北京航空航天大学学报
Journal of Beijing University of Aeronautics and Astronautics
ISSN 1001-5965,CN 11-2625/V

《北京航空航天大学学报》网络首发论文

题目：视觉 SLAM 方法综述
作者：王朋，郝伟龙，倪翠，张广渊，巩慧
DOI：10.13700/j.bh.1001-5965.2022.0376
收稿日期：2022-05-18
网络首发日期：2022-10-10
引用格式：王朋，郝伟龙，倪翠，张广渊，巩慧．视觉 SLAM 方法综述[J/OL]．北京航空航天大学学报. https://doi.org/10.13700/j.bh.1001-5965.2022.0376

视觉 SLAM 方法综述

王朋[1,2]，郝伟龙[1,✉]，倪翠[1]，张广渊[1]，巩慧[1]

（1. 山东交通学院 信息科学与电气工程学院，济南 250357；2. 山东省科学院 自动化研究所，济南 250013）

*通信作者 E-mail： 2174698790@qq.com

摘要 SLAM（simultaneous localization and mapping）技术通过搭载特定传感器，使移动机器人在没有任何环境先验条件下，通过在运动过程中自主建立环境模型来计算自身位姿，能够大幅提高移动机器人的自主导航能力，以及对不同应用环境的适应性，并有助于后续动态路径规划、实时避障与多机器人协作的实现。视觉 SLAM 是指以相机作为外部传感器，通过采集周围环境信息来创建地图并实时估计机器人自身位姿。论文对具有代表性的经典视觉 SLAM 方法以及与深度学习相结合的视觉 SLAM 方法进行介绍，并针对视觉 SLAM 中所采用的不同特征检测方法、后端优化、闭环检测，以及在动态环境下视觉 SLAM 的应用进行了论述与分析，就视觉 SLAM 在未来的热点研究方向和发展前景展开探讨，最后对文中所讨论的视觉 SLAM 问题进行了总结。

关键词 视觉 SLAM；深度学习；特征检测；位姿估计；闭环检测

中图分类号 TP391.4 **文献标识码**：A DOI: 10.13700/j.bh.1001-5965.2022.0376

An overview of visual SLAM methods

WANG Peng[1,2], HAO Weilong[1,✉], NI Cui[1], ZHANG Guangyuan[1], GONG Hui[1]

（1. School of Information Science and Electrical Engineering, Shandong Jiao Tong University, Jinan 250357, China；

2. Institute of Automation, Shandong Academy of Sciences, Jinan 250013, China）

*Tel.: 15963663085 E-mail: 2174698790@qq.com

Abstract SLAM(Simultaneous localization and Mapping) enables mobile robots to calculate their position and pose by independently building an environment model during movement without any environmental prior conditions by carrying specific sensors. It can greatly improve the autonomous navigation ability of mobile robots and their adaptability to different application environments, and contribute to the subsequent implementation of dynamic path planning, real-time obstacle avoidance and multi-robot collaboration. Visual SLAM refers to using the camera as an external sensor to collect ambient information to create a map and estimate the robot's own position in real time. Paper introduces the typical classical visual SLAM methods and the visual SLAM methods combined with deep learning, and discusses and analyzes the different feature detection methods, back-end optimization, loop closure detection, and the application of visual SLAM in dynamic environment. This paper discusses the hot research direction and development prospect of visual SLAM in the future, and finally summarizes the problems of visual SLAM discussed in this paper.

Key words visual SLAM; deep learning; feature detection; pose estimation; loop closure detection

当移动机器人被放置在一个未探测过的环境中时，首先要做的是要知道自己所在的环境地图以及自己在该环境地图中的位置，该问题于 1986 年提出并命名为 SLAM 问题，由于它具有非常重要的理

收稿日期：2022-05-18

基金项目：中国博士后科学基金（基金号 2021M702030）；山东省交通运输厅科技计划项目（基金号 2021B120）

Fund：China Postdoctoral Science Foundation(2021M702030); Science and Technology Project of Shandong Provincial Department of Transportation (2021B120)

网络首发时间：2022-10-10 10:03:48 网络首发地址：https://kns.cnki.net/kcms/detail/11.2625.V.20221009.1040.009.html

论价值和很高的应用价值，SLAM 迅速成为了热点研究方向，被认为是实现全自主移动机器人的关键。早期的 SLAM 主要是激光 SLAM，拥有精度高，解决方案较为成熟的优点，但是激光 SLAM 主要适用于室内环境，且存在体积大、价格昂贵和感知信息少的缺点，而视觉 SLAM 使用相机作为主传感器，用拍摄的视频流作为输入，实现实时定位与建图，能够同时适用于室内和室外环境，且成本较低[1]。随着计算机视觉的快速发展，视觉 SLAM 以其信息获取及时、信息量丰富等优势逐渐成为研究热点。自 21 世纪以来，不断有学者提出新的视觉 SLAM 方法，扩充视觉 SLAM 领域中的相关技术，为该领域提供了更加广阔的研究思路。

