

电子科技大学

专业学位研究生学位论文中期考评表

攻读学位级别： ☐ 博士 ☒ 硕士

培养方式： ☒ 全日制 ☐ 非全日制

专业学位类别及领域： 电子信息

学 院： 信息与软件工程学院

学 号： 202222090508

姓 名： 陈佐均

论 文 题 目： 基于 RGB-D 相机的物体级语义
SLAM 算法研究

校内指导教师： 王治国

校外指导教师： 梁熙

填 表 日 期： 2024 年 9 月 15 日

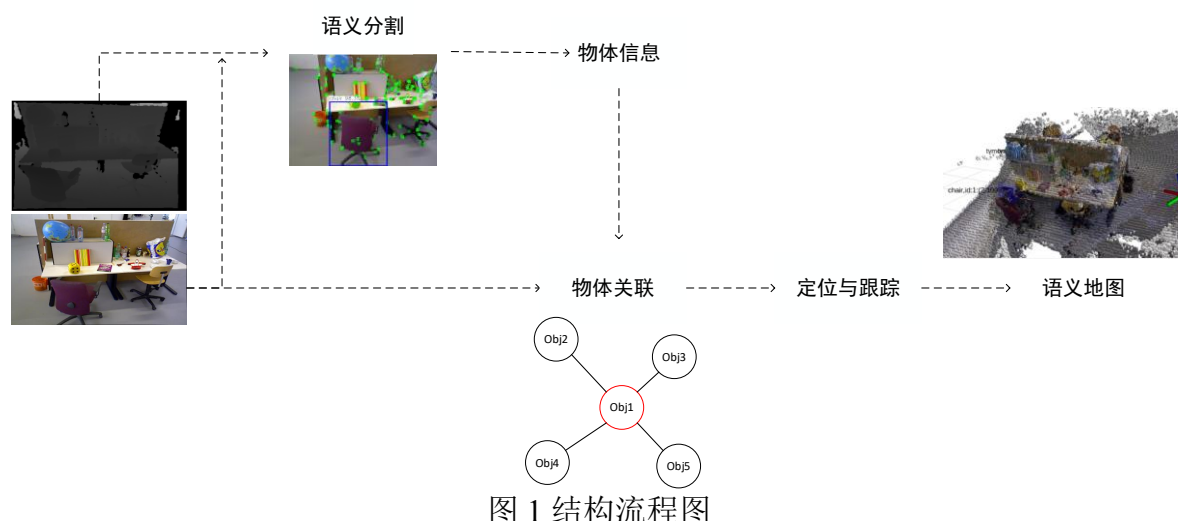
电子科技大学研究生院

一、已完成的主要工作

1. 开题报告通过时间： 2023 年 12 月 21 日	
2. 课程学习情况	
是否已达到培养方案规定的学分要求	<input type="checkbox"/> 是 <input checked="" type="checkbox"/> 否
3. 论文研究进展	
<p>从理论分析或计算部分、实验实践（或实证）工作等方面进行总结（可续页）</p> <p>1. 理论分析</p> <p>在视觉 SLAM 系统中，位姿估计与地图构建密切相关，其中最重要的一个优化步骤是 Bundle Adjustment (BA)。BA 通过最小化观测到的重投影误差来优化相机位姿和地图点的位置。然而，SLAM 系统中的 BA 通常分为三种：Motion-Only BA、Local BA 和 Full BA。这些优化方法在一定程度上提升了跟踪和建图的精度。然而，随着环境复杂度的增加，系统在处理不规则场景时的表现仍有提升空间。</p> <p>Motion-Only Bundle Adjustment (MOBA) 通常用于追踪阶段（tracking stage），即每一帧的相机位姿估计过程中。MOBA 的主要作用是通过优化相机的位姿，在保持地图点不变的前提下，最小化观测到的重投影误差，从而提高跟踪的精度和系统的鲁棒性。</p> $\{\mathbf{R}, \mathbf{t}\} = \underset{\mathbf{R}, \mathbf{t}}{\operatorname{argmin}} \sum_{i \in \mathcal{X}} \rho \left(\ \mathbf{x}_{(\cdot)}^i - \pi_{(\cdot)}(\mathbf{R}\mathbf{X}^i + \mathbf{t})\ _{\Sigma}^2 \right) \quad (1)$ <p>其中 ρ 是稳健的 Huber 损失函数，Σ 是与关键点尺度相关的协方差矩阵。$\pi_{(\cdot)}$ 是投影函数分为单目 π_m 和校正双目 π_s。</p> <p>Local Bundle Adjustment (Local BA) 主要用于局部地图优化阶段（local mapping stage）。在这个阶段，Local BA 通过同时优化当前关键帧、与其相连的共视关键帧及其关联的地图点，从而提高局部地图的精度。它的作用是减少局部区域中的累积误差，保证局部关键帧的位姿和地图点的位置更加一致和精确。