度学习等相关技术融合视觉和激光特征的回环检测方法在提高 SLAM 系统的性能上表现出了显著优势。

4 使用视觉和激光信息的地图构建

环境地图在无人车辆、智能机器人等移动平台自主定位与导航等方面发挥着重要的作用。结合激光雷达和视觉传感器优势能获得更高精度的位姿估计，有助于在地图构建过程中进行准确的特征匹配，从而提高地图的细节质量。同时，结合二者的感知观测优势构建环境地图能提供全面和准确的环境信息，进一步提高地图构建的质量和可靠性，为各种应用需求提供更加强大和灵活的解决方案。根据具体的需求和环境特点，机器人使用视觉和激光信息构建的常用地图类型可分为语义点云地图和语义栅格地图。在需要高精度导航和避障的应用中，会优先考虑使用点云地图；而在计算资源有限或者环境相对简单的情况下，栅格地图更加适合。如表 2 所示总结了 2 种地图类型的优势和局限性。

表 2 语义点云地图和语义栅格地图对比分析

地图类型	优势	局限性
语义点云地图	<ul style="list-style-type: none">• 精度高• 包含环境的详细信息	<ul style="list-style-type: none">• 依赖传感器的分辨率• 存储数据量大
语义栅格地图	<ul style="list-style-type: none">• 资源利用率高• 易于理解和处理	<ul style="list-style-type: none">• 动态环境中的表达能力弱• 丢失环境细节信息

4.1 语义点云地图

结合激光点云数据和视觉语义信息，语义点云地图在提供环境空间几何信息的同时，包括了物体和场景的类别信息，有助于机器人理解环境，更直观地进行交互任务。

文献[56]构造空间多尺度神经网络实时进行图像语义分割，将分割的语义信息对应到三维点云的相应空间位置，实现三维语义点云地图重建。为提高三维语义点云地图的质量，文献[57]和文献[58]在实时构建三维语义地图的基础上，利用语义信息动态调整建图过程的损失函数，对建图结果进一步优化。文献[59]除了对点云和图像语义信息进行融合外，使用欧式聚类 and 语义信息约束最小化语义分割和建图误差，减少了建图过程中的细节信息丢失。与此同时，相关学者更倾向于解决语义点云建图过程的实时性和动态物体带来的影响。文献[60]采用对建图精度和计算效率进行衡量并自适应调整语义融合参数的策略，以平衡计算资源受限情况下的建图精度和效率。文献[61]采用运动一致性检测动态物体，结合图像的深度不连续性和点云基于几何的分割方法对动态物体进行剔除，构建仅含静态物体的语义点云地图（如图 9 所示）。

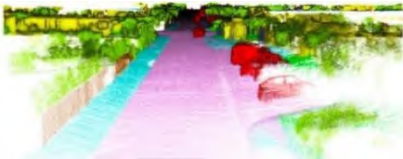


图 9 语义点云地图

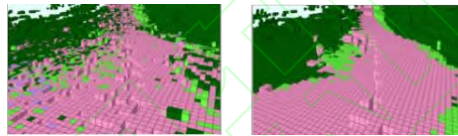
场景语义分割是构建语义点云地图的基本环节，分割结果影响着建图的精确性，实际应用中，构建语义点云地图大多须同时执行语义分割和环境建图 2 项任务，系统的实时性有待提升。

另外,语义点云地图在室内或简单环境的任务中表现了出优良的性能,但当应用于存在遮挡或动态物体干扰等大型复杂室外场景时,对计算存储资源和建图模型的环境适应性提出了更高的要求,在此环境下,要求模型能高效处理海量数据,并能从噪声和干扰中提取出有用信息,以确保建图的准确性和可靠性。

4.2 语义栅格地图

语义栅格地图生成具有语义理解的栅格地图。与语义点云地图类似,激光数据用于确定物体的位置和距离,而视觉数据则用于识别物体的类别和属性。不同的是,通过表示场景中所有栅格的占用状态和语义信息,语义栅格地图对遮挡目标、不规则形状目标具有更强泛化能力。

