

室内目标级三维语义地标的不确定性融合方法

童鑫, 司海飞*

金陵科技学院智能科学与控制工程学院, 江苏 南京

收稿日期: 2026年8月18日; 录用日期: 2026年9月11日; 发布日期: 2026年9月18日

摘要

针对室内目标级语义建图中的深度污染、同类实例关联歧义和连续帧误差相关问题, 本文提出一种不确定性融合方法。方法以可见表面统计代表点描述地标, 利用收缩感兴趣区域内的中位数和中位绝对偏差估计深度, 并由距离、深度离散度和有效深度比例构建各向异性观测协方差; 在类别一致约束下执行一对一门控关联, 对时空邻近观测降权。固定锚点ROS 2节点级实验包含60次主实验、15次门限敏感性实验和45次补充消融实验, 每次接收240帧并评价预热后的220帧。补充消融表明, 鲁棒深度估计在15组配对实验中稳定降低深度RMSE、MAE和P95; 各向异性协方差在部分位置与关联指标的样本均值上呈有限改善, 但NEES和95.45%覆盖率仍明显偏离理论目标。30个严格配对中, 相关降权使平均NEES全部下降、95.45%覆盖率全部提高, 位置RMSE在29个配对中下降, 但中扰动(moderate)和强扰动(difficult)下仍存在过度自信。Mahalanobis相对Euclidean在18个配对中降低地标级错误合并比例, 同时在25个配对中降低F1, 并在全部配对中增加ID switch和建图节点单帧更新耗时。结果支持将相关降权视为抑制信息重复计数的工程近似; 统计门限还需与协方差尺度及地标生命周期联合设计。结论限于当前合成观测的节点级受控条件。

关键词

目标级语义建图, 不确定性建模, 数据关联, 相关观测降权, ROS 2

Uncertainty-Aware Fusion Method for Indoor Object-Level 3D Semantic Landmarks

Xin Tong, Haifei Si*

College of Intelligent Science and Control Engineering, Jinling Institute of Technology, Nanjing Jiangsu

Received: August 18, 2026; accepted: September 11, 2026; published: September 18, 2026

Abstract

Indoor object-level semantic mapping is affected by depth contamination, ambiguous associations

*通讯作者。

文章引用: 童鑫, 司海飞. 室内目标级三维语义地标的不确定性融合方法[J]. 建模与仿真, 2026, 15(9): 75-88.

DOI: 10.12677/mos.2026.159135

among same-class instances, and correlated errors across consecutive frames. This paper proposes an uncertainty-aware fusion method. Landmarks are represented by visible-surface statistical representative points. Depth is estimated using the median and median absolute deviation within a contracted region of interest, and an anisotropic observation covariance is constructed from range, depth dispersion, and valid-depth ratio. One-to-one gated association is performed under class-consistency constraints, while observations close in time and viewpoint are downweighted. The fixed-anchor ROS 2 node-level experiments included 60 main runs, 15 gate-sensitivity runs, and 45 supplementary ablation runs. Each run received 240 frames, with 220 frames evaluated after warm-up. Supplementary ablations showed that robust depth estimation consistently reduced depth RMSE, MAE, and P95 across 15 paired experiments. Anisotropic covariance yielded limited improvements in the sample means of some position and association metrics, whereas NEES and 95.45% coverage remained clearly inconsistent with their theoretical targets. Across 30 strictly paired comparisons, correlation discounting reduced mean NEES and increased 95.45% coverage in every pair, and reduced position RMSE in 29 pairs; however, overconfidence remained under moderate and difficult disturbances. Relative to Euclidean association, Mahalanobis association reduced the landmark-level false-merge proportion in 18 pairs, but reduced F1 in 25 pairs and increased ID switches and per-frame mapper update time in all pairs. The results support correlation discounting as an engineering approximation for suppressing repeated counting of information; statistical gating still needs to be jointly designed with covariance scale and landmark lifecycle. The conclusions are limited to the present node-level controlled setting with synthetic observations.

Keywords

Object-Level Semantic Mapping, Uncertainty Modeling, Data Association, Correlated Observation Discounting, ROS 2

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

室内移动机器人在巡检、仓储和服务作业中需同时表达环境几何、物体类别、实例位置及其可信程度。深度学习与视觉 SLAM 已由前端特征和位姿估计扩展至语义建图与不确定性估计[1], 多源融合则将多传感器、几何和语义信息统一纳入定位与地图表达[2]。动态目标会破坏传统视觉 SLAM 的静态假设[3]; 针对室内 RGB-D 场景, 自适应语义分割[4]以及深度筛选和空间关联[5]可抑制其影响并更新地图点。Kimera 以 3D 动态场景图组织度量、语义、对象及空间关系[6], 长期家庭服务研究也以物体空间特征、语义和置信度维护实例级环境信息[7]。然而, 稳定获得二维检测区域的三维代表位置、维持同类实例身份并量化其可信程度, 仍需专门的目标级三维地标建模。

RGB-D 的颜色和深度互补, 多模态融合可改善室内语义分割和目标区域表征[8]。Hydra 实现了 3D 场景图的在线构建与优化[9]; 同类实例的持续维护依赖帧间关联, 3D 多目标跟踪表明检测与关联紧密耦合[10]。将语义先验和关联不确定度引入视觉 SLAM 可提升复杂动态环境中点与地图点关联的鲁棒性[11], ConceptGraphs 通过多视图关联构建开放词汇 3D 场景图[12]。相关综述指出地图构建、状态估计和多源融合应协同考虑[13]。本文关注连续观测相关性对目标级地标后验一致性的影响: 过严的 Mahalanobis 门限会拒绝正确观测并触发新地标, 而将连续帧视为独立信息会造成协方差过快收缩。

本文以可见表面统计代表点限定目标级地标的位置语义, 构建由鲁棒深度估计、质量自适应各向异

性协方差、类别一致一对一关联和相关观测降权组成的处理链。方法不假定有限视角下的代表点等于物体物理中心, 也不将当前经验协方差模型视为传感器与位姿误差的联合传播模型。

本文给出可见表面统计代表点的操作性定义、基于检测区域深度质量的观测协方差模型, 以及结合类别门控、帧内线性指派和连续观测信息降权的地标更新方法。实验采用 ROS 2 Humble 节点级受控环境, 主实验为 60 次、门限敏感性实验为 15 次, 并补充 45 次模块消融实验; 结论限于固定锚点与合成观测条件。除汇总统计外, 本文还报告来自真实导出记录的代表性空间可视化, 用于解释关联与融合现象。

2. 问题描述与总体框架

室内目标级语义建图需要把不同时刻的二维检测、深度测量和相机位姿转换为可持续维护的三维地标。检测区域可能包含无效深度、背景和遮挡, 连续帧又常共享视点与误差状态。本文据此将处理链分为观测建模、候选关联和递推融合, 并以统一状态连接三个环节。

整体流程见图 1: 检测区域、深度和给定相机位姿首先形成带协方差的三维语义观测; 随后通过类别一致门控、距离计算与一对一指派完成目标-地标关联; 对匹配观测执行递推融合, 对未匹配观测初始化新地标, 最终得到持续更新的目标级三维语义地标。

