

周铖君<sup>1</sup> 陈炜峰<sup>1</sup> 尚光涛<sup>1</sup> 王曦杨<sup>1</sup> 徐崇辉<sup>1</sup> 李振雄<sup>1</sup>

# 基于异构数据融合的 SLAM 研究综述

## 摘要

激光与视觉 SLAM 技术经过几十年的发展,目前都已经较为成熟,并被广泛应用于军事和民用领域.单一传感器的 SLAM 技术都存在局限性,如激光 SLAM 不适用于周围存在大量动态物体的场景,而视觉 SLAM 在低纹理环境中鲁棒性差,但两者融合使用具有巨大的取长补短的潜力,激光与视觉甚至是更多传感器融合的 SLAM 技术将会是未来的主流方向.本文回顾了 SLAM 技术的发展历程,分析了激光雷达与视觉的硬件信息,给出了一些经典的开源算法与数据集.根据融合传感器所使用的算法,从传统基于不确定度、基于特征以及基于深度学习的角度详细介绍了多传感器融合方案,概述了多传感器融合方案在复杂场景中的优异性能,并对未来发展作出了展望.

## 关键词

同时定位与地图构建(SLAM);激光 SLAM;视觉 SLAM;多传感器融合;移动机器人

中图分类号 TP391.41;TP242

文献标志码 A

收稿日期 2022-10-02

## 作者简介

周铖君,男,硕士生,研究方向为智能机器人、深度学习.nazejun0331@163.com

陈炜峰(通信作者),男,博士,教授,研究方向为智能机械、机电控制.002021@nuist.edu.cn

## 0 引言

SLAM(Simultaneous Localization and Mapping,同时定位与地图构建)技术能够提供空间定位信息,在 GPS 信号弱的环境中<sup>[1]</sup>,同时估计自身所处的位置与地图和虚拟场景的构建<sup>[2]</sup>.其中,定位是指确认自身和周围物体在世界坐标系下的位姿,建图是指建立机器人所感知周围环境的地图<sup>[3]</sup>.

学者们从不同方面对 SLAM 技术进行了综述,例如:王霞等<sup>[4]</sup>介绍了主流的视觉 SLAM(VSLAM)架构,并总结了该领域的标志性成果和研究进展;胡凯等<sup>[5]</sup>归类分析了视觉里程计(Visual Odometry,VO)技术的研究成果,并预测了未来的发展趋势;危双丰等<sup>[6]</sup>对基于滤波器和基于图优化的两种激光 SLAM 框架进行了分析.但已有综述未见详细介绍激光与视觉甚至是更多传感器融合的 SLAM 方案.

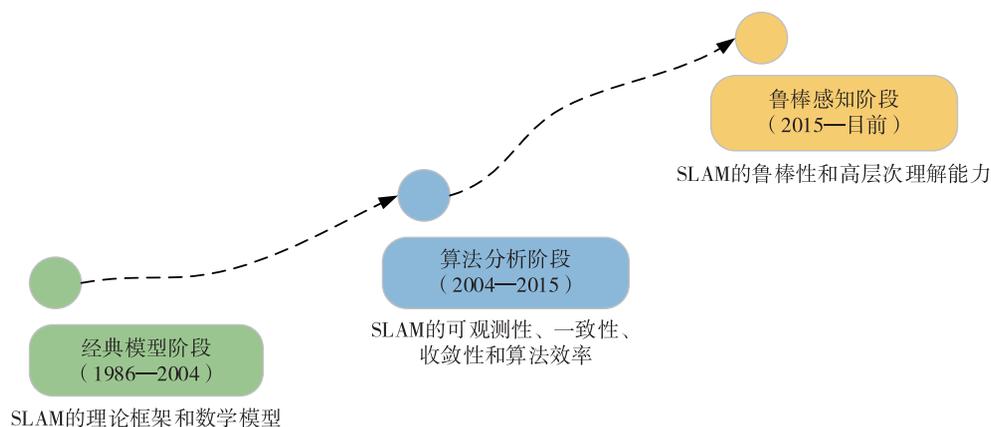
本文首先回顾了 SLAM 的发展历史和最近几十年在该领域的重要成果,阐述了 VSLAM、激光 SLAM 以及它们的衍生方案,然后介绍了 SLAM 领域所使用的传感器,重点分析了视觉、激光以及更多传感器 SLAM 的相互融合方案,最后是总结和展望.

## 1 传统激光与视觉 SLAM 概述

2016 年,Cadena 等<sup>[7]</sup>将 SLAM 的发展分为经典模型、算法分析和高级鲁棒感知三个阶段,如图 1 所示.早期的研究主要是基于激光雷达的 SLAM 方案,其框架大多是基于卡尔曼滤波及其衍生算法,例如:Neira 等<sup>[8]</sup>指出标准形式的 EKF(Extended Kalman Filter,扩展卡尔曼滤波)方法无法筛选出错误的地标观测点,因此提出了在闭环阶段进行全局位姿优化,并且 EKF 要求确定的运动和观测模型以及模型的线性假设;Julier 等<sup>[9]</sup>研究了非线性模型对于应用 EKF 的性能影响;Aulinas 等<sup>[10]</sup>将已有的过滤策略分类为 KF(卡尔曼滤波)、IF(信息滤波)、UKF(无迹卡尔曼滤波)和 CKF(容积分卡尔曼滤波)等,比较了它们在不同场景和不同地标数量的地图上的优缺点,并给出了适用范围;尹姝等<sup>[11]</sup>在线性卡尔曼滤波的基础上,提出一种结合自适应卡尔曼滤波和 BVIRE(边界虚拟参考标签)的移动机器人定位方法,较传统方法具有更高的定位精度和更强的鲁棒性.

随着计算机视觉技术的快速发展,研究人员发现相机(摄像机)能获得比激光雷达更丰富的纹理信息,并且摄像机价格低廉、结构简

1 南京信息工程大学 自动化学院,南京,210044

图1 SLAM发展阶段<sup>[7]</sup>Fig. 1 SLAM development stages<sup>[7]</sup>

单,安装的方式也比较多元化<sup>[12]</sup>.早期的SLAM基于滤波理论,其非线性的误差模型和巨大的计算量成了它实用落地的障碍.基于特征的方法解决了误差累积和计算复杂度高的问题,例如: Dissanayake等<sup>[13]</sup>系统分析了基于特征的SLAM基本特性,总结了不同版本SLAM的可观测性以及收敛相关的问题; Fraundorfer等<sup>[14]</sup>介绍了视觉里程计(VO)从只能离线工作到可实时工作的发展历史,讨论了特征匹配、健壮性和应用程序,综述了VO中常用的特征检测器和不同的异常点抑制方案,提出了随机抽样一致性和设计的策略; Mur-Artal等<sup>[15]</sup>提出一个全新的基于特征的ORB-SLAM系统,与当时最先进的单目SLAM方法相比,取得了前所未有的性能,它可在室内和室外环境中实时运行,对严重的运动杂波具有鲁棒性,允许宽基线环路关闭和重新定位,并包括全自动初始化.

此时的SLAM理论已经比较完善,框架也已基本确定.但随着深度学习技术的兴起,研究人员发现,在计算机视觉中应用深度学习方法可以极大地缓解传统方法难以解决的问题,如庞大的计算量和对环境的适应能力不强等.Li等<sup>[16]</sup>指出:传统方法需要手动设计特征提取和匹配算法,深度学习技术可以自动学习适合当前环境的特征表示和匹配方式;传统的SLAM方法通常假设场景是静态或存在刚性变换,对于非刚性变换(如扭曲、形变等)的处理相对较困难,而深度学习技术能够更好地处理非刚性变换,通过学习变换模型或生成模型来更好地理解 and 估计场景的变化;深度学习对于大规模数据的需求较高,而现代SLAM系统中通常配备了多种传感器(如相机、激光雷达等),能够产生大量包含丰富信息的数据,深度学习

技术可以利用这些数据进行训练,从而提高SLAM系统的性能和鲁棒性. Bescos等<sup>[17]</sup>开发了一种基于ORB-SLAM的具备动态对象检测和背景修复功能的可视化系统,通过多视图几何、深度学习或两者结合起来检测运动目标,尤其适用于高度动态的场景中.Han等<sup>[18]</sup>利用CNN来进行图像的三维重建,使用深度学习技术从单个或多个RGB图像估计一般物体的3D形状.Li等<sup>[19]</sup>将RCNN应用于配备有二维激光和惯性测量单元的移动机器人上,解决了旋转角度很大时准确性大幅下降的问题.

2007年以前,学者们认为打造SLAM系统需要一个双目摄像头套件,而Davison等<sup>[20]</sup>提出的MonoSLAM系统打破了当时人们的认知.MonoSLAM系统首次展示了如何使用单目摄像头构建SLAM系统,在概率框架内在线建立自然地标的稀疏且持续的地图.2007年,牛津大学提出PTAM<sup>[21]</sup>算法,它是最早提出将追踪和建图分开作为两个线程的一种SLAM算法,也是在VSLAM中首次区分出前后端的概念.该算法在后端部分采用非线性优化,而不是主流的EKF等滤波方法,具有里程碑意义.2015年提出的ORB-SLAM<sup>[15]</sup>可以看作是PTAM算法的一个延伸,它是当时最完整的基于特征点法的VSLAM,该系统框架包括跟踪、建图和闭环三个线程.ORB-SLAM<sup>[15]</sup>的后续版本有ORB-SLAM2<sup>[22]</sup>和ORB-SLAM3<sup>[23]</sup>.