文献[62]利用神经网络从二维图像中计算像素语义分布并将其转移到三维网格空间,利用考虑了更全面依赖关系的高阶条件随机场(higher order conditional random fields, CRFs)模型来加强地图中视觉和点云信息的语义一致性。文献[63]采用后期动态的融合策略整合视觉和激光点云信息,利用图像语义信息对点云在局部和全局上的几何关系进行特征编码以优化栅格的占有概率。除侧重于提高栅格地图的语义一致性和准确性外,栅格地图研究同样关注建图过程的效率提升和增强地图在动态环境中的表达能力。文献[64]提出冗余栅格过滤模型,通过类熵为每个语义栅格分配合适的核长度,最后采用多熵核推理从图像和点云观测值中估计出连续的语义栅格地图,提高了建图效率。文献[65]在建图过程中考虑了语义预测的不确定性,并将语义不确定性纳入到不确定性感知的贝叶斯核推理模型中,以提高感知模糊的动态场景中的建图鲁棒性和准确性。建图结果如图 10 所示。



(a) 未考虑语义不确定度 (b) 考虑语义不确定度

图 10 语义占有率地图

语义栅格地图的分辨率可根据需要调整,侧重于在机器人任务中提供环境占用状态和导航信息,但需要复杂的更新机制,例如状态标记、概率模型等来反映潜在的动态环境变化。与语义点云地图相比,语义栅格地图通常需要较少的计算和存储资源,可通过数据压缩技术来减少存储和传输所需的数据量,在多机器人协同作业中易于实现数据共享。但对于复杂的三维场景,仍须对大量的点云数据或多视图图像信息进行处理和分析,以提取和更新占用状态信息。

5 未来技术发展趋势

自动驾驶、智能机器人等领域是 SLAM 技术发展的有利需求牵引,视觉和激光信息融合提升了 SLAM 各个环节的精度与鲁棒性,在测距和场景理解等方面具有显著优势,但目前融合方法的发展仍面临诸多挑战:融合大量激光点云和图像数据需要复杂的算法设计和较高的计算资源,在大范围、长时导航中的定位应用受限;遮挡环境、天气或光照变化等对激光雷达和相机的影响程度不同,融合系统的环境适应性仍有提升空间。从实际需求来看,视觉和激光融合 SLAM 未来的发展趋势主要聚焦在以下几个方面:

1) 多传感器协同感知。当前依赖激光雷达和相机进行融合感知的手段在极端环境中仍存在不足,高动态、复杂光照条件下效果不佳,借助其他传感器辅助视觉和激光信息能实现更全面的环境感知和理解。文献[66]利用热红外相机的优势,在多尘、低纹理或完全黑暗的环境中为激光雷达提供辅助信息。文献[67]利用 IMU 减少视觉和激光融合 SLAM 的累积误差,提高系统在高动态环境中的定位精度。目前多传感器之间的精确同步过程比较复杂,不同传感器具有不同的采样率和时间延迟,结合高效精确的同步方法可以克服多传感器融合的限制性,进一步提高激光雷达和相机系统的环境感知能力,为自动驾驶、智能机器人等领域提供更加安全、可靠的感知定位

解决方案。

2) 轻量级的 SLAM 模型。在自动驾驶、智能机器人等需要快速响应的应用中, 轻量级的 SLAM 模型更能满足实际实时性和可扩展性的关键要求。文献[68]提出下文引导网络(context guided network, CGNet)以实现轻量级语义分割。文献[69]将中央处理器(central processing unit, CPU)级的轻量匹配算法贝叶斯密集逆搜索(Bayesian dense inverse searching, BDIS)集成到 SLAM 框架中, 能在有限的计算资源中实时运行。轻量级的 SLAM 模型具有易于维护和部署、节省资源等优势。借助更高效的滤波算法、数据关联方法和优化策略或通过剪枝、量化和知识蒸馏等技术, 减少模型大小和计算复杂度, 使视觉和激光信息融合 SLAM 技术应用不再受限于设备的计算资源, 以便根据不同的应用场景和需求进行调整。

