arXiv:2303.01155v2 [cs.RO] 7 Apr 2023

# 利用分层表示的基于标记的视觉同步定位与建图

Ali Tourani[1], Hriday Bavle[1], Jose Luis Sanchez-Lopez[1],Rafael Mu˜noz Salinas[2],and Holger Voos[1]

**摘要——基准标记能够编码丰富的环境信息，并有助于视觉同步定位与建图（VSLAM）方法重建具有实用语义信息的地图。当前基于标记的视觉同步定位与建图方法主要利用标记来改善低特征环境中的特征检测，和/或融入闭环约束，但生成的仅是低层级环境几何地图，在复杂环境中容易出现误差。为弥补这一不足，本文提出一种视觉同步定位与建图方法，结合单目相机与基准标记，在提升相机位姿估计的同时生成环境的分层表示。该方法可检测周围环境中的语义实体，包括标记中编码的墙壁、走廊和房间，并在这些实体之间适当地添加拓扑约束。在机器人采集的真实世界数据集上进行的实验结果表明，通过引入新的约束并构建更完善的地图表示，所提出的方法在精度方面优于传统的基于标记的视觉同步定位与建图基线方法。此外，与采用激光雷达同步定位与建图方法重建的地图相比，该方法重建的地图质量也令人满意。**

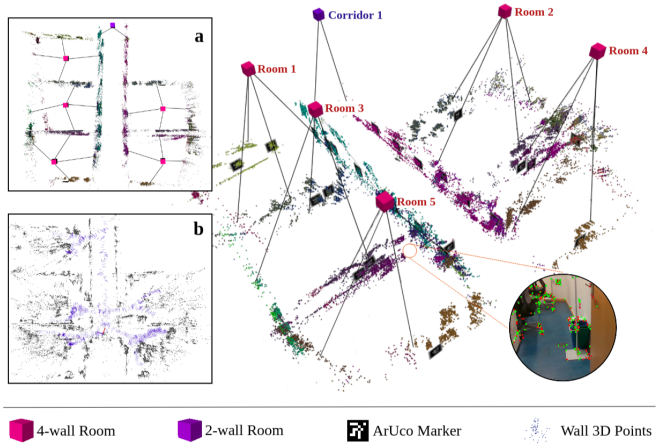



图1：采用所提出的方法，以层次化表示构建的环境最终重建地图。因此，主要地图实体是从环境中提取的已检测三维点和已访问的 ArUco 标记。三维点的不同颜色表示不同的墙面，连接这些点与立方体的线条则表示它们所属的房间，包括走廊或四面墙房间：a）重建地图的二维俯视图；b）用于地图重建的关键点和机器人轨迹记录。

## I. 引言

视觉同步定位与建图（VSLAM）系统可采用多种视觉传感器，例如单目、双目、红绿蓝深度（RGB-D）、全景及事件相机，在估计环境地图的同时定位摄像机 [1]。视觉传感器的主要优势在于，只需低成本硬件，便能从周围环境中获取丰富的视觉和语义信息，用于执行各种任务 [2]。语义数据是指从环境中获取的高层信息，它能增强 VSLAM 任务的鲁棒性，并拓展可使用重建地图的应用范围 [1], [3]。例如，机器人可能需要识别场景中的物体用于路径规划和动态物体移除，这些问题可以通过语义数据 [4]加以解决。在这方面，利用基准标记是将语义信息编码到环境中的一种可行方法，因为它们具有丰富且定义明确的特征 [5]。此外，凭借独特的图案，它们可以提供精准的位姿估计，从而辅助视觉同步定位与建图框架；在低纹理环境中提供可靠的特征，并依据标记标识符实现闭环检测。虽然大多数基准标记肉眼可见，因此会影响环境的视觉观感，但我们最近提出了新型不可见基准标记 [6] ，它们有望无缝融入环境。不过，为求简便，本文采用可见基准标记，并将不可见基准标记的使用留待未来研究。近期研究（如 [7]和 [8] ）提出了采用基准标记的视觉同步定位与地图构建方法，但未赋予这些标记有意义的语义信息，而是生成纯几何地图表示；在复杂环境中，这会导致相机位姿估计和所生成环境地图出现偏差。

为了充分发挥基准标记的潜力，准确识别语义元素及其拓扑

[1]作者来自卢森堡大学安全、可靠性与信任跨学科中心（SnT）的自动化与机器人研究组，地址为卢森堡。Holger Voos 也隶属于卢森堡大学科学、技术与医学学院，地址为卢森堡。{ali.tourani, hriday.bavle, joseluis.sanchezlopez, holger.voos}@uni.lu [2]作者来自西班牙科尔多瓦大学拉巴纳莱斯校区计算机科学与数值分析系。rmsalinas@uco.es本研究由卢森堡大学高级研究院（IAS）的 TRANSCEND 项目、欧盟委员会“地平线2020”研究与创新计划（资助协议编号 101017258，SESAME）以及卢森堡国家研究基金（FNR）的 5G-SKY 项目（编号 C19/IS/13713801）资助。为实现开放获取，作者已将知识共享署名（CC BY）公共版权许可应用于因本次投稿而产生的作者接受稿版本。作者谨向 Deniz Is¸ınsu Avs¸ar 和 Jan P.F. Lagerwall，他们均来自卢森堡大学物理与材料科学系（DPHYMS），感谢他们为将不可见基准标记用作未来场景所付出的努力并提出宝贵意见。