1 经典视觉 SLAM 方法

视觉 SLAM 工作流程主要可以分为五大部分：信息采集、前端视觉里程计、后端优化、闭环检测、建图，如图 1 所示。前端视觉里程计通过提取、匹配图像数据中的特征点从而得到路标点并估计出相机当前的位姿，实现实时定位。这期间存在的累积误差由后端优化部分根据帧间的共视关系等方法进行消除，提高跟踪与建图的精度。在位姿估计期间如果有两帧图像的相似度达到一定标准则认为形成了闭环，对两帧图像之间的所有路标点和位姿进行一次优化，经过优化处理过的路标点最终用于建图，建图的精度将直接影响后续动态路径规划、实时避障等工作的效果。

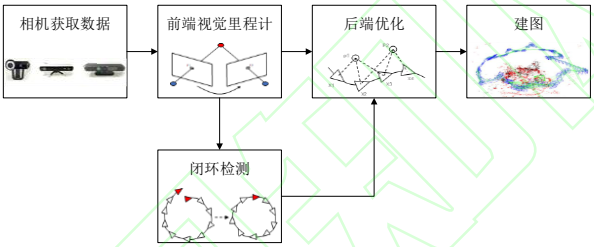

图 1 视觉 SLAM 流程图
Fig.1 Flow chart of Visual SLAM

经典的视觉 SLAM 常用的传感器主要有单目、双目和 RGB-D 相机。单目视觉 SLAM 基于单目相机完成，这种相机结构简单，成本低，相机与物体进行相对运动时，与相机距离不同的物体在图像序列中产生的位移也不同，据此物体在相机视野中形成了视差，依据视差就能判断相机距离不同物体的远近关系，但是难以获得相机与物体的真实距离。为了解决无法得到图像深度的问题，人们提出了双目视觉 SLAM 和 RGB-D SLAM。双目视觉 SLAM 由两个单目相机构成，两个相机之间的距离称为基线，基线是已知的，通过对两个相机同一场景的拍摄情况进行对比计算能够得到相机到每一个像素之间的距离，双目视觉 SLAM 通过两相机形成的视差计算得到，能够同时适应室内和室外环境，受环境因素的影响较小，但是其测量到的深度范围大小与基线的大小呈正相关，因此通常情况下传感器的体积较大，而且计算视差的过程较为复杂，导致实时性较低，因此降低计算量是双目视觉 SLAM 的重点问题之一。RGB-D SLAM 使用激光传感器，利用红外结构光原理或飞行时间法（Time of Flight，TOF），向物体投射激光并接收返回的激光来测量相机到物体的距离，由于该距离是直接由激光传感器物理测量得到，因此不需要像双目视觉 SLAM 一样消耗大量计算资源，但是存在成本高、测量范围小、噪声大、易受光照强度等外界环境影响的问题。

1.1 MonoSLAM

2007 年 Andrew Davison 提出的 MonoSLAM[2]通过在每帧图像中提取稀疏的特征点来进行位姿估计，能够在概率框架内在线运行，这是第一个使用单目相机完成的实时视觉 SLAM 系统，是基于扩展卡尔曼滤波器（extended Kalman filter, EKF）方法的视觉 SLAM，有着良好的实时性。实验表明，在小规模运动中，MonoSLAM 往往能将误差限制在 1~2 cm 之间。MonoSLAM 的诞生标志着视觉 SLAM 的研究由理论转向了实际，具有里程碑的意义。但是 MonoSLAM 仅用了单个线程，特征点的提取和匹配、相机位姿的估计和建图工作都是逐帧进行的，而且更新的计算复杂度很高，为了提高

SLAM 速度，每帧图像中只能处理其中的大约 10 个特征点，且特征点非常容易丢失。

1.2 PTAM

2007 年 Georg Klein 和 David Murray 提出了 PTAM[3]（parallel tracking and mapping），为了解决 MonoSLAM 计算复杂度高的问题，PTAM 对视觉里程计部分做出了优化，不再逐帧提取特征点，而是在一部分图像序列中选出最具有代表性的一帧作为关键帧，只处理关键帧中的特征点，降低了计算的复杂度，首次提出了 Tracking、Mapping 的前后端双线程架构，如图 2 所示。Tracking 线程使用 FAST[4]（features from accelerated segment test）作为特征提取的方法，只需要更新相机位姿，不受建图线程计算成本的影响，因此易于实现实时计算，为了提高计算速度，PTAM 还在该线程中使用粗细两轮求解来估计相机位姿，第一轮粗求解只从图像金字塔中的第一层图像提取少量特征点，将估计出的相机位姿作为初始值传入第二轮的细求解中，第二轮细求解会在整个图像金字塔中匹配特征点，有效的保证了特征点的尺度不变性，提高了跟踪精度。Mapping 线程使用 BA（bundle adjustment）进行优化，与滤波法相比，BA 是一种基于非线性优化的算法，这种算法在单位时间内可处理的特征点数量远大于滤波法，因此相比于滤波法有着更高的精度。