3) 端到端的 SLAM 框架。深度学习的灵活性和强大表征能力推动着视觉和激光信息融合的端到端 SLAM 框架发展。特斯拉“FSD META V12”版本系统采用端到端的自动驾驶模型, 统一自动驾驶(unified autonomous driving, UniAD)^[70]则是以最终任务为导向, 采用相机和激光雷达多模块联合优化的自动驾驶方案, 该方案的技术框架如图 11 所示。文献[71]直接从视觉图像中估计运动姿态。文献[72]基于投影感知的三维点云表示, 提出了端到端激光里程计框架。文献[73]提出基于自回归网络的端到端在线高清地图生成模型, 利用视觉信息直接生成矢量地图。但端到端的模型内部的决策过程不透明, 导致追踪和理解每一步的决策逻辑困难。借助可视化神经网络中的激活图或特征图等手段, 或引入注意力机制、开发新的框架或算法, 结合模拟学习、自监督学习等技术进一步提升端到端模型的综合性能, 使模型的决策过程更加透明, 增强端到端的视觉和激光融合 SLAM 的泛化能力。同时, 随着端到端 SLAM 相关软件、算法和应用生态的逐步建立和完善, 端到端 SLAM 专用芯片将在性能、应用范围和生态体系建设方面取得更大的突破。

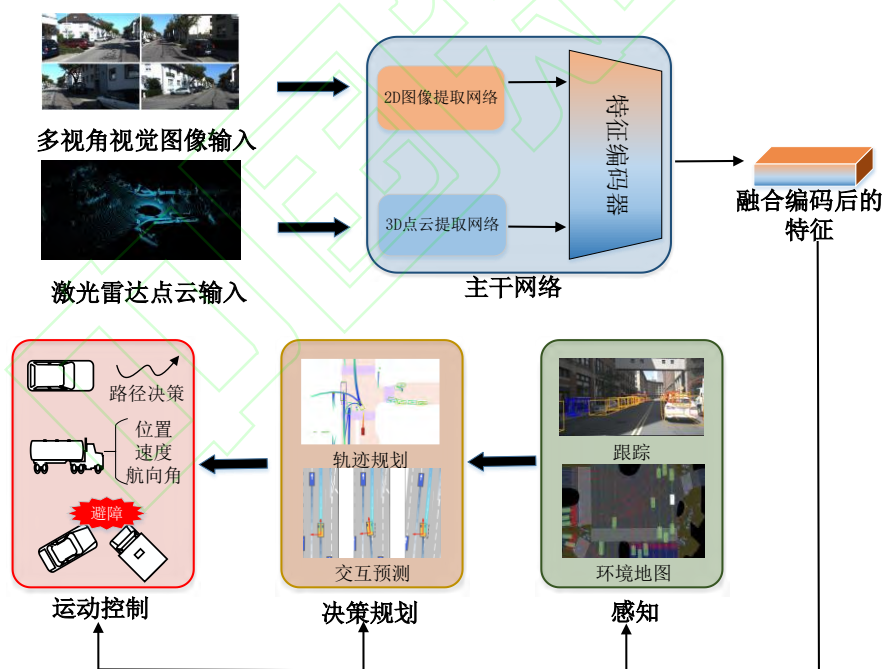


图 11 UniAD 技术框架

6 结束语

本文首先对自动驾驶、智能机器人等移动平台依托的 SLAM 技术及其采用的感知传感器方案进行分类和评述, 关注视觉和激光信息融合的 SLAM 技术途径, 分析了 2 类传感器优势及局限性; 然后从弥补单一传感器缺陷的角度出发, 重点阐述了视觉和激光信息在里程计、回环检测和建图方面的融合应用与发展现状; 最后讨论了视觉和激光信息融合 SLAM 系统存在的挑战, 从多传感器协同感知、轻量级的 SLAM 模型和端到端的 SLAM 框架 3 个方面展望了该技术未来