关系，同时生成准确且有意义的地图，本文提出了一种面向单目相机的视觉同步定位与建图框架，利用基准标记中编码的数据增强地图重建，并在最终地图中加入抽象语义元素及其拓扑关系。该系统基于 UcoSLAM [7]，构建，采用 ArUco [9] 标记和从自然路标中提取的视觉关键点，通过关键点和标记重建环境的几何地图。此外，受*S-Graphs* [10], [11],利用高层次层次化表示进行定位与建图的启发，本文提出的方法根据附着于环境中的 ArUco 标记所编码的数据，以墙壁、走廊和房间的形式添加丰富信息。图1展示了该方法重建的示例地图及其层次化表示。

本文的主要贡献总结如下：

- 将基于标记的视觉同步定位与建图扩展至重建环境地图，并利用标记编码墙面、走廊和房间等高级语义特征，
- 设计了新颖的几何约束，即标记到墙壁 和墙壁到房间，以提升地图质量并减少相机定位误差，
- 使用真实世界的室内数据集验证所提出的方法，结果表明其性能优于基于标记的视觉同步定位与建图基线方法。

本文其余部分安排如下：第二节探讨视觉同步定位与建图领域的相关研究，重点关注利用语义数据改进重建地图。第三节详细介绍所提出的方法及其各个模块。第四节给出评估标准，并讨论所提出方法的有效性。最后，本文在第五节作出总结。

## II. 相关工作

过去几年中，视觉同步定位与建图方法日趋成熟，许多研究人员通过提出有效方案来应对该领域尚未解决的挑战，为这一领域的发展作出了贡献。本文作者还针对多种最新技术水平的视觉同步定位与建图研究进行了全面综述，并探讨了它们的权衡取舍 [12]。因此，许多近期方法都着重改进视觉同步定位与建图系统的前端线程，以实现更好的位姿估计和视觉里程计 [13]。 [7]中的作者提出了 UcoSLAM，该方法利用环境中的自然地标和人工地标来重建地图。它支持仅关键点模式、仅标记模式和混合模式；当相机视野中检测到 ArUco标记时，还能改进全局地图构建器和回环检测器模块。Pfrommer 和 Daniilidis 提出了另一种基于标记的方法 TagSLAM [8]，该方法利用 AprilTag [14] 基准标记执行同步定位与建图任务。尽管这些方法在许多场景中表现良好，但它们无法从环境中获取语义信息来改进地图表示。Gomez-Ojeda等人 [15]提出了 PL-SLAM，该方法使用定向 FAST 与旋转 BRIEF（ORB）和线段检测器（LSD）算法，从双目视觉相机中提取点和线的视觉特征。然而，该方法在特征跟踪以及从结构线中提取数据方面计算开销较大。Bruno 和 Colombini [16] 提出了 LIFT-SLAM，该方法融合了基于深度学习的特征描述子与传统几何特征描述子。尽管其特征检测和跟踪性能稳健，但该方法使用了未经优化的卷积神经网络（CNN），因此只能实现近实时执行。

利用语义数据是近年来备受关注的另一个研究方向。Gonzalez 等人 [17]提出了 $S^3$LAM，这是一种基于 ORB-SLAM 2.0[18]的单目方法，利用通用物体和结构的语义分割。该方法针对每个先验平面簇，使用三角测量得到的三维点拟合平面，并采用名为*Detectron2* [19]的基于卷积神经网络的全景分割框架。同样，YOLO-SLAM [20] 也是一种基于卷积神经网络的方法，它将几何约束与 Darknet19-YOLOv3 目标检测器结合，以生成语义信息。然而，这些方法在识别小型或不规则物体和平面时，性能可能会下降。QuadricSLAM [21] 是另一种方法，它采用二次曲面表示作为目标描述，对环境中的物体进行建模，并利用目标检测边界框估计二次曲面参数。这种单目视觉同步定位与建图方法能够表示物体的大小、位置和朝向，以进行地图重建。QuadricSLAM 的主要挑战在于初始化阶段，因为这需要从多个观察角度观测物体。Hosseinzadeh等人 [22] 提出了另一种单目视觉同步定位与建图系统，该系统将基于深度学习的目标检测器检测到的已访问物体，以及使用平面地标检测器识别出的场景主导结构，添加到地图中。尽管该方法性能稳健，但估计物体的先验形状会带来较高的计算复杂度。另一种名为 Blitz-SLAM [23] 的语义视觉同步定位与建图方法，结合物体的原始掩码和深度信息，并消除动态物体形成的噪声块。文献 [24] 提出了名为 SOLO-SLAM 的并行框架，该框架利用地图点的语义属性和几何约束来过滤动态物体，并重建改进后的地图。其主要挑战是在低动态场景中会丢失大量信息。MISD-SLAM [25] 是另一种语义方法，它配备了基于预训练卷积神经网络的实时实例分割和动态特征移除功能。然而，该算法会移除仅在部分帧中静止的物体。

与上述研究不同，本文提出的方法利用标记中编码的语义数据创建环境地图，而不使用任何目标检测器。因此，根据关键点和标记数据进行数学建模，从而估计墙壁和房间约束，对系统性能的影响并不显著。此外，

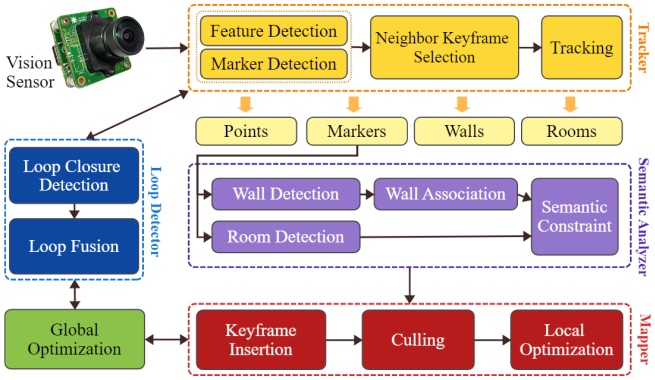



图2：所提方法的整体流程。该方法的主要贡献是增加语义分析器模块，并修改跟踪器的输出，为其提供适当的输入。此外，还修改了全局/局部优化器模块，以实现更好的地图重建。

