

**人工智能课程报告**

****

**题目：** 动态环境下的视觉SLAM

**姓名：** 潘青

**学号：** 2021171242

**专业：** 电子信息 053班

**学院：** 计算机与信息学院

目录

[1研究方向及背景介绍 2](#_Toc418534147)

[2相关问题及解决方案 3](#_Toc2079204909)

[2.1传统的slam算法 3](#_Toc1528416516)

[2.2几何的方法 4](#_Toc71787682)

[2.3.结合深度学习的方法 6](#_Toc784636015)

[3研究展望 8](#_Toc479747813)

# 1研究方向及背景介绍

SLAM(Simultaneous Localization and Mapping)即同时定位与建图，也就是机器人在未知环境中根据传感器数据来估计自身位姿并建立相应地图。分为激光SLAM和视觉SLAM。而视觉SLAM与激光SLAM的区别在于视觉SLAM以摄像头作为传感器，而激光SLAM以激光雷达作为传感器。我的研究方向为视觉SLAM。图1-1为SLAM的经典框架。

视觉里程计接收到相机的图像数据，通过两帧图像的位置变换来估计相机的运动，仪器的测量就免不了受到噪声的影响，后端优化主要就是指处理过程中的噪声问题，除了解决从图像中恢复出相机运动的问题，我们还要关心计算出的结果受到多大噪声的影响，进而尽可能地消除误差。回环检测判断机器人是否回到了先前经过的位置，如果检测到回环，它会把信息传递给后端进行优化处理。回环是一个比后端更加紧凑、准确的约束，这一约束条件可以形成一个拓扑一致的轨迹地图。如果能够检测到闭环，并对其优化，就可以让结果更加准确。建图模块根据估计的轨迹和传感器数据，建立与任务要求相对应的地图，是对环境的描述，而地图的保存或者展现方式则根据SLAM的具体应用而定。

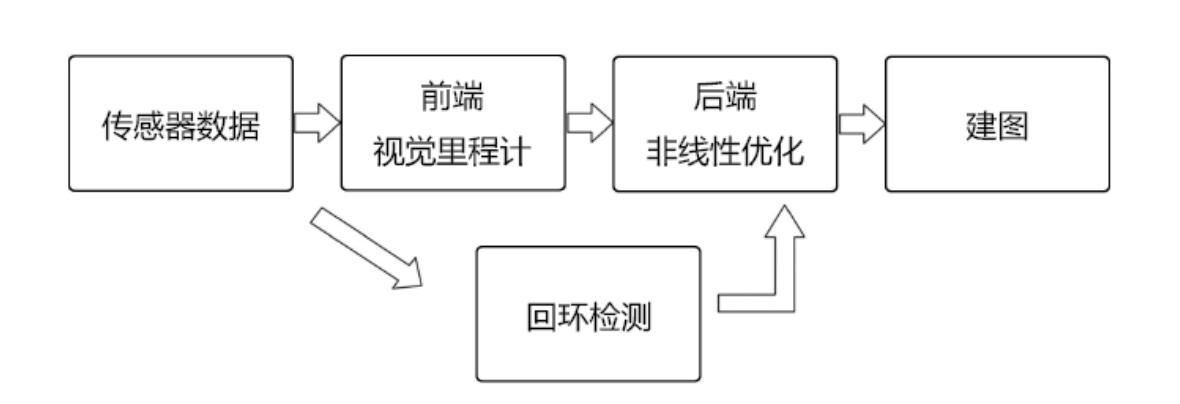


图1-1 SLAM框架

在视觉SLAM中，传感器数据可以来自三种不同的相机： 单目、双目和RGBD相机。单目相机在每一时刻只能获取单张图像，无法直接获得深度信息，需要两帧图像通过三角化等计算来获取深度信息，尺度不确定性成了单目SLAM的特点；双目视觉SLAM利用几何约束的原理匹配左右两幅图像的特征，从而能够在当前帧速率的条件下直接提取完整的特征数据，因此应用比较广泛，但是成本较高，且视角范围受到一定限制，缺乏灵活性； RGBD相机可以在获得RGB图像的同时获取深度图像，但是有效探测距离太短，应用环境有限。