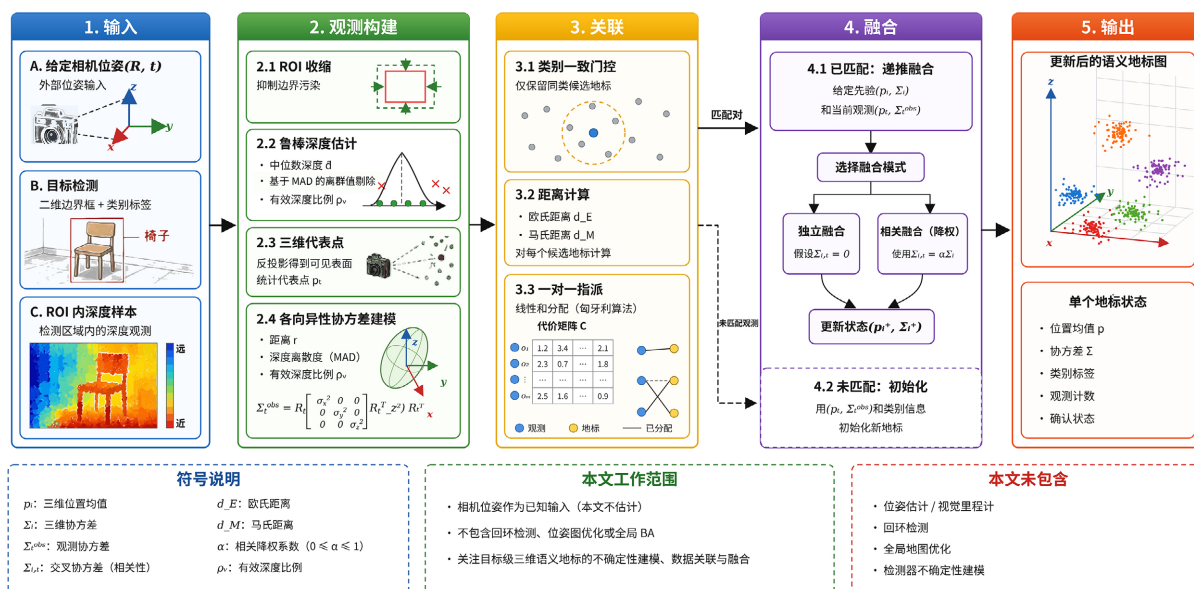


Figure 1. Overall process of uncertainty-aware fusion for object-level 3D semantic landmarks

图 1. 面向目标级三维语义地标的的多不确定性融合总体流程

状态、代表点与视角依赖

为同时支持三维定位、统计门控和相关性判断, 将第 i 次目标观测定义为

$$\mathcal{O}_i = \{z_i, R_i, \pi_i, q_i, c_i, t_i\}, \quad (1)$$

式(1)中, z_i 与 R_i 是地图坐标系下的三维代表位置和观测协方差; π_i 为类别概率; q_i 汇总深度质量; c_i 与 t_i 分别为相机光心和时间戳。 q_i 只描述质量, 不替代位置协方差。

第 j 个持久地标维护几何、语义、观测次数及最近关联的时空信息:

$$\mathcal{L}_j = \{\mu_j, \Sigma_j, p_j, n_j, c_j^{\text{last}}, t_j^{\text{last}}\}, c_j = \arg \max_k p_{j,k}, \quad (2)$$

式(2)中, μ_j 和 Σ_j 为地标位置的后验均值与协方差, p_j 为类别概率, n_j 为累计观测次数; c_j^{last} 与 t_j^{last} 用于相关性判断, 类别 c_j 由 p_j 的最大项派生。

检测框中心像素与 ROI 深度统计得到当前视点下的可见表面代表点。不同视角可能对应不同表面, 故跨视角观测写为

$$z_i = h(c_i, \mathcal{L}_j) + \varepsilon_i. \quad (3)$$

式(3)中, $h(c_i, \mathcal{L}_j)$ 表示视点和目标几何共同决定的可见表面统计位置, ε_i 为剩余测量误差。本文不显式求解 $h(\cdot)$; μ_j 因而只表示观测历史下的统计汇总, 不是物体质心、包围盒中心或模型原点的无偏估计。

3. 深度估计与三维观测不确定性

为降低检测框边缘污染, 本文先收缩 ROI, 再执行有效值筛选和 MAD 离群过滤; 代表深度经针孔模型反投影后用于构造三维协方差。

3.1. 鲁棒深度估计与反投影

检测框四边按比例 0.2 收缩。区域内有限、严格大于 0 且不超过 10.0 m 的深度构成 $\mathcal{D}_i$, 随后按式(4)~(7)估计代表深度:

$$m_i = \text{median}(\mathcal{D}_i), \quad (4)$$

$$\text{MAD}_i = \text{median}_{d \in \mathcal{D}_i} |d - m_i|, \quad (5)$$

$$\mathcal{D}'_i = \{d \in \mathcal{D}_i : |d - m_i| \leq \tau_{\text{MAD}} \max(1.4826 \text{MAD}_i, 10^{-6} \text{ m})\}, \quad (6)$$

$$\hat{d}_i = \text{median}(\mathcal{D}'_i), \tau_{\text{MAD}} = 3.5. \quad (7)$$

式(4)~(5)给出中位数 m_i 与中位绝对偏差 MAD_i ; 式(6)以 1.4826MAD_i 近似高斯鲁棒尺度, 用 $\tau_{\text{MAD}} = 3.5$ 过滤离群值, 并以 10^{-6} m 防止零尺度退化; 式(7)对过滤集合再取中位数得到 $\hat{d}_i$ 。

过滤后样本数与收缩 ROI 像素数之比为 r_i 。当 r_i 较低时, 后续模型增大观测协方差; 初始或过滤后集合为空时拒绝该观测。

以原检测框中心 (u_i, v_i) 为代表像素, 针孔模型给出相机坐标点, 并由相机到地图的刚体变换得到位置观测:

$$p_i^c = \begin{bmatrix} (u_i - c_x) \hat{d}_i / f_x \\ (v_i - c_y) \hat{d}_i / f_y \\ \hat{d}_i \end{bmatrix}, z_i = C_{mc} p_i^c + t_{mc}. \quad (8)$$

式(8)中, f_x, f_y, c_x, c_y 为相机内参, p_i^c 为相机坐标点, C_{mc} 和 t_{mc} 为相机到地图的旋转与平移。本模型尚未联合传播像素定位、内参和位姿协方差。

3.2. 质量自适应协方差

统计门控与递推融合还需要与深度质量相匹配的观测协方差。本文将距离、ROI 深度离散度和有效样本比例映射为经验轴向标准差:

$$\sigma_{d,i} = \frac{\sqrt{(a + b \hat{d}_i^2)^2 + (1.4826 \text{MAD}_i)^2}}{\sqrt{\max(r_i, r_{\min})}}, a = 0.01 \text{ m}, b = 0.003 \text{ m}^{-1}, r_{\min} = 0.05. \quad (9)$$

