

弱纹理场景下结合表面法向量约束的单目SLAM位姿优化算法研究

张磊¹, 魏国亮^{2*}, 简晓富¹

¹上海理工大学光电信息与计算机工程学院, 上海

²上海理工大学管理学院, 上海

收稿日期: 2025年9月10日; 录用日期: 2025年9月18日; 发布日期: 2025年10月27日

摘要

为解决传统单目SLAM系统中特征点法在弱纹理场景下因有效特征点不足, 导致局部位姿优化阶段过度依赖点的重投影误差约束而出现精度不足的问题, 本研究提出了一种结合深度学习模型的鲁棒的单目SLAM系统框架。在预处理阶段, 利用深度学习模型提取输入图像帧稠密表面法向量, 通过聚类算法获取帧内代表方向。随后基于相邻帧代表方向之间的匹配关系推导出两种新的几何约束, 将其作为额外的图优化因子融入到后续的局部位姿优化过程中, 以弥补局部位姿优化中仅依赖点重投影误差约束的不足。在TUM数据集和ICL NUIM数据集上的测试结果表明, 所提出的系统在弱纹理场景中的位姿估计精度与运行鲁棒性均得到有效提升。

关键词

单目SLAM, 弱纹理场景, 深度学习, 表面法向量

Research on Monocular SLAM Pose Optimization Algorithm Integrating Surface Normal Constraints in Weak-Texture Scenes

Lei Zhang¹, Guoliang Wei^{2*}, Xiaofu Jian¹

¹School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

²Business School, University of Shanghai for Science and Technology, Shanghai

Received: September 10, 2025; accepted: September 18, 2025; published: October 27, 2025

*通讯作者。

文章引用: 张磊, 魏国亮, 简晓富. 弱纹理场景下结合表面法向量约束的单目 SLAM 位姿优化算法研究[J]. 建模与仿真, 2025, 14(10): 329-337. DOI: 10.12677/mos.2025.1410627

Abstract

To address the problem that the feature-point-based method in traditional monocular SLAM systems has insufficient accuracy in the local pose optimization stage due to over-reliance on the reprojection error constraint caused by insufficient effective feature points in weak texture scenes, this study proposes a robust monocular SLAM system framework combined with a deep learning model. In the preprocessing stage, a deep learning model is used to extract dense surface normal vectors of input image frames, and a clustering algorithm is adopted to obtain representative directions within the frames. Subsequently, two new geometric constraints are derived based on the matching relationship between the representative directions of adjacent frames, and they are incorporated into the subsequent local pose optimization process as additional graph optimization factors to make up for the deficiency of only relying on the reprojection error constraint in local pose optimization. The test results on the TUM dataset and ICL NUIM dataset demonstrate that the proposed system effectively enhances both the pose estimation accuracy and running robustness in weak texture scenes.

Keywords

Monocular SLAM, Weak Texture Scenes, Deep Learning, Surface Normal

Copyright © 2025 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

视觉同步定位与建图技术作为机器人自主导航和增强现实应用的核心技术,近年来取得了显著进展。在众多 SLAM [1] 应用场景中,单目相机因其硬件配置简单、成本低廉成为室内环境的主流传感器。目前主流的单目 SLAM 系统主要遵循特征点法和直接法这两大技术路线。ORB-SLAM [2] 系列作为特征点法的代表,通过高效的 ORB 特征提取与匹配机制,结合基于词袋模型的回环检测,在纹理丰富的场景中可实现高精度的实时定位。直接法 SLAM [3] 作为另一主流技术路线,通过直接利用图像像素的光度信息构建最小化光度误差的优化问题,避免了对特征提取与匹配的依赖性,从而适用于更广泛的场景。根据使用像素数量的不同,直接法可进一步分为稀疏法[4]、半稠密法[5]和稠密法[6]三类。然而,室内环境普遍存在的弱纹理区域,如墙面、地面等,给现有方法带来了严峻挑战。对于特征点法而言,点的重投影误差约束是局部位姿优化的核心,当弱纹理区域有效特征点数量不足时,约束信息的匮乏将直接导致位姿估计精度下降,进而引发跟踪漂移甚至跟踪失败,严重影响了系统的精度和鲁棒性。直接法则受限于光度一致性假设,在光照突变或存在反射表面的场景中仍难以避免较大漂移。上述技术路线因在复杂场景下适应性不足,急需新的技术手段突破瓶颈。