</p> $\{\mathbf{X}^i, \mathbf{R}_l, \mathbf{t}_l i \in \mathcal{P}_L, l \in \mathcal{K}_L\} = \underset{\mathbf{X}^i, \mathbf{R}_l, \mathbf{t}_l}{\operatorname{argmin}} \sum_{k \in \mathcal{K}_L \cup \mathcal{K}_F} \sum_{j \in \mathcal{X}_k} \rho(E_{kj}) \quad (2)$ $E_{kj} = \ \mathbf{x}_{(\cdot)}^j - \pi_{(\cdot)}(\mathbf{R}_k \mathbf{X}^j + \mathbf{t}_k)\ _{\Sigma}^2 \quad (3)$ <p>其中 $\mathbf{R} \in \text{SO}(3)$ 表示相机的方向 和 $\mathbf{t} \in \mathbb{R}^3$ 是相机位置，最小化世界坐标中匹配的 3D 点 $\mathbf{X}^i \in \mathbb{R}^3$ 和关键点 $\mathbf{x}_{(\cdot)}^i$ 之间的重投影误差，无论是单目 $\mathbf{x}_m^i \in \mathbb{R}^2$ 还是双目 $\mathbf{x}_s^i \in \mathbb{R}^3$ 都是在关键帧中匹配上的点。</p> <p>随着时间的推移和轨迹的扩展，系统在构建地图时会不可避免地会累积一些小的误差。虽然局部 BA 可以对局部地图进行优化，但它无法纠正整个地图的全局误差。当系统检测到回环时，意味着机器人或相机已经回到了之前的某个位置，这时可以利用这个约束对全局地图进行一次完整的优化，从而调整关键帧和地图点的位置，使地图更准确、更一致。Full Bundle Adjustment (Full BA) 是一种全局优化方法，主要用于系统检测到回环（loop closure）之后的阶段。其目的是对整个地图中的所有关键帧和地图点进行联合优化，以最小化全局范围内的重投影误差。</p> <p>传统的 SLAM 中 BA 主要依赖于关键帧和地图点的优化，而忽略了场景中更高层次的结构信息。通过在优化过程中结合物体级别的语义信息，SLAM 系统可以更好地理解场景中的物体与相机之间的关系，改善地图构建的准确性。引入物体的几何和语义信息，可以在不同尺度上为系统提供更多约束，尤其是在纹理缺乏或重复场景中，物体的存在可以作为稳定的特征参考，增强系统的鲁棒性。此外，物体的识别和跟踪能够帮助系统识别有意义的场景结构，改进地图的语义表达，从而提高系统在复杂环境下的表现。这种物体感知不仅有助于提高相机位姿估计的精度，还能够为后续的任务（如导航或对象交互）提供更加可靠的环境</p>	

模型。

本课题设计的 SLAM 系统中引入了物体的概念。在 SLAM 系统中加入物体，并且将使用物体信息对 SLAM 系统的 BA 优化、回环检测等添加更多的约束信息。通过关联好的物体对 SLAM 系统的精度和鲁棒性进行一定的提升，并且构建一个物体语义地图，图 1 是具体的结构图：