除了基于特征的方法,直接法在SLAM中也同样占有重要的地位.相关的算法包括LSD-SLAM<sup>[24]</sup>、SVO<sup>[25]</sup>、DSO<sup>[26]</sup>.同样地,RGB相机也有相关算法,Whelan等<sup>[27]</sup>充分利用了RGB-D相机的颜色和深度信息,通过ICP(Iterative Closest Point,迭代最近点)估计位姿变化,并且通过不断地迭代优化来提高摄

像机的位姿估计精度.类似的算法还有 DTAM<sup>[28]</sup>、DVO<sup>[29]</sup>、RTAB-Map<sup>[30]</sup>和 RGBD-SLAM-V2<sup>[31]</sup>.

随着激光 SLAM 领域的不断发展,也出现了许多优秀的算法.2002 年,Montemerlo 等<sup>[32]</sup>提出 FastSLAM 算法,它采用粒子滤波来估计机器人位姿,是第一种能够实时输出栅格地图的激光 SLAM 方案.但在大尺度环境下,FastSLAM 算法需要大量粒子来表示机器人的位姿,严重消耗内存,并且随着不断地进行重采样,粒子耗散问题会逐渐加剧到无法忽视的程度.Gmapping 算法<sup>[33]</sup>基于 FastSLAM 进行优化,将粒子的数量保持在一个比较小的数值,对预测分布采样,它基于优化扫描匹配来优化位姿,解决了内存消耗严重的问题;同时,减少重采样次数,即当预测分布与真实分布相差很大时,才进行重采样,解决了粒子耗散问题.但该方案非常依赖里程计.Kohlbrecher 等<sup>[34]</sup>提出的 Hector-SLAM 将高斯牛顿

方法用于解决前端扫描匹配问题,并且不依赖里程计数据,但在机器人速度过快和强旋转时,会发生漂移现象.Cartographer 算法<sup>[35]</sup>则增加了闭环检测进程,在前端扫描匹配过程结合 CSM(Correlation Scan Match)与梯度优化,但该算法会带来巨大的计算量.

在 3D 激光 SLAM 领域,Zhang 等<sup>[36]</sup>提出一种 LOAM 方案,利用 3D 激光雷达采集数据,基于特征点的扫描匹配,采用非线性优化方法进行运动估计,能够实时运行且精确度高.随后,他们又推出了改进版本 V-LOAM<sup>[37]</sup>,该方案利用视觉里程计以高频率估计位姿变换,利用激光里程计以低频率优化运动估计,并校准漂移,当光照变化明显时仍具有较好的鲁棒性.多传感器融合需要相匹配的算法,如 LVIO<sup>[38]</sup>、LeGO-LOAM<sup>[39]</sup>和 LIO-mapping<sup>[40]</sup>等.表 1 和表 2 分别为视觉 SLAM 和激光 SLAM 领域常用算法汇总.

表 1 视觉 SLAM 领域的常用算法

Table 1 Commonly used algorithms in VSLAM

算法	传感器形式	网址
MonoSLAM <sup>[20]</sup>	单目	<a href="https://github.com/hanmekim/SceneLib2">https://github.com/hanmekim/SceneLib2</a>
PTAM <sup>[21]</sup>	单目	<a href="https://www.robots.ox.ac.uk/~gk/PTAM/">https://www.robots.ox.ac.uk/~gk/PTAM/</a>
ORB-SLAM2 <sup>[22]</sup>	单目/双目/RGBD	<a href="https://github.com/raulmur/ORB_SLAM2">https://github.com/raulmur/ORB_SLAM2</a>
LSD-SLAM <sup>[24]</sup>	单目为主	<a href="https://vision.in.tum.de/research/vslam/lsdslam">https://vision.in.tum.de/research/vslam/lsdslam</a>
SVO <sup>[41]</sup>	单目	<a href="https://github.com/uzh-rpg/rpg_svo">https://github.com/uzh-rpg/rpg_svo</a>
DTAM <sup>[28]</sup>	RGB-D	<a href="https://github.com/anuranbaka/OpenDTAM">https://github.com/anuranbaka/OpenDTAM</a>
DVO <sup>[29]</sup>	RGB-D	<a href="https://github.com/tum-vision/dvo_slam">https://github.com/tum-vision/dvo_slam</a>
DSO <sup>[26]</sup>	单目	<a href="https://github.com/JakobEngel/dso">https://github.com/JakobEngel/dso</a>
RTAB-Map <sup>[30]</sup>	双目/RGB-D	<a href="https://github.com/introlab/rtabmap">https://github.com/introlab/rtabmap</a>
Elastic Fusion <sup>[27]</sup>	RGB-D	<a href="https://github.com/mp3guy/ElasticFusion">https://github.com/mp3guy/ElasticFusion</a>

表 2 激光 SLAM 领域的常用算法

Table 2 Commonly used algorithms in lidar SLAM

开源方案	传感器(激光)	优点	缺点
FastSLAM <sup>[32]</sup>	2D	实时输出栅格地图	在大场景下,会严重消耗内存,会产生粒子耗散问题
Gmapping <sup>[33]</sup>	2D	运行速度快;对激光雷达频率要求低;鲁棒性高	严重依赖里程计;大的场景、粒子较多的情况下,消耗资源大
Hector-SLAM <sup>[34]</sup>	2D	不依赖里程计;可以适应空中及地面不平坦的区域	对于雷达帧率要求高;优化算法容易陷入局部最小值
LOAM <sup>[36]</sup>	3D	不依赖里程计	无回环检测
V-LOAM <sup>[37]</sup>	3D	精度高;算法鲁棒性好	无回环检测
Cartographer <sup>[35]</sup>	2D	累计误差较低;对雷达性能要求不高	计算量大
LVIO <sup>[38]</sup>	3D	漂移量小;在光照、旋转和结构退化环境下具有良好的鲁棒性	特征匹配较差时效果不好
LeGO-LOAM <sup>[39]</sup>	3D	地面点丰富时相对稳定;得到的地图较为稀疏	缺少接地点时容易坍塌
LIO-mapping <sup>[40]</sup>	3D	具有紧密耦合多线激光雷达和 IMU 框架	计算量大;无回环检测部分;累积误差无法消除
LIO-SAM <sup>[42]</sup>	3D	稳定性强;有回环检测	没有在全局范围内进行扫描匹配

## 2 SLAM 领域常用的传感器

### 2.1 视觉 SLAM 中的硬件

视觉 SLAM 中使用的传感器通常包括单目相机、双目相机和 RGB-D 相机。单目相机指的是仅使用一个摄像头进行轨迹估计和建图,其结构简单、成本低,便于标定和识别。双目相机由两个单目相机组成。单目与双目相机测量物体相对深度的方式都是通过对极几何约束来实现的<sup>[43]</sup>。与单目视觉不同的是,双目视觉既可以在运动时估计深度,也可在静止时估计深度。RGB-D 相机又被称为深度相机,通过该相机能检测出拍摄空间的景深距离,这也是与普通摄像头最大的区别。通过深度相机获取到的数据,人们能准确知道图像中每个点离摄像头的距离,不需要额外的计算量就可以还原真实场景,实现场景建模,在室内场景中具有很高的精度<sup>[44]</sup>。但在室外环境中,RGB-D 相机极易受到光照变化和运动模糊的影响,且长距离跟踪会导致较大的累积误差和尺度偏移。目前主流的单目、双目与 RGB-D 相机的相关硬件信息如表 3 所示。

### 2.2 激光 SLAM 中的硬件

如今的市场逐渐演化为纯视觉和激光雷达两条不同的路线。其中,激光 SLAM 又分为 2D 和 3D 激光,2D 激光雷达一般用于室内机器人(如扫地机器人),而 3D 激光雷达大多用于无人驾驶领域<sup>[45]</sup>。制作激光雷达的国内厂商有禾赛科技、速腾聚创、镭神智能、北科天绘和华为等,国外厂商有 Velodyne LiDAR、IBEO、Quanergy、Aeva、Cepton 等。表 4 介绍了不同型号激光雷达的输入输出数据。图 2 展示了 2D 和 3D 激光雷达生成的点云地图。

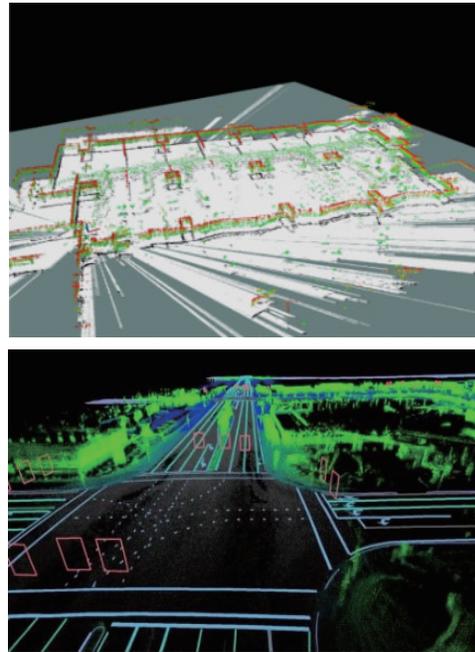


图 2 二维(上)与三维点云地图(下)