# 2相关问题及解决方案

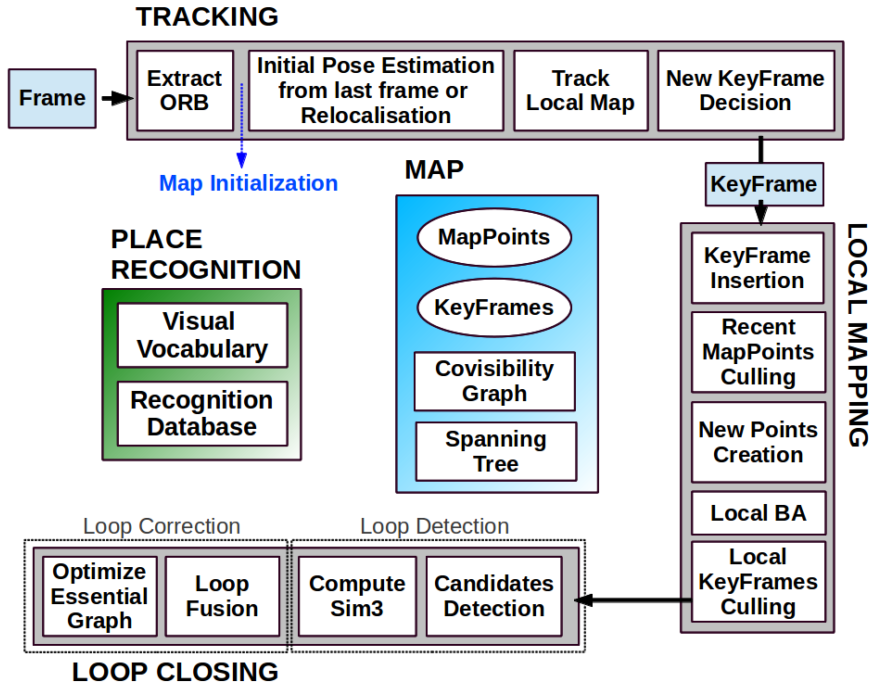
我目前主要针对的是视觉SLAM在实际的应用场景中。因为在真实的环境中，存在许多动态物体，比如人、车辆和动物。这些动态物体在定位时会导致错误的位姿估计，同时在建立的地图中会把这些动态物体包含在内。所以如何去除动态物体，使位姿估计更加准确，保证算法的鲁棒性，同时保证构建的地图的准确性和可复用性成了目前需要解决的问题。

解决方案

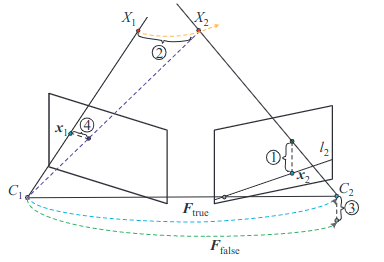
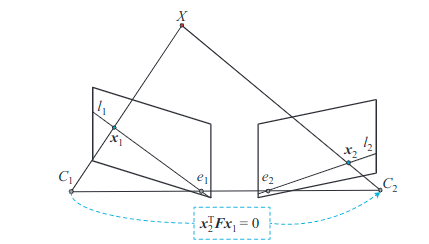
## 2.1传统的slam算法