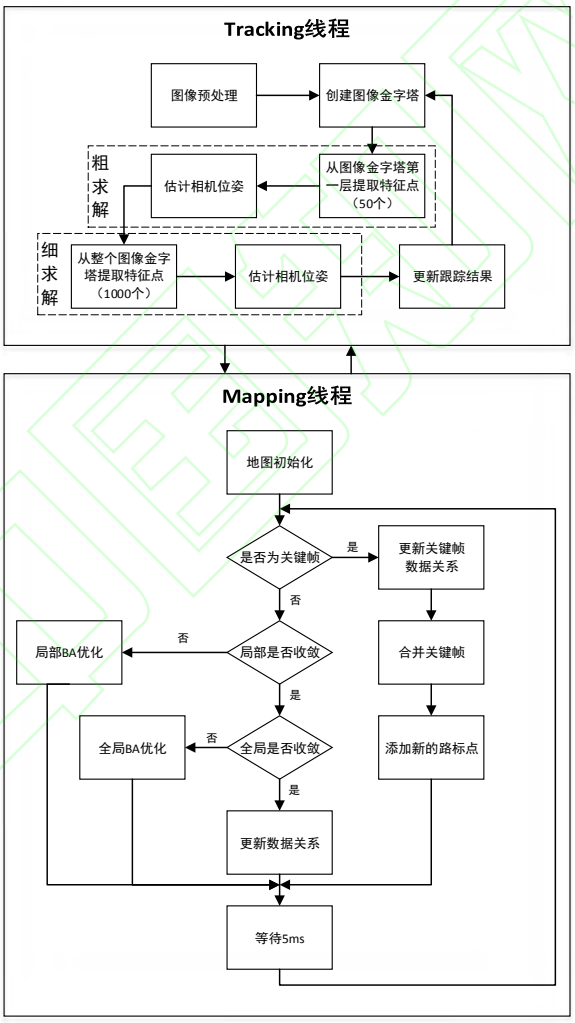

图 2 PTAM 框架

Fig.2 PTAM framework

PTAM 是第一个使用非线性优化的方法，也是第一个将 BA 与视觉 SLAM 相结合的方法。在 PTAM 中，首次出现了前后端的概念，这也为后来出现的视觉 SLAM 方法提供了良好思路。但是 PTAM 缺少闭环检测模块，无法对相机位姿和地图进行更好地优化。

1.3 DTAM

2011 年 Newcombe 等人提出了一个单目视觉 SLAM 系统 DTAM[5]（dense tracking and mapping），该方法属于直接稠密法，其依然延续了处理关键帧的思想，但是传统的视觉 SLAM 方法在关键帧上提取特征点需要消耗大量计算资源，这有利于提高系统的鲁棒性但不利于系统的实时性，而 DTAM 直接对关键帧中的每一个像素的信息进行分析计算，通过最小化全局空间规范能量函数来计算关键帧实时重建三维模型，根据图像灰度信息、最小化光度误差（Photometric error）来优化相机运动。DTAM 不需要单独对特征点和描述子进行详细的计算，既减少了特征点的计算时间，也避免了弱纹理场景中特征点缺失导致的跟踪失败问题。DTAM 有着稳定的定位与建图效果，但是其工作都是在默认光照环境不变的条件下进行的，因此对光照强度变化比较敏感。

1.4 LSD-SLAM

2014 年 Jakob Engel 等人提出了 LSD-SLAM[6]（large-scale direct monocular SLAM），是一个半稠密的单目视觉 SLAM，后来又将这种方法成功的移植到了双目和 RGB-D 相机上，它不使用特征点，而是将关注点放到了像素梯度与直接法的关系上，提取梯度较为明显的像素，通常情况下是物体的边缘线条，使用方差归一化后的光度误差（variance-normalized photometric error）对图像的梯度进行跟踪生成半稠密深度图。LSD-SLAM 通过对当前帧和当前关键帧进行匹配得到当前帧的位姿，然后判断当前帧与当前关键帧之间的距离是否大于某一阈值来决定是否将当前帧设置为当前关键帧，如果当前帧被设置为了新的关键帧，则将上一个关键帧的深度图投影到当前帧来为新关键帧初始化深度图，最后通过闭环检测进行位姿和地图的优化。实验表明，在稳定的运动环境下，LSD-SLAM 能将相机轨迹的绝对误差限制在 2 cm 内。但目前来说半稠密法获得的点云数量仍然不足，还存在着尺度漂移，对相机曝光参数非常敏感，相机快速移动时容易丢失目标等问题，且没有基于直接法的闭环检测，依然需要以特征点法的方式进行闭环检测。