近年来,深度学习技术[7]凭借强大的特征学习与推理能力,为突破传统单目 SLAM 的技术瓶颈提供了新路径。在单目场景中,深度学习模型可有效弥补三维信息缺失的缺陷:单目深度估计模型[8]通过学习图像外观与深度值的映射关系,能为 SLAM 系统提供额外深度约束;而表面法向量作为描述物体表面朝向的核心几何属性,不仅能反映局部结构的姿态,更可通过正交性、平行性等约束关系提供全局结构线索,且对尺度变化具有更强鲁棒性。现有表面法向量估计模型[9]已实现从单张 RGB 图像直接输出逐像素法向量,为构建新型几何约束奠定了基础。

针对传统单目 SLAM 中特征点法在弱纹理场景下因有效特征点不足, 导致局部位姿优化阶段约束信息匮乏的核心问题, 本文提出一种融合深度学习表面法向量约束的鲁棒单目 SLAM 系统。通过引入深度学习模型提取稠密表面法向量并构建几何约束, 补充局部位姿优化过程中的约束信息, 旨在提升系统在弱纹理场景下的位姿估计精度与运行鲁棒性。

2. 系统框架

本研究旨在构建一个适用于弱纹理场景的单目 SLAM 系统。系统框架如图 1 所示。本研究以 ORB-SLAM2 为基础, 引入了深度学习模型 DSINE [10]提取输入图像帧的稠密表面法向量, 通过聚类算法提炼出图像中具有代表性的主要方向, 称为主导方向, 作为全局结构特征的简化表达。系统完成初始化后, 借助 ORB-SLAM2 原有跟踪线程生成的初始位姿估计值, 建立相邻帧主导方向之间的匹配关系, 进而推导出两种新的几何约束作为图优化因子, 融入到后续的局部位姿图优化过程中。有效弥补了局部位姿优化过程中仅依赖点重投影误差约束的不足, 显著提升了系统的鲁棒性和精度。