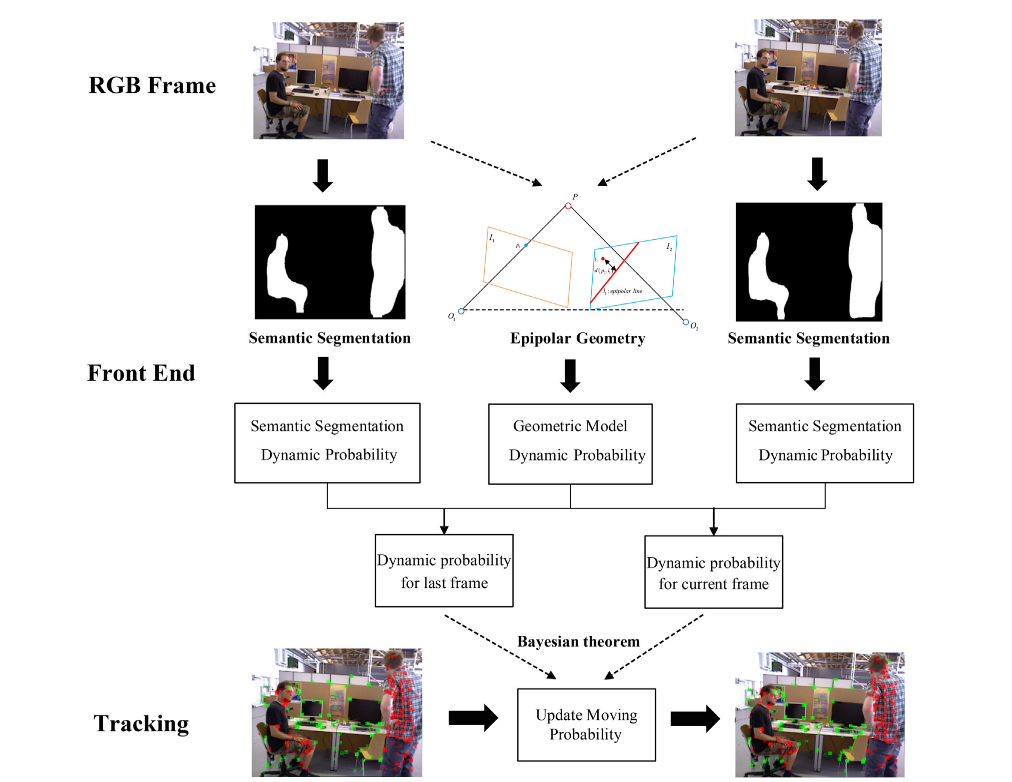
MonoSLAM[1]首个实时单目视觉SLAM 。PTAM[2]，是最早提出将Track和Map分开作为两个线程的一种SLAM算法，是一种基于关键帧的单目视觉SLAM算法

 LSD-SLAM[3]系统是2014年提出的基于直接法的大范围单目SLAM方法，它提出了一种新颖的基于相似变换空间对应的李代数上的直接跟踪法，能够明确的检测到尺度漂移，同时采用概率方法对图像跟踪过程中的噪声进行处理。单目直接法SLAM方案，可以生成半稠密地图。ORB-SLAM2是一个SLAM的经典框架,如图2-1所示。 ORB-SLAM[4]是一个基于特征点的实时单目SLAM系统，在大规模的、小规模的、室内室外的环境都可以运行。该系统对剧烈运动也很鲁棒，支持宽基线的闭环检测和重定位，包括全自动初始化。该系统包含了所有SLAM系统共有的模块：跟踪（Tracking）、建图（Mapping）、重定位（Relocalization）、闭环检测（Loop closing）。由于ORB-SLAM系统是基于特征点的SLAM系统，故其能够实时计算出相机的轨线，并生成场景的稀疏三维重建结果。ORB-SLAM2在ORB-SLAM的基础上，还支持标定后的双目相机和RGB-D相机。但是对于动态环境ORB-SLAM易产生错误的位姿估计，影响定位精度。许多研究都是基于ORB-SLAM的框架来进行的。