1.5 SVO

2014 年 Forster 等人提出了 SVO[7]（semi-direct monocular visual odometry），它结合了特征点法和直接法的优点，是稀疏直接法视觉里程计，并使用提出的深度滤波器估计特征点的位置。首先在图像中找到一些稀疏的特征点，然后将特征点周围一定范围内的像素用直接法去进行特征点匹配，估计相机的运动轨迹及位姿，由于只寻找了稀疏的特征点，且不需要对特征描述子进行计算，因此 SVO 的速度很快，能够达到 100 Hz，后续 2016 年的续作 SVO2.0 能够达到 400 Hz，因此它适用于计算平台受限的情况，如无人机和低端计算平台，但是 SVO 也存在着明显的局限性，由于该方法的目标应用平台是应用于无人机的俯视相机，相机的主要运动方式是水平平移，且没有考虑相机视角旋转的情况，因此在其它常见的应用场景中表现不佳，由于缺少了后端优化和闭环检测功能，对相机位姿的估计具有较为明显的累积误差，非常依赖于位姿估计的准确性，一旦跟踪失败很难重新定位，因此 SVO 的建图功能非常有限，不算一个完整的 SLAM。

1.6 ORB-SLAM

2015 年 Raúl Mur-Artal 等人提出了基于单目相机的视觉 SLAM 系统 ORB-SLAM[8]，2017 年又提出了 ORB-SLAM2[9]，在上一代的基础上增加了对双目和 RGB-D 相机的支持，具有广泛的适用性，在 2020 年又提出了 ORB-SLAM3[10]，实现了相机与惯性测量单元（inertial measurement unit，IMU）的结合，通过外加传感器提升了视觉 SLAM 的性能。ORB-SLAM 在先前传统的 PTAM 框架下添加了闭环检测模块，使用了 Tracking、Local Mapping、Loop Closing 三线程并行的方案。Tracking 线程负责提取图像中的 ORB[11]（oriented FAST and rotated BRIEF）特征点，然后与上一关键帧进行比较，计算特征点的位置来估计相机当前的位姿，并适时的根据一定的规则选取新的关键帧，每一个 ORB 特征点都有可能对应一个路标点，而同一个路标点也可能会对应多个帧中不同的 ORB 特征点，关键帧和路标点组成了完整的地图。Local Mapping 线程负责接收 Tracking 线程选出的关键帧并进行处理，将冗余的局部关键帧和质量较差的路标点删除，并通过三角化生成新的路标点，然后进行 BA 优化。Loop Closing 线程负责接收 Local Mapping 线程筛选过后的关键帧，与其他所有关键帧进行闭环检测，然后更新相机位姿和地图。ORB-SLAM 围绕 ORB 特征进行计算，大大提高了计算的效率，并

且拥有良好的尺度不变性和旋转不变性，ORB 特征点中拥有描述子，在相机进行大范围运动时也能较好的完成闭环检测和重定位。不同于之前的单目视觉 SLAM 方法，ORB-SLAM 在单目条件下能够自动选择合适的初始化方案，而在双目和 RGB-D 条件下初始化速度很快，且原地不动就能完成初始化。闭环检测大大提高了定位和建图的准确性。在关键帧的选取方面先采用较为宽松的选取方案，提高了跟踪定位过程的鲁棒性，然后再删除效果不佳的关键帧，以提高闭环检测的精度和 BA 优化的效率。在静态环境下定位准确且系统鲁棒性强，能够达到实时效果。经过多组实验证明，ORB-SLAM 平均的关键帧轨迹均方根误差为 1.57 cm，且在相机快速移动的状态下依然能够保持良好的跟踪状态，鲁棒性远大于之前的 PTAM 和 LSD-SLAM，但是单目跟踪在缺少闭环条件的大运动场景中会产生较大的漂移量[12]。ORB-SLAM 对每一帧图像都进行 ORB 特征点提取非常耗时，地图由稀疏特征点组成精度较低，且对动态环境较为敏感，在动态环境下容易跟踪失败。

1.7 DSO

2016 年 Jakob Engel 提出了 DSO[13]（direct sparse odometry），是一个基于单目相机的稀疏直接法视觉里程计，无论是在鲁棒性、准确性还是在速度上都优于 LSD-SLAM 和 ORB-SLAM。不同与特征点法需要找到某一特征点与哪些帧中的点相匹配，DSO 把每一个点投影到所有帧中，计算它在各个帧中的残差，将残差在同一合理范围内的点视作同一个点的投影，而并不关心这些点之间的对应关系，通常情况下系统的运行速度很快，能够生成非常稠密的点云。后端是一个由多个关键帧组成的滑动窗口，需要对滑动窗口和与优化相关的结构进行维护。为了降低光照对视觉 SLAM 性能造成的影响，DSO 提出了光度标定，能够降低光照对跟踪的影响，有效的提高了系统的鲁棒性。但是，虽然在速度上 DSO 相比之前的方法有很大的优势，但是同 SVO 一样，DSO 没有闭环检测，跟踪失败后难以重新定位。