的发展趋势，以期为相关技术人员的研究工作提供参考。

参考文献

- [1] 王巍, 邢朝洋, 冯文帅. 自主导航技术发展现状与趋势[J]. 航空学报, 2021, 42(11): 18-36.
- [2] ZHU N, MARAIS J, BÉTAILLE D, et al. GNSS position integrity in urban environments: A review of literature[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 19(9): 2762-2778.
- [3] 任伟建, 高强, 康朝海, 等. 移动机器人同步定位与建图技术综述[J]. 计算机测量与控制, 2022, 30(2): 1-10,37.
- [4] DEBEUNNE C, VIVET D. A review of visual-LiDAR fusion based simultaneous localization and mapping[J]. Sensors, 2020, 20(7): 2068.
- [5] 周铨君, 陈炜峰, 尚光涛, 等. 基于异构数据融合的 SLAM 研究综述[J]. 南京信息工程大学学报, 2024, 16(4): 490-503.
- [6] 鱼瑜, 吴明亮, 张来喜, 等. 多传感器融合定位方法探讨[J]. 导航定位学报, 2023, 11(5): 151-163.
- [7] CHEN W, ZHOU C, SHANG G, et al. SLAM overview: From single sensor to heterogeneous fusion[J]. Remote Sensing, 2022, 14(23): 6033.
- [8] 毛军, 付浩, 褚超群, 等. 惯性/视觉/激光雷达 SLAM 技术综述[J]. 导航定位与授时, 2022, 9(4): 17-30.
- [9] 熊超, 乌萌, 刘宗毅, 等. 激光雷达与相机联合标定进展研究[J]. 导航定位学报, 2024, 12(2): 155-166.
- [10] 马争光, 赵永国, 刘成业, 等. 激光和视觉融合 SLAM 方法研究综述[J]. 计算机测量与控制, 2019, 27(3): 1-6.
- [11] BAVLE H, SANCHEZ-LOPEZ J L, CIMARELLI C, et al. From SLAM to situational awareness: Challenges and survey[J]. Sensors, 2023, 23(10): 4849.
- [12] ZHANG J, SINGH S. Visual-LiDAR odometry and mapping: low-drift, robust, and fast[C]// 2015 IEEE international conference on robotics and automation (ICRA). [S.l.]: IEEE, 2015: 2174-2181.
- [13] WANG J, RUNZ M, AGAPITO L. DSP-SLAM: Object oriented SLAM with deep shape priors[C]// 2021 International Conference on 3D Vision (3DV). [S.l.]: IEEE, 2021: 1362-1371.
- [14] CHOU C C, CHOU C F. Efficient and accurate tightly-coupled visual- LiDAR SLAM[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(9): 14509-14523.
- [15] SHAN T, ENGLLOT B, RATTI C, et al. Lvi-sam: Tightly-coupled LiDAR-visual-inertial odometry via smoothing and mapping[C]// 2021 IEEE international conference on robotics and automation (ICRA). [S.l.]: IEEE, 2021: 5692-5698.
- [16] LIN J, ZHANG F. R³LIVE++: A robust, real-time, radiance reconstruction package with a tightly-coupled LiDAR-inertial-visual state Estimator[J]. arXiv preprint arXiv: 2209.03666, 2022.
- [17] ZHENG C, ZHU Q, XU W, et al. Fast-livo: Fast and tightly-coupled sparse-direct LiDAR-inertial-visual odometry[C]// 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). [S.l.]: IEEE, 2022: 4003-4009.
- [18] MKACHER F, DUDA A. Calibrating NTP[C]// 2019 IEEE International Symposium on Precision Clock Synchronization for Measurement, Control, and Communication (ISPCS). Portland, OR, USA: IEEE, 2019: 1-6.
- [19] BELTRÁN J, GUINDEL C, ESCALERA A D L, et al. Automatic extrinsic calibration method for LiDAR and camera sensor setups[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(10): 17677-17689.
- [20] ZUÑIGA-NOËL D, RUIZ-SARMIENTO J R, GOMEZ-OJEDA R, et al. Automatic multi-sensor extrinsic calibration for mobile robots[J]. IEEE Robotics and Automation Letters, 2019, 4(3): 2862-2869.
- [21] SHI J, ZHU Z, ZHANG J, et al. Calibrcnn: Calibrating camera and LiDAR by recurrent convolutional neural network and geometric constraints[C]// 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). [S.l.]: IEEE, 2020: 10197-10202.