## 2.2几何的方法

在静态场景假设的前提下，空间中的 3 维特征点才能满足多视图几何的投影关系．该假设同时也是相机帧间位姿估计和空间位姿图优化算法的基本前提．依赖几何约束技术处理动态 SLAM 问题，则是利用对极几何特性来分割静态和动态特征。动态要素将违反多视图几何中为静态场景定义的标准约束，如图 1(a) 所示，标准约束为 ，其中 、 为连续 2 帧图像的匹配点对位置，F 为基础矩阵．约束可以从极线、三角剖分、基础矩阵估计或重投影误差等式中得出，动态场景下违反几何约束的具体情况如图 1(b) 所示，其中：①点 在 帧中的投影点距离极线 太远；②由相机光心与投影点相连的反投影射线不能相交于一点；③动态特征的出现会导致基础矩阵估计错误；④点 在帧 中的重投影特征与观测特征之间的距离过大



wang等人提出了一种[5] 算法。如图所示首先对图像进行预处理，RGB图像提取特征点，进行匹配，计算基础矩阵。同时深度图进行聚类。动态物体检测和去除基于两个假设 （1）静态物体中有足够多的匹配（多于20对），（2）两帧之间视差小。经过 两次基础矩阵约束，消除不匹配的点。同时对于满足以下条件的特征点视为动态特征点。动态物体上的动态特征点占有一定比例动态物体上的动态特征点占比大于一定阈值 动态物体上剔除动态特征点后，其与特征点与图片中所有特征点比例将减小。若某个深度聚类上的特征点满足以上三条，则视其为动态区域，并剔除其上所有特征点

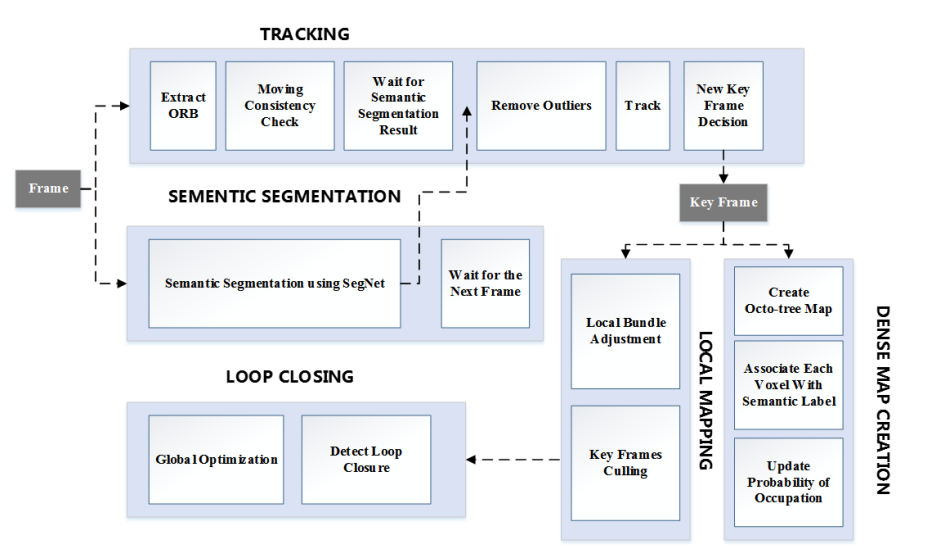
Kundu[6] 等 通过定义 2 个几何约束来检测动态特征．第 1 个约束来自对极几何，该约束指出在后续视图中的匹配点应位于相应的对极线上．如果跟踪的特征离极线太远，则很可能是动态特征．第2 个约束是 FVB（flow vector bound），其目的是检测 3D 点沿对极线移动的情况．对特征点的位移设置上下边界，检测结果中的那些位于边界外的特征点则很可能是动态特征．最后，通过贝叶斯滤波器来确定特征是否为动态特征

## 2.3.结合深度学习的方法

近年来，基于深度学习的图像语义分割和目标检测方法在效率和准确性上都获得了极大的发展．许多研究人员试图通过语义标记或目标检测预处理来去除潜在的动态物体，进而解决动态 SLAM 问题．这些方法在处理特定场景中的特定动态物体时非常有效，但是，当出现未知动态物体时，它们的鲁棒性可能会下降．