式(9)中, $a + b\hat{d}_i^2$ 为距离相关误差项, 1.4826MAD_i 为鲁棒离散尺度, 分母中的有效样本比例项用于调整有效深度不足的影响。受控实验取 $a = 0.01 \text{ m}$ 、 $b = 0.003 \text{ m}^{-1}$ 和 $r_{\min} = 0.05$ 。各项先统一为长度量再合成, 故 $\sigma_{d,i}$ 的量纲为长度。该式是质量到协方差尺度的经验映射, 不是特定相机的完整物理噪声模型。

以式(9)的轴向标准差为基准, 用固定横向比例构造相机坐标系各向异性协方差, 再旋转至地图坐标系:

$$\mathbf{R}_i^c = \text{diag}(k_x^2 \sigma_{d,i}^2, k_y^2 \sigma_{d,i}^2, \sigma_{d,i}^2), \mathbf{R}_i = \mathbf{C}_{mc} \mathbf{R}_i^c \mathbf{C}_{mc}^T, \quad (10)$$

式(10)中, $k_x = 0.55$ 、 $k_y = 0.65$ 为无量纲横向比例; $\mathbf{R}_i^c$ 和 $\mathbf{R}_i$ 分别是相机与地图坐标系下的协方差。相机系三轴暂按不相关处理, 三轴方差及旋转后特征值采用 $2.5 \times 10^{-5} \text{ m}^2$ 下限。

一致性评价针对融合后的地标状态, 而非单帧观测。对已确认地标 j 在预热后第 k 个评价帧, 以生成器定义的固定语义锚点 $\mathbf{x}_j^{\text{ref}}$ 为参照, 定义

$$\begin{aligned} \mathbf{e}_{j,k} &= \boldsymbol{\mu}_{j,k} - \mathbf{x}_j^{\text{ref}}, \\ \text{NEES}_{j,k} &= \mathbf{e}_{j,k}^T \boldsymbol{\Sigma}_{j,k}^{-1} \mathbf{e}_{j,k}. \end{aligned} \quad (11)$$

式(11)中, $\mathbf{e}_{j,k}$ 、 $\boldsymbol{\mu}_{j,k}$ 和 $\boldsymbol{\Sigma}_{j,k}$ 分别为位置误差、融合后均值和后验协方差。评价先排除前 20 帧, 只保留活动且累计观测不少于 5 次的确认地标, 并按类别一致、距离不超过 0.75 m 的一对一匹配形成地标-帧样本。每次运行的平均 NEES 和 95.45% 覆盖率分别在全部有限样本上求均值和满足 $\text{I}[\text{NEES}_{j,k} \leq \chi_3^2(0.9545)]$ 的比例, 再对主实验和门限敏感性实验同一实验单元的 5 次运行, 以及补充消融实验同一实验单元的 15 次运行, 分别报告均值与样本标准差; $\chi_3^2(0.9545) = 8.0249$, 三维高斯模型下单样本期望为 3。未确认地标、未匹配地标和虚警不产生 NEES 样本。

4. 类别一致关联与相关观测融合

4.1. 一对一关联

对类别一致的观测-地标候选, 创新协方差同时包含地标后验不确定性和当前观测不确定性。当前实现忽略二者的交叉协方差, 因此

$$\mathbf{S}_{ij} = \boldsymbol{\Sigma}_j + \mathbf{R}_i. \quad (12)$$

利用创新协方差归一化位置残差, 得到平方 Mahalanobis 距离

$$d_{ij}^2 = (\mathbf{z}_i - \boldsymbol{\mu}_j)^T \mathbf{S}_{ij}^{-1} (\mathbf{z}_i - \boldsymbol{\mu}_j). \quad (13)$$

式(13)在创新协方差定义的统计尺度下度量残差。候选条件为 $d_{ij}^2 \leq \tau_m$, 受控实验默认 $\tau_m = 11.345$; Euclidean 对照为 $\|\mathbf{z}_i - \boldsymbol{\mu}_j\|_2 \leq 0.45 \text{ m}$ 。这些门限不外推为其他相机、帧率或场景的最优值。

类别不一致或门限外元素赋无效代价, 线性指派(linear sum assignment)求解帧内一对一匹配。未匹配观测初始化新地标, 故严格门控可能通过“拒绝-新建”路径增加重复建图; 评价同时报告地标级错误合并比例、目标级重复建图比例和 ID switch。

4.2. 相关观测的信息降权

连续 RGB-D 帧常共享视点、可见表面和误差状态。若按独立观测完整累加, 其均值变化可能有限而协方差持续收缩。本文据时间间隔和相机光心位移设置几何信息权重:

$$\gamma_i = \begin{cases} \gamma_c, & 0 \leq t_i - t_j^{\text{last}} < \tau_t \text{ and } \|\mathbf{c}_i - \mathbf{c}_j^{\text{last}}\|_2 < \tau_c, \tau_t = 0.15 \text{ s}, \tau_c = 0.08 \text{ m}. \\ 1, & \text{otherwise} \end{cases} \quad (14)$$

当 $0 \leq t_i - t_j^{\text{last}} < \tau_i = 0.15 \text{ s}$ 且 $\|\mathbf{c}_i - \mathbf{c}_j^{\text{last}}\|_2 < \tau_c = 0.08 \text{ m}$ 时取 $\gamma_i = \gamma_c$; 视点不可用或任一条件不满足时均取 $\gamma_i = 1$ 。相关降权模式(correlated)采用 $\gamma_c = 0.175$, 独立融合模式(independent)恒取 1。该规则是相关性的工程近似, 不等价于显式跨帧协方差估计。

对已关联观测, 记更新前的先验为 $(\boldsymbol{\mu}_j^-, \boldsymbol{\Sigma}_j^-)$, 更新后的后验为 $(\boldsymbol{\mu}_j^+, \boldsymbol{\Sigma}_j^+)$, 新观测为 $(\mathbf{z}_i, \mathbf{R}_i)$ 。信息形式更新将新观测的信息矩阵和信息向量同时乘以 γ_i :

$$\boldsymbol{\Sigma}_j^+ = \left[(\boldsymbol{\Sigma}_j^-)^{-1} + \gamma_i \mathbf{R}_i^{-1} \right]^{-1}, \quad (15)$$

$$\boldsymbol{\mu}_j^+ = \boldsymbol{\Sigma}_j^+ \left[(\boldsymbol{\Sigma}_j^-)^{-1} \boldsymbol{\mu}_j^- + \gamma_i \mathbf{R}_i^{-1} \mathbf{z}_i \right]. \quad (16)$$

式(15)~(16)对观测信息矩阵与信息向量使用相同权重。 $\gamma_i = 1$ 时退化为独立高斯融合; $0 < \gamma_i < 1$ 等价于将观测协方差放大为 $\mathbf{R}_i / \gamma_i$, 从而减缓后验协方差收缩。

权重仅作用于几何信息; 类别概率仍按检测置信度和累计次数更新。固定 γ_c 不能保证所有扰动条件下达到理论一致性, 也不能替代完整相关噪声滤波。

4.3. 初始化与生命周期

未匹配观测以 $n_j = 1$ 初始化, 累计 5 次观测后确认并发布; 活动地标连续 20 s 未更新则失效。当前实现没有重激活和重复地标合并, 因此严格门控拒绝正确观测后, 新旧地标可能长期并存。该机制为解释地图碎片化提供实现依据, 但不是独立因果验证。