- [22] GRAETER J, WILCZYNSKI A, LAUER M. Limo: LiDAR-monocular visual odometry[C]// 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). [S.l.]: IEEE, 2018: 7872-7879.
- [23] 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). [S.l.]: IEEE, 2018: 5144-5151.
- [24] ZHU Y, ZHENG C, YUAN C, et al. Camvox: A low-cost and accurate LiDAR -assisted visual SLAM system[C]// 2021 IEEE International Conference on Robotics and Automation (ICRA). [S.l.]: IEEE, 2021: 5049-5055.
- [25] YAN M, WANG J, LI J, et al. Loose coupling visual- LiDAR odometry by combining VISO2 and LOAM[C]// 2017 36th Chinese Control Conference (CCC). [S.l.]: IEEE, 2017: 6841-6846.
- [26] YUAN Z, WANG Q, CHENG K, et al. Sdv-loam: Semi-direct visual- LiDAR odometry and mapping[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(9): 11203-11220.
- [27] 刘净轩, 王亮, 李和平, 等. 融合双目视觉和 2D 激光雷达的室外定位[J]. 控制与决策, 2023, 38(7): 1861-1868.
- [28] 徐兴, 刘琼, 黄开坤. 动态场景下仓储机器人的视觉定位与建图[J]. 导航定位学报, 2023, 11(6): 110-118.
- [29] QI C R, LIU W, WU C, et al. Frustum pointnets for 3D object detection from RGB-D data[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: [s.n.], 2018: 918-927.
- [30] LOU L, LI Y, ZHANG Q, et al. SLAM and 3D semantic reconstruction based on the fusion of LiDAR and monocular vision[J]. Sensors, 2023, 23(3): 1502.
- [31] KIM A, OŠEP A, LEAL-TAIXÉ L. Eagermot: 3D multi-object tracking via sensor fusion[C]// 2021 IEEE International Conference on Robotics and Automation (ICRA). [S.l.]: IEEE, 2021: 11315-11321.
- [32] KU J, MOZIFIAN M, LEE J, et al. Joint 3D proposal generation and object detection from view aggregation[C]// 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). [S.l.]: IEEE, 2018: 1-8.
- [33] LIU P, BI Y, SHI J, et al. Semantic-assisted LiDAR tightly coupled SLAM for dynamic environments[J]. IEEE Access, 2024: 34042-34053.
- [34] 张岩, 王红旗, 刘群坡, 等. 一种特征点权重自适应优化的动态 SLAM 算法[J]. 导航定位学报, 2024, 12(3): 165-175.
- [35] CHEN X, MILIOTO A, PALAZZOLO E, et al. Suma++: Efficient LiDAR-based semantic SLAM[C]// 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). [S.l.]: IEEE, 2019: 4530-4537.
- [36] WANG F, WANG Z, YAN F, et al. A novel real-time semantic-assisted LiDAR odometry and mapping system[C]// 2019 Tenth International Conference on Intelligent Control and Information Processing (ICICIP). [S.l.]: IEEE, 2019: 44-50.
- [37] CAO Q, LIAO Y, FU Z, et al. An iterative closest point method for LiDAR odometry with fused semantic features[J]. Applied Sciences, 2023, 13(23): 12741. DOI:10.3390/app132312741.
- [38] MA T, JIANG G, OU Y, et al. Semantic geometric fusion multi-object tracking and LiDAR odometry in dynamic environment[J]. Robotica: International Journal of Information, Education and Research in Robotics and Artificial Intelligence, 2024, 42(3): 891-910.
- [39] 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. [S.l.]: [s.n.], 2018: 53-58.
- [40] LUO L, CAO S Y, HAN B, et al. Bvmatch: LiDAR-based place recognition using bird's-eye view images[J]. IEEE Robotics and Automation Letters, 2021, 6(3): 6076-6083.
- [41] 包加桐, 杨圣奥, 朱润辰, 等. 视觉与激光结合的室内移动机器人重定位方法[J]. 中国测试, 2021(11): 1-7.