Fig. 2 2D point cloud map (up) and 3D point cloud map (down)

表 4 激光雷达输入输出数据及相关硬件信息

Table 4 Lidar input and output data and related hardware information

点云地图	输入	输出
2D	IMU 数据、里程计数据、2D 激光雷达数据	IMU 数据、里程计数据、3D 激光雷达数据
3D	覆盖栅格地图、机器人的轨迹或输出位姿的轨迹图	3D 点云图、机器人的轨迹或输出位姿的轨迹图

表 3 单目、双目和 RGB-D 相机的相关硬件信息

Table 3 Hardware information for monocular, binocular and RGB-D cameras

相机类别	工作原理	国内厂商	国外厂商	参数	价格
RGB-D 相机	ToF(根据脉冲光从相机到目标物体的飞行时间来测量距离)	海康威视、舜宇光学、乐行天下	德州仪器、意法半导体、微软	分辨率低,一般低于 640×640; 帧率可达上百帧;测量范围为 0.1~100 m	根据不同的测量范围和帧率大小,价格从几千到几百万不等
	结构光(相机根据返回的结构光图案,计算物体与自身之间的距离)	奥比中光、华杰艾米、图漾科技	Intel、LeapMotion、Stereolabs	分辨率高,可达 2K 分辨率;帧率一般为 30 帧;测量范围受基线限制,距离越远,误差越大,一般在 2 m 以内	几千元
双目/单目相机	视差法(通过两个相机之间的距离估计每个像素的空间位置)	图漾科技、纵目科技、凌云光科技	PrimeSense、Intel、Heptagon	1 080×720;帧率一般为 30 帧;测量范围为 0.1~10 m	1 mm 级精度千元级,0.1 mm 级精度万元级,0.01 mm 级精度几十万元级

### 3 多传感器SLAM融合方案

尽管激光SLAM与视觉SLAM算法都比较成熟,但仅依靠单一传感器的SLAM方案很容易受到外界因素的影响.因此,学者们逐渐将研究重心转移到两者的融合方案.视觉SLAM在纹理丰富的动态环境中工作稳定,可以为激光雷达SLAM提供非常精确的点云匹配,而激光雷达提供的精确方位和距离信息将对正确匹配的点云发挥更大的作用.在光照严重不足或缺乏纹理的环境中,激光雷达SLAM的定位工作使视觉可以在几乎没有信息的情况下记录场景.不同传感器的融合,本质上是不同算法的融合.在现有融合方案的基础上,本文将从基于不确定度、基于传统特征和基于深度学习的方法分析多传感器SLAM融合方案.

#### 3.1 基于不确定度的融合方案

基于不确定度的方案通常应用于2D激光与视觉融合的SLAM中,目前主流方案有三种:卡尔曼滤波器(KF)、粒子滤波器和基于图形以及它们的衍生算法.KF和粒子滤波器是贝叶斯滤波器的两种不同实现方案,KF主要负责预测和更新两个部分,但它无法满足非线性问题的需求,为此研究人员开发出了EKF的方法,它在中小型场景中取得了良好的效果,但涉及到大型地图时,计算量巨大.UKF(无迹卡尔曼滤波)是解决非线性问题的一种很好的方法.然而,KF及其变体只能处理高斯分布的情况,并且在面对任意分布的情况时,使用KF会带来更大的误差.基于粒子滤波器的方法很好地解决了多个样本任意分布的情况.在该方法中,粒子数越多的区域,概率就越高.基于图形的SLAM则通过最小化平方差之和来找到位姿与位姿之间的关系.

##### 3.1.1 基于KF与粒子滤波器的融合方案

2006年,Newman等<sup>[46]</sup>首次将激光雷达和相机同时安装在移动机器人上,激光雷达获取环境的局部几何形状,用于增量构建工作空间的三维点云地图.他们使用来自摄像机的图像序列来检测闭环事件(不参考车辆位置的内部估计),在闭环检测过程中,使用相机序列进行检测,随后再次使用局部激光扫描处理闭环检测的图像,有效地消除了闭环检测过程中产生的误差,但由此产生的巨大计算量仍然难以解决.2010年,Sun等<sup>[47]</sup>假设传感器噪声服从高斯分布,使用EKF对系统状态进行最小均方差估计,将具有相同角点特征的视觉数据和激光数据进

行融合,采取主动检测策略,提高了SLAM的精度,得到了更多的三维地图信息.

KF与粒子滤波器的融合方案已经逐渐成熟,一些学者将其应用于移动机器人、自动驾驶汽车和无人机,在算法不断改进的过程中,获得了单一传感器无法实现的优异性能.

2007年,Iocchi等<sup>[48]</sup>使用粒子滤波器来估计大型室内环境的映射问题和局部映射之间的位移,他们主要使用双目相机测量平面位移,辅以2D激光雷达数据,并与高精度IMU传感器合作,成功构建了低成本的3D地图.Aycard等<sup>[49]</sup>将激光雷达和双目视觉系统的数据作为输入,应用于复杂的交叉路口场景,将车辆和其他动态物体的风险水平作为输出,利用粒子滤波器解决自身的定位问题(每个粒子对应一个车辆位置,利用激光雷达数据计算每个粒子的概率),取得了良好的检测效果.Collier等<sup>[50]</sup>在FAB-MAP算法的基础上,分别使用SIFT和VD-LSD来训练激光与视觉传感器的词袋模型,在无人机上执行位置识别,无论是光照条件差还是纹理场景低,它都具有良好的召回率和准确率,但当无人机飞行速度过快时容易导致特征跟踪失败.Magree等<sup>[51]</sup>则使用视觉和激光辅助导航,导航架构基于EKF滤波器,为无人机提供传感器更新,在扫描和点对应级别进行耦合,这减少了无人机快速飞行产生模糊几何形状的影响.Wang等<sup>[52]</sup>将改进后的蒙特卡罗定位方法应用于机器人的位姿估计步骤,提出一种基于2D激光雷达的定位算法,从三维点云图中提取出二维结构使得2D激光数据与地图对齐,可在机器人定位自身的同时定位局部地图的比例尺.为解决计算量大、建图过程复杂、定位精度低等问题,研究人员对后端优化和闭环检测步骤进行了大量改进.Yin等<sup>[53]</sup>提出一种基于3D-2D联合对应掩码(CoMask)的端到端的全自动方法,可以高精度地直接估计外部参数,将遗传算法与Levenberg-Marquardt方法相结合,可在不做任何初始估计的情况下求解全局优化问题.图3给出了文献[53]的大致框架,本文采用不同的颜色表示不同的步骤,方便读者更好地理解.

Shin等<sup>[54]</sup>提出一种结合光检测、激光雷达测距和单目相机进行稀疏深度测量的直接视觉-激光雷达SLAM框架(图4),联合优化了多个关键帧下的每个测量值,实现了对稀疏深度的直接利用,即使是极稀疏深度测量(8射线)下,该方法也能获得稳健的SLAM结果,但不适用于光线较差的情况.

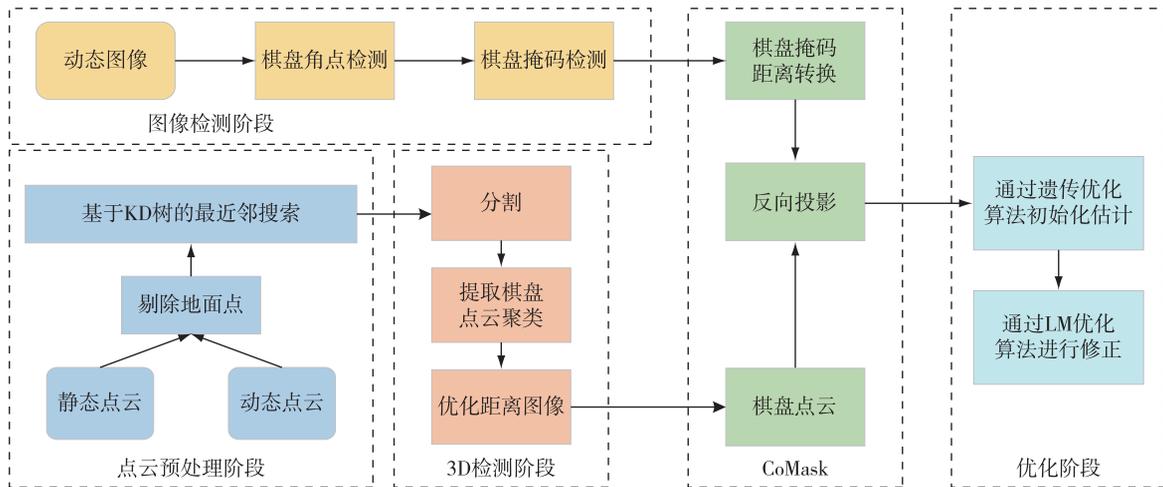


图3 基于3D-2D联合对应CoMask的端到端框架<sup>[53]</sup>

Fig. 3 End-to-end frame diagram based on 3D-2D joint corresponding CoMask<sup>[53]</sup>