最终环境重建地图的精度与基于激光雷达（LiDAR）的方法相当，展现了所提方法的潜力。

## III. 所提方法

所提方法的流程如图2所示。单目相机采集的帧会被送入系统进行处理。跟踪器检测 ORB视觉特征和 ArUco标记，作为系统的主要输入。系统会选择关键帧并持续跟踪，输出包括三维点、检测到的标记、墙壁，以及场景中遍历过的房间。语义分析器模块利用跟踪阶段检测到的标记位姿，执行墙面和房间检测任务。此外，建图器模块负责向地图中添加新的顶点和边，并改进最终地图，以实现优化的重建。最后，基于标记的闭环检测器模块会检查当前帧中的标记是否曾被观测到，并触发全局优化器来改善重建地图。

为了开发一种能够重建出更丰富地图的视觉同步定位与建图框架，本文提出的方法以 UcoSLAM 为基线方法，并对其组件进行修改，使其具备语义数据分析流程。上述修改使系统能够利用 ArUco 标记检测环境中的墙壁和不同类型的房间这两种语义概念。此外，由于这些基准标记用于相机位姿估计，因此还可以存储额外编码的语义数据；凭借其独特的识别特性，它们成为理想的替代方案。该方法还旨在利用视觉数据，通过一种与 *S-Graphs* [26]相似的方法，在单个可优化图中表示环境和机器人的位姿。因此，本文提出的方法不采用 *S-Graphs* 和*S-Graphs+* [11]所引入的里程计读数及利用激光雷达扫描提取平面表面的方式，而是计算墙面的位置以及利用附着在墙面上的基准标记来确定房间的变化。

### *A.* 概述

所提出方法在时刻 $t$ 的处理流程涉及四个主要坐标系：里程计参考坐标系 $O$、相机坐标系 $C_t$、标记坐标系 $M_t$和全局坐标系 $G_t$。作为系统的主要传感器，单目相机采集一组帧 $\mathbf{F} = \{\mathbf{f}\}$，其中帧为 $\mathbf{f} = \{t, \mathbf{T}\ \delta\}$；$\in SE(3)$ 是通过将 $C_t$ 变换到 $G_t$而得到的相机位姿，$\delta$ 为相机内参集合。使用 ORB 关键点检测器和特征提取器处理每一帧相机图像后，可提取一组关键点 $\mathbf{g} = \{l, u, \mathbf{d}\}$，其中 $l$ 为图像的下采样层级，$u$ 为相对于第一层上采样后的像素坐标，$\mathbf{d}$ $(d_1 ... d_n)|d_i \in [0, 1]$ 为长度为 $n$的描述子向量。因此，最终构建的环境地图 $\mathbf{E}$表示如下：

$$\mathbf{E} = \{\mathbf{K}, \mathbf{P}, \mathbf{M}, \mathbf{W}, \mathbf{R}\} \tag{1}$$

其中，$\mathbf{K} = \{k\} \subset \mathbf{F}$ 是关键帧集合，$\mathbf{P} = \{\mathbf{p}\}$ 表示从环境中提取的特征点 $\mathbf{p} = \{\mathbf{x}, \mathbf{v}, \hat{}\mathbf{d}\}$ 集合，以及它们对应的三维位置 $\mathbf{x} \in \mathbb{R}^3$、视线方向 $\mathbf{v} \in \mathbb{R}^3$和描述子 $\hat{}\mathbf{d}$。此外，$\mathbf{M} = \{\mathbf{m}\}$ 是环境中检测到的 ArUco标记集合，其中每个标记 $\mathbf{m} = \{s, \mathbf{p}, \mathbf{cc}\}$包含标记尺寸（即边长）$s \in \mathbb{R}$、标记位姿 $\mathbf{p} \in SE(3)$（根据 $M_t$ 到 $G_t$计算得出），以及角点坐标 $\mathbf{cc} = (c_1 ... c_4)|c_i \in \mathbb{R}^3$ 值。从环境中检测到的墙壁集合由 $\mathbf{W} = \{\mathbf{w}\}$表示，其中每面墙 $\mathbf{w} = \{\mathbf{q}, \mathbf{m_w}\}$ 包含墙面方程 $\mathbf{q} \in \mathbb{R}^4$ 和附属标记列表 $\mathbf{m_w} \subset \mathbf{M} = (m_1 ... m_n)|m_i \in \mathbb{N}$，其中 $m_i$ 表示 ArUco 标记ID。同样，$\mathbf{R} = \{\mathbf{r}\}$ 表示环境中发现的房间集合，其中每个房间 $\mathbf{r} = \{\mathbf{r_c}, \mathbf{r_w}\}$ 包含房间中心点 $\mathbf{r_c} \in \mathbb{R}^3$ 以及构成该房间的墙壁列表 $\mathbf{r_w} \subset \mathbf{W} = (w_1 ... w_n)|w_i \in \mathbb{N}$。

### *B.* 语义实体

所提出的方法添加了两种语义实体，以构建更丰富的地图，包括墙壁 和 房间。由于检测上述实体的方式各不相同，本节将详细讨论相关流程。

**墙壁。** 通过检测到的 ArUco 标记提供的位姿信息，主要将标记及其周边特征所在的平面识别为平面形实体，从而确定墙壁。该方法假设所有基准标记都直接放置在墙壁上，而非散热器或抽屉柜等其他物体上。因此，墙壁的方程根据所附标记的位姿确定。每当近期关键帧中访问到某个标记时，便会在场景中执行墙壁检测流程，从而更高效地在特定时刻检测并建图墙壁。

