

<https://doi.org/10.3799/dqkx.2026.057>



# 野外岩石取样机器人多模态融合环境感知方法

何佳琪<sup>1,3</sup>, 王 猛<sup>2</sup>, 程 春<sup>2</sup>, 曹卫华<sup>1,3</sup>, 黎育朋<sup>1,3</sup>, 张之武<sup>4</sup>, 陈泳桥<sup>2</sup>, 李 岩<sup>2</sup>, 王英超<sup>2</sup>

1. 中国地质大学(武汉)人工智能与自动化学院, 湖北武汉 430074

2. 中国冶金地质总局矿产资源研究院, 北京 101300

3. 复杂系统先进控制与智能自动化湖北省重点实验室, 湖北武汉 430074

4. 中国冶金地质总局, 北京 100025

**摘 要:** 岩石取样机器人是地质勘探智能化的核心装备, 但机器人在实际作业中面临着野外非结构化环境、剧烈光照波动以及岩性复杂多变等挑战。因此, 结合自研的野外岩石取样机器人平台, 提出了一种多模态融合的野外环境感知方法, 实现从野外岩石区域的高精度分割和取样位姿的精准解算。针对野外地形分布无规律的问题, 采用多传感器融合的环境建图方法, 构建几何结构准确且纹理色彩丰富的三维地图。在此地图基础上, 融合点云的几何结构与视觉纹理信息, 设计了多模态协同约束的无监督岩石区域分割算法, 利用物理先验有效克服了野外光照条件变化剧烈的干扰, 并解决了岩性动态多变导致的标注数据匮乏的问题。结合机器人取样作业的执行约束, 提出一种取样点位姿自动生成策略, 通过岩石区域的局部表面特征分析与机器人运动学约束校正, 选取最优岩石取样点, 并实现取样点从环境空间到机器人作业空间的精准映射。在野外实测中, 岩石分割精度达 89.57%, 较传统区域生长法有显著提升; 且取样点位姿估计位置误差为 0.696 cm, 法向量误差为 1.44°, 满足自动化岩石取样的精度需求。

**关键词:** 岩石取样机器人; 环境感知; 岩石分割; 取样位姿估计; 计算机视觉。

中图分类号: TD41

文章编号: 1000-2383(2026)08-2927-13

收稿日期: 2026-01-13

## Multimodal Fusion Perception for Rock Sampling Robot in Field Environments

He Jiaqi<sup>1,3</sup>, Wang Meng<sup>2</sup>, Chen Chun<sup>2</sup>, Cao Weihua<sup>1,3</sup>, Li Yupeng<sup>1,3</sup>,  
Zhang Zhiwu<sup>4</sup>, Chen Yongqiao<sup>2</sup>, Li Yan<sup>2</sup>, Wang Yingchao<sup>2</sup>

1. School of Artificial Intelligence and Automation, China University of Geosciences (Wuhan), Wuhan 430070, China

2. Institute of Mineral Resources Research, China Metallurgical Geology Bureau, Beijing 101300, China

3. Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems, Wuhan 430074, China

4. China Metallurgical Geology Bureau, Beijing 100025, China

**Abstract:** Rock-sampling robots are essential equipment for the intelligence of field geological exploration. However, in actual operations, robots face severe challenges such as unstructured field environments, intense illumination fluctuations, and highly variable rock lithology. Therefore, relying on a self-developed field rock-sampling robot platform, this paper proposes a

**基金项目:** 中国冶金地质总局未来勘探系统一期工程——踏勘机器人项目(No. CMGBKYW202504).

**作者简介:** 何佳琪(2001—), 女, 博士研究生, 研究方向为野外地质踏勘机器人的环境智能感知与行为优化. ORCID:0009-0001-1259-3160. E-mail:jjiaqihe@cug.edu.cn

\* **通讯作者:** 王猛, ORCID:0009-0009-7389-7027. E-mail:13759975467@163.com

**引用格式:** 何佳琪, 王猛, 程春, 曹卫华, 黎育朋, 张之武, 陈泳桥, 李岩, 王英超, 2026. 野外岩石取样机器人多模态融合环境感知方法. 地球科学, 51(8):2927-2939.

**Citation:** He Jiaqi, Wang Meng, Chen Chun, Cao Weihua, Li Yupeng, Zhang Zhiwu, Chen Yongqiao, Li Yan, Wang Yingchao, 2026. Multimodal Fusion Perception for Rock Sampling Robot in Field Environments. *Earth Science*, 51(8):2927-2939.