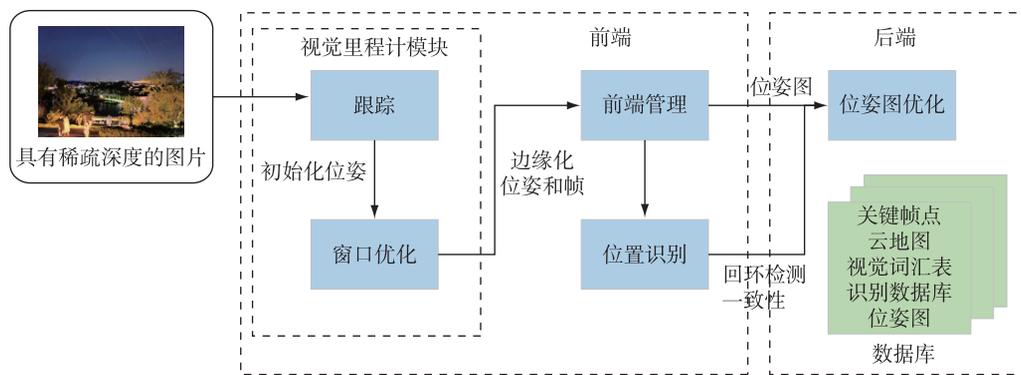


图4 DVL-SLAM框架<sup>[54]</sup>

Fig. 4 Block diagram of direct visual-lidar SLAM<sup>[54]</sup>

### 3.1.2 基于图优化的融合方案

Majdik 等<sup>[55]</sup>将 SURF 特征视为环境标志,并跟踪这些标志在机器人不同位置之间的位移.视觉映射和激光雷达映射系统的交叉使用实现了高效的定位,并且可自主地过滤掉错误检测的地标. Houben 等<sup>[56]</sup>将不同阶段的激光 SLAM 抽象成一个仅与地图构建过程相连的薄界面,提出一种快速简单的标记方法,可有效地检测和解码,并提供一种可以无缝连续地将其位置信息集成在地图中的图优化方法. Jiang 等<sup>[2]</sup>提出一种新的基于图优化的 SLAM 框架,该框架考虑了廉价的激光雷达和视觉传感器数据的融合,设计了一个成本函数来处理扫描数据和图像数据,并在闭环阶段引入具有视觉特征的词袋模型,生成了包含视觉特征和障碍物的 2.5D 地图,比传统的网格地图更快. Mu 等<sup>[57]</sup>基于 UKF,将激光雷达、RGB-D 相机、编码器和 IMU 四个传感器组合进行联

合定位,有效地提高了闭环检测的精度,可构建出更加精细的地图. Chen 等<sup>[58]</sup>研究了激光雷达和视觉 SLAM 的后端,构建了一种基于闭环检测和全局图优化的方法,在主阶段利用激光雷达的几何特征和视觉特征,在辅助阶段构建描述视觉相似度的词袋模型,显著提高了系统的稳定性.

### 3.2 基于传统特征的融合方案

传统的基于特征的方法在多传感器融合领域同样发挥着至关重要的作用,目前的融合方案主要基于 ORB-SLAM 框架. ORB-SLAM 及其后续版本因其优异的实时性能和鲁棒性,已成为应用最广泛的视觉 SLAM 解决方案之一. 但 ORB-SLAM 系列在很大程度上依赖于环境特征,在没有纹理特征的环境中很难获得足够的特征点,多传感器的融合方案恰好能够解决这些问题,它可以为 ORB-SLAM 系统提供充足的信息.

3.2.1 基于ORB-SLAM框架的融合方案

2016年,Liang等<sup>[59]</sup>利用ORB特征和词袋特征进行闭环检测,将识别良好的LRGC SLAM框架和SPA优化算法应用于SLAM,在环境中引入视觉信息,成功解决了大规模的激光SLAM闭环检测问题,但在ORB特征缺失的情况下很容易失败.同年,Lv等<sup>[60]</sup>利用激光雷达精确地获取距离信息,改进了ORB-SLAM算法的地图初始化过程,通过计算平均深度来估计绝对尺度,实现了未知环境下的精确定位.Zhang等<sup>[61]</sup>使用激光信息辅助视觉优化位姿,整体框架基于ORB-SLAM2.首先将精确环境信息从激光雷达传递到视觉部分,然后转化为视觉跟踪线程的姿态来优化初始值,并且系统可以自适应地改变两种传感器的权重进行融合,但在生成轨迹的过程中可能会出现波动.

此后,又有研究人员尝试使用点线特征<sup>[62]</sup>和激光点云<sup>[63]</sup>代替ORB特征.Huang等<sup>[62]</sup>在姿态估计中引入点和线的特征,并将ORB特征作为点线特征(点线特征不易受噪声、宽视角和运动模糊的影响),与传统的仅基于点的激光雷达-视觉测距方法相比,提高了环境结构信息的利用率和姿态估计的准确性.Ali等<sup>[63]</sup>提出一种基于特征的SLAM算法,

首先使用相机参数矩阵将三维点云栅格转化为图像,随后将图像导入ORB特征检测器.该方法可估计机器人的六自由度姿态,在各种环境下都具有优异的性能,但环境中的动态对象会影响系统的性能.为此,Kang等<sup>[64]</sup>提出一种范围增强的全景视觉同时定位和映射系统(RPV-SLAM),该系统以全景相机为主要传感器,利用倾斜激光雷达获得的距离信息来增强视觉特征并输出度量尺度,从激光雷达传感器中获得深度信息的初始范围,在此范围内提取ORB特征,从稀疏深度测量中恢复密集深度图,这在复杂的室外条件下仍然是稳健的.Chou等<sup>[65]</sup>受到ORB-SLAM2框架的启发,提出一种紧密耦合的视觉激光雷达SLAM,在前端两者独立运行,后端则融合所有的激光雷达与视觉信息,提出一种新的激光雷达残差压缩方法(图5).

3.2.2 其他融合方案

除了目前主流的基于ORB-SLAM框架进行多传感器融合的方法外,还有许多优秀的融合方案值得借鉴和研究.Radmanesh等<sup>[66]</sup>提出一种基于光探测与激光测距的单目SLAM方法来提供深度信息,该方法使用相机数据以无监督的方式处理未知物体,使用视觉检测到的特征作为地标特征,并将其与激

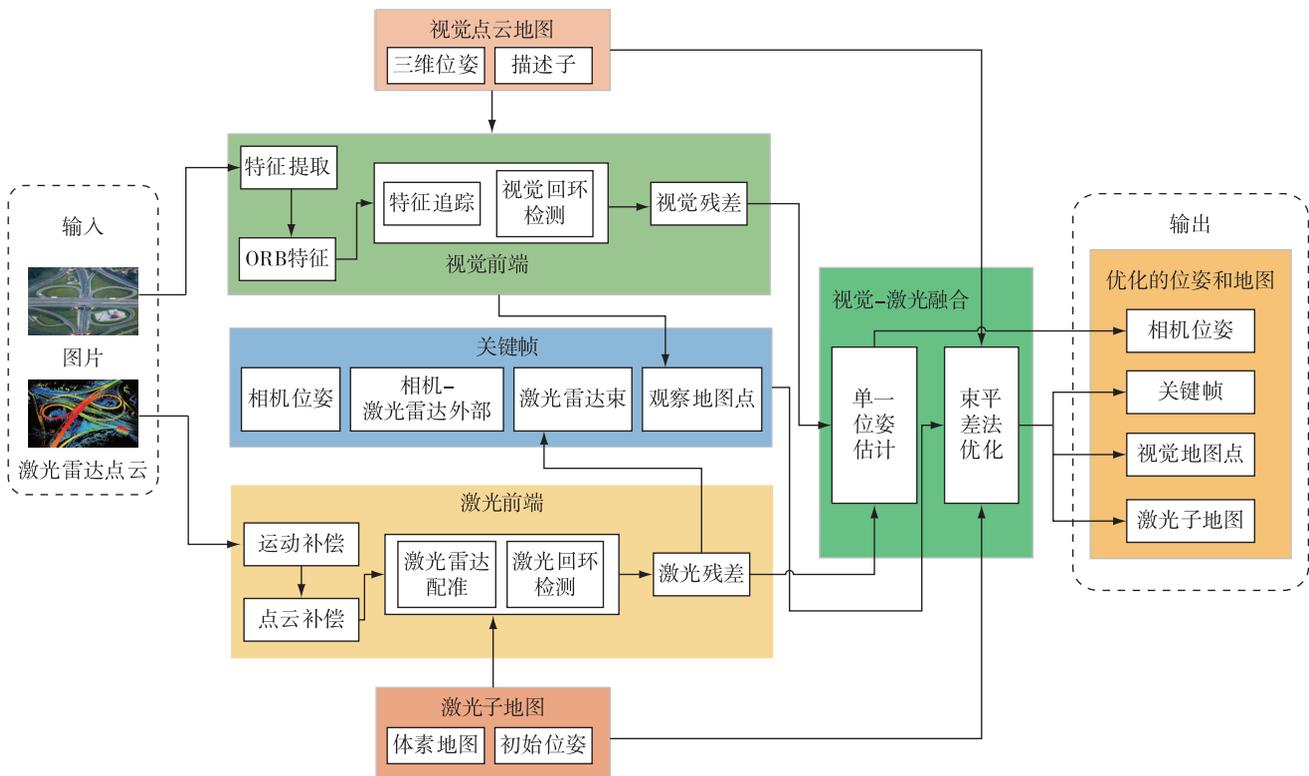


图5 基于特征的视觉-激光紧耦合融合框架<sup>[65]</sup>

Fig. 5 Block diagram of feature-based tightly coupled visual-lidar SLAM<sup>[65]</sup>

光雷达传感器数据融合.在计算效率和精度方面,该方法优于目前仅由激光雷达生成的地图.Wang 等<sup>[67]</sup>提出一种双层优化策略,在局部估计层,通过激光雷达里程计和视觉惯性里程计获得相对姿态,并在全局优化层引入 GPS 信息来校正累积漂移,从而在没有全局漂移的情况下实现精确的绝对定位.Yi 等<sup>[68]</sup>优化了 ORB-SLAM3,提出一种行为树框架,该框架可以从视觉特征、激光雷达地标和 GPS 中智能地选择最佳的全球定位方法,形成一个长期可用的特征图,可以自动校正比例,最大限度地减少全局漂移,满足了复杂大规模场景的需求.