为将墙面添加到最终地图中，会在全局坐标系中提取每个墙面 $\boldsymbol{w}_i$ ${}^G\boldsymbol{w}_i = \begin{bmatrix} {}^G\boldsymbol{n}_i & {}^G d \end{bmatrix}$

其法向量为 ${}^{G}\boldsymbol{n}_i = \begin{bmatrix} n_x & n_y & n_z \end{bmatrix}^T$。墙面的顶点节点在图中表示为 ${}^{G}\boldsymbol{w_i} = [{}^{G}\phi, {}^{G}\theta, {}^{G}d]$，其中 ${}^{M}\phi$ 和 ${}^{M}\theta$ 分别表示墙面在 $G_t$ 中的方位角和仰角。对于附着在墙面上的每个标记 ${}^{G}\boldsymbol{m}_i$，可将代价函数 ${}^{G}\boldsymbol{w}_i$ 定义为：

$$c_{w_i}({}^{G}\boldsymbol{w}_i, {}^{G}\boldsymbol{m}_i) = \|[{}^{M}\delta\phi_{w_{i_{m_i}}}, {}^{M}\delta\theta_{w_{i_{m_i}}}, {}^{M}d_{w_i}]^T\|^2_{\boldsymbol{\Lambda}_{\tilde{\boldsymbol{w}}_i}} \tag{2}$$

其中，${}^{M}\delta\phi_{w_{i_{m_i}}}$ 墙面的方位角与 $w_i$ 其标记转换至标记坐标系后的方位角 $m_i$ 之差为$M_i$，${}^{M}\delta\theta_{w_{i_{m_i}}}$ 仰角之差，而${}^{M}d_{w_i}$则是墙面与标记之间的垂直距离；对于给定的标记-墙面配对，该距离应为零。

**房间**。本文考虑的另一种语义实体是房间。由于房间的感知可能会受到各种布局和结构的影响而变得困难，所提出的方法利用附着在房间墙壁上的 ArUco标记中编码的数据来检测这一语义实体。为此，提供了一个字典，其中包含环境中的房间以及附着在其墙壁上的基准标记，以供该框架使用。请注意，字典中唯一编码的信息是与房间对应的标记*ID*，无需编码任何额外的位姿信息。所提出的方法对*S-Graphs+* [11]中介绍的房间分割流程进行了修改，其中标记发挥着至关重要的作用。因此，本文考虑了两种房间类型，即“双墙房间”和“四墙房间”：

双墙房间（走廊）：在此情况下，房间中仅有两面平行的墙壁标记有基准标记。该场景适用于检测场景中墙壁不可检测或无法到达的走廊或房间。因此，房间 ${}^{G}\mathbf{r}_x = [{}^{G}\mathbf{w}^{x_{a_1}}, {}^{G}\mathbf{w}_{x_{b_1}}]$ 包含与 $x$-轴平行的 $x$-墙面平面，而 ${}^{G}\mathbf{r}_y = [{}^{G}\mathbf{w}^{y_{a_1}}, {}^{G}\mathbf{w}_{y_{b_1}}]$ 包含与 $y$- 轴平行的 $y$- 墙面平面。

为了计算双墙房间 ${}^{G}\mathbf{r}_{x_i}$ 的中心，使用两面 $x$-墙面平面方程以及标记$\mathbf{m}_i$ 的中心点${}^{G}\mathbf{c}_i$，如下所示：

$$\begin{aligned} {}^{G}\mathbf{k}_{x_i} &= \tfrac{1}{2}\left[|{}^{G}d_{x_{a_1}}| \cdot {}^{G}\mathbf{n}_{x_{a_1}} - |{}^{G}d_{x_{b_1}}| \cdot {}^{G}\mathbf{n}_{x_{b_1}}\right] + |{}^{G}d_{x_{b_1}}| \cdot {}^{G}\mathbf{n}_{x_{b_1}} \\ {}^{G}\boldsymbol{\eta}_{x_i} &= {}^{G}\hat{\mathbf{k}}_{x_i} + \left[{}^{G}\mathbf{c}_i - [\ {}^{G}\mathbf{c}_i \cdot {}^{G}\hat{\mathbf{k}}_{x_i}]\ \cdot {}^{G}\hat{\mathbf{k}}_{x_i}\right] \end{aligned} \tag{3}$$

其中，${}^{G}\boldsymbol{\eta}_{x_i}$是双墙房间 ${}^{G}\mathbf{r}_{x_i}$的中心点，${}^{G}\hat{\ }\mathbf{k}_{x_i}$由 ${}^{G}\hat{\ }\mathbf{k}_{x_i} = {}^{G}\mathbf{k}_{x_i}/\|{}^{G}\mathbf{k}_{x_i}\|$获得。标记的中心点 ${}^{G}\mathbf{c}_i$ 通过帧 $G$ 中的标记位姿获得。需要注意的是，沿 $y$ 方向的双墙房间中心也可以通过类似方式计算。

双墙房间节点以房间中心点进行初始化，最小化双墙房间节点及其对应墙面平面的代价函数定义如下：

$$\begin{aligned} &c_{\boldsymbol{r}_{x_i}}({}^{G}\boldsymbol{r}_{x_i}, [{}^{G}\mathbf{w}_{x_{a_1}}, {}^{G}\mathbf{w}_{x_{b_1}}, {}^{G}\mathbf{c}_i]) \\ &= \sum_{t=1,i=1}^{T,K} \|{}^{G}\hat{\boldsymbol{\eta}}_{x_i} - f({}^{G}\tilde{\mathbf{w}}_{x_{a_1}}, {}^{G}\tilde{\mathbf{w}}_{x_{b_1}}, {}^{G}\mathbf{c}_i)\|^2_{\boldsymbol{\Lambda}_{\tilde{\boldsymbol{r}}_{i,t}}} \end{aligned} \tag{4}$$