multimodal fusion perception method for field environments to achieve high-precision segmentation of field rock regions and precise calculation of sampling poses. Addressing the issue of irregular terrain distribution, a multi-sensor fusion mapping method is employed to construct a 3D map with accurate geometric structures and rich texture colors. Based on this map, by integrating the geometric structures of point clouds with visual texture information, a multimodal collaborative constrained unsupervised rock segmentation algorithm is designed. This algorithm leverages physical priors to effectively overcome interference from drastic light changes and resolves the problem of insufficient annotated data caused by dynamic lithological variations. Furthermore, in light of the execution constraints of robotic sampling, an automated sampling-pose generation strategy is proposed. Through the parallel processes of local surface feature analysis of the rock region and kinematic constraint correction of the robot, the optimal sampling point is selected, enabling a precise mapping of the sampling point from the environmental space to the robotic operational space. Field experiments demonstrate that the rock segmentation accuracy reaches 89.57%, a significant improvement over the traditional region-growing method. Furthermore, the sampling-pose estimation achieves a mean position error of 0.696 cm and a normal vector error of  $1.44^\circ$ , which fulfills the precision requirements for autonomous rock sampling in the field.

**Key words:** rock sampling robot; environmental perception; rock segmentation; sampling pose estimation; computer vision.

## 0 引言

固体矿产资源勘探是发现和评价地下重要矿产资源的基础工作,岩石取样是其中关键的环节之一.通过获取具有代表性的岩石样本,地质人员能够判断矿体属性、识别矿化迹象,并支撑后续的矿体推断与资源评价.近年来,随着资源开发逐步向偏远山区、复杂地貌等极端区域推进,传统依赖人工的取样方式暴露出诸多难以规避的局限.不仅作业效率低下、人力成本高昂,更面临地形险峻、气候恶劣带来的极高人身安全风险;同时,人工取样精度受主观经验影响较大,取样位置选择缺乏一致性和可重复性,难以满足现代地质工作对安全性、规范性以及智能化的迫切需求.

近年来,随着智能机器人与信息科学技术的发展,机器人成为替代人工作业的核心解决方案,其自主化、智能化水平直接决定取样任务的可行性与可靠性(姚晓峰等,2024).机器人替代人工作业实现自主取样的核心挑战在于机器人如何在复杂野外环境中准确区分岩石区域与非岩石区域,并进一步从不规则岩石表面选取最佳作业位姿.这要求机器人首先需要具备精确的环境感知能力,作业空间映射能力以及专用的取样装置.其中,高可靠的环境感知是机器人自主取样的前提与基石,负责构建周围环境地图,并从复杂的野外环境中精确识别成岩石区域;专用的取样装置则是机器人执行取样动作的硬件基础,提供必要的物理交互手段,帮助机器人进行岩石破碎与收集;作业空间映射能力则机器人执行取样的关键,即机器人可以自主在识别到的岩石区域选择满足作业条件的取样点,并生成对

应的取样位姿.

由于野外环境存在强非结构化特征与剧烈光照波动,使得稳定、精确的环境感知成为制约机器人自主作业的首要瓶颈.环境感知技术可以分为环境三维建图与语义分割两个部分(Xie *et al.*, 2025; Xu *et al.*, 2025),前者通过融合多传感器数据,还原野外场景的三维空间结构与外观特征,为机器人提供高精度的时空基准(Li *et al.*, 2021; Liu *et al.*, 2025);后者则基于三维地图实现岩石目标的精准识别(Bar *et al.*, 2020),明确作业对象的空间范围,是后续取样决策的认知基础.目前常用的环境三维建图方法根据传感器类型及其感知特性,可分为基于视觉、基于激光雷达以及多传感器融合3类(Cadena *et al.*, 2016; Qin *et al.*, 2018).基于视觉的方法利用相机捕捉丰富的纹理与颜色信息,无论是通过特征点匹配(Mur-Artal *et al.*, 2015)还是基于像素梯度的直接法(Engel *et al.*, 2017),在光照适宜且纹理丰富的室内场景中均表现优异.但在野外地质踏勘场景中,视觉感知面临严峻挑战,如强烈的阳光直射、树木投下的斑驳阴影、以及四足机器人运动带来的剧烈晃动与图像运动模糊,极易导致视觉特征追踪失败,引发定位漂移甚至建图中断(Campos *et al.*, 2021).相比之下,基于激光雷达的方法(Xu *et al.*, 2022; Liu *et al.*, 2024)利用激光测距的物理特性,具备极高的测量精度与抗光照干扰能力,在几何结构恢复方面具有先天优势(Zhang *et al.*, 2025).但激光雷达生成的点云本质上是稀疏的几何离散点,缺乏颜色与纹理等语义线索.另一方面,在植被、碎石坡与裸岩混合分布的复杂野外区域,单纯依赖几何特征难以有效区分具有相似空间