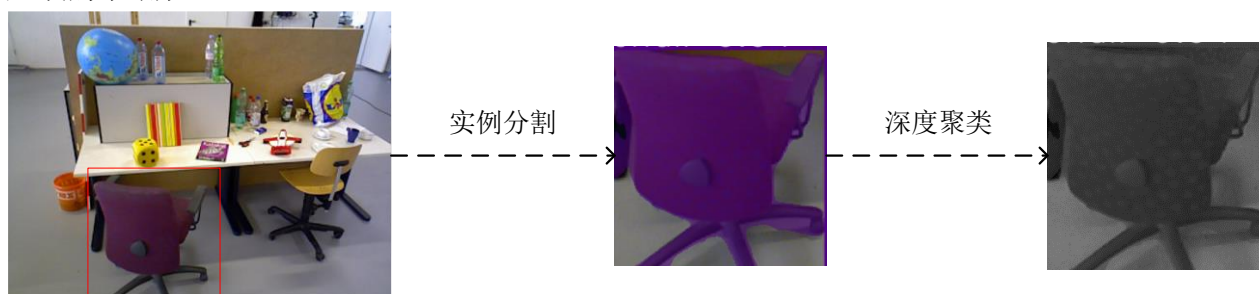


2.研究进展（一）

在本课题中，使用 ROS（Robot Operating System）作为进程间通信的框架，实现 SLAM 与 YOLOv8 实例分割网络的协同工作。在 SLAM 线程中，通过 ROS 节点发布需要处理的图像作为话题，并在图像到达时将对应的标志设置为 true，从而通知其他线程开始处理图像数据。与此同时，在一个独立的线程中运行 YOLOv8 实例分割网络，对发布的话题图像进行分析，检测预训练好的物体类别，并提取相应的 mask。

YOLOv8 在实时场景下具有很高的检测效率，能快速识别和分割图像中的目标对象。然而，单纯依赖 YOLOv8 生成的 mask 在物体边界上可能存在一定的不精确，尤其是在物体边缘复杂或背景干扰较大的情况下，物体与背景的分割可能不够清晰。这种现象在机器人实际应用中会影响 SLAM 的准确性，进而对机器人的环境感知和导航产生负面影响。

为了解决这一问题，本课题对 YOLOv8 生成的 mask 进行了进一步处理。具体来说，本课题使用聚类算法对 mask 的边界区域进行细化处理。由于物体与背景之间通常存在一定的深度差异，可以通过对深度信息进行聚类分析来优化物体边界。深度聚类不仅可以帮助识别物体的真实轮廓，还能有效区分复杂场景中物体与背景的过渡区域，从而提升 mask 在边缘区域的准确性。



通过这种方法，可以提升 SLAM 系统对环境中的物体的感知精度。该处理流程通过将 YOLOv8 的实例分割与深度聚类相结合，实现了对物体边界的更精确描绘，从而提高对观测到的物体的表现。

3.研究进展（二）

在本课题中，构建了一个专门的物体数据库，用于存储过程中相机观测到的已识别物体。该数据库不仅用于保存物体的相关信息，还用于处理物体的关联问题，即如何正确地将新观测到的物体与数据库中的物体进行匹配和关联。本课题将物体关联问题分为两类：同一物体的关联和不同物体之间的关联。

首先，当相机观测到一个物体时，系统会将其与数据库中已有的物体进行匹配。如果在数据库中找到了与当前观测相符的物体，认为这是同一个物体的重复观测。为了确保关联的准确性，系统会根据已有的观测数据，应用一定的策略来进一步优化同一物体的关联过程。这包括了物体的点云信息如点云数量、点云的 3D 包围框等。

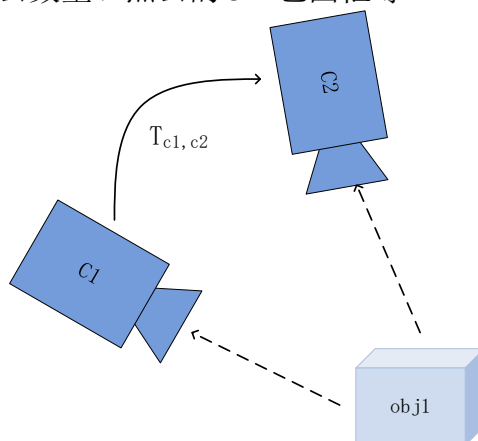


图 3 两次观测到同一个物体示意

如图 3 所示，当相机运动到两个位置 C_1 和 C_2 时观测到同一个物体 $Obj1$ ，需要在这个 $Obj1$ 物体与数据库匹配后，使用当前的 $Obj1$ 的点云数据与匹配的物体点云进行融合更新。