其中，$f({}^{G}\tilde{\ }\mathbf{w}^{x_{a_1}}, {}^{G}\tilde{\ }\mathbf{w}^{x_{b_1}}, {}^{G}\mathbf{c}_i)$ 使用式 3 将墙面平面映射到房间的中心点。

四面墙房间：此场景指由基准标记标记的四面墙房间。在这种情况下，四面墙房间由四个墙面平面组成，即 ${}^{G}\mathbf{r}_i = [{}^{G}\mathbf{w}_{x_{a_1}}\ {}^{G}\mathbf{w}_{x_{b_1}}\ {}^{G}\mathbf{w}_{y_{a_1}}\ {}^{G}\mathbf{w}_{y_{b_1}}]$。此类房间的中心点可通过以下公式计算：

$$\begin{aligned} {}^{G}\mathbf{q}_{x_i} &= \tfrac{1}{2}\left[|{}^{G}d_{x_{a_1}}| \cdot {}^{G}\mathbf{n}_{x_{a_1}} - |{}^{G}d_{x_{b_1}}| \cdot {}^{G}\mathbf{n}_{x_{b_1}}\right] + |{}^{G}d_{x_{b_1}}| \cdot {}^{G}\mathbf{n}_{x_{b_1}} \\ {}^{G}\mathbf{q}_{y_i} &= \tfrac{1}{2}\left[|{}^{G}d_{y_{a_1}}| \cdot {}^{G}\mathbf{n}_{y_{a_1}} - |{}^{G}d_{y_{b_1}}| \cdot {}^{G}\mathbf{n}_{y_{b_1}}\right] + |{}^{G}d_{y_{b_1}}| \cdot {}^{G}\mathbf{n}_{y_{b_1}} \\ {}^{G}\boldsymbol{\rho}_i &= {}^{G}\boldsymbol{q}_{x_i} + {}^{G}\boldsymbol{q}_{y_i} \end{aligned} \tag{5}$$

其中，${}^{G}\boldsymbol{\rho}_i$ 是四面墙房间的中心点 ${}^{G}\mathbf{r}_i$。还应注意，当 $|d_{x_1}| > |d_{x_2}|$ 时，式 5 仍然成立。用于最小化四面墙房间节点及其对应墙面平面集合的代价函数与双墙房间类似，但有细微差别：

$$\begin{aligned} &c_{\boldsymbol{\rho}}({}^{G}\boldsymbol{\rho}, [{}^{G}\mathbf{w}_{x_{a_i}}, {}^{G}\mathbf{w}_{x_{b_i}}, {}^{G}\mathbf{w}_{y_{a_i}}, {}^{G}\mathbf{w}_{y_{b_i}}]) \\ &= \sum_{t=1,i=1}^{T,S} \|{}^{G}\hat{\boldsymbol{\rho}}_i - f({}^{G}\tilde{\mathbf{w}}_{x_{a_i}}, {}^{G}\tilde{\mathbf{w}}_{x_{b_i}}, {}^{G}\tilde{\mathbf{w}}_{y_{a_i}}, {}^{G}\tilde{\mathbf{w}}_{y_{b_i}})\|^2_{\boldsymbol{\Lambda}_{\tilde{\boldsymbol{\rho}}_{i,t}}} \end{aligned} \tag{6}$$

其中，$f({}^{G}\tilde{\ }\mathbf{w}^{x_{a_i}}, {}^{G}\tilde{\ }\mathbf{w}^{x_{b_i}}, {}^{G}\tilde{\ }\mathbf{w}^{y_{a_i}}, {}^{G}\tilde{\ }\mathbf{w}^{y_{b_i}})$ 使用式 5 将四个估计的墙面平面映射到四面墙房间的中心点。上述代价函数维持了构成房间的四面墙之间的结构一致性。

### *C.* 最终图

图3展示了所提出方法生成的最终语义图结构。因此，系统提取的关键帧是信息的主要来源，其中包含视觉特征点及其对应的三维坐标，以及场景中访问过的 ArUco标记。需要注意的是，关键帧、特征点和基准标记之间定义的约束用于计算里程计和检测闭环。墙面-标记约束通过公式2引入。图的最高层级保留了利用标记 *ID* 在环境中检测到的房间，以及承载这些标记的墙壁；对于双墙房间，采用根据公式4获得的约束，对于四面墙房间，则采用根据公式6获得的约束。

## IV. 评估

本节将从准确性和性能方面介绍所提出方法的评估流程，并将其与最先进的方法进行比较。为此，使用所提出的

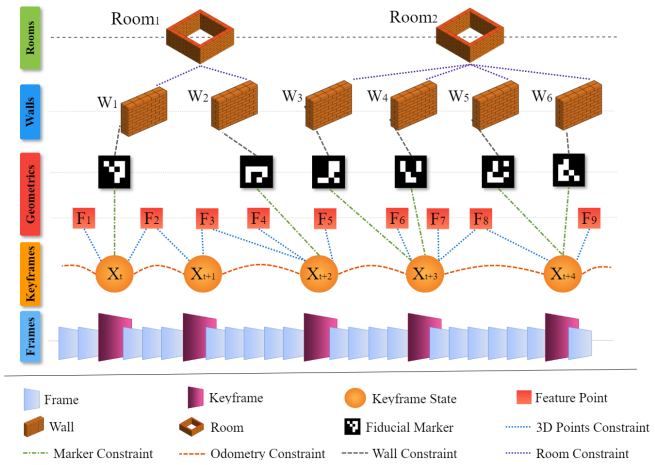



图3：所提出方法的分层架构图表示。在此方面，地图采用了新的语义约束，包括墙壁和房间，以及几何约束，以实现更好的地图重建。




图4：用于评估该方法的采集数据集：a）用于采集数据的足式机器人；b）数据集中的一些实例。

方法、作为基线方法的 UcoSLAM [7]，以及*S-Graph+* [11]（一种基于激光雷达的方法）进行了各种真实场景测试，以提供真实值测量结果。与单目视觉传感器相比，激光雷达传感器的噪声更小，因此可以放心地将其视为真实值。