5. 实验设计

5.1. 实验平台、节点链路与有效性审计

实验环境为 Ubuntu 22.04、ROS 2 Humble 和 Python 3.10。合成观测节点发布类别、三维位置、协方差、相机光心及评估标识; 建图节点执行类别一致门控、线性指派和信息形式融合; 评估节点记录逐帧关联、位置误差、统计一致性及建图节点单帧更新耗时。场景包含 6 个静态目标和 3 个类别, 每个类别设置两个空间相近实例以形成同类关联歧义。评估参照为生成器定义的固定语义锚点, 不等于物体质心。

主实验 60/60 次、门限实验 15/15 次和补充消融实验 45/45 次均完成; 各次实验均接收 240/240 帧, 预热 20 帧后评价 220 帧。配置、逐帧指标、关联记录和汇总文件完整, 核心指标为有限值且同帧关联满足一对一约束。确认前尚未定义的逐帧指标不计入评价。

5.2. 主实验设计与评价指标

主实验采用低扰动(clean)、中扰动(moderate)和强扰动(difficult) 3 个复合扰动档位、2 种关联方法、2 种融合方式和 5 个配对随机种子的全因子设计, 共 60 次运行。三档配置均直接取自观测生成程序: 漏检概率分别为 0.03、0.10 和 0.20; 虚警数服从逐帧 Poisson 分布。位置误差不是固定标准差白噪声, 而是先由有效深度比例和 MAD 生成式(9)~(10)的报告协方差, 再以噪声尺度调整实际协方差, 并按一阶自回归过程引入时间相关。离群点在随机方向附加 0.16~0.42 m 位移; 误分类时从其余两类等概率选择。虚警位置在 $[-1.5, 1.9] \times [-1.2, 1.7] \times [0.45, 1.55] \text{ m}^3$ 内均匀生成, 协方差为 $\text{diag}(0.10^2, 0.10^2, 0.14^2) \text{ m}^2$ 。因此档位差异只能解释为组合扰动效应。

同一档位和随机种子下的四种方法组合严格配对。门限敏感性实验固定中扰动、Mahalanobis 和相关降权, 在 7.815、11.345 和 16.000 三个平方门限下各运行 5 个配对种子; 其种子与主实验中扰动组不同,

仅用于比较门限变化的相对趋势。
主实验设计、扰动配置及评价指标汇总见表 1。

Table 1. Design and evaluation metrics of the ROS 2 node-level controlled experiments
表 1. ROS 2 节点级受控实验设计与评价指标

项目	设置
主实验矩阵	3 个扰动档位 × 2 种关联 × 2 种融合 × 5 个配对种子 = 60 次
关联方法	Euclidean; Mahalanobis (默认平方门限 $\tau_m = 11.345$)
融合方式	independent; correlated ($\gamma_c = 0.175$)
门限实验(τ_m)	7.815、11.345、16.000, 各 5 次, 共 15 次
帧数	每次 240 帧; 预热 20 帧; 评价 220 帧
clean 扰动	漏检/误分类/离群概率: 0.03/0.01/0.01; 虚警 $\lambda = 0.015$; $r = 0.93 \pm 0.035$; MAD 基值/标准差: 0.004/0.0015m; 噪声尺度/AR 系数: 0.90/0.88
moderate 扰动	漏检/误分类/离群概率: 0.10/0.04/0.045; 虚警 $\lambda = 0.045$; $r = 0.78 \pm 0.08$; MAD 基值/标准差: 0.010/0.0035m; 噪声尺度/AR 系数: 1.20/0.90
difficult 扰动	漏检/误分类/离群概率: 0.20/0.09/0.09; 虚警 $\lambda = 0.09$; $r = 0.62 \pm 0.12$; MAD 基值/标准差: 0.018/0.006m; 噪声尺度/AR 系数: 1.60/0.92
正文报告指标	位置 RMSE; 平均 NEES; 95.45%覆盖率; F_1 ; 地标级错误合并比例; 目标级重复建图比例; ID switch; 建图节点单帧更新耗时

注: r 从相应正态分布采样后截断至 $[0.05, 1]$; MAD 均值还含 $0.001d^2$ m 距离项并截断为非负。虚警 λ 为每帧 Poisson 均值。

评价节点每帧仅保留活动且已确认地标, 按类别一致、与固定锚点距离不超过 0.75 m 的线性指派得到 TP; 其余已确认地标为 FP, 未匹配的 6 个真值目标为 FN。逐帧 $F_1 = 2PR/(P+R)$, 运行级 F_1 为预热后 220 帧的算术平均。位置 RMSE 汇总同期全部有效匹配的三维欧氏误差后计算; 未确认地标不进入逐帧 TP、FP 或误差样本, 未匹配真值计入 FN。

地标级错误合并比例(landmark-level false-merge proportion)是发生跨真实对象合并的地标 ID 数, 占所有接收过真实对象观测的地标 ID 数的比例。目标级重复建图比例(object-level duplicate-mapping proportion)是被映射为两个及以上地标 ID 的真实对象数, 占全部 6 个真实对象的比例。ID switch 按每个真实对象的关联事件序列累计相邻地标 ID 变化次数。三项身份指标均由完整运行期内的真实对象关联记录计算, 包括预热期和地标未确认阶段; 虚警观测不参与统计。

注: 建图节点单帧更新耗时(per-frame mapper update time)从观测回调进入内部处理开始, 至输出消息组装和发布前结束; 包含观测转换、失效检查、代价矩阵构造、线性指派、融合及新地标初始化, 平均值使用预热后的 220 帧。该指标不包含检测、RGB-D 处理、TF、SLAM、DDS 传输、消息发布、日志、磁盘分析或绘图。

5.3. 补充消融实验设计

补充消融固定中扰动(moderate)、Mahalanobis 关联和 correlated 融合, 其余门限与生命周期参数保持不变, 并使用 15 个相同配对随机种子(20261717~20261731)。鲁棒深度消融仅比较同一收缩 ROI 内有效

深度的直接均值与本文 $\text{median} + \text{MAD}$ (阈值 3.5) 筛选后中位数; 协方差消融仅比较迹匹配各向同性模型 (方差取各向异性协方差迹的 $1/3$) 与式(10)的各向异性模型。每组均运行 240 帧并评价预热后的 220 帧, 以隔离两个模块的独立贡献。

5.4. 代表性运行可视化设置

可视化仅用于补充汇总统计对空间布局和后验不确定性的解释, 不替代跨种子统计。图由真实 ROS 2 话题级 CSV 生成, 固定为 moderate、seed 20261717 的四种关联-融合组合; 真值锚点为场景生成器定义的语义参考点, 不代表物体几何质心。