结构的障碍物,严重制约了后续的语义分析与取样操作。

由于单一模态在野外非结构化环境中的局限性,近年来多传感器融合的建图方法成为研究主流。根据融合深度的不同,可分为松耦合与紧耦合两类(Zhang and Singh, 2018; 郭云飞等, 2025)。松耦合系统(Zhang *et al.*, 2017)虽然结构简单,但各传感器独立估算状态(Forster *et al.*, 2016),忽略了传感器间的内在约束,在剧烈运动或特征退化区域容易引入累积误差(Shin *et al.*, 2020)。为了满足野外勘探对高精度与高鲁棒性的需求,紧耦合融合架构(Sun *et al.*, 2018)成为更优的选择。紧耦合方法在原始数据层面将激光雷达的几何约束、视觉的光度约束与IMU的惯性预积分进行联合优化(Lin *et al.*, 2021)。这种方式不仅有效弥补了激光雷达在无结构区域的退化问题,也修正了视觉传感器在剧烈运动下的尺度漂移。特别是直接法驱动的激光—视觉—惯性紧耦合系统,能够直接利用彩色图像像素为激光点云着色,在保证厘米级定位精度的同时,构建出几何结构准确且纹理色彩丰富的彩色点云地图。

然而,上述环境建图方法输出的本质上仍是带有颜色信息的几何离散点集,缺乏对环境内容的语义理解。机器人仅知“障碍物在哪里”,却无法识别“哪里是可取样的岩石”,这使得仅凭建图结果无法直接指导机器人执行岩石取样工作。

在构建三维环境地图的基础上,如何从复杂野外环境中可靠地分割出岩石区域,是实现自动化取样作业的关键步骤。现有语义分割方法大体可分为有监督与无监督两类(Pan *et al.*, 2020)。以深度学习为代表的有监督方法通过学习带标签数据构建判别模型实现精细识别其中深度学习方法凭借自动特征提取能力在结构化数据集上表现优异(Zhao *et al.*, 2023)。根据数据处理方式的不同,这些方法主要可分为基于原始点、基于体素以及基于投影3类(Hu *et al.*, 2021; Wang *et al.*, 2024)。为了克服单一模态语义分割的局限性,近年来的研究采用球面投影和传感器标定信息来实现多源信息的深度融合语义分割。El Madawi *et al.* (2019)通过将RGB图像转换为极坐标网格映射形式,实现了颜色信息与激光雷达深度信息的早期融合,显著增强了感知信息的利用率。Lin *et al.* (2025)针对传统球面投影方法在视觉映射中的畸变问题,提出基于透视投影

的融合策略,并结合主干网络保留了原始密集视觉信息。此外,针对多模态数据标注困难的问题,Yan *et al.* (2022)提出了一种名为2DPASS的方法,利用多尺度融合与知识蒸馏模块,在训练过程中将二维图像网络的先验知识迁移至三维网络,仅依赖点云标签即可实现高性能分割。

然而,尽管上述基于深度学习的有监督方法在公开数据集上取得了优异性能,但在面向野外岩石取样的实际应用中仍面临严峻挑战。核心问题在于,深度模型高度依赖大规模、高质量的标注数据来驱动特征学习。而野外地质场景岩性类别繁多、风化程度各异,且受季节更替与光照变化影响显著,构建覆盖跨区域、跨季节特性的稳定训练集成本极为高昂。一旦缺乏完备的标注数据,深度模型的泛化能力将显著下降,难以直接迁移至未知的踏勘环境中。相比之下,无监督方法(Pang *et al.*, 2024)不依赖大量人工标注,通过点云几何特征(Wang and Sun, 2021)或颜色信息(Yang *et al.*, 2022)等先验进行区域分割,如基于法向量、曲率、粗糙度、凸性、局部高差等规则(Gorai *et al.*, 2021; 朱合华等, 2023)。上述这类方法多依赖于物理意义明确的几何或外观特征,其可解释性强,对模型训练和标注数据的依赖性极低。然而,在野外实际环境中,单一模态的无监督方法存在明显局限。仅依赖几何特征难以区分平坦的岩石表面与硬质土路,而仅依赖视觉特征则极易受到树木阴影、动态光照的干扰,导致严重的误分割。因此,如何融合几何与视觉信息,在无监督条件下实现鲁棒的岩石分割,是当前亟待解决的问题。