### *A.* 评估设置

为评估所提出方法在真实环境中的性能，我们将*Intel® RealSenseTM* 深度相机 *D435*作为单目传感器安装在*Boston DynamicsSpot®* 机器人上，并从室内环境中采集数据。机器人在两栋不同大学建筑的不同办公区域内运行，这些区域包含各种走廊和房间布局，墙壁上贴有印刷的 $17cm \times 17cm$ ArUco 标记。环境中所放置 ArUco 标记的标记*ID* 以及房间的唯一标签也存储在数据库文件中，并输入系统。图4展示了用于采集数据的机器人及采集数据集中的一些实例。此外，数据集的特征见表I。

评估使用一台配备第*11*代 *Intel® CoreTM i9 @2.60GHz*的计算机进行。

表I：所采集室内数据集的特征。

| 序列* | 时长 | #标记 | 描述 |
|---|---|---|---|
| *Seq-01* | 08分53秒 | 35 | 两条走廊通过一个没有机器人旋转的平台连接 |
| *Seq-02* | 09分29秒 | 35 | 两条走廊通过一个平台及机器人旋转 |
| *Seq-03* | 16分05秒 | 26 | 一条走廊和五个房间 |
| *Seq-04* | 10分25秒 | 14 | 一条走廊和两扇门 |
| *Seq-05* | 09分37秒 | 20 | room 连接至主走廊的四条过道 |
| *Seq-06* | 23分34秒 | 22 | 走廊 三条走廊和三个房间 |

*数据以*rosbag*文件包的形式存储。

处理器和32吉字节内存。此外，为了更直观地展示环境中检测到的物体，包括三维点、ArUco标记、墙壁和房间，并更准确地评估所提出的方法，以便与其他最先进的方法进行比较，研究采用了一套机器人操作系统（ROS）工具。

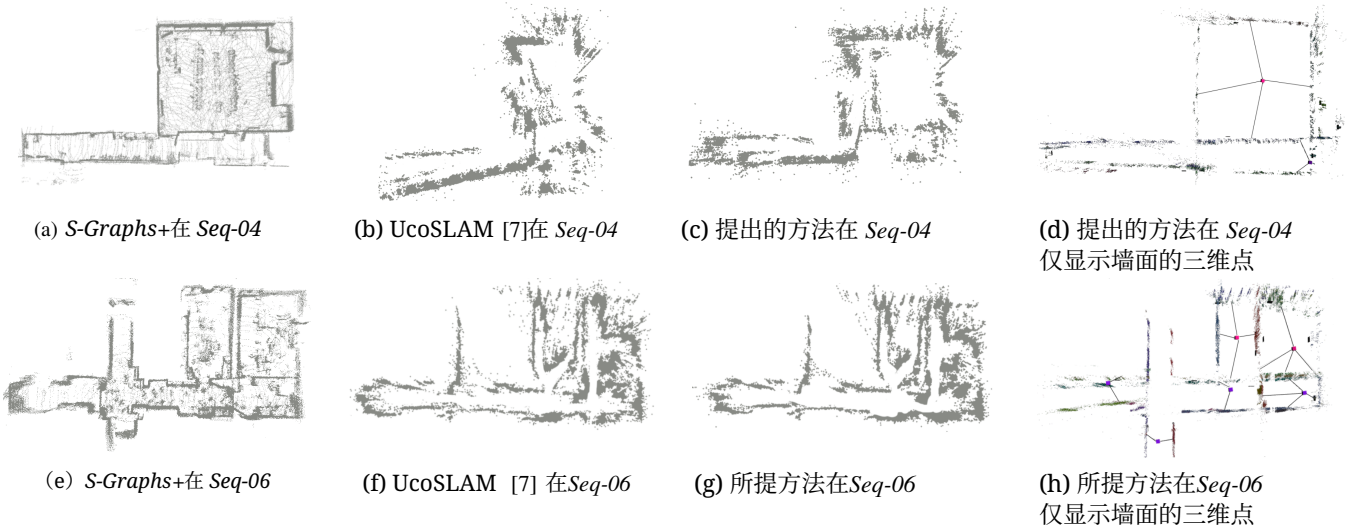
### *B.* 实验结果

为展示所提出方法相较于基线方法和真实值的准确性，本文采用绝对轨迹误差（ATE）进行测量。相应地，将所提出方法和基线方法的均方根偏差（RMSE）及标准差（STD）与真实值进行比较，数值较小的方法被认为具有更高的准确性。

根据表 II 中给出的评估结果，所提出的方法在大多数情况下优于基线方法。这一改进的主要原因是，该方法能够向地图中添加新的约束，并利用语义实体之间的关联来增强最终地图的重建效果。图 6 中也可以看到上述改进。此外，所提出的方法会利用数据关联结果；在访问先前观测过的墙面时，除了基于标记的闭环检测外，还会执行局部语义闭环。在这方面，如果当前可见墙面的方程先前已测量过，即该墙面先前已被检测到，则会再次执行全局优化，以改进重建地图。