如果观测到的物体在数据库中未能匹配到任何已有物体，就将其视为一个新物体。在这种情况下，该物体将被添加到数据库中，并赋予唯一的标识。同时，系统还会对新物体与数据库中其他物体的关系进行更新，这意味着不仅要新物体纳入数据库，还需要评估其与已有物体之间的潜在关联，例如基于物体的空间位置、类别或功能进行进一步的分析和分类。

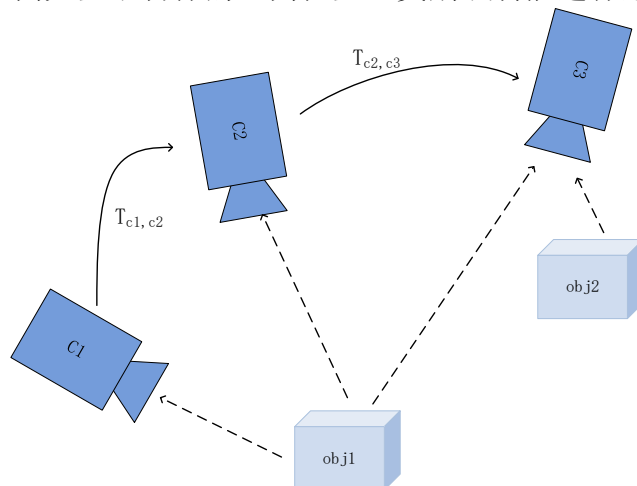


图 4 观测到新物体示意图

如图 4 所示，当相机运动到 C_3 位置时，观测到了新的物体 $Obj2$ ，这时 $Obj2$ 在物体数据库中检测不到对应的物体，则新增一个物体信息。

图 5 是一个物体和其他邻近物体的关联关系图。对于一个物体需要维护其和另外 k 个（这里设置的 4）最近物体的关联，使用的是物体之间的欧氏距离进行计算，维护每个物体最近的 5 个物体。

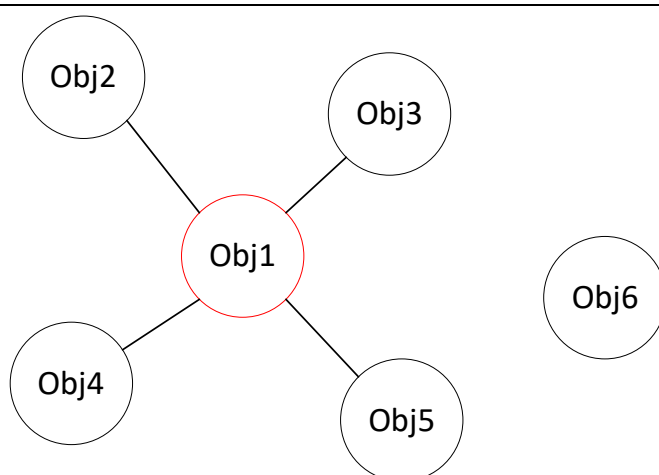


图 5 不同物体的关联

最后对于构建的物体数据库最后会输出一个语义地图，其中包含了识别到的物体信息。对于其中某一帧图片成为关键帧之后就会将其中的物体进行关联判断，其中一帧图片如图 6 所示。

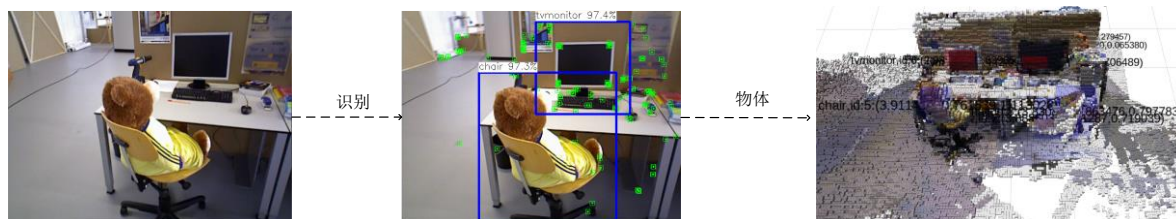


图 6 物体构建示意图

图 6 所示，这一帧在 YOLOv8 线程中检测到了有椅子和电视两个类别的物体，会在数据库中更新对应的物体点云或者是新增物体点云，并且在最后输出的物体点云地图上，对其使用 3D 的包围框、物体 label 以及物体点云中心位置等信息进行区分标识。