图 2 给出代表性运行的受控合成场景。六个静态语义目标分属三个类别, 每类包含两个空间相近实例; 相机光心沿预设视点轨迹采样。

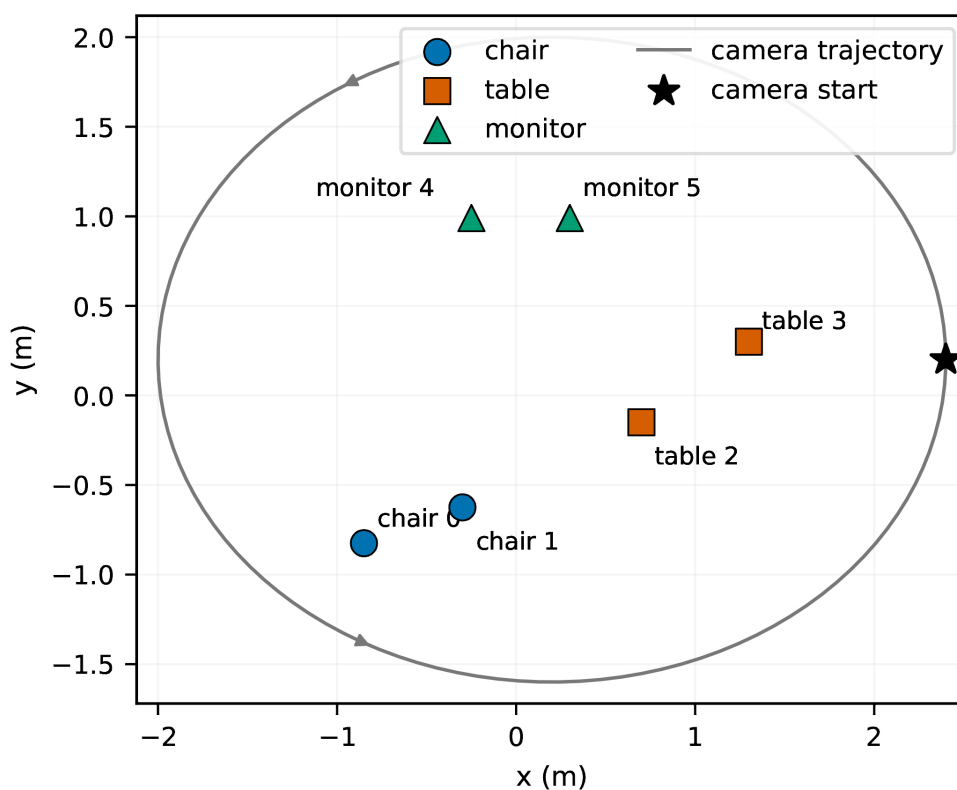


Figure 2. Controlled synthetic scene with six semantic anchors and the camera viewpoint trajectory
图 2. 受控合成场景、六个语义锚点与相机视点轨迹

6. 结果与分析

6.1. 数据关联结果

严格配对结果显示, Mahalanobis 相对 Euclidean 的 F1 在 25/30 个配对中降低、5/30 不变; 地标级错误合并比例在 18/30 个配对中降低、10/30 不变、2/30 增加, 平均差值为 -0.0349 。ID switch 和建图节点单帧更新耗时在全部 30 个配对中增加, 平均分别增加 163.47 次和 1.72 ms。当前结果体现地标级错误合并抑制与关联连续性之间的权衡, 不能概括为 Mahalanobis 全面优于 Euclidean。

主实验各关联-融合组合的数据关联结果见表 2。

Table 2. Association results of the main experiment (Mean $\pm$ sample SD, $n = 5$)
表 2. 主实验关联结果(均值 $\pm$ 样本标准差, $n = 5$)

扰动	关联	融合	F1	地标级错误合并比例	ID switch	单帧更新耗时/ms
clean	E	I	1.000 $\pm$ 0.000	0.000 $\pm$ 0.000	24.6 $\pm$ 9.6	0.88 $\pm$ 0.01
clean	E	C	1.000 $\pm$ 0.000	0.000 $\pm$ 0.000	24.6 $\pm$ 9.6	0.91 $\pm$ 0.02
clean	M	I	0.997 $\pm$ 0.006	0.000 $\pm$ 0.000	50.2 $\pm$ 13.2	1.57 $\pm$ 0.14
clean	M	C	0.961 $\pm$ 0.043	0.000 $\pm$ 0.000	60.6 $\pm$ 18.1	1.63 $\pm$ 0.11
moderate	E	I	0.900 $\pm$ 0.056	0.040 $\pm$ 0.023	101.4 $\pm$ 13.9	0.88 $\pm$ 0.02
moderate	E	C	0.900 $\pm$ 0.056	0.040 $\pm$ 0.023	101.4 $\pm$ 13.9	0.94 $\pm$ 0.03
moderate	M	I	0.652 $\pm$ 0.054	0.010 $\pm$ 0.006	292.8 $\pm$ 25.6	2.80 $\pm$ 0.13
moderate	M	C	0.645 $\pm$ 0.053	0.010 $\pm$ 0.005	305.4 $\pm$ 28.2	2.82 $\pm$ 0.09
difficult	E	I	0.719 $\pm$ 0.019	0.108 $\pm$ 0.072	201.6 $\pm$ 15.8	0.89 $\pm$ 0.03
difficult	E	C	0.716 $\pm$ 0.023	0.114 $\pm$ 0.058	203.0 $\pm$ 19.9	0.93 $\pm$ 0.02
difficult	M	I	0.420 $\pm$ 0.031	0.038 $\pm$ 0.013	466.2 $\pm$ 18.5	3.49 $\pm$ 0.06
difficult	M	C	0.431 $\pm$ 0.027	0.035 $\pm$ 0.014	462.2 $\pm$ 20.4	3.45 $\pm$ 0.07

注：表 2、表 3 中，E：Euclidean；M：Mahalanobis；I：independent；C：correlated。

6.2. 位置精度与统计一致性

30 个严格配对中，相关降权使位置 RMSE 在 29 个配对中降低、1 个配对中升高，平均差值为-3.96 mm；平均 NEES 在 30/30 个配对中降低，95.45%覆盖率在 30/30 个配对中提高，平均差值分别为-20.09 和 0.224。图 3 给出各档位的均值与样本标准差。中、强扰动下的 NEES 仍高于 3，覆盖率低于 0.9545，故只能认为过度自信有所缓解，不能声称完全校准。各组合的位置精度与统计一致性结果见表 3。

Table 3. Position accuracy and statistical consistency results (Mean $\pm$ sample SD, $n = 5$)
表 3. 位置精度与统计一致性结果(均值 $\pm$ 样本标准差, $n = 5$)