2 结合深度学习的视觉 SLAM 方法

深度学习通过模拟人脑神经元的结构构建复杂的神经网络模型，利用大量数据进行训练，模拟人的学习过程。在早期，大部分视觉 SLAM 都是围绕着静态环境进行工作，动态场景下视觉 SLAM 的正常使用相当困难。但是通过深度学习能够实现计算机对物体的识别，还可以完成像素级的语义分割，这对 SLAM 识别动态物体有着极大的帮助，使得视觉 SLAM 在动态场景中也能够正常高效的工作。通过深度学习，计算机甚至还能学会估计相机位姿等更为复杂的问题。因此，视觉 SLAM 与深度学习的结合也成为了近年来比较热门的研究方向。

2.1 UnDeepVO

2018 年 Li 发表的论文中提出了 UnDeepVO[14]（monocular visual odometry through unsupervised deep learning），该单目视觉里程计是基于无监督深度学习设计的。系统主体由位姿估计器和深度估计器两部分构成，位姿估计器本质上是一个基于 VGG 架构的卷积神经网络，深度估计器主要基于编码-解码来得到图像深度，相比利用视差计算深度的方法，这种方法能使系统更容易收敛。首先通过立体影像及其尺度来训练 UnDeepVO，然后利用连续的单目影像进行测试，不仅能够估计相机的位姿，而且还能得到真实尺度的深度图。通过多组实验可知，在 100~800 m 之间，UnDeepVO 的平均平移均方根误差为 4.07%，平均旋转均方根误差为 2.02°/100m，而作为对照组，ORB-SLAM 在不使用闭环检测功能时，两项数据分别为 27.05%和 10.23°/100m。

2.2 DynaSLAM

2018 年 Berta Bescos 等人提出了 DynaSLAM[15]（tracking，mapping and inpainting in dynamic scenes），该系统以 ORB-SLAM2 为基础，利用实例分割和多视图几何的方法识别动态的区域并提取出来，有效的提升了 SLAM 在动态场景下的性能。DynaSLAM 是一个 SLAM 与深度学习相结合产生的经典的动态视觉 SLAM 系统，使用 Mask R-CNN 网络进行实例分割，检测出有潜在运动可能的物体，利用多视图几何的方法检测动态的特征点，通过区域生长法进一步得到动态区域，将二者融合以后便可以检测出那些有潜在运动可能的物体[16]，在后续的 SLAM 工作中便可以剔除动态的区域，只

使用静态区域的 ORB 特征点进行跟踪与建图，最后进行背景修复，将相邻关键帧的图像投影到当前关键帧缺失的动态区域上进行修补。但是 DynaSLAM 将所有具有潜在运动可能的物体剔除，如静止的汽车、椅子等，当图像中的优质特征点过少时将影响位姿估计的效果。

2.3 DSP-SLAM

2021 年 Wang 等提出了 DSP-SLAM[17]（object oriented SLAM with deep shape priors），这是一个以 ORB-SLAM 为骨架的面向对象的 SLAM 系统，在单目、双目、RGB-D 相机下能够达到 10 Hz 的工作速度，其创新点在于将地图分为了实例对象和背景两部分，通过深度学习的语义分割将实例对象检测出来，并通过一种新的二阶优化算法，以特定类别的深度形状作为先验估计目标实例的形状和位姿，即使对新见到的物体也能够做到精确且完整的重建，背景环境通过稀疏特征点来构建，然后在联合因子图中对二者进行优化，这样不仅能保证环境地图建立的实时性，还为前景目标对象构建了一个丰富而精确的稠密 3D 模型。