在 *Seq-04* 这类情形中，上述影响更加明显：机器人从走廊出发，经房间的两扇门之一进入房间，再从另一扇门离开，继续沿同一走廊行进，因此无法利用关键点和标记进行闭环。相比之下，我们的方法通过创建墙面和房间约束，能够正确识别走廊及其墙壁，从而取得更好的结果。图 5 清楚地展示了这一改进：与基线方法相比，我们的方法能够更准确地重建环境地图。此外，它还能从环境中提取有意义的语义信息和拓扑信息。

**局限性。** 在测试基于标记的闭环检测被触发多次的环境时，所提出的方法可能会表现出与基线方法 UcoSLAM 相似的结果。



(a) *S-Graphs+*在 *Seq-04* (b) UcoSLAM [7]在 *Seq-04* (c) 提出的方法在 *Seq-04* (d) 提出的方法在 *Seq-04* 仅显示墙面的三维点

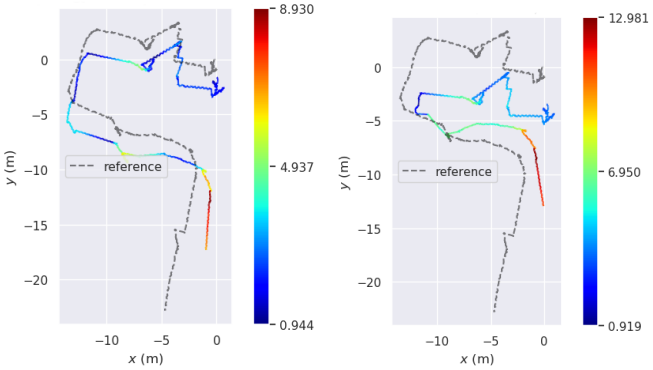
(e) *S-Graphs+*在 *Seq-06* (f) UcoSLAM [7] 在*Seq-06* (g) 所提方法在*Seq-06* (h) 所提方法在*Seq-06* 仅显示墙面的三维点

图 5：使用所提方法、基线方法和真实值生成的重建地图。可以看出，所提方法能够生成更精确的地图，其细节和高层语义实体均与真实值相似。

表 II：使用均方根偏差（RMSE）误差（单位为米）和标准差（STD）对所提方法在采集的数据集上的结果进行评估。最佳结果以粗体标出。

| | 方法 | | | |
|---|---|---|---|---|
| | 本文方法 | | UcoSLAM [7] | |
| **序列** | RMSE | STD | RMSE | STD |
| *Seq-01* | **8.038** | 3.058 | 8.130 | **3.035** |
| *Seq-02* | **6.883** | **3.598** | 6.930 | 3.633 |
| *Seq-03* | **2.266** | **0.894** | 2.687 | 1.335 |
| *Seq-04* | **3.787** | **1.848** | 5.822 | 2.726 |
| *Seq-05* | 1.255 | 0.763 | **1.238** | **0.751** |
| *Seq-06* | **2.676** | **1.524** | 2.720 | 1.536 |

*Seq-06* 在图 5 中展示了最终地图的一个示例：在机器人多次重访这些位置（从而多次重访标记）的环境中，所提出的方法与 UcoSLAM 生成了质量相近的地图。即使在此类场景下存在上述局限性，我们的方法仍能够从环境中提取更多有意义的信息，同时保持与基线方法相近的精度。




(a) 提出的方法 (b) UcoSLAM [7]

图 6：不同方法在 米 单位下的 *Seq-04* 平移部分绝对轨迹误差（绝对轨迹误差）。可以看出，所提方法构建的房间与真实值更为相似。

## V. 结论

本文提出了一种视觉同步定位与建图方法，该方法采用单目相机作为传感器，并利用周围布设的基准标记中编码的数据，实现精确的位姿估计和语义分割。所提出的方法能够检测环境中的三类实用语义实体，包括墙壁、走廊和房间，并生成包含高层语义数据的分层图。此外，通过加入更多拓扑约束并利用场景中所发现的上述语义实体之间的关系，该方法能够重建出更精确的环境地图

。足式机器人在真实环境条件下采集的数据集上的评估结果表明，与基线方法相比，该方法的精度有所提升；与基于激光雷达的最新技术水平相比，也取得了令人满意的结果。

本文提出的方法是一个项目的一部分，该项目利用基准标记的优势执行各种任务，同时确保这些标记对人眼不可见、但可被机器人识别。关于未来工作，作者计划使用 [6]中介绍的上述不可见基准标记对环境进行编码，并为机器人配备适当的传感器配置以读取这些标记，从而使所提方法的改进版本能够借助这种新型标记完成同步定位与建图任务。此外，提升所提系统的性能，使其能够在线实时运行，也是另一项