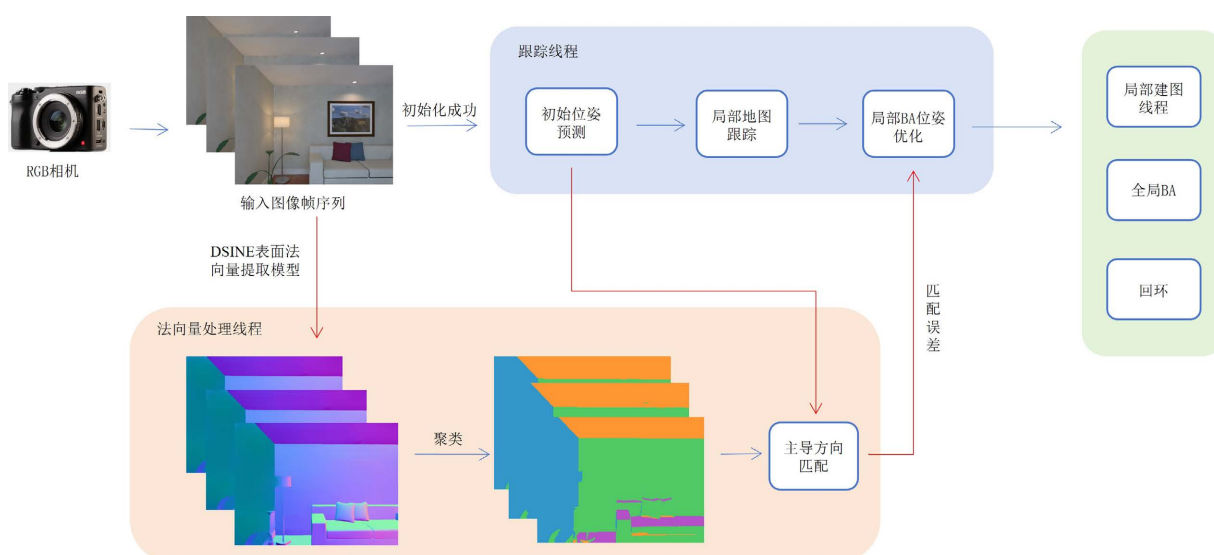


Figure 1. System framework

图 1. 系统框架

2.1. 法向量处理线程

2.1.1. 图像帧预处理

当输入图像帧传入跟踪线程和法向量处理线程后, 并行进行处理。其中稠密表面法向量提取利用的是深度学习模型 DSINE。DSINE 通过融入像素级光线方向约束与相邻像素法线的相对旋转建模这两种几何归纳偏差, 并结合光线方向激活函数, 实现了高精度表面法向量提取, 既能精准捕捉局部细节, 又能保证物理连续性与边缘清晰度。该方法展现出强大的泛化能力, 即便面对任意分辨率、宽高比的复杂图像, 仍能生成细节丰富的预测结果。并且尽管 DSINE 训练数据集规模仅为同类方法的十分之一, 但在定量与定性指标上均超越了近期基于视觉 Transformer 的最先进模型[11]。

因 SLAM 系统需要兼顾实时性, 无法直接处理单帧图像中数十万级像素的稠密法向量, 一张 640×480 分辨率图像含 307,200 个法向量。因此, 本研究引入 MeanShift 聚类算法[12], 通过聚类算法提炼出图像中具有代表性的主要方向, 称为主导方向, 为后续位姿估计构建简洁约束。具体流程如下: 任意一

个法向量都可以被视为单位球面上的一个点，而传统的欧氏距离无法准确描述其在球面几何空间中的分布特性。为此，通过利用点积结果等于夹角余弦值的特性，度量两个法向量在球面上的朝向差异。公式如下：

$$\theta(u, v) = \arccos([u \cdot v]_{-1}^1) \quad (1)$$

其中， $u \cdot v$ 表示两向量的点积， $[x]_a^b$ 为区间截断算子。当 $x < a$ 时取 a ， $x > b$ 时取 b ，否则保留 x 。

通过截断算子确保结果落在余弦函数的定义域内，该公式直接输出两法向量在单位球面上的夹角 θ ，精准度量方向差异，为后续聚类中邻域筛选、核加权平均等操作提供符合球面几何特性的核心度量标准。对于单个法向量 P ，需在全体法向量集合中执行迭代漂移，挖掘局部区域内占主导地位的方向，具体步骤以球面夹角公式为核心依据展开。首先是邻域筛选，计算法向量 P 与所有其他法向量的球面夹角：

$$\theta_i = \theta(p, n_i) \quad (2)$$

其中， n_i 表示图像帧其余法向量，筛选出满足 $\theta_i < \theta_b$ 这个局部邻域范围的法向量集合， θ_b 是控制局部范围的超参数，它决定了聚类时会把哪些方向的法向量视为相近，进而影响聚类结果的紧密程度和鲁棒性。

这一步让聚类聚焦于与 p 方向相近的法向量，提取场景局部几何特征。其次是核加权平均，对邻域法向量，通过高斯核函数赋予权重，公式为：

$$W = \exp\left(-\frac{\theta_i^2}{\theta_b^2}\right) \quad (3)$$

距离 p 越近， θ_i 越小的法向量，权重越高，在聚类中心更新中占据主导地位。基于这些权重对邻域法向量加权求和后，再经归一化得到新的聚类中心：

$$N_c = \frac{\sum_i weights_i \cdot n_i}{\left| \sum_i weights_i \cdot n_i \right|} \quad (4)$$