- [42] SHAN T, ENGLLOT B, DUARTE F, et al. Robust place recognition using an imaging LiDAR [C]// 2021 IEEE International Conference on Robotics and Automation (ICRA). [S.l.]: IEEE, 2021: 5469-5475.
- [43] CHEN S, ZHOU B, JIANG C, et al. A LiDAR/visual SLAM backend with loop closure detection and graph optimization[J]. Remote Sensing, 2021, 13(14): 2720.
- [44] WANG W, LIU J, WANG C, et al. DV-LOAM: Direct visual LiDAR odometry and mapping[J]. Remote Sensing, 2021, 13(16): 3340.
- [45] COLLIER J, SE S, KOTAMRAJU V. Multi-sensor appearance-based place recognition[C]// 2013 International Conference on Computer and Robot Vision. [S.l.]: IEEE, 2013: 128-135.
- [46] QIAN Z, FU J, XIAO J. Towards accurate loop closure detection in semantic SLAM with 3D semantic covisibility graphs[J]. IEEE Robotics and Automation Letters, 2022, 7(2): 2455-2462.
- [47] LI L, KONG X, ZHAO X, et al. SA-LOAM: Semantic-aided LiDAR SLAM with loop closure[C]// 2021 IEEE International Conference on Robotics and Automation (ICRA). [S.l.]: IEEE, 2021: 7627-7634.
- [48] 佟国峰, 杨宇航, 彭浩, 等. 基于视觉语义与激光点云交融构建的 SLAM 算法[J]. 控制与决策, 2024, 39(1): 103-111.
- [49] WANG G, GAO S, DING H, et al. LIO-CSI: LiDAR inertial odometry with loop closure combined with semantic information[J]. Plos One, 2021, 16(12): 0261053-0261053.
- [50] XIE S, PAN C, PENG Y, et al. Large-scale place recognition based on camera-LiDAR fused descriptor[J]. Sensors, 2020, 20(10): 2870.
- [51] HUI F, LIU Y. A loop closure detection SLAM system by encoding visual and point cloud features[C]// 2023 38th Youth Academic Annual Conference of Chinese Association of Automation (YAC). [S.l.]: IEEE, 2023: 433-438.
- [52] LEE A J, SONG S, LIM H, et al. (LC)²: LiDAR-camera loop constraints for cross-modal place recognition[J]. IEEE Robotics and Automation Letters, 2023, 8(6): 3589-3596.
- [53] ZENG Y, ZHANG D, WANG C, et al. Lift: Learning 4D LiDAR image fusion transformer for 3D object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: [s.n.], 2022: 17172-17181.
- [54] LAI H, YIN P, SCHERER S. Adafusion: Visual-LiDAR fusion with adaptive weights for place recognition[J]. IEEE Robotics and Automation Letters, 2022, 7(4): 12038-12045.
- [55] YUE H, CAO D, LIU Z, et al. Cross fusion of point cloud and learned image for loop closure detection[J]. IEEE Robotics and Automation Letters, 2024, 9(3): 2965-2972.
- [56] 马淼, 刘培敏, 潘海鹏. 结合语义信息与 3D 点云技术的未知环境地图构建方法[J]. 中国图象图形学报, 2023, 28(8): 2432-2446.
- [57] LIANG S, LI W, DUAN G, et al. Semantic map construction based on LiDAR and vision fusion[C]// 2023 IEEE 98th Vehicular Technology Conference (VTC2023-Fall). [S.l.]: IEEE, 2023: 1-5.
- [58] CHENG X, GENG K, YIN G, et al. Semantic mapping optimization based on LiDAR and camera data fusion for autonomous vehicle[C]// 2022 6th CAA International Conference on Vehicular Control and Intelligence (CVCI). [S.l.]: IEEE, 2022: 1-6.
- [59] JEONG J, YOON T S, PARK J B. Towards a meaningful 3D map using a 3D LiDAR and a camera[J]. Sensors, 2018, 18(8): 2571.
- [60] WEBB A M, BROWN G, LUJÁN M. ORB-SLAM-CNN: Lessons in adding semantic map construction to feature-based SLAM[C]// Towards Autonomous Robotic Systems: 20th Annual Conference, TAROS 2019, London, UK, July 3–5, 2019, Proceedings, Part I 20. [S.l.]: Springer International Publishing, 2019: 221-235.
- [61] RÜNZ M, AGAPITO L. Co-fusion: Real-time segmentation, tracking and fusion of multiple objects[C]// 2017 IEEE International Conference on Robotics and Automation (ICRA). [S.l.]: IEEE, 2017: 4471-4478.