DS\_SLAM[7]在ORB-SLAM2的基础上提出了动态环境中的完整语义SLAM系统，加入一个实时语义分割网络放在一个独立的线程中，它将语义分割与运动一致性检测方法结合起来，过滤掉场景的动态部分。语义分割用来分割RGB图像中的人等动态物体，然后去除掉动态物体上的特征点，防止它们用来估计相机的运动以至于产生错误的位姿估计，运动一致性计算光流金字塔，得到当前帧中匹配的特征点。然后如果匹配对太接近图像的边缘或匹配对中心的像素差太大，匹配对将被丢弃。利用基本矩阵计算当前坐标系下的外极线。最后，确定匹配点到对应极线的距离是否小于某一阈值。如果距离大于阈值，则判定匹配点正在移动。同时 DS-SLAM创建了一个单独的线程来构建稠密的语义八叉树地图。 稠密的语义三维八叉树地图采用优势对数计分法滤除不稳定体素并更新这些体素的语义。在动态场景中，提升了定位模块和建图模块的稳定性和鲁棒性。。

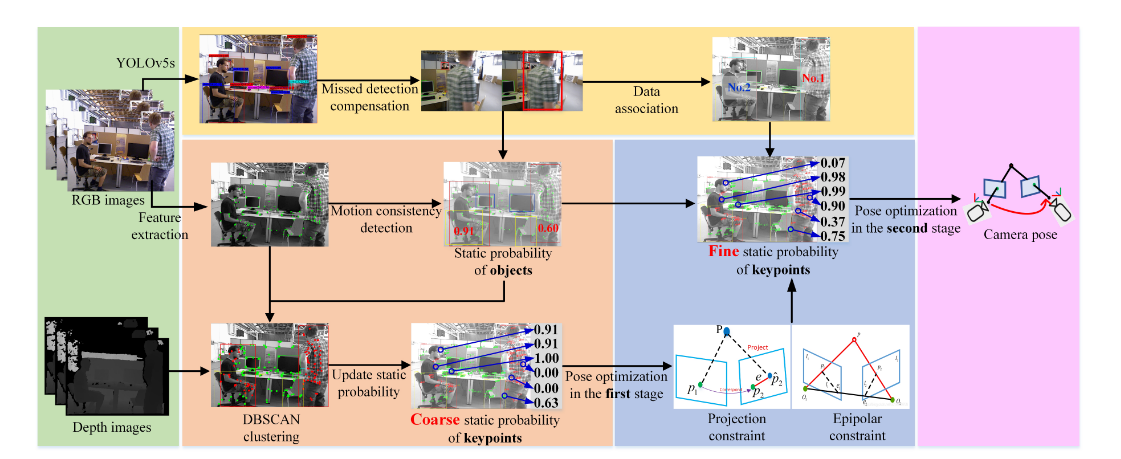
DynaSLAM[8]系统 同 样 也 是在 ORB-SLAM2 系统上进行改进，在使用单目或双目相机时，仅利用实例分割网络 Mask R-CNN 逐像素分割帧中的先验动态物体（例如人和汽车），令 SLAM 算法 不会在该物体上提取特征．而在使用 RGB-D 相机 的情况下，除了利用 Mask R-CNN 网络来分割先验 动态物体以外，还利用多视图几何法来检测潜在移 动物体，比如被人移动的椅子．利用多视图几何法 来判断关键点是否为动态点，并添加了背景修复功能，但是耗时严 重，实时性差。



但是语义分割线程有时会存在漏分割，分割错误，而且有的分割网络实时性不够。在 ORB-SLAM2 系统的基础上改进的方案还有文将目标检测方法作为附加线程引入到SLAM 系统．Detect-SLAM [10] 系统结合语义先验知识并使用目标检测网络 SSD 对关键帧进行动态物体检测，对于动态物体上的特征点，则通过特征匹配和扩展影响区域的形式进行运动概率传播，在跟踪线程中去除动态特征点． 与此同时，将 SLAM 构建的实例级语义地图投影到图像上得到候选区域，改善了目标检测算法，以便在更具挑战性的条件下，例如当存在不寻常的视点、光照条件较差和运动模糊时，有效地检测和识别物体．该方法使 SLAM 系统与目标检测网络优势互补