这一步通过强化密集分布区域的法向量影响，推动聚类中心向法向量最集中的方向移动，自然聚焦于场景中局部占优的几何趋势，为后续位姿估计提炼出最具代表性的方向线索。最后计算新聚类中心 N_c 与原中心 p 的球面夹角 $\theta_{new} = \theta(p, N_c)$ ，若 $\theta_{new} < E_{ps}$ ，则停止迭代，反之，以 θ_{new} 作为新的中心，继续执行迭代操作，直至满足收敛条件。 E_{ps} 是判定聚类中心是否收敛的阈值，用于表示聚类中心的最小可接受变化量。通过上述步骤，最终可以得到稳定可靠的聚类中心，能够精准反映局部区域内法向量的主导方向。

2.1.2. 主导方向匹配

通过上述步骤得到主导方向后，因传统 ORB-SLAM2 系统中的跟踪线程会提供当前帧初始位姿估计值，故可以基于这个初始位姿建立相邻帧主导方向之间的匹配关系。世界坐标系到上一帧相机坐标系的位姿变化记为 T_{pw} ，世界坐标系到当前帧相机坐标系的位姿变换记为 T_{cw} 。则上一帧到当前帧的相对位姿变化为：

$$T_{cp} = T_{cw} \cdot T_{pw}^{-1} \quad (5)$$

其旋转分量 R_{cp} 描述两帧间的相对旋转关系。对于上一帧主导方向集合 $\{n_p^k | k=1, 2, \dots, n\}$ ，通过 R_{cp} 转换至当前帧坐标系，转换关系如下：

$$n_p^{kc} = R_{cp} \cdot n_p^k \quad (6)$$

对于当前帧主导方向集合 $\{n_c^m | m=1,2,\dots,n\}$ ，计算转换后的主导方向和当前帧主导方向的球面夹角：

$$\theta_{km} = \arccos\left(\left[n_p^{kc} \cdot n_c^m\right]_{-1}^1\right) \quad (7)$$

若 $\theta_{km} < \tau$ ， τ 为匹配阈值，则判定 n_p^k 与 n_c^m 为匹配对。

2.2. 多约合优化框架

在 ORB-SLAM2 系统中，Motion-only BA 是跟踪线程中的核心步骤，用于经跟踪线程得到的当前帧位姿估计值的优化，保证系统跟踪的连续性。但是在弱纹理场景下，ORB 特征稀缺，点重投影误差边数量不足，位姿优化约束薄弱，从而导致初始位姿估计不准确引发的跟踪漂移甚至丢失。为增强 SLAM 系统中位姿优化的鲁棒性与精度，利用提取的主导方向信息，构建出新的平滑性约束与一致性约束，作为传统点特征重投影误差约束的补充。平滑性约束与一致性约束在约束维度、度量逻辑上存在明显差异，却又在保障位姿优化鲁棒性的层面形成深度互补。平滑性约束以刚体运动连续性为理论基础，通过欧氏距离平方量化相邻帧间同一几何结构主导方向经旋转后的空间向量差异，聚焦于相邻帧位姿在时间序列上的平滑过渡，弥补缺纹理下 ORB 特征稀缺导致的相邻帧局部约束不足，避免跟踪线程位姿优化出现帧间跳跃；一致性约束则基于跨视角几何方向一致性推导，借助单位向量点积与余弦夹角的关系，以余弦相似度平方度量全局坐标系下相邻视角主导方向的角度偏差，致力于多视角下全局几何结构的稳定锚定，解决弱纹理场景中全局约束薄弱引发的长期跟踪漂移。二者从局部到全局、时间到空间维度形成有机互补，既保障了相邻帧位姿优化的连续性，又强化了全局位姿估计的稳定性，有效提升了弱纹理环境下 SLAM 系统位姿优化的鲁棒性与精度。以下是基于刚体运动连续性与跨视角几何方向一致性推导约束的理论依据与误差函数形式。