此外,现有的感知研究多局限于环境的语义解析,即侧重于实现目标区域的分割与识别,却缺乏建立从环境语义空间到机器人构型空间的有效映射。由于野外岩石表面普遍存在显著的非结构化特征与复杂的几何形貌,单纯的语义分割结果无法直接指导取样作业。如何在已分割的岩石区域内,基于局部表面平整度与机械臂运动学约束,自适应筛选出满足表面法向量约束与机械臂逆运动学解算的可操作域,构建全自主岩石取样机器人系统,是实现从环境感知向自主作业转化过程中的关键。

尽管上述技术为岩石取样机器人的智能作业提供了理论基础,但在面向野外岩石取样的实际应用中仍存在挑战。一方面,现有的语义分割方法难以适应野外非结构化场景。有监督方法高度依赖大

规模标注数据,而野外地质场景岩性繁多、风化显著且跨域差异巨大,导致模型在缺乏完备训练集时泛化能力受限;无监督方法虽无需标注,但单一模态特征极其脆弱:视觉特征易受光照与阴影干扰,几何特征在平坦区域缺乏辨识度,单一维度的感知难以满足实际踏勘的稳定性要求.另一方面,现有研究多止步于环境的语义解释,割裂了“感知”与“执行”的联系.机器人仅获得岩石的语义标签,却无法感知岩石表面的微地貌特征是否满足取样器的接触约束,导致从语义理解到机器人取样作业动作生成之间存在逻辑断层.

针对上述挑战,本文提出一种面向野外岩石取样机器人的多模态融合感知方法,并在实际应用场景中进行了验证,具体贡献如下:(1)基于几何—纹理协同约束的岩石区域分割方法.针对野外场景缺乏标注数据且单一特征不稳定的问题,本文充分利用环境建图算法输出的彩色点云,构建了融合激光点云几何特征与相机视觉特征的联合生长准则.通过多模态协同约束,有效克服了单一模态下的阴影干扰与地物混淆问题,实现了岩石区域的精准提取.(2)结合机器人取样作业的执行约束,提出一种取样点位姿自动生成策略,通过岩石区域的局部表面特征分析与机器人运动学约束校正,选取最优岩石取样点,并实现取样点从环境空间到机器人作业空间的精准映射.

## 1 概述

本章结合机器人在野外环境下实现自主取样的难点,对本文所提的岩石取样机器人的结构设计以及多模态融合的环境感知方法的总体框架分别进行阐述.

### 1.1 岩石取样机器人结构设计

针对野外地质踏勘场景中地形险峻、地表高度非结构化的特点,本文设计了一套具备高通过性与物理交互能力的自主岩石取样机器人系统,整体结构设计如图1所示.

在移动平台方面,选用具备复杂地形适应能力的四足机器人作为载体,以保障机器人在碎石坡等复杂地貌下的平稳行进.并在机器人机身前部集成激光雷达与光学相机,这两类感知传感器结合本文提出的多模态融合的环境感知方法,为机器人在野外环境中的生存能力以及取样认知提供基础.在取样装置方面,系统搭载了一台七自由度机械臂作为