扰动	关联	融合	位置 RMSE/mm	平均 NEES	95.45%覆盖率
clean	E	I	19.87 $\pm$ 1.75	14.52 $\pm$ 2.63	0.424 $\pm$ 0.036
clean	E	C	14.60 $\pm$ 1.55	4.75 $\pm$ 0.62	0.834 $\pm$ 0.037
clean	M	I	18.81 $\pm$ 1.67	12.96 $\pm$ 2.35	0.435 $\pm$ 0.047
clean	M	C	14.26 $\pm$ 1.68	4.43 $\pm$ 0.80	0.860 $\pm$ 0.050
moderate	E	I	32.70 $\pm$ 2.56	38.57 $\pm$ 6.61	0.103 $\pm$ 0.043
moderate	E	C	27.84 $\pm$ 2.47	18.34 $\pm$ 2.50	0.338 $\pm$ 0.030
moderate	M	I	29.59 $\pm$ 2.52	30.18 $\pm$ 4.48	0.113 $\pm$ 0.041
moderate	M	C	26.35 $\pm$ 1.93	15.10 $\pm$ 1.04	0.333 $\pm$ 0.107
difficult	E	I	53.94 $\pm$ 3.92	85.56 $\pm$ 7.41	0.041 $\pm$ 0.020
difficult	E	C	49.69 $\pm$ 3.53	47.32 $\pm$ 3.38	0.067 $\pm$ 0.040
difficult	M	I	57.05 $\pm$ 11.42	68.21 $\pm$ 30.09	0.049 $\pm$ 0.053
difficult	M	C	55.47 $\pm$ 11.08	39.54 $\pm$ 16.20	0.080 $\pm$ 0.079

注：三维高斯模型下 NEES 期望为 3，覆盖率目标为 0.9545；固定锚点为生成器语义参考，不代表物体中心。

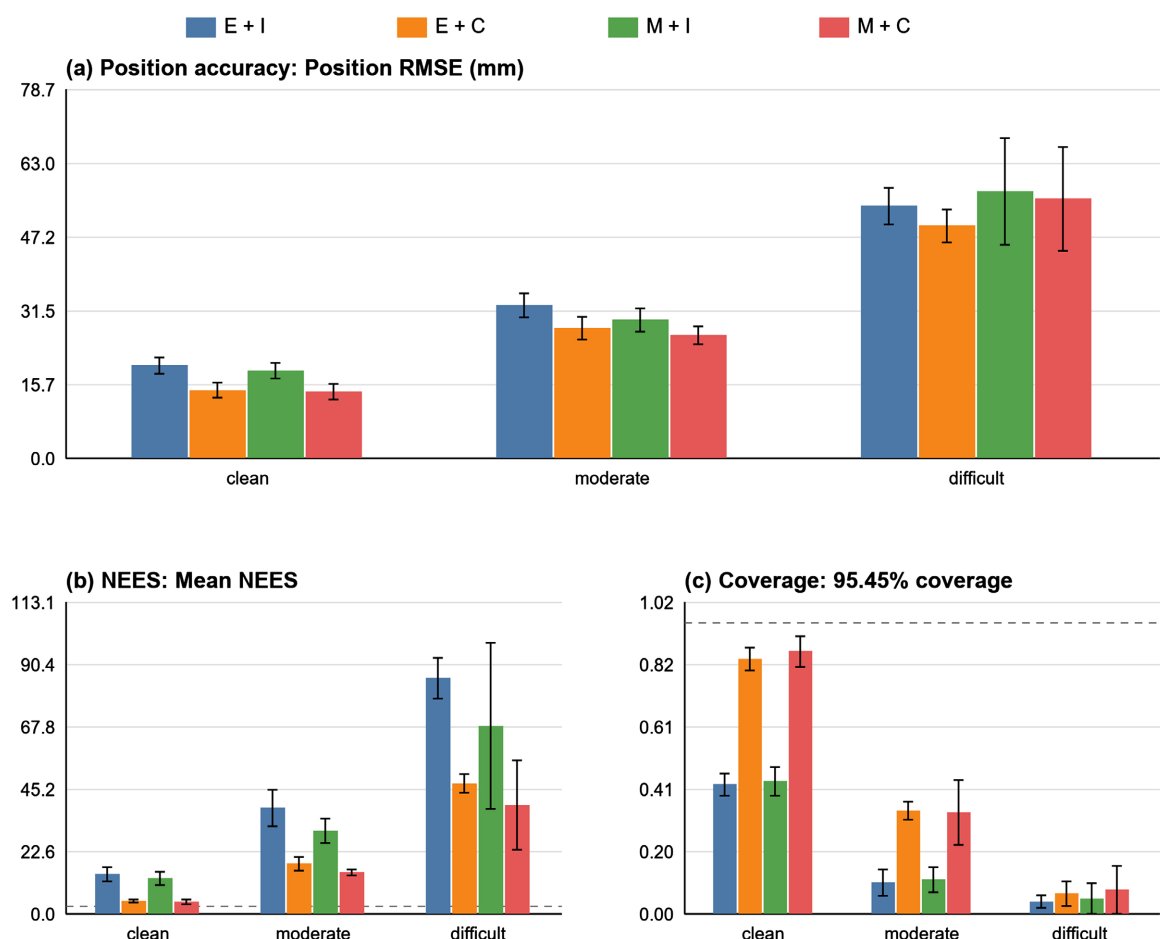


Figure 3. Position accuracy and fusion consistency across different disturbance profiles
图 3. 不同扰动档位下的位置精度与融合一致性

6.3. Mahalanobis 门限敏感性

平方门限由 7.815 放宽至 16.000 时, 平均 F1 由 0.5542 增至 0.7839, ID switch 由 353.2 次降至 229.6 次, 地标级错误合并比例由 0.0020 增至 0.0148。三个门限下目标级重复建图比例均为 1.0, 即 6 个真实对象均曾对应两个及以上地标 ID。门限实验使用另一组五个配对种子, 只用于判断相对趋势。

不同平方 Mahalanobis 门限下的敏感性结果见表 4。

Table 4. Sensitivity to the squared Mahalanobis gate (Mean $\pm$ sample SD, $n = 5$)
表 4. Mahalanobis 平方门限敏感性(均值 $\pm$ 样本标准差, $n = 5$)

平方门限 τ_m	F1	地标级错误合并比例	目标级重复建图比例	ID switch	单帧更新耗时/ms
7.815	0.554 $\pm$ 0.011	0.0020 $\pm$ 0.0045	1.000 $\pm$ 0.000	353.2 $\pm$ 24.5	2.96 $\pm$ 0.14
11.345	0.686 $\pm$ 0.027	0.0104 $\pm$ 0.0108	1.000 $\pm$ 0.000	294.4 $\pm$ 21.7	2.68 $\pm$ 0.13
16.000	0.784 $\pm$ 0.066	0.0148 $\pm$ 0.0143	1.000 $\pm$ 0.000	229.6 $\pm$ 23.4	2.44 $\pm$ 0.12

6.4. 模块消融结果

表 5 汇总固定 moderate、Mahalanobis 和 correlated 设置下的 45 次补充消融实验。三种配置各使用相

同的 15 组 paired seeds, 每次接收 240 帧, 并在预热 20 帧后评价 220 帧。鲁棒深度消融仅将 Mean 替换为 median + MAD rejection; 协方差消融仅将 trace-matched isotropic 替换为 anisotropic, 其余参数固定。

Table 5. Paired ablation results from the 240-frame supplementary experiments (Mean $\pm$ sample SD, $n = 15$)
表 5. 240 帧补充实验的配对消融结果(均值 $\pm$ 样本标准差, $n = 15$)