2.2.1. 约束构建

首先是平滑性约束，根据刚体运动连续性假设：短时间间隔内，相邻帧观测到的同一几何结构，如墙面、地面，其主导方向经帧间旋转后应保持一致。设匹配主导方向对为 $\{n_{ci}, n_{pi}\}$ ， n_{ci} 为当前帧主导方向， n_{pi} 为上一帧主导方向，平滑性约束误差构建函数如下：

$$\mathcal{E}_{\text{smooth}} = \sum_{i,j \in x} \rho\left(\|R_{cp} n_{pj} - n_{ci}\|^2\right) \quad (8)$$

其中， $\|\cdot\|^2$ 为欧氏距离平方，表示度量旋转后方向与当前帧方向的差异， $\rho(\cdot)$ 为鲁棒核函数，用于抑制异常值，如误匹配主导方向对优化的干扰。

该式通过最小化上一帧主导方向经旋转后，与当前帧主导方向的欧氏距离，强制相邻帧局部几何结构平滑变化，符合刚体运动连续特性。其次是相邻视角方向一致性推导出的一致性约束。对于相邻视角的同一几何结构，其主导方向在全局坐标系下保持方向一致性。因主导方向是由图像帧稠密表面法向量聚类而来，而法向量在本研究中都归一化为单位向量，所以主导方向都是单位向量，采用余弦相似度度量方向差异，误差函数如下：

$$\mathcal{E}_{\text{consist}} = \sum_{i,j \in x} \rho\left(\left(1 - (R_{cq} n_{pj})^T n_{ci}\right)^2\right) \quad (9)$$

因单位向量的点积满足：

$$a^T b = \cos \theta \quad (10)$$

因此 $1 - (R_{cq} n_{pj})^T n_{ci}$ 直接反映相邻帧主导方向匹配对的方向差异，通过最小化该值的平方，使得相邻

视角主导方向匹配对保持一致性，从而保障全局几何结构的稳定性。

2.2.2. Motion-only BA 阶段因子图重构

ORB-SLAM2 中，Motion-only BA 的因子图结构单一，仅依赖 ORB 点特征的重投影误差边，利用相机模型最小化 3D 地图点 X_i 经当前帧位姿 $\{R, t\}$ 投影后的像素坐标 $\pi(RX_i + t)$ 与实际观测特征点 u_i 的差异约束位姿：

$$\mathcal{E}_{\text{proj}} = \sum_i \rho(\|\pi(RX_i + t) - u_i\|^2) \tag{11}$$

其中 $\pi(\cdot)$ 为相机投影函数。本研究通过因子图重构，新增上述两类几何约束边，通过多源约束协同强化位姿优化，解决对单一特征的依赖。点重投影误差边、平滑性约束边和一致性约束边三者通过加权融合形成多约束联合优化目标：

$$\mathcal{E}_{\text{total}} = \lambda_1 \mathcal{E}_{\text{proj}} + \lambda_2 \mathcal{E}_{\text{smooth}} + \lambda_3 \mathcal{E}_{\text{consist}} \tag{12}$$

其中 λ_1 、 λ_2 、 λ_3 为权重参数。根据不同约束的物理意义与场景适配性，自适应分配不同的权重。在纹理丰富场景，点特征匹配稳定， λ_1 会占据主导， λ_2 、 λ_3 则相应减小，同理在弱纹理场景下， λ_2 、 λ_3 的值则会提升。该设计在 Motion-only BA 位姿优化环节，将点特征的精确性与主导方向的几何一致性协同，填补弱纹理场景的约束空白，实现从单一特征依赖到多源约束协同的升级，直接增强位姿估计鲁棒性，为后续跟踪与建图提供更可靠基础。