3 视觉 SLAM 的主要问题

3.1 特征提取

相比激光 SLAM，虽然视觉 SLAM 有着成本低、获取信息量大、适用性广等天然优势，但是仍存在着不可忽略的问题。MonoSLAM 的提出首次解决了视觉 SLAM 的实时性问题，使视觉 SLAM 的实际应用成为了可能，但是其使用的特征点法无法保证特征点跟踪的准确性和稳定性。因此，对特征点的优化就成为了特征点法的关键问题。在视觉 SLAM 发展的早期，许多特征点提取算法相继被提出，通常是提取图像中的角点作为特征点，例如 Harris 角点、FAST 角点、GFTT[18]（good features to track）角点等。但是在真正的实际应用当中，单纯的角点可能无法满足定位与建图的要求，由于相机分辨率的限制，当相机到角点的距离发生改变时，远处的角点可能无法被提取到，或者当相机视角旋转时，同一个角点在不同视角下可能会发生很大的变化，这样就会导致两个相同的角点无法匹配，降低了 SLAM 的精度。不过随着计算机视觉领域技术的快速发展，越来越多的特征点概念被相继提出，如具有代表性的 SIFT[19]（scale variant feature transform）、SURF[20]（speeded up robust features）、ORB 等，他们往往都具有良好的性质，如通过构建图像金字塔，在每层不同尺度的金字塔图像中提取特征点来保证特征点具有尺度不变性，使用灰度质心法，连接图像块质心与图像块几何中心得到特征点的主方向，保证特征点具有旋转不变性[21,22]。除此之外相比于普通的角点，这些人工设计的特征点有着更好的可重复性、可区别性、本地性和高效性。其中，SIFT 充分考虑了相机在运动的过程中图像出现的尺度、旋转、光照等变化，提取出的特征点具有尺度不变性和旋转不变性，对光照变化也有较好的鲁棒性。但是，SIFT 的描述子维度高达 128 维，导致计算的时间复杂度高，因此使用 SIFT 进行 SLAM 工作很难保证系统的实时性。与 SIFT 相比，SURF 也具有尺度不变性和旋转不变性，在计算的时间复杂度方面也有所优化，是与 SIFT 相类似的特征点。相比于前两者，ORB 特征在拥有尺度不变性和旋转不变性的同时，结合了 FAST 计算速度快的优势，同时又改进了 FAST 不具有方向性的缺点，并且使用计算速度极快的二进制描述子 BRIEF（binary robust independent elementary features），使得整个提取图像特征的环节速度大大提高，有效降低了计算的时间复杂度。因此，ORB 特征能够在满足 SLAM 系统良好鲁棒性的前提下，同时满足实时性要求，是目前视觉 SLAM 中最常用的一种特征点。

虽然 ORB 可以满足许多视觉 SLAM 场景的使用，但是想要得到更好的表达效果依然需要进一步去优化特征点。当前处于研究热点的语义 SLAM 采用了深度学习的方法，但其本质仍然是优化提取图像特征。因此，图像特征的提取与表示依然是视觉 SLAM 领域亟待解决的重要问题之一[23]。

3.2 位姿与地图的优化

SLAM 的后端优化任务主要是解决前端视觉里程计位姿估计和建图带来的误差，防止形成较大的误差累积。前端视觉里程计中采用计算相邻两帧之间关系的方法来估计相机的位姿和建图，这种链式

计算方法随着时间的推移会产生较大的累积误差。因此，后端优化的思路往往是拓宽约束帧之间的时间和空间跨度，如在全局范围内选取一些具有代表性的关键帧，利用这些关键帧建立一个全局约束来优化之前通过相邻帧计算得到的不够准确的位姿和地图。

在相机运动的过程中，有时会重复观测到之前观测过的场景，而由于前端链式计算带来的累积误差，此时的建图结果可能和之前的建图结果并不重合，因此想要建出较为精确的地图需要添加闭环检测功能。实现闭环检测的方式有很多，如对两帧之间的特征点进行匹配，但诸如此类的方法受光照等外界环境变化的影像较大。实现闭环检测最常用的算法是词袋模型，是将图像中包含的实体抽象出来，对应于词袋模型中的“单词”。由许多单词组合成“字典”，然后通过查询字典的方式，使用单词出现的情况来描述整幅图像，这样就把一幅图像转换成了一个向量来描述，通过向量的相似度即可估计图像的相似度。无监督机器学习中的聚类问题与字典的生成问题非常相似，通过 K-means 算法可以得到有效的解决，当然，将深度学习与闭环检测联系起来也是当下的研究热点，在不同关键帧中进行语义分割，然后通过特定的算法比较当前帧和关键帧的相似度来判断是否形成闭环也是比较常用的方法。

3.3 动态环境下的视觉 SLAM

早期的视觉 SLAM 基本都是基于静态环境的，环境中动态的物体往往会对 SLAM 系统的定位与建图工作产生较大的影响，因此，想要使视觉 SLAM 能够更稳定的应用于实际生活中，提高系统在动态环境下的鲁棒性是必须要解决的重大问题。当前动态环境下的视觉 SLAM 问题的主流解决思路是将影响位姿估计和建图的动态部分剔除，仅使用环境中剩余的静态部分进行位姿估计和建图。使用运动向量来表示图像特征点的运动状态，当相机进行水平方向的移动时，大多数特征点的运动向量应该是相同的，不同的运动向量表示的特征点就很可能是正在移动的点，但是在相机视角旋转，尤其是相机视角绕光轴旋转时，环境中属于静态区域的特征点会产生许多方向不同的运动向量，这样判断动态区域的问题就会变得非常复杂。利用 Delaunay 三角剖分法，将图像序列中的特征点相连接形成多个三角形区域，比较相邻两帧中空间位置相同的两特征点相连的几条边，如果边的长度发生了较大的变化，说明这两个特征点不属于同一个目标，将所有符合条件的边全部剔除后，计算由三角形组成的连通区域的面积，其中面积最大的部分即为静态区域。在特定的环境中还可以使用深度学习利用具有先验知识的语义信息来提取环境中主要的移动物体，如行人和车辆。当然，直接将动态区域剔除可能会导致图像中的特征点过少，这同样不利于 SLAM 工作，而增加传感器能够解决这一问题，如 IMU 能够在短时间内较为准确地测量相机自身的角速度和加速度，在有动态物体干扰的情况下也可以比较准确的估计相机的位姿。