### 3.3 基于深度学习的融合方案

目前深度学习技术中使用的神经网络主要分为 CNN、RNN 和 DNN<sup>[69]</sup> 三类.图 6 展示了融合方案使用的经典神经网络框架.由于图像中存在固有的局部模式(如人的嘴巴、眼睛和鼻子),局部特征图像的识别通常具有更快的速度和更高的准确率<sup>[70]</sup>,因此研究人员将图像处理和神经网络相结合,产生了 CNN.DNN 具有与多层感知机相似的结构,可以克服梯度消失问题,避免陷入局部最优解<sup>[71]</sup>.然而,DNN 无法对时间序列上的变化进行建模,因此出现了 RNN,它可以处理上下文相关的数据类型<sup>[72]</sup>,但 RNN 网络在多传感器融合领域并没有得到太多尝试.CNN 具有分类、识别、预测和决策能力的特点,并且在 DNN 拟合范围内处理,可以更快地达到局部最优解.它们都可以融合 SLAM 的多个传感器模块一起使用<sup>[73]</sup>.

#### 3.3.1 基于 CNN 的融合方案

近年来,CNN 在图像处理方面的优势已被广泛应用于单一传感器 SLAM 方案<sup>[74]</sup>中,它使得单目相机也可以获得可靠的深度信息.Ma 等<sup>[75]</sup>考虑从低分辨率稀疏深度数据中预测密集深度数据,以获得具有更高鲁棒性和准确性的地图.他们使用 CNN 来学习深度图像预测的深度回归模型,并将该模型用作稀疏 SLAM、视觉惯性里程计算法和激光雷达测量超分辨率的插件,提高了准确率.Kang 等<sup>[76]</sup>将激光雷达数据与 RGB-D 点云对齐,以生成相应场景的连续视频帧,并使用 CNN 网络进行训练,采用 PoseNet 神经网络实现运动恢复和系统自动初始化.在大规模室内和复杂场景中的实验结果表明,与传统的 SLAM 算法相比,闭环检测的累计误差减少,整体鲁棒性高于 ORB-SLAM2.图 7 流程<sup>[76]</sup>主要由 5 个部分组成:1) 收集环境信息和激光雷达数据;2) 跟踪过程中的精确自动初始化 RGB-D SLAM 算法,提取出所有关键帧;3) 消除冗余的关键帧;4) 使用 ICP 算法确定相机姿态并选择正确的关键帧;5) 回环检测.

#### 3.3.2 基于 DNN 的融合方案

CNN 由于其出色的图像处理性能而受到了大量研究人员的青睐,相比之下,基于 DNN 的多传感器融合方法相对较少,目前更多是将 DNN 与 CNN 进行部分融合.An 等<sup>[77]</sup>提出一种新的无监督多通道视觉-激光雷达 SLAM 方法(MVL-SLAM),充分结合了激光雷达和视觉传感器的优点,并将 RCNN 应用于

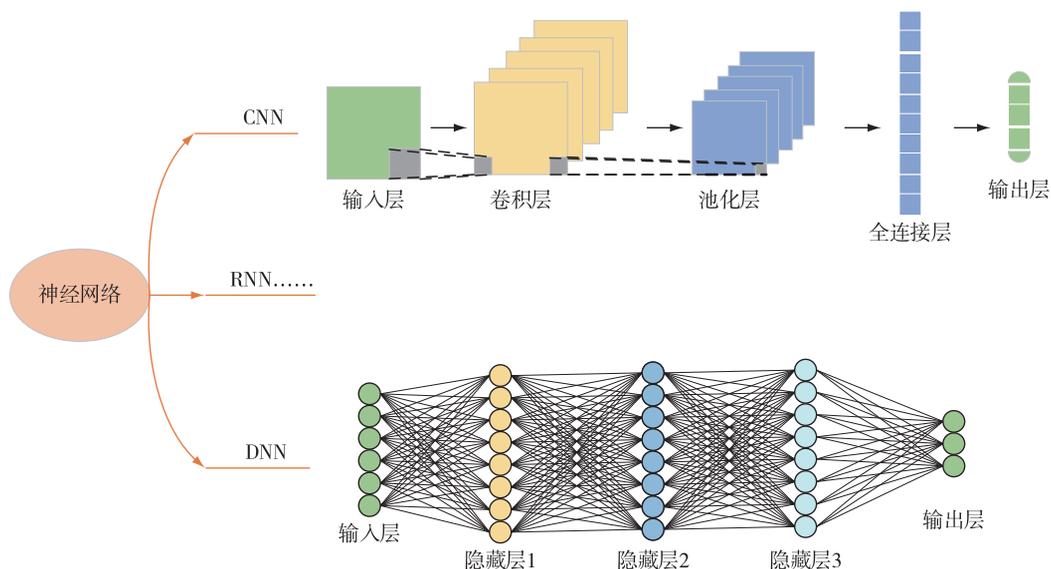


图 6 CNN 和 DNN 的结构框图

Fig. 6 Block diagram of CNN and DNN

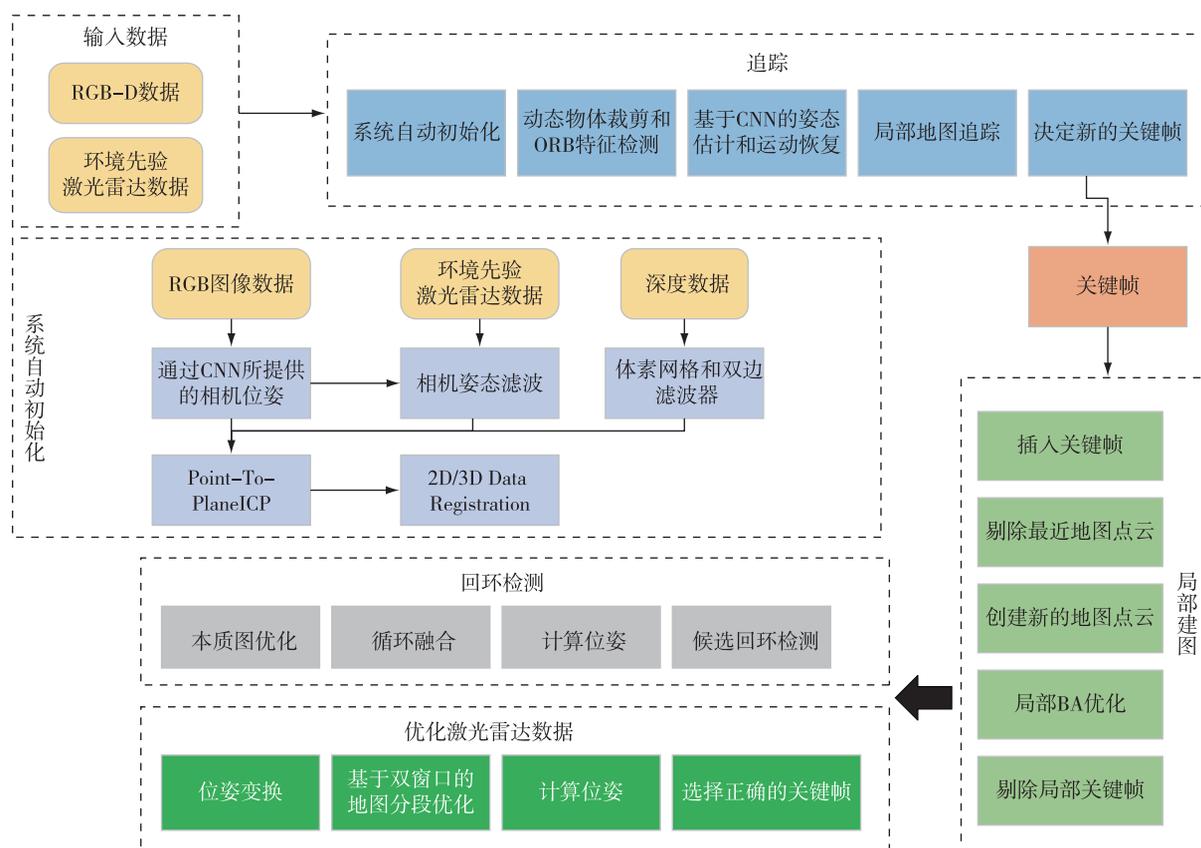


图7 基于CNN的激光雷达与深度相机融合方案流程

Fig. 7 Flowchart of CNN based lidar and depth camera fusion scheme

融合方案组件,使用DNN的特征作为闭环检测组件.该方法不需要制作预训练数据,可直接从3D制图组件构建环境3D地图.Cattaneo等<sup>[78]</sup>使用激光雷达地图执行全局视觉定位,利用DNN创建共享嵌入空间,该空间包含图像和激光雷达地图信息,允许图像到3D激光雷达位置识别.他们利用DNN和CNN分别提取激光雷达点云和图像信息,给定权重后在共享嵌入空间中进行融合,以实现精确的位置识别(图8).