(a) 鲁棒深度估计消融			
指标	Baseline	Proposed	Δ (95% CI)
Depth RMSE/mm	12.80 $\pm$ 0.34	5.58 $\pm$ 0.12	-7.22 [-7.41, -7.03]
Depth MAE/mm	9.81 $\pm$ 0.25	4.36 $\pm$ 0.10	-5.45 [-5.60, -5.31]
Depth P95/mm	25.91 $\pm$ 0.80	11.21 $\pm$ 0.28	-14.69 [-15.17, -14.21]
Position RMSE/mm	26.65 $\pm$ 4.28	26.30 $\pm$ 4.33	-0.34 [-0.80, 0.11]
(b) 协方差模型消融			
指标	Baseline	Proposed	Δ (95% CI)
Position RMSE/mm	27.01 $\pm$ 3.79	26.30 $\pm$ 4.33	-0.71 [-1.76, 0.34]
Association F1	0.6355 $\pm$ 0.0336	0.6547 $\pm$ 0.0471	+0.0192 [-0.0002, 0.0386]
Mean NEES	15.58 $\pm$ 3.70	15.21 $\pm$ 2.89	-0.36 [-1.72, 0.99]
95.45% Coverage	32.10% $\pm$ 11.75%	32.32% $\pm$ 10.37%	+0.22 [-5.63, 6.06] pp

注：结果为 15 组相同随机种子的配对实验，均采用 moderate、Mahalanobis 和 correlated 设置。各向同性模型采用迹匹配形式 $\Sigma_{iso} = \text{trace}(\Sigma_{aniso})/3 \times I$ 。 Δ 定义为 Proposed-Baseline；对误差指标 $\Delta < 0$ 表示改善，对 F1 和 Coverage 则 $\Delta > 0$ 表示改善。括号为基于逐 seed 配对差的双侧 t 型 95% CI (df = 14)，Coverage 差值以百分点表示。

相对 Mean，鲁棒深度将 Depth RMSE 从 12.80 $\pm$ 0.34 mm 降至 5.58 $\pm$ 0.12 mm， $\Delta = -7.22$ mm (95% CI [-7.41, -7.03])；Depth MAE 从 9.81 $\pm$ 0.25 mm 降至 4.36 $\pm$ 0.10 mm， $\Delta = -5.45$ mm (95% CI [-5.60, -5.31])；Depth P95 从 25.91 $\pm$ 0.80 mm 降至 11.21 $\pm$ 0.28 mm， $\Delta = -14.69$ mm (95% CI [-15.17, -14.21])。三项深度误差均为 15/15 seed 改善。Position RMSE 仅由 26.65 $\pm$ 4.28 mm 降至 26.30 $\pm$ 4.33 mm， $\Delta = -0.34$ mm (95% CI [-0.80, 0.11]；10/15 改善)；该区间跨越 0，因此不据此声称稳定的位置精度提升。

相对迹匹配各向同性模型，各向异性协方差的 Position RMSE $\Delta = -0.71$ mm (95% CI [-1.76, 0.34]；11/15 改善)、Association F1 $\Delta = +0.0192$ (95% CI [-0.0002, 0.0386]；12/15 改善)、mean NEES $\Delta = -0.36$ (95% CI [-1.72, 0.99]；8/15 改善)和 95.45% Coverage $\Delta = +0.22$ 个百分点(95% CI [-5.63, 6.06]；7/15 改善)均跨越 0。因此，其样本均值在部分位置和关联指标上呈有限改善，而跨 seed 证据较弱；mean NEES 和 Coverage 仍明显偏离理论目标(约 3%和 95.45%)，不能认为统计一致性已经得到充分校准。

6.5. 代表性空间可视化分析

图 4 仅展示固定种子下的代表性空间结果，用于辅助解释关联与融合现象，不替代跨种子统计。
图 4 上部对比 Euclidean/Mahalanobis 与 independent/correlated 的最终已确认地标分布，下部进一步比较同一真实对象在独立融合与相关降权模式下的 95% XY 边缘协方差椭圆面积。面积变化仅反映后验不确定性尺度，需与位置 RMSE、NEES 和覆盖率联合解释；相关降权在当前实现中的合理性由 NEES 和覆盖率的配对改善共同支持，不能据此单独声称完全校准。

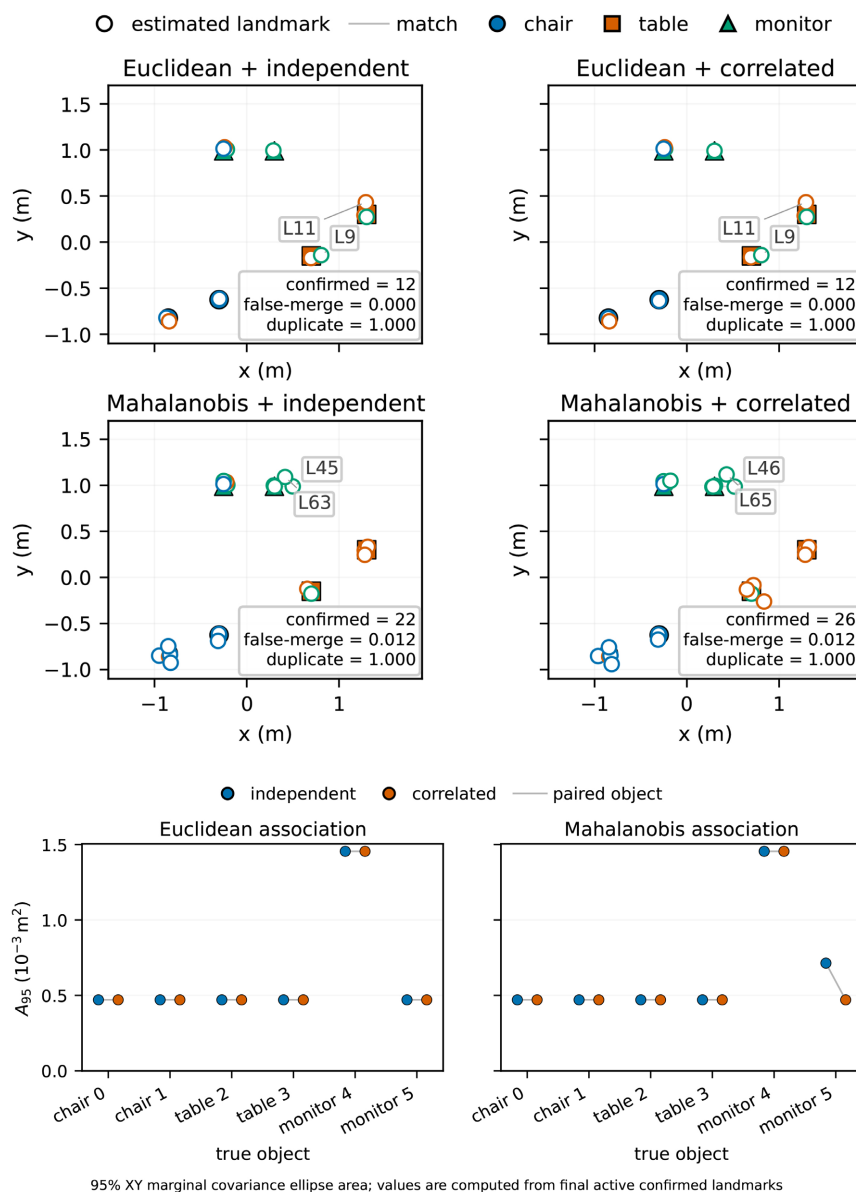


Figure 4. Representative spatial results for moderate, seed 20261717: final landmark estimates and paired 95% XY marginal covariance ellipse areas