4 视觉 SLAM 的未来发展

相比激光 SLAM 而言，视觉 SLAM 除了具有成本低、获取信息量大、适用性广等特点外，还有一个关键性的优势，视觉 SLAM 的工作方式更加接近人类认识世界的习惯，这也注定了其未来在智能机器人领域内将具有重要的地位。另一方面，虽然现在实际生产中使用更多的是激光 SLAM，但是在未来多机器人合作领域，激光 SLAM 由于存在缺乏闭环检测手段、难以消除累积误差等关键问题，很难得到进一步的快速发展与应用，而视觉 SLAM 在消除累积误差方面的技术更加完善，且依靠深度学习和多传感器融合的帮助，越来越多视觉 SLAM 本身难以攻克的难题也得到了解决，因此视觉 SLAM 在未来有更大的发展空间。

4.1 与深度学习结合

近年来，计算机视觉领域技术的飞速发展很大程度上推动了视觉 SLAM 的发展，尤其是深度学习，在图像处理方面取得了优异的成绩，在图像识别的精度上超过了以往的其他方法。深度学习能够自动的提取图像中难以观测到的有效特征信息，而视觉 SLAM 也需要提取图像中有效的特征点，因此越来越多的学者使用深度学习来完成视觉 SLAM 的工作。Kishore Konda[24]在 2015 年发表的论文是

基于深度学习的 SLAM 领域内的早期作品，将视觉里程计描述为一个分类问题，并使用卷积神经网络以图像序列作为输入，预测相机位姿和速度的离散变化，结果表明，就估计精度而言该方法还无法与主流的视觉里程计方法相比较，但是该论文也证明了将深度学习与视觉 SLAM 结合推向实际应用的可能性，为后来视觉 SLAM 与深度学习的结合提供了宝贵的思路。经过众多学者的实践证明，基于深度学习的 SLAM 比传统的视觉 SLAM 具有更好的鲁棒性。深度学习与闭环检测相结合也能够获得很好的效果[25,26]，相关实验结果表明，基于深度学习的闭环检测方法能够为系统带来更高的准确性和鲁棒性。此外，深度学习还可以与建图环节相结合来创建语义地图。语义地图可以使机器人突破对环境地图单纯的几何理解，有利于提高机器人的自主性和功能的扩展性。例如在无人驾驶领域，应用语义地图可以使汽车能够应对更加复杂的情况，提高汽车行驶时的安全性和稳定性；在家用机器人中使用语义地图能使机器人更加精确的分辨和寻找物体，甚至完成其它更加高级的工作。如果能够通过深度学习赋予环境更多的深层信息，未来还能使机器人的工作方式更加接近人类，使机器人更加智能化。

4.2 多传感器融合

相机能够捕捉环境中丰富的特征信息，但是在剧烈运动、光照条件不佳等特殊场景下，相机能捕捉到的特征信息会变少，从而导致 SLAM 难以正常工作。在目前计算机性能和算法上很难有重大突破的情况下，可以添加传感器克服类似困难场景。2019 年香港科技大学开源了视觉-惯性 SLAM 系统 VINS-Fusion[27]，使用单目、双目、IMU、GPS（global positioning system）的组合，IMU 能够在短时间内对相机的移动状态进行较为准确的估计，在无法很好提取足够特征点的情况下发挥了关键作用。同时，视觉 SLAM 中的闭环检测又能很好地解决 IMU 累积误差的问题，起到了良好的补充作用，GPS 也大大增加了整个 SLAM 系统的鲁棒性。试验结果表明，在相同条件下，多传感器融合的 SLAM 系统无论精度还是鲁棒性都要优于纯视觉 SLAM。虽然视觉 SLAM 向多传感器融合方向发展已经成为共识，但是多传感器结合会导致移动机器人的体积增大，这就需要根据具体应用情况人为的寻找一个合适的度。此外，多传感器结合还会造成信息冗余，如何将短时间内接收到的大量信息集中在一起进行合理的处理，并做出正确的决策也是一个重要的问题[28]。因此，多传感器融合是 SLAM 领域未来的研究重点。

5 结论

早期的视觉 SLAM 通过较为简单的特征点来实现定位与建图。为了提高 SLAM 系统的实时性，人们提出了关键帧的处理方式。同时，为了解决特征点难以匹配的困难，许多优秀的人造特征点也被设计出来，如 SIFT、SURF、ORB 等。为了解决累积误差问题，在后端优化的基础上又添加了闭环检测模块来优化移动机器人的位姿及地图。但是，目前仍存在若干难以彻底解决的问题，如动态场景下的 SLAM 等。随着过去几年深度学习的飞速发展，视觉 SLAM 所面对的种种困难似乎都有了解决思路，传统的功能模块中也都加入了深度学习的影子，无论是系统质量还是效率都得到了较大提升。然而，视觉 SLAM 的研究还没有到达完善的地步，还需要继续提高系统的准确性和鲁棒性来适应更多、更复杂的应用场景。未来，视觉 SLAM 将会在无人驾驶、智能机器人等领域得到快速发展与广泛应用。