### 3.4 评估工具和数据集

近几十年来,出现了许多优秀的SLAM算法,并在自主导航、移动机器人和AR/VR领域得到了成熟的应用.每种算法都有其独特的改进方法,不同的算法需要不同的时间,可以达到不同的精度和应用场合,因此需要一个统一的评估工具来测试该算法在数据集上的性能.准确度是研究人员评估SLAM算法的最重要指标,SLAM算法包括绝对轨迹误差(Absolute Trajectory Error, ATE)和相对姿态

误差(Relative Pose Error, RPE).RPE用于计算两个相同时间戳的位姿变化差异,适用于估计系统漂移.ATE直接计算相机姿态的真实值与SLAM系统的估计值之间的差.EVO(<https://github.com/MichaelGrupp/evo>, accessed 21 June 2022)是评估改进SLAM算法的常用工具,可用于各种数据集,还可以绘制测试算法与真实轨迹的对比图.SLAMBench2(<https://github.com/MichaelGrupp/evo>, accessed 26 September 2022)是一个公开的软件框架,通过可扩展的数据集列表评估当前和未来的SLAM系统.它包括开放源代码和封闭源代码,同时使用可比较的和指定的性能指标列表.它支持多种现有的SLAM算法和数据集,如Elastic Fusion<sup>[27]</sup>、ORB-SLAM2<sup>[22]</sup>和OKVIS<sup>[79]</sup>.

用于测试视觉SLAM与激光SLAM的常用数据集有KITTI、Oxford、ASL Kinect、ASL RGB-D、TUM RGB-D、ICL-NUIM、VaFRIC、EuRoC、TUM VI和TUM monoVO.具体情况如表5所示.

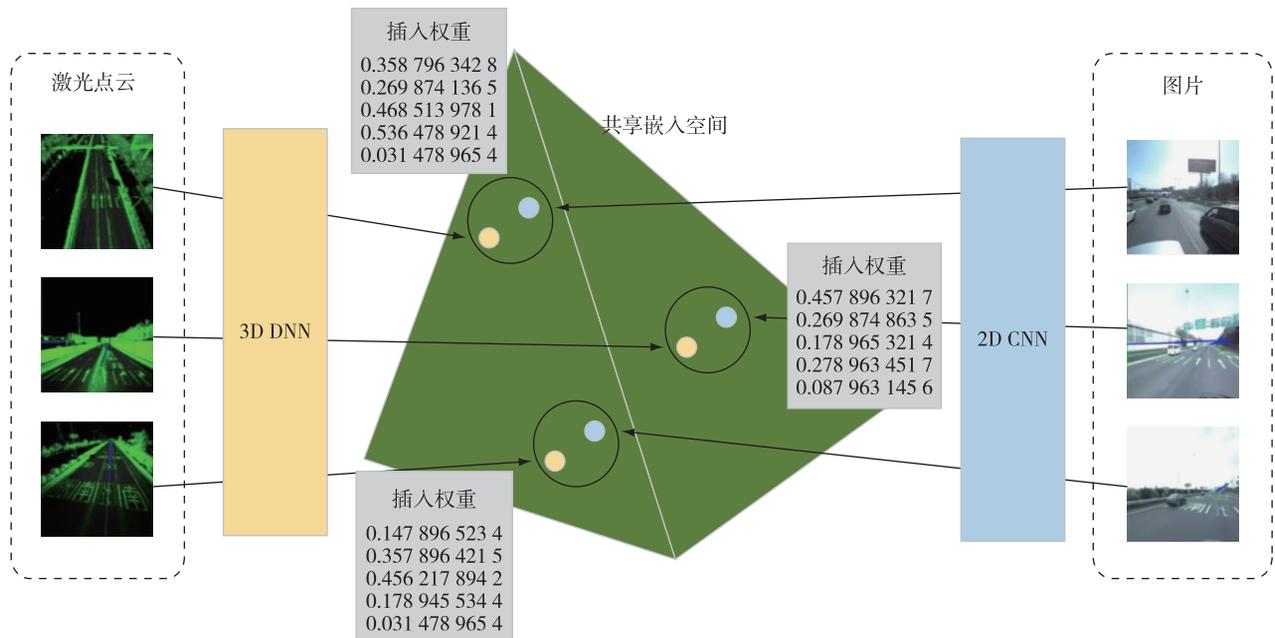


图 8 基于 DNN 的激光雷达与相机融合方案流程<sup>[78]</sup>

Fig. 8 Flowchart of DNN-based lidar and camera fusion scheme<sup>[78]</sup>

表 5 SLAM 领域的常用数据集

Table 5 Common datasets in SLAM area

数据集	制作单位	特点	传感器	环境
KITTI	卡尔斯鲁厄理工学院	主要包含自动驾驶场景	RGB-D+激光雷达+GPS+IMU	室外
Oxford	牛津大学	主要包含自动驾驶场景	RGB-D+激光雷达+GPS+IMU	室外
ASL Kinect	苏黎世联邦理工学院	包含深度信息的 SLAM 与三维重建数据集	RGB-D	室内
ASL RGB-D	苏黎世联邦理工学院	包含深度信息的 SLAM 与三维重建数据集	RGB-D+激光雷达	室内
TUM RGB-D	慕尼黑科技大学	包含深度信息的 SLAM 与三维重建数据集	RGB-D	室内
ICL-NUIM	帝国理工学院	包含深度信息的 SLAM 与三维重建数据集	RGB-D	室内
VaFRIC	帝国理工学院	包含深度信息的 SLAM 与三维重建数据集	RGB-D	室内
EuRoC	欧洲机器人协会	不包含深度信息的 SLAM 数据集	IMU+双目	室内
TUM VI	慕尼黑科技大学	不包含深度信息的 SLAM 数据集	IMU+双目	室内/室外
TUMmonoVO	慕尼黑科技大学	不包含深度信息的 SLAM 数据集	单目	室内/室外

#### 4 总结与展望

本文回顾了视觉与激光 SLAM 的发展历史,对该领域的研究做出了总结.多传感器融合 SLAM 可以显著提高机器人的自主交互能力,并且在鲁棒性、精度和高级感知等方面远超前于单一 SLAM 技术.结合这些研究,本文对多传感器融合 SLAM 的未来作出了以下展望:

1) 发展历史及工程应用.尽管近年来多传感器 SLAM 的融合方案取得了一些突出成就,但与传统的纯视觉与激光 SLAM 相比,仍处于发展阶段,而更

多传感器的参与,意味着需要更庞大的计算能力和更完善的系统来消除无用信息,它将会严重干扰 SLAM 的实时性能.这种情况将随着算法的不断发展和软硬件的不断更新而改善.

2) 理论支持.基于不确定度和基于传统特征的融合方案并非贯穿于整个 SLAM 过程,大多数融合方案都针对前端、后端、回环检测和建图 4 过程中的一个.因此,如何将融合方案应用于整个 SLAM 过程仍然是一个巨大的挑战.同时,深度学习技术提取的特征缺乏直观意义和理论依据,人们无法知道计算机使用什么标准来提取特征.目前,基于 RNN 的

多传感器融合领域仍处于空白阶段,传统方案仍具有很大优势.

3) 人机交互能力.多传感器带来更丰富的环境信息,在实际场景中,系统筛选出的冗余信息是否与人们需要获取的信息不同,将丰富的信息应用于机器人的自主交互过程仍然是一项艰巨的任务.