图 4. 中扰动、seed 20261717 下的代表性空间结果：最终地标估计与 95% XY 边缘协方差椭圆面积配对

7. 讨论

7.1. 相关观测降权的作用与残余过度自信

相关降权使 30/30 个配对的平均 NEES 下降、95.45%覆盖率提高，并使位置 RMSE 在 29/30 个配对中下降。式(14)降低时空邻近观测的信息贡献，这一实现与信息重复累加受到抑制的解释一致。图 4 下部的 95% XY 边缘协方差椭圆面积配对图与这一机制相符，但不构成完整相关噪声模型的验证。位置 RMSE 的配对方向表明，一致性改善并非仅表现为后验协方差增大，多数配对中的递推均值误差也有所降低。像素定位、相机内参、位姿不确定性、跨帧相关和代表点的视角依赖尚未显式建模，因此强扰动下仍存在高 NEES 和低覆盖率；相关降权只能视为当前受控条件下的工程近似，不是相关噪声滤波的完整模型。

7.2. 统计门控与生命周期机制的耦合

Mahalanobis 在部分配对中降低地标级错误合并比例, 同时降低 F1, 并在全部配对中增加 ID switch 和建图节点单帧更新耗时。这一现象可由当前门控与地标生命周期逻辑解释: 偏严格的门限可能拒绝部分真实匹配, 未匹配观测随后初始化新地标; 活动地标数量增加又可能扩大后续代价矩阵。门限敏感性实验中, 宽松门限提高 F1、降低 ID switch, 同时增加地标级错误合并比例; 该趋势与上述权衡解释一致。由于三个门限下目标级重复建图比例均为 1.0, 现有结果不能排除固定生命周期和缺少重激活、删除及重复合并机制的影响。上述解释与当前实现逻辑一致, 但本实验未对门控拒绝、新地标初始化和代价矩阵规模进行独立因果消融。

7.3. 适用边界与局限性

证据来自固定锚点合成观测的 ROS 2 节点级实验, 使用给定相机位姿以隔离评价目标级地标的确定性建模、关联与融合行为, 不包含真实检测器或真实 RGB-D 数据。代表点对应可见表面而非物体中心; 协方差尚未联合传播像素定位、相机内参、位姿及交叉相关误差; 门限、相关权重和生命周期规则采用固定设置。低扰动、中扰动与强扰动均为组合扰动档位, 其差异不能归因于单一噪声因素。

8. 结论

本文提出了一种面向室内目标级三维语义地标的的多不确定性融合方法。45 次补充消融表明, 鲁棒深度估计在 15 组配对实验中稳定降低 Depth RMSE、MAE 和 P95; 最终 Position RMSE 仅小幅变化, 配对 95% CI 跨越 0。各向异性协方差在部分位置和关联指标的样本均值上呈有限改善, 但 Position RMSE、Association F1、mean NEES 和 Coverage 的配对区间均跨越 0, 且 NEES 和 Coverage 仍明显偏离理论目标。主实验进一步表明, correlated fusion 在 30/30 个配对中降低 mean NEES、在 30/30 个配对中提高 95.45% Coverage, 并在 29/30 个配对中降低 Position RMSE; Mahalanobis 关联体现了错误合并抑制与关联连续性之间的权衡。上述结果表明, 所提出方法在当前固定锚点、合成观测和给定位姿的受控目标级语义建图条件下, 对深度观测质量、地标递推融合和关联行为具有相应作用。

基金项目

教育部“春晖计划”科研项目“基于深度学习的工业机器人多传感主体融合智能引导机理与关键技术研究”(HZKY20220122)。

参考文献

- [1] 黄泽霞, 邵春莉. 深度学习下的视觉 SLAM 综述[J]. 机器人, 2023, 45(6): 756-768.
- [2] 王金科, 左星星, 赵祥瑞, 吕佳俊, 刘勇. 多源融合 SLAM 的现状与挑战[J]. 中国图象图形学报, 2022, 27(2): 368-389.
- [3] 王柯赛, 姚锡凡, 黄宇, 等. 动态环境下的视觉 SLAM 研究评述[J]. 机器人, 2021, 43(6): 715-732.
- [4] 王梦瑶, 宋薇. 动态场景下基于自适应语义分割的 RGB-D SLAM 算法[J]. 机器人, 2023, 45(1): 16-27.
- [5] 牛珉玉, 黄宜庆. 基于动态耦合与空间数据关联的 RGB-D SLAM 算法[J]. 机器人, 2022, 44(3): 333-342.
- [6] Rosinol, A., Violette, A., Abate, M., Hughes, N., Chang, Y., Shi, J., et al. (2021) Kimera: From SLAM to Spatial Perception with 3D Dynamic Scene Graphs. *The International Journal of Robotics Research*, 40, 1510-1546. <https://doi.org/10.1177/02783649211056674>
- [7] 杨峻峰, 田国会, 邵旭阳. 基于在线信息体构建与维护的长期自治家庭服务机器人环境理解方法[J]. 机器人, 2025, 47(2): 269-282.
- [8] 张健, 陈烨恒, 朱世强, 等. 一种基于邻域度量关系的 RGB-D 融合语义分割算法[J]. 机器人, 2023, 45(2): 156-165.

- [9] Hughes, N., Chang, Y. and Carlone, L. (2022) Hydra: A Real-Time Spatial Perception System for 3D Scene Graph Construction and Optimization. *Proceedings of Robotics: Science and Systems XVIII*, New York, 27 June-1 July 2022, 1-13. <https://doi.org/10.15607/rss.2022.xviii.050>
- [10] 王光明, 宋亮, 沈玥伶, 等. 基于目标检测和场景流估计联合优化的 3D 多目标跟踪[J]. 机器人, 2024, 46(5): 554-561.
- [11] 张玮奇, 王嘉, 张琳, 等. SUI-SLAM: 一种面向室内动态环境的融合语义和不确定度的视觉 SLAM 方法[J]. 机器人, 2024, 46(6): 732-742.
- [12] Gu, Q., Kuwajerwala, A., Morin, S., Jatavallabhula, K.M., Sen, B., Agarwal, A., *et al.* (2024) ConceptGraphs: Open-Vocabulary 3D Scene Graphs for Perception and Planning. 2024 *IEEE International Conference on Robotics and Automation (ICRA)*, Yokohama, 13-17 May 2024, 5021-5028. <https://doi.org/10.1109/icra57147.2024.10610243>
- [13] 曾庆化, 罗怡雪, 孙克诚, 李一能, 刘建业. 视觉及其融合惯性的 SLAM 技术发展综述[J]. 南京航空航天大学学报, 2022, 54(6): 1007-1020.