参考文献（References）

[1] 权美香，朴松昊，李国．视觉 SLAM 综述[J]．智能系统学报，2016，11（6）：768-776.
Quan M X,Piao S H,Li G.An overview of visual SLAM [J].CAAI transactions on intelligent systems,2016,11(6):768-776(in Chinese).
[2] 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.
[3] 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.
[4] Rosten E.Machine learning for very high-speed corner detection[C].Proceedings of the ECCV.6.
[5] 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.
[6] Engel J,Schöps T,Cremers D.LSD-SLAM:Large-scale direct monocular SLAM[C].European conference on computer

vision.Springer,Cham,2014:834-849.
[7] 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.
[8] 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.
[9] Mur-ArtalRa úl,D Tard ósjuan.ORB-SLAM2:An Open-Source SLAM System for Monocular, Stereo,and RGB-D Cameras[J]. IEEE Transactions on Robotics,33(5):1255-1262.
[10] Campos C,Elvira R,Rodr guez 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.
[11] Rublee E,Rabaud V,Konolige K,et al.ORB:An efficient alternative to SIFT or SURF[C].2011 International conference on computer vision.Ieee, 2011:2564-2571.
[12] Şımşek B,Satir S,Bılge H Ş.Performance Comparison of Direct and Feature Based vSLAM Algorithms[C].2021 29th Signal Processing and Communications Applications Conference (SIU).IEEE, 2021:1-4.
[13] Engel J,Koltun V,Cremers D.Direct sparse odometry[J]. IEEE transactions on pattern analysis and machine intelligence,2017,40(3):611-625.
[14] Li R,Wang S,Long Z,et al.Undeepvo:Monocular visual odometry through unsupervised deep learning[C].2018 IEEE international conference on robotics and automation (ICRA).IEEE,2018:7286-7291.
[15] Bescos B,F ácil 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.
[16] Kang S,Gao Y,Li K,et al.A Visual SLAM Algorithm Based on Dynamic Feature Point Filtering[C].2021 IEEE International Conference on Robotics and Biomimetics (ROBIO).IEEE,2021:1968-1973.
[17] Wang J,R ünz M,Agapito L.DSP-SLAM: Object Oriented SLAM with Deep Shape Priors[C].2021 International Conference on 3D Vision (3DV).IEEE, 2021:1362-1371.
[18] Shi J.Good features to track[C].1994 Proceedings of IEEE conference on computer vision and pattern recognition.IEEE,1994:593-600.
[19] Lowe D G.Object recognition from local scale-invariant features[C].Proceedings of the seventh IEEE international conference on computer vision. Ieee,1999,2:1150-1157.
[20] Bay H,Tuytelaars T,Gool L V.Surf:Speeded up robust features[C].European conference on computer vision.Springer,Berlin,Heidelberg,2006:404-417.
[21] Sun C Z,Zhang B,Wang J K,et al.A Review of Visual SLAM Based on Unmanned Systems[C].2021 2nd International Conference on Artificial Intelligence and Education (ICAIE).IEEE,2021:226-234.
[22] Jian M,Wang J,Yu H,et al.Visual saliency detection by integrating spatial position prior of object with background cues[J].Expert Systems with Applications,2021,168:114219.
[23] 汤一明，刘玉菲，黄鸿．视觉单目标跟踪算法综述[J]．测控技术，2020，39（8）：21-34．
Tang Y M,Liu Y F,Huang H.Overview of visual single target tracking algorithm[J].Measurement and Control Technology,2020, 39(8):21-34(in Chinese).
[24] Konda K R,Memisevic R.Learning visual odometry with a convolutional network[C].VISAPP (1).2015: 486-490.
[25] Hou Y,Zhang H,Zhou S.Convolutional neural network-based image representation for visual loop closure detection[C].2015 IEEE international conference on information and automation. IEEE, 2015:2238-2245.
[26] Zhang X,Su Y,Zhu X.Loop closure detection for visual SLAM systems using convolutional neural network[C].2017 23rd International Conference on Automation and Computing (ICAC).IEEE,2017:1-6.
[27] Qin T,Cao S,Pan J,et al.A general optimization-based framework for global pose estimation with multiple sensors[J].arXiv preprint arXiv:1901.03642,2019.
[28] Gautam A,Mahangade S,Gupta V I,et al.An experimental comparison of visual SLAM systems[C]. 2021 International Symposium of Asian Control Association on Intelligent Robotics and Industrial Automation (IRIA).IEEE,2021:13-18.