- [62]YANG S, HUANG Y, Scherer S. Semantic 3D occupancy mapping through efficient high order CRFs[C]// 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). [S.l.]: IEEE, 2017: 590-597.
- [63]XIA Z, LIU Y, LI X, et al. Scpnnet: Semantic scene completion on point cloud[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: [s.n.], 2023: 17642-17651.
- [64]DENG Y, WANG M, YANG Y, et al. SEE-CSOM: Sharp-edged and efficient continuous semantic occupancy mapping for mobile robots[J]. IEEE Transactions on Industrial Electronics, 2023, 71(2): 1718-1728.
- [65]KIM J, SEO J, MIN J. Evidential semantic mapping in off-road environments with uncertainty-aware Bayesian kernel inference[J]. arXiv preprint arXiv: 2403.14138, 2024.
- [66]CHEN W, WANG Y, CHEN H, et al. EIL - SLAM: Depth - enhanced edge - based infrared - LiDAR SLAM[J]. Journal of Field Robotics, 2022, 39(2): 117-130.
- [67]LIU Z, LI Z, LIU A, et al. LVI-Fusion: A robust LiDAR-visual-inertial SLAM scheme[J]. Remote Sensing, 2024, 16(9): 1524.
- [68]WU T, TANG S, ZHANG R, et al. Cgnet: A light-weight context guided network for semantic segmentation[J]. IEEE Transactions on Image Processing, 2020, 30: 1169-1179.
- [69]SONG J, ZHANG R, ZHU Q, et al. BDIS-SLAM: A lightweight CPU-based dense stereo SLAM for surgery[J]. International Journal of Computer Assisted Radiology and Surgery, 2024, 19(5): 811-820.
- [70]HU Y, YANG J, CHEN L, et al. Planning-oriented autonomous driving[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: [s.n.], 2023: 17853-17862.
- [71]WANG S, CLARK R, WEN H, et al. DeepVO: Towards end-to-end visual odometry with deep recurrent convolutional neural networks[C]// 2017 IEEE International Conference on Robotics and Automation (ICRA). Singapore: IEEE, 2017: 2043-2050.
- [72]WANG G, WU X, JIANG S, et al. Efficient 3D deep LiDAR odometry[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(5): 5749-5765.
- [73]LIU Y, YUAN T, WANG Y, et al. Vectormapnet: End-to-end vectorized HD map learning[C]// International Conference on Machine Learning. [S.l.]: PMLR, 2023: 22352-22369.

第一作者简介: 曾瑞琪(2002—), 女(壮族), 广西百色人, 博士研究生, 研究方向为视觉导航定位、多传感器融合定位。

通信作者简介: 纪新春(1987—), 男, 内蒙古呼伦贝尔人, 高级工程师, 研究方向为视觉导航定位、车载多源融合导航等。