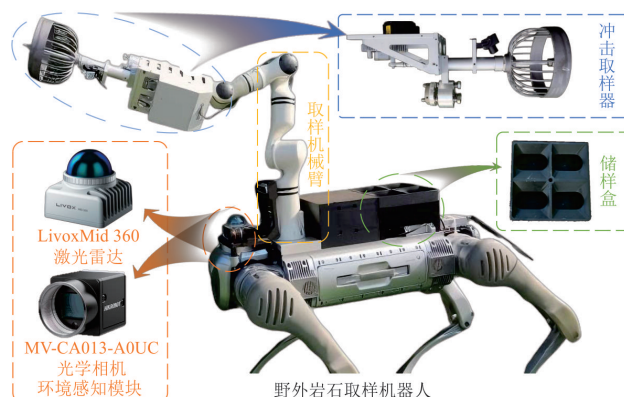


图1 岩石取样机器人实物图

Fig.1 Photograph of the rock-sampling robot

取样作业的灵活支撑,并在机械臂末端部署了自主研发的冲击取样器.该取样器利用电机驱动连杆机构进行蓄能,并通过弹性元件的瞬时释能驱动冲击头破碎岩石,实现了低功耗驱动下的高功率输出.此外,取样器配有带有过滤网孔的网兜结构用于岩石收集与初步筛选,最终采集的岩石样本存放在机身背部的储样盒内,以满足多点连续作业的样本分类存储需求.

### 1.2 多模态融合的环境感知方法框架

面向野外岩石取样机器人的多模态融合感知方法的总体框架如图2所示,该方法框架通过激光雷达与相机不同模态数据之间的协同约束,将非结构化的野外环境逐步解析为可供机器人交互的结构化信息,实现复杂野外环境下岩石目标的精准识别与自主取样.

具体而言,该方法主要包含3个核心处理环节.首先,多传感器融合的环境建图模块作为感知的时空基准,负责输出的全局一致彩色点云,并提供野外场景的实时三维几何结构与表观纹理信息.其次,岩石区域语义分割模块利用多模态自适应种子生长算法,将彩色点云中的岩石目标与背景环境分离,生成具备语义标签的点云簇.最后,取样点位姿自动生成模块则基于对岩石区域的几何结构特征的量化分析与机械臂运动学约束校正,从语义点云中推理出最佳作业位姿.在此过程中,岩石实体作为核心感知对象与机械臂的执行空间紧密相关.前者提供岩石表面的曲率和法向量属性,进一步细化了可取样区域的特征,并能通过平整度指标对候选点进行优选分级;后者则提供了取样器作业所需的位置与姿态等详细参数,最终生成的关键位姿信息传输至取样机械臂的运动控制器,支持后续的路径

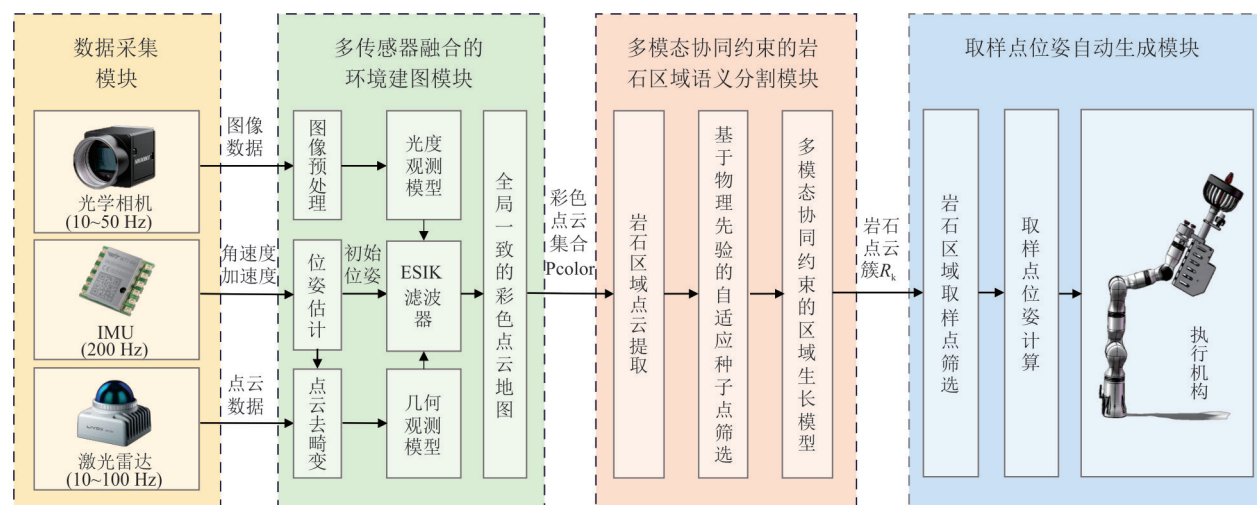


图2 方法总体框架

Fig. 2 Overall framework of the method

规划与动作执行。

基于此感知方法框架,岩石取样机器人可实现如下核心功能:

①环境建图:融合激光与视觉数据,实时刻画包含颜色与几何信息的环境三维特征;