3. 实验与分析

3.1. 实验条件

本节将在 TUM RGB-D 数据集[13]与 ICL NUIM[14]数据集上验证所提出系统的性能，并与主流单目 SLAM 算法进行比较分析，包括 ORB-SLAM2、CNN-SLAM、LSD-SLAM、GeoNet、Structure-SLAM。本文实验硬件配置 CPU 为 AMD R9-7945HX，16 核心 32 线程，GPU 为 GeForce RTX 4060，内存为 16 g，系统为 ubuntu 20.04。

3.2. TUM RGB-D 数据集实验

Table 1. Translation root mean square error (RMSE) results of the monocular SLAM algorithm on the TUM RGB-D dataset (unit: m)

表 1. 单目 SLAM 算法在 TUM RGB-D 数据集上的平移均方根误差(RMSE)结果(单位: m)

单目 SLAM 算法	fr1-desk	fr1-desk2	fr1-xyz	s-t-far	s-t-near
LSD-SLAM	0.158	0.096	0.164	0.214	-
CNN-SLAM	0.056	0.058	0.062	0.037	-
GeoNet	0.048	0.068	0.068	0.024	0.107
Structure-SLAM	-	0.124	-	0.026	0.015
ORB-SLAM2	0.046	0.056	0.056	0.024	0.032
Ours	0.029	0.037	0.052	0.020	0.020

TUM RGB-D 数据集是由德国慕尼黑工业大学发布的大型数据集，被广泛用于测试 SLAM 算法在不同场景下的鲁棒性与精度。fr1-desk 和 fr1-desk2 这两个序列属于手持 SLAM 场景，以办公室环境中的办公桌为核心观测对象，用于测试算法在有结构有纹理场景下的表现。fr1-xyz 序列常用于初始调试和基础

位姿估计验证。s-t-far 和 s-t-near 这两个序列属于 TUM RGB-D 中 Structure vs. Texture 的核心测试集，专门用于评估算法在结构-纹理差异场景下的性能。这些序列通过结构和纹理相组合以及不同观测距离的设计，为 SLAM 算法提供了多层次的测试场景，可系统性验证 SLAM 系统在不同场景下的表现。测试结果如表 1 所示。

表 1 以平移均方根误差为评价指标，对单目 SLAM 算法的平移跟踪精度开展测试。由于单目 SLAM 存在真实尺度未知的问题，因此为确保 RMSE 结果能真实反映算法位姿预测精度，实验中对 ORB-SLAM2、LSD-SLAM 等对比算法以及本文算法的估计轨迹，均采用相似变换的方法[2]，与数据集提供的地面真值轨迹进行对齐，通过同步实现轨迹的尺度缩放、旋转与平移调整，消除尺度不确定性带来的系统性误差。结果显示，在 TUM RGB-D 数据集的 fr1-desk、fr1-desk2、fr1-xyz、s-t-far 及 s-t-near 序列中，本文算法性能均显著更优；-表示无法获得有效的实验数据。其中在具有结构纹理的场景下，fr1-desk 序列中本文算法 RMSE 较 LSD-SLAM 降低 81.6%、较 ORB-SLAM2 降低 36.9%，同类型场景的 fr1-desk2 序列中其 RMSE 较 LSD-SLAM 降低 61.5%、ORB-SLAM2 降低 33.9%；在基础位姿测试场景的 fr1-xyz 序列中，本文算法 RMSE 较 ORB-SLAM2 降低 7.1%，而在结构纹理存在差异的场景下，s-t-far 序列中本文算法 RMSE 较 ORB-SLAM2 降低 16.7%，s-t-near 序列中其 RMSE 较 ORB-SLAM2 降低 37.5%。尽管 Structure-SLAM 在 s-t-near 序列的 RMSE 稍低于本文算法，但本文算法在全部测试序列中始终保持精度的稳定性与优越性，充分彰显其在多场景下的强适应性与鲁棒性。