## 参考文献

### References

- [ 1 ] 赵乐文,任嘉倩,丁杨.基于GNSS的空间环境参数反演平台及精度评估[J].南京信息工程大学学报(自然科学版),2021,13(2):204-210  
ZHAO Lewen, REN Jiaqian, DING Yang. Platform for GNSS real-time space environment parameter inversion and its accuracy evaluation[J]. Journal of Nanjing University of Information Science & Technology ( Natural Science Edition ), 2021, 13(2):204-210
- [ 2 ] Jiang G L, Yin L, Jin S K, et al. A simultaneous localization and mapping (SLAM) framework for 2.5D map building based on low-cost lidar and vision fusion [J]. Applied Sciences, 2019, 9(10):2105
- [ 3 ] Smith R C, Cheeseman P. On the representation and estimation of spatial uncertainty [J]. The International Journal of Robotics Research, 1986, 5(4):56-68
- [ 4 ] 王霞,左一凡.视觉SLAM研究进展[J].智能系统学报,2020,15(5):825-834  
WANG Xia, ZUO Yifan. Advances in visual SLAM research[J]. CAAI Transactions on Intelligent Systems, 2020, 15(5):825-834
- [ 5 ] 胡凯,吴佳胜,郑翥,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, ZHENG 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
- [ 6 ] 危双丰,庞帆,刘振彬,等.基于激光雷达的同时定位与地图构建方法综述[J].计算机应用研究,2020,37(2):327-332  
WEI Shuangfeng, PANG Fan, LIU Zhenbin, et al. Survey of lidar-based SLAM algorithm [J]. Application Research of Computers, 2020, 37(2):327-332
- [ 7 ] 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
- [ 8 ] Neira J, Tardos J D. Data association in stochastic mapping using the joint compatibility test [J]. IEEE Transactions on Robotics and Automation, 2001, 17(6):890-897
- [ 9 ] Julier S J, Uhlmann J K. A counter example to the theory of simultaneous localization and map building [C] // Proceedings 2001 IEEE International Conference on Robotics and Automation ( Cat. No. 01CH37164 ). May 21-26, 2001, Seoul, Korea ( South ). IEEE, 2003: 4238-4243
- [ 10 ] Aulinas J, Petillot Y, Salvi J, et al. The SLAM problem: a survey [J]. Artificial Intelligence Research and Development, 2008:363-371
- [ 11 ] 尹姝,陈元桦,仇翔.基于RFID和自适应卡尔曼滤波的室内移动目标定位方法[J].南京信息工程大学学报(自然科学版),2018,10(6):749-753  
YIN Shu, CHEN Yuanyuan, QIU Xiang. Indoor moving-target localization using RFID and adaptive Kalman filter [J]. Journal of Nanjing University of Information Science & Technology ( Natural Science Edition ), 2018, 10(6):749-753
- [ 12 ] 李晓飞,宋亚男,徐荣华,等.基于双目视觉的船舶跟踪与定位[J].南京信息工程大学学报(自然科学版),2015,7(1):46-52  
LI Xiaofei, SONG Yanan, XU Ronghua, et al. Tracking and positioning of ship based on binocular vision [J]. Journal of Nanjing University of Information Science & Technology ( Natural Science Edition ), 2015, 7(1):46-52
- [ 13 ] Dissanayake G, Huang S D, Wang Z, et al. A review of recent developments in simultaneous localization and mapping [C] // 2011 6th International Conference on Industrial and Information Systems. August 16-19, 2011, Kandy, Sri Lanka. IEEE, 2011:477-482
- [ 14 ] Fraundorfer F, Scaramuzza D. Visual odometry: part II: matching, robustness, optimization, and applications [J]. IEEE Robotics & Automation Magazine, 2012, 19(2):78-90
- [ 15 ] Mur-Artal R, Montiel J M M, Tardós J D. ORB-SLAM: a versatile and accurate monocular SLAM system [J]. IEEE Transactions on Robotics, 2015, 31(5):1147-1163
- [ 16 ] Li R H, Wang S, Gu D B. DeepSLAM: a robust monocular SLAM system with unsupervised deep learning [J]. IEEE Transactions on Industrial Electronics, 2021, 68(4):3577-3587
- [ 17 ] Bescos B, Fàcil 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
- [ 18 ] Han X F, Laga H, Bennamoun M. Image-based 3D object reconstruction: state-of-the-art and trends in the deep learning era [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(5):1578-1604
- [ 19 ] Li C, Wang S, Zhuang Y, et al. Deep sensor fusion between 2D laser scanner and IMU for mobile robot localization [J]. IEEE Sensors Journal, 2021, 21(6):8501-8509
- [ 20 ] 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
- [ 21 ] 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. November 13-16, 2007, Nara, Japan. IEEE, 2008: 225-234
- [ 22 ] 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
- [23] Campos C, Elvira R, Rodríguez J J G, et al. ORB-SLAM3: an accurate open-source library for visual, visual-inertial, and multimap SLAM [J]. IEEE Transactions on Robotics, 2021, 37(6): 1874-1890
- [24] Engel J, Schöps T, Cremers D. LSD-SLAM: large-scale direct monocular SLAM [M] // Computer Vision – ECCV 2014. Cham: Springer International Publishing, 2014: 834-849
- [25] Forster C, Pizzoli M, Scaramuzza D. SVO: fast semi-direct monocular visual odometry [C] // 2014 IEEE International Conference on Robotics and Automation (ICRA). May 31 – June 7, 2014, Hong Kong, China. IEEE, 2014: 15-22
- [26] Engel J, Koltun V, Cremers D. Direct sparse odometry [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(3): 611-625
- [27] 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
- [28] Newcombe R A, Lovegrove S J, Davison A J. DTAM: dense tracking and mapping in real-time [C] // 2011 International Conference on Computer Vision. November 6–13, 2011, Barcelona, Spain. IEEE, 2012: 2320-2327
- [29] Kerl C, Sturm J, Cremers D. Dense visual SLAM for RGB-D cameras [C] // 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. November 3–7, 2013, Tokyo, Japan. IEEE, 2014: 2100-2106
- [30] Labbé M, Michaud F. RTAB-Map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation [J]. Journal of Field Robotics, 2019, 36(2): 416-446
- [31] Endres F, Hess J, Sturm J, et al. 3-D mapping with an RGB-D camera [J]. IEEE Transactions on Robotics, 2014, 30(1): 177-187
- [32] Montemerlo M, Thrun S, Koller D, et al. FastSLAM: a factored solution to the simultaneous localization and mapping problem [C] // AAAI-02 Proceedings, 2002: 593-598
- [33] Grisetti G, Stachniss C, Burgard W. Improved techniques for grid mapping with Rao-Blackwellized particle filters [J]. IEEE Transactions on Robotics, 2007, 23(1): 34-46
- [34] Kohlbrecher S, von Stryk O, Meyer J, et al. A flexible and scalable SLAM system with full 3D motion estimation [C] // 2011 IEEE International Symposium on Safety, Security, and Rescue Robotics. November 1–5, 2011, Kyoto, Japan. IEEE, 2011: 155-160
- [35] Hess W, Kohler D, Rapp H, et al. Real-time loop closure in 2D lidar SLAM [C] // 2016 IEEE International Conference on Robotics and Automation (ICRA). May 16–21, 2016, Stockholm, Sweden. IEEE, 2016: 1271-1278
- [36] Zhang J, Singh S. LOAM: lidar odometry and mapping in real-time [C] // Robotics: Science and Systems Conference. Robotics: Science and Systems Foundation, 2014: 109-111
- [37] Zhang J, Singh S. Visual-lidar odometry and mapping: low-drift, robust, and fast [C] // 2015 IEEE International Conference on Robotics and Automation (ICRA). May 26–30, 2015, Seattle, WA, USA. IEEE, 2015: 2174-2181
- [38] Zhang J, Singh S. Laser-visual-inertial odometry and mapping with high robustness and low drift [J]. Journal of Field Robotics, 2018, 35(8): 1242-1264
- [39] Shan T X, Englot B. LeGO-LOAM: lightweight and ground-optimized lidar odometry and mapping on variable terrain [C] // 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). October 1–5, 2018, Madrid, Spain. IEEE, 2019: 4758-4765
- [40] Ye H Y, Chen Y Y, Liu M. Tightly coupled 3D lidar inertial odometry and mapping [C] // 2019 International Conference on Robotics and Automation (ICRA). May 20–24, 2019, Montreal, QC, Canada. IEEE, 2019: 3144-3150
- [41] Forster C, Zhang Z C, Gassner M, et al. SVO: semidirect visual odometry for monocular and multicamera systems [J]. IEEE Transactions on Robotics, 2017, 33(2): 249-265
- [42] Shan T X, Englot B, Meyers D, et al. LIO-SAM: tightly-coupled lidar inertial odometry via smoothing and mapping [C] // 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). October 24–January 24, 2021, Las Vegas, NV, USA. IEEE, 2021: 5135-5142
- [43] 张钊源. 基于双目视觉的机器人导航系统的研究与实现 [D]. 重庆: 重庆大学, 2019
- ZHANG Zhaoyuan. Research and implementation of mobile robot navigation system based on stereo vision [D]. Chongqing: Chongqing University, 2019
- [44] Esper I D E, Smolkin O, Manko M, et al. Evaluation of RGB-D multi-camera pose estimation for 3D reconstruction [J]. Applied Sciences, 2022, 12(9): 4134
- [45] 项志宇. 基于激光雷达的机器人障碍检测和自定位 [D]. 杭州: 浙江大学, 2002
- XIANG Zhiyu. Obstacle detection and self-localization for mobile robots by using laser range finders [D]. Hangzhou: Zhejiang University, 2002
- [46] Newman P, Cole D, Ho K. Outdoor SLAM using visual appearance and laser ranging [C] // Proceedings 2006 IEEE International Conference on Robotics and Automation. May 15–19, 2006, Orlando, FL, USA. IEEE, 2006: 1180-1187
- [47] Sun F C, Zhou Y, Li C, et al. Research on active SLAM with fusion of monocular vision and laser range data [C] // 2010 8th World Congress on Intelligent Control and Automation. July 7–9, 2010, Jinan, China. IEEE, 2010: 6550-6554
- [48] Iocchi L, Pellegrini S, Tipaldi G D. Building multi-level planar maps integrating LRF, stereo vision and IMU sensors [C] // 2007 IEEE International Workshop on Safety, Security and Rescue Robotics. September 27–29, 2007, Rome, Italy. IEEE, 2007: 1-6
- [49] Aycard O, Baig Q, Bota S, et al. Intersection safety using lidar and stereo vision sensors [C] // 2011 IEEE Intelligent Vehicles Symposium (IV). June 5–9, 2011, Baden-Baden, Germany. IEEE, 2011: 863-869
- [50] Collier J, Se S, Kotamraju V. Multi-sensor appearance-