基于扩展卡尔曼滤波和匈牙利算法补偿漏检，用DBSCAN聚类算法区分前景点和背景点

CFP\_SLAM[9]采用yolov5目标检测和几何约束获得高动态物体和低动态物体先验信息。计算特征点的静态移动概率从粗到细。两步计算概率先用yolov5进行目标检测，对于漏检问题，采用EKF和匈牙利算法补偿漏检，关联相邻帧之间的框特征点提取后进行运动一致性检测（光流和对极约束）检测特征点静态概率；深度图聚类来区分前景点和背景点，得到粗略的特征点静态概率，再通过重投影约束和极线约束进行细化得到相机位姿。

这种方法采用目标检测网络找出候选框，速度更快，然后采用聚类的方式把动态物体分割出。

但是语义分割线程有时会存在漏分割，分割错误，而且有的分割网络实时性不够，而且分割产生的掩码会包括静态部分，导致这些点被剔除掉无法用于位姿估计，并且如果人只有一部分在运动但是分割网络处理后会把人身上的全部特征点去除，但是人身上的这部分静态特征点可以用来进行位姿估计，提高定位精度

# 3研究展望

动态环境下的SLAM有以上几种解决方法，但是具体如何操作，创新部分不相同，我目前还在学习中，目前主流方案是借助语义分割网络来处理图像中的动态物体，目前想法是在语义分割之后如何使用这些动态物体上的静态特征点用来位姿估计当前

利用深度学习来处理动态 SLAM 问题，优点主要在于目标检测、语义分割等方面的深度神经网络方法能以较高的准确率和效率来识别、分割物体．利用深度神经网络提供的物体语义信息，结合运动先验知识将场景中的物体分为动态、静态和潜在移动物体（MaskFusion 方法中将与人体接触的物体判定为移动物体）．基于深度学习的动态 SLAM方法在处理特定场景中特定动态物体时获得了非常有效的结果．然而，单一地利用深度学习和物体先验知识来识别移动物体，会受限于预先训练数据集中的物体类别结果，不属于训练类别中的运动物体不会被检测到，并会被错误地融合到背景地图中，导致定位精度低和建图效果差，这也是基于学习的方法的一个重要的实际限制．而且，此类方法只能检测先验动态物体，并不能识别任意移动物体

我目前还在学习中，目前主流方案是借助语义分割网络来处理图像中的动态物体，目前想法是在语义分割之后如何使用这些动态物体上的静态特征点用来位姿估计当前

参考文献：

[1]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.

[2 Klein G , Murray D . Parallel Tracking and Mapping for Small AR Workspaces[C]// IEEE & Acm International Symposium on Mixed & Augmented Reality. ACM, 2008.

[3] Engel J , Schps T , Cremers D . LSD-SLAM: Large-scale direct monocular SLAM[C]// European Conference on Computer Vision. Springer, Cham, 2014.

[4] 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.

[5] Wang R , Wan W , Wang Y , et al. A New RGB-D SLAM Method with Moving Object Detection for Dynamic Indoor Scenes[J]. Remote Sensing, 2019, 11(10):1143-.

[6] Kundu A , Krishna K M , Sivaswamy J . Moving Object Detection by Multi-View Geometric Techniques from a Single Camera Mounted Robot[C]// Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on. IEEE, 2009.

[7] Yu C , Liu Z , Liu X , 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.

[8]B Bescós, JM Fácil, Civera J , et al. DynSLAM: Tracking, Mapping and Inpainting in Dynamic Scenes[C]// 2018:1-1.

[9] Zhong F , Sheng W , Zhang Z , et al. Detect-SLAM: Making Object Detection and SLAM Mutually Beneficial[C]// 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2018.

[10]Xinggang , Yunzhou , Zhenzhong，et al. CFP-SLAM: A Real-time Visual SLAM Based on Coarse-to-Fine Probability in Dynamic Environments //arXiv e-printsComputer Science - Robotics 2022,month = febarXiv:2202.01938