作为未来工作的目标。

## 参考文献


[1] C. Cadena、L. Carlone、H. Carrillo、Y. Latif、D. Scaramuzza、J. Neira、I. Reid 和 J. J. Leonard，《同步定位与建图的过去、现在与未来：迈向鲁棒感知时代》，IEEE机器人汇刊，第 32 卷，第 6 期，第 1309–1332 页，2016 年。[2] M. Filipenko 和 I. Afanasyev，《室内环境中移动机器人各类同步定位与建图系统的比较》，载于 2018 年国际智能系统会议（IS）。IEEE，2018 年，第 400–407 页。[3] S. Han 和 Z. Xi，《基于语义分割的动态场景语义同步定位与建图》，IEEE Access，第 8 卷，第 43 563–43 570 页，2020 年。[4] S. Wen、P. Li、Y. Zhao、H. Zhang、F. Sun 和 Z. Wang，《动态环境中的语义视觉同步定位与建图》，Autonomous Robots，第 45 卷，第 4 期，第 493–504 页，2021 年。[5] J. Zhao、Y. Huang、X. He、S. Zhang、C. Ye、T. Feng 和 L. Xiong，《用于自主室内停车的基于视觉语义地标的鲁棒建图与定位》，Sensors，第 19 卷，第 1 期，第 161 页，2019 年。[6] H. Agha、Y. Geng、X. Ma、D. I. Avs.ar、R. Kizhakidathazhath、Y.-S. Zhang、A. Tourani、H. Bavle、J.-L. Sanchez-Lopez、H. Voos 等，《利用胆甾相球形反射器的全向手性布拉格衍射实现不可克隆、对人眼不可见的机器视觉标记》，Light: Science & Applications，第 11 卷，第 1 期，第 1–19 页，2022 年。[7] R. Mu˜noz-Salinas 和 R. Medina-Carnicer，《UcoSLAM：通过融合关键点与方形平面标记实现的同步定位与建图》，模式识别，第 101 卷，第 107193 页，2020 年。[8] B. Pfrommer 和 K. Daniilidis，《TagSLAM：采用基准标记的鲁棒同步定位与建图》，2019 [ ] ˜ 1910 00679 。在线。[9] 可访问：https://arxiv.org/abs/ 。S. Garrido-Jurado、R. Munoz-Salinas、F. J. Madrid-Cuevas 和 M. J. Mar´ın-Jim´enez，《遮挡条件下高可靠性基准标记的自动生成与检测》，模式识别，第 47 卷，第 6 期，第 2280–2292 页，2014 年。[10] H. Bavle、J. L. Sanchez-Lopez、M. Shaheer、J. Civera 和 H. Voos，《用于结构化室内环境机器人导航的情境图》，2022 年。 [在线]。可访问：https://arxiv.org/abs/2202. 12197[11] ——，《S-graphs+：利用分层表示实现实时定位与建图》，arXiv 预印本 arXiv:2212.11770，2022 年。[12] A. Tourani、H. Bavle、J. L. Sanchez-Lopez 和 H. Voos，《视觉同步定位与建图：当前趋势是什么？未来又将如何发展？》，Sensors，第 22 卷，第 23 期，第 9297 页，2022 年。[13] I. A. Kazerouni 、 L. Fitzgerald 、 G. Dooly 和 D. Toal，《视觉同步定位与建图最新技术水平综述》，Expert Systems with Applications，第 117734 页，2022 年。[14] E. Olson，《AprilTag：一种鲁棒且灵活的视觉基准标记系统》，载于 2011 年 IEEE国际机器人与自动化会议。IEEE，2011 年，第 3400–3407 页。[15] R. Gomez-Ojeda、F.-A. Moreno、D. Zuniga-No¨el、D. Scaramuzza 和 J. Gonzalez-Jimenez，《PL-SLAM：通过结合点与线段实现的双目同步定位与建图系统》，IEEE机器人汇刊，第 35 卷，第 3 期，第 734–746 页，2019 年。[16] H. M. S. Bruno 和 E. L. Colombini，《LIFT-SLAM：一种基于深度学习特征的单目视觉同步定位与建图方法》，Neurocomputing，第 455 卷，第 97–110 页，2021 年。[17] M. Gonzalez、E. Marchand、A. Kacete 和 J. Royan，《S3LAM：结构化场景同步定位与建图》，arXiv 预印本 arXiv:2109.07339，2021 年。[18] R. Mur-Artal 和 J. D. Tard´os，《ORB-SLAM2：一种适用于单目、双目和 RGB-D相机的开源同步定位与建图系统》，IEEE机器人汇刊，第 33 卷，第 5 期，第 1255–1262 页，2017 年。[19] Y. Wu、A. Kirillov、F. Massa、W.-Y. Lo 和 R. Girshick，《Detectron2》，https://github.com/facebookresearch/detectron2，2019 年。[20] W. Wu、L. Guo、H. Gao、Z. You、Y. Liu 和 Z. Chen，《YOLO-SLAM：一种面向动态环境、带有几何约束的语义同步定位与建图系统》，Neural Computing and Applications，第 34 卷，第 8 期，第 6011–6026 页，2022 年。[21] L. Nicholson、M. Milford 和 N. S¨underhauf，《QuadricSLAM：将目标检测结果中的对偶二次曲面作为面向对象的同步定位与建图中的地标》，IEEE Robotics and Automation Letters，第 4 卷，第 1 期，第 1–8 页，2018 年。[22] M. Hosseinzadeh、K. Li、Y. Latif 和 I. Reid，《实时单目感知物体模型的稀疏同步定位与建图》，载于 2019 年国际机器人与自动化会议（ICRA）。IEEE，2019 年，第 7123–7129 页。[23] Y. Fan、Q. Zhang、Y. Tang、S. Liu 和 H. Han，《Blitz-SLAM：一种用于动态环境的语义同步定位与建图方法》，模式识别，第 121 卷，第 108225 页，2022 年。

[24] L. Sun、J. Wei、S. Su 和 P. Wu，“Solo-slam：一种适用于动态场景的并行语义 SLAM 算法”，*Sensors*，第 22 卷，6977，2022 年 18，第 p. .[25] Y. You、P. Wei、J. Cai、W. Huang、R. Kang 和 H. Liu，“Misd-slam：面向动态环境的多模态语义 SLAM”，*Wireless Communications and Mobile Computing*，第 2022 卷，2022 年。[26] H. Bavle、J. L. Sanchez-Lopez、M. Shaheer、J. Civera 和 H. Voos，“用于结构化室内环境中机器人导航的情境图”，*IEEE Robotics and Automation Letters*，第 7 卷，第 4 期，第 9107–9114 页，2022 年。