3.3. ICL NUIM 数据集实验

ICL NUIM 数据集广泛用于测试 SLAM 算法在动态干扰、光照变化及结构纹理差异场景下的鲁棒性与精度。其中，lr-kt1、lr-kt2、lr-kt3 三个序列同属客厅场景测试集，均以客厅为核心观测对象，涵盖常规有结构有纹理环境、基础结构稳定且含局部动态干扰与光照变化的场景，以及有结构但低纹理且伴随相机快速运动的场景，共同用于验证算法在不同客厅场景条件下的位姿估计、建图精度与特征跟踪能力。of-kt1、of-kt2 两个序列则为办公室场景测试集，以复杂办公室环境为核心观测对象，用于评估算法在这类复杂办公环境中的适应性与抗干扰性能。这些序列通过结构纹理的密度调整、动态干扰梯度设计及不同观测距离的搭配，为 SLAM 算法提供了丰富的测试场景。本研究算法及近年经典单目 SLAM 系统算法在 ICL NUIM 数据集的测试结果如表 2 所示：

Table 2. Translation root mean square error (RMSE) results of the monocular SLAM algorithm on the ICL NUIM dataset (unit: m)

表 2. 单目 SLAM 算法在 ICL NUIM 数据集上的平移均方根误差(RMSE)结果(单位: m)

单目 SLAM 算法	lr-kt1	lr-kt2	lr-kt3	of-kt1	of-kt2
LSD-SLAM	0.059	0.323	-	0.157	0.213
CNN-SLAM	0.540	0.211	-	0.790	0.172
GeoNet	0.040	0.036	0.100	0.050	0.040
Structure-SLAM	0.026	0.045	0.035	0.045	0.030
ORB-SLAM2	0.024	0.061	0.046	0.569	0.031
Ours	0.019	0.032	0.033	0.128	0.026

表 2 展示了各单目 SLAM 算法在 ICL NUIM 数据集不同序列下的平移均方根误差。整体来看，本文算法的精度在多数序列中优于其他对比算法：在 lr-kt1 序列中，其 RMSE 较 Structure-SLAM 降低了

26.9%, 较 ORB-SLAM2 降低了 20.8%; lr-kt2 序列下, 较 Structure-SLAM 降低了 28.9%, 较 ORB-SLAM2 降低了 47.5%; lr-kt3 序列中, 较 Structure-SLAM 降低了 5.7%, 较 ORB-SLAM2 降低约 28.3%; of-kt1 序列下, 较 LSD-SLAM 降低了 18.5%; of-kt2 序列下, 较 Structure-SLAM 降低 13.3%, 较 ORB-SLAM2 降低了 16.1%。综上, 本文算法在 ICL NUIM 数据集的各测试序列中, 平移跟踪精度普遍优于 LSD-SLAM、CNN-SLAM、GeoNet 等算法, 对 Structure-SLAM、ORB-SLAM2 也展现出不同程度的精度优势。

4. 结论

本研究提出了一种结合深度学习模型的鲁棒的单目 SLAM 系统框架, 通过深度学习模型提取图像帧的稠密表面法向量, 经聚类得到帧内代表方向; 进一步基于相邻帧代表方向的匹配关系推导新型几何约束, 并将其作为额外优化因子融入局部位姿图优化过程, 有效弥补了仅依赖点重投影误差的约束缺陷。在 TUM RGB-D 数据集与 ICL NUIM 数据集上的测试结果表明, 所提系统显著提升了弱纹理场景下位姿估计的精度与运行鲁棒性, 为解决单目 SLAM 在弱纹理环境中的性能瓶颈提供了有效途径。未来可进一步探索该框架与多传感器的融合方案, 或针对动态场景开展适应性改进, 以拓展其应用范围。