- based place recognition [C]//2013 International Conference on Computer and Robot Vision. May 28–31, 2013, Regina, SK, Canada. IEEE, 2013: 128-135
- [51] Magree D, Johnson E N. Combined laser and vision-aided inertial navigation for an indoor unmanned aerial vehicle [C]//2014 American Control Conference. June 4–6, 2014, Portland, OR, USA. IEEE, 2014: 1900-1905
- [52] Wang S, Kobayashi Y, Ravankar A A, et al. A novel approach for lidar-based robot localization in a scale-drifted map constructed using monocular SLAM [J]. *Sensors*, 2019, 19(10): 2230
- [53] Yin L, Luo B, Wang W, et al. CoMask: corresponding mask-based end-to-end extrinsic calibration of the camera and lidar [J]. *Remote Sensing*, 2020, 12(12): 1925
- [54] Shin Y S, Park Y S, Kim A. DVL-SLAM: sparse depth enhanced direct visual-lidar SLAM [J]. *Autonomous Robots*, 2020, 44(2): 115-130
- [55] Majdik A L, Szoke I, Tamas L, et al. Laser and vision based map building techniques for mobile robot navigation [C]//2010 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR). May 28–30, 2010, Cluj-Napoca, Romania. IEEE, 2010: 1-6
- [56] Houben S, Droschel D, Behnke S. Joint 3D laser and visual fiducial marker based SLAM for a micro aerial vehicle [C]//2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI). September 19–21, 2016, Baden-Baden, Germany. IEEE, 2017: 609-614
- [57] Mu L L, Yao P T, Zheng Y C, et al. Research on SLAM algorithm of mobile robot based on the fusion of 2D lidar and depth camera [J]. *IEEE Access*, 2020, 8: 157628-157642
- [58] Chen S B, Zhou B D, Jiang C H, et al. A lidar/visual SLAM backend with loop closure detection and graph optimization [J]. *Remote Sensing*, 2021, 13(14): 2720
- [59] Liang X, Chen H Y, Li Y J, et al. Visual laser-SLAM in large-scale indoor environments [C]//2016 IEEE International Conference on Robotics and Biomimetics (ROBIO). December 3–7, 2016, Qingdao, China. IEEE, 2017: 19-24
- [60] Lv Q, Ma J Y, Wang G S, et al. Absolute scale estimation of ORB-SLAM algorithm based on laser ranging [C]//2016 35th Chinese Control Conference (CCC). July 27–29, 2016, Chengdu, China. IEEE, 2016: 10279-10283
- [61] Zhang Y C, Zhang H, Xiong Z H, et al. A visual SLAM system with laser assisted optimization [C]//2019 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM). July 8–12, 2019, Hong Kong, China. IEEE, 2019: 187-192
- [62] Huang S S, Ma Z Y, Mu T J, et al. Lidar-monocular visual odometry using point and line features [C]//2020 IEEE International Conference on Robotics and Automation (ICRA). May 31–August 31, 2020, Paris, France. IEEE, 2020: 1091-1097
- [63] Ali W, Liu P L, Ying R D, et al. A feature based laser SLAM using rasterized images of 3D point cloud [J]. *IEEE Sensors Journal*, 2021, 21(21): 24422-24430
- [64] Kang J, Zhang Y J, Liu Z, et al. RPV-SLAM: range-augmented panoramic visual SLAM for mobile mapping system with panoramic camera and tilted lidar [C]//2021 20th International Conference on Advanced Robotics (ICAR). December 6–10, 2021, Ljubljana, Slovenia. IEEE, 2022: 1066-1072
- [65] Chou C C, Chou C F. Efficient and accurate tightly-coupled visual-lidar SLAM [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(9): 14509-14523
- [66] Radmanesh R, Wang Z Y, Chipade V S, et al. LIV-LAM: lidar and visual localization and mapping [C]//2020 American Control Conference (ACC). July 1–3, 2020, Denver, CO, USA. IEEE, 2020: 659-664
- [67] Wang K, Cao C, Ma S, et al. An optimization-based multi-sensor fusion approach towards global drift-free motion estimation [J]. *IEEE Sensors Journal*, 2021, 21(10): 12228-12235
- [68] Yi S Q, Worrall S, Nebot E. Integrating vision, lidar and GPS localization in a behavior tree framework for urban autonomous driving [C]//2021 IEEE International Intelligent Transportation Systems Conference (ITSC). September 19–22, 2021, Indianapolis, IN, USA. IEEE, 2021: 3774-3780
- [69] 胡凯, 郑翡, 卢飞宇, 等. 基于深度学习的行为识别算法综述 [J]. *南京信息工程大学学报(自然科学版)*, 2021, 13(6): 730-743  
HU Kai, ZHENG Fei, LU Feiyu, et al. A survey of action recognition algorithms based on deep learning [J]. *Journal of Nanjing University of Information Science & Technology (Natural Science Edition)*, 2021, 13(6): 730-743
- [70] Mumuni A, Mumuni F. CNN architectures for geometric transformation-invariant feature representation in computer vision: a review [J]. *SN Computer Science*, 2021, 2(5): 1-23
- [71] Covington P, Adams J, Sargin E. Deep neural networks for YouTube recommendations [C]//Proceedings of the 10th ACM Conference on Recommender Systems. New York, NY, USA: ACM, 2016: 191-198
- [72] Ma R B, Wang R, Zhang Y B, et al. RNN-SLAM: reconstructing the 3D colon to visualize missing regions during a colonoscopy [J]. *Medical Image Analysis*, 2021, 72: 102100
- [73] Chen W F, Shang G T, Ji A H, et al. An overview on visual SLAM: from tradition to semantic [J]. *Remote Sensing*, 2022, 14(13): 3010
- [74] Ai Y B, Rui T, Lu M, et al. DDL-SLAM: a robust RGB-D SLAM in dynamic environments combined with deep learning [J]. *IEEE Access*, 2020, 8: 162335-162342
- [75] Ma F C, Karaman S. Sparse-to-dense: depth prediction from sparse depth samples and a single image [C]//2018 IEEE International Conference on Robotics and Automation (ICRA). May 21–25, 2018, Brisbane, QLD, Australia. IEEE, 2018: 4796-4803
- [76] Kang X J, Li J, Fan X T, et al. Real-time RGB-D simultaneous localization and mapping guided by terrestrial lidar point cloud for indoor 3-D reconstruction and camera pose estimation [J]. *Applied Sciences*, 2019, 9(16): 3264

- [77] An Y, Shi J, Gu D B, et al. Visual-lidar SLAM based on unsupervised multi-channel deep neural networks [J]. Cognitive Computation, 2022, 14(4): 1496-1508
- [78] Cattaneo D, Vaghi M, Fontana S, et al. Global visual localization in lidar-maps through shared 2D-3D embedding space [C]//2020 IEEE International Conference on Robotics and Automation (ICRA). May 31– August 31, 2020, Paris, France. IEEE, 2020: 4365-4371
- [79] 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

## Overview of SLAM research based on heterogeneous data fusion

ZHOU Chengjun<sup>1</sup> CHEN Weifeng<sup>1</sup> SHANG Guangtao<sup>1</sup> WANG Xiyang<sup>1</sup> XU Chonghui<sup>1</sup> LI Zhenxiang<sup>1</sup>

<sup>1</sup> School of Automation, Nanjing University of Information Science & Technology, Nanjing 210044, China

**Abstract** Laser SLAM (Simultaneous Localization and Mapping) and visual SLAM have been fully developed and widely used in military and civil fields. However, single sensor SLAM has limitations, for instance, laser SLAM is not suitable for scenes with a large number of dynamic objects around it, while visual SLAM has poor robustness in low-texture environments. Therefore, fusion of the two technologies has great potential to compensate each other, and it can be prospected that SLAM technology combining laser and vision or even more sensors will be the mainstream direction in the future. Here, we review the development of SLAM technology, analyze the hardware information of lidar and camera, and introduce some classical open-source algorithms and datasets. Furthermore, the multi-sensor fusion schemes are detailed from perspectives of uncertainty, feature and novel deep learning. The excellent performance of multi-sensor fusion schemes in complex scenes are summarized, and the future development trend of multi-sensor fusion is prospected.

**Key words** simultaneous localization and mapping (SLAM); lidar SLAM; VSLAM; multi-sensor fusion; mobile robot