4. 阶段性研究成果

按《研究生学位论文撰写格式规范》的格式要求分类填写与学位论文相关的阶段性研究成果，例如期刊论文、会议论文、科研获奖、专利、制定标准等，限填第一作者或导师为第一作者时的第二作者成果，其中已录用、已投稿或拟投稿的在括号内注明（可续页）

二、存在的主要问题和解决办法

1. 未按开题计划完成的研究工作，研究工作存在的原理性、技术性难题以及在实验条件等方面的限制（可续页）

1. 物体 mask 的准确获取问题

使用聚类的方式对 YOLOv8 分割得到的 mask 进行更进一步的分割可能对于不同的物体，结果上不一定有很好的提升，而且在物体复杂或者边缘很不规则时仅使用深度信息可能会导致过度分割的结果，得到错误的结果。

2. 物体如何加入后端优化问题

准确分割表示的物体对于 SLAM 系统也是一个可用的信息，可以用于估计相机位姿中的优化项，也能够 SLAM 系统中对于回环检测过程添加新的约束项，不同物体之间的关联关系能有效的对于一个场景中的物体进行检测，判断场景中的相似性。但是如何加入到后端的优化当中，并且能发挥出相应的作用，还需要设计一个可用的优化策略。

3. 物体点云的准确性的优化问题

同一个物体在不同角度的观测中如何被认定为是同一物体，需要设计一个相应的策略，让物体的判断更加准确。

2. 针对上述问题采取何种解决办法，对学位论文的研究内容及所采取的理论方法、技术路线和实施方案的进一步调整，以及下一步的研究研究计划（可续页）

针对上述的问题，后续的解决方案路线：

1. 对于物体的 mask 边缘分割需要使用一些其他的信息，比如颜色信息，或者更换处理 mask 边缘的策略，比如尝试使用 canny 边缘检测策略、全链接条件随机场对物体 mask 的边缘进行进一步的分割。

2. 物体位姿加入优化项的最小二乘中，作为一个可优化的项。并且认为相同的场景下，获取到的物体他们对于其他物体的关联关系是一致的，所以可以通过物体的关联关系匹配上相同的场景，检测出回环情况。

3. 融合同一物体的不同观测角度时，需要考虑到物体点云的位置，以及大小，这需要用到物体大小的先验知识，并且对于判断物体位置的差异阈值需要在实验当中选定优化。

下一步研究计划：

起止年月	完成内容
2024. 09-2024. 11	完成物体级语义 SLAM 算法，完成专利 1 篇。
2024. 11-2025. 01	完成 SLAM 算法的优化改进以及语义 SLAM 系统开发。
2025. 01-2025. 02	撰写并完成硕士学位论文初稿。
2025. 02-2025. 05	完成硕士学位论文修改提升。
2025. 05-2025. 06	完成硕士学位论文答辩。

三、中期考评审查意见

1.导师对工作进展及研究计划的意见：
进展符合预期，研究计划合理可行。

校内导师（组）签字：  2024 年 9 月 20 日

校外导师签字：  2024 年 9 月 20 日

2. 中期考评专家组意见

考评日期	2024.09.20	考评地点	国际创新中心 B 栋一楼会议室 B105 腾讯会议:479-687-638
考评专家	杨远望、庄杰、李耶		
考评成绩	合格____票 基本合格____票 不合格____票		
结 论	<div><input type="checkbox"/>通过 <input type="checkbox"/>原则通过 <input type="checkbox"/>不通过</div> <div>通 过：表决票均为合格</div> <div>原则通过：表决票中有 1 票为基本合格或不合格，其余为合格和基本合格</div> <div>不 通 过：表决票中有 2 票及以上为不合格</div>		

对学位论文工作进展以及下一步研究计划的建议，是否适合继续攻读学位：
研究工作进展正常，计划可行，适合继续攻读学位。

专家组签名：

年 月 日

3.学院意见：

负责人签名：

年 月 日