参考文献

- [1] Abaspor Kazerouni, I., Fitzgerald, L., Dooly, G. and Toal, D. (2022) A Survey of State-of-the-Art on Visual Slam. *Expert Systems with Applications*, **205**, Article ID: 117734. <https://doi.org/10.1016/j.eswa.2022.117734>
- [2] Mur-Artal, R., Montiel, J.M.M. and Tardos, J.D. (2015) ORB-SLAM: A Versatile and Accurate Monocular SLAM System. *IEEE Transactions on Robotics*, **31**, 1147-1163. <https://doi.org/10.1109/tro.2015.2463671>
- [3] Engel, J., Schöps, T. and Cremers, D. (2014) LSD-SLAM: Large-Scale Direct Monocular SLAM. *Proceedings of the 13th European Conference on Computer Vision*, Zurich, 6-12 September 2014, 834-849. https://doi.org/10.1007/978-3-319-10605-2_54
- [4] Wang, R., Schworer, M. and Cremers, D. (2017) Stereo DSO: Large-Scale Direct Sparse Visual Odometry with Stereo Cameras. 2017 *IEEE International Conference on Computer Vision (ICCV)*, Venice, 22-29 October 2017, 3923-3931. <https://doi.org/10.1109/iccv.2017.421>
- [5] Forster, C., Zhang, Z., Gassner, M., Werlberger, M. and Scaramuzza, D. (2017) SVO: Semidirect Visual Odometry for Monocular and Multicamera Systems. *IEEE Transactions on Robotics*, **33**, 249-265. <https://doi.org/10.1109/tro.2016.2623335>
- [6] Newcombe, R.A., Lovegrove, S.J. and Davison, A.J. (2011) DTAM: Dense Tracking and Mapping in Real-Time. 2011 *International Conference on Computer Vision*, Barcelona, 6-13 November 2011, 2320-2327. <https://doi.org/10.1109/iccv.2011.6126513>
- [7] Liu, J., Zhang, T. and Sun, S. (2024) Review of Deep Learning Algorithms in Molecular Simulations and Perspective Applications on Petroleum Engineering. *Geoscience Frontiers*, **15**, Article ID: 101735. <https://doi.org/10.1016/j.gsf.2023.101735>
- [8] Yang, L., Kang, B., Huang, Z., Xu, X., Feng, J. and Zhao, H. (2024) Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 10371-10381. <https://doi.org/10.1109/cvpr52733.2024.00987>
- [9] Hu, M., Yin, W., Zhang, C., Cai, Z., Long, X., Chen, H., et al. (2024) Metric3D v2: A Versatile Monocular Geometric Foundation Model for Zero-Shot Metric Depth and Surface Normal Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **46**, 10579-10596. <https://doi.org/10.1109/tpami.2024.3444912>
- [10] Bae, G. and Davison, A.J. (2024) Rethinking Inductive Biases for Surface Normal Estimation. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 9535-9545. <https://doi.org/10.1109/cvpr52733.2024.00911>
- [11] Kar, O.F., Yeo, T., Atanov, A. and Zamir, A. (2022) 3D Common Corruptions and Data Augmentation. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 18941-18952. <https://doi.org/10.1109/cvpr52688.2022.01839>
- [12] Choi, S., Kang, D. and Cho, M. (2024) Contrastive Mean-Shift Learning for Generalized Category Discovery. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 23094-23104.

-
- <https://doi.org/10.1109/cvpr52733.2024.02179>
- [13] Firman, M. (2016) RGBD Datasets: Past, Present and Future. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Las Vegas, 26 June-1 July 2016, 661-673.
<https://doi.org/10.1109/cvprw.2016.88>
- [14] Handa, A., Whelan, T., McDonald, J. and Davison, A.J. (2014) A Benchmark for RGB-D Visual Odometry, 3D Reconstruction and SLAM. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Hong Kong, 31 May-7 June 2014, 1524-1531.