②岩石语义分割:利用多模态特征,在无监督条件下实现岩石区域与非岩石区域的自动分类;

③取样点选取:结合局部曲率与法向量分析,筛选满足接触式作业要求的平整区域,保证机器人人作业过程的一致性和可重复性;

④机械臂位姿解算:利用手眼标定参数与逆运动学解算,将环境语义目标转化为机械臂关节控制指令。

在该框架下,机器人系统依托环境建图模块的实时三维点云地图,基于激光几何与视觉纹理的多模态约束构建岩石区域语义分割模型;随后,以作业可行性为核心指标,结合岩石区域点云分析与机械臂运动学模型,构建面向执行的位姿生成模型,最终实现从语义空间到动作空间的精准映射,实现机器人自主取样的作业流程全自动化。

## 2 方法

本章将阐述面向野外岩石取样机器人的多模态融合感知方法,环境建图是系统的感知基石,旨在构建三维环境地图;岩石区域语义分割则承担环境认知任务,用于提取目标岩石区域;取样点位姿自动生成则是连接语义空间与机器人作业空间的桥梁,负责选取取样点并将其解析为机器人可执行

的最佳作业位姿。在接下来的小节中,将依次对上述各模块的详细技术路线进行论述。

### 2.1 多传感器融合的环境建图

野外岩石取样机器人的作业环境通常表现出高度的非结构化与动态时变特征,一方面,野外光照条件极不稳定,强烈的直射光与岩石阴影交替,加之四足机器人行进间的高频震动,极易导致视觉图像出现过曝光或运动模糊,使传统基于特征点提取的视觉里程计失效;另一方面,在地势平坦或植被覆盖区域,激光雷达面临几何特征退化的问题,单一模态难以维持长时间的定位精度。此外,松耦合的融合方案忽略了传感器间的内在约束,在剧烈运动下易产生累积漂移。为了克服上述问题,构建可同时表征野外环境几何结构与表面纹理特征的高精度三维地图,本文采用紧耦合的激光雷达—视觉—惯性里程计(LiDAR-visual-inertial odometry, LVIO)框架。该框架基于误差状态迭代卡尔曼滤波(error-state iterative kalman filter, ESIKF),通过多模态数据的协同优化,为机器人后续取样作业任务提供高精度的时空基准。

构建三维环境地图的高精度几何结构是实现机器人自主取样的前提,其核心在于利用几何一致性约束,来确保构建的地图可以准确反映周围环境的几何拓扑形态,为后续机械臂的取样点选取与位姿计算提供精确的几何基准。

定义: $\{G\}$ 为世界坐标系, $\{I\}$ 为机器人(IMU)坐标系, $\{L\}$ 为激光雷达坐标系, $\{C\}$ 为相机坐标系。假设当前时刻 $k$ 机器人的状态预测值为 $x_k$ ,其中

包含从机器人本体坐标系  $\{I\}$  到世界坐标系  $\{G\}$  的旋转矩阵  ${}^G R_{I_k}$  和平移向量  ${}^G p_{I_k}$ . 对于当前帧激光雷达观测到的第  $j$  个点在雷达坐标系下的原始坐标为  ${}^L p_j$ , 利用激光雷达与 IMU 的刚性外参 ( ${}^L R_L, {}^L p_L$ ) 和当前机器人状态  $x_k$ , 将点  $p_j$  投影至世界坐标系  ${}^G p_j$ :

$${}^G p_j = {}^G R_{I_k} ({}^L R_L {}^L p_j + {}^L p_L) + {}^G p_{I_k}. \quad (2.1)$$

在全球地图中搜索点  ${}^G p_j$  所在的体素, 若该体素内已构建完整的局部平面特征 (由法向量  $n_j$ , 中心点  $q_j$  描述), 则构建点  ${}^G p_j$  到平面的几何残差  $r_{\mathcal{L}}$ :

$$r_{\mathcal{L}}(x_k) = n_j^T ({}^G p_j - q_j). \quad (2.2)$$

通过最小化该几何残差优化机器人的位姿, 并根据新观测点不断更新体素内的平面参数  $n$  与  $q$ , 实现对环境几何结构的精细刻画.

在此基础上, 引入视觉观测模型, 利用光度一致性约束将高分辨率的纹理信息精确映射至几何表面. 假设同一物理点在短时间内的不同视角下的灰度观测应当保持一致. 基于此假设, 将传统的多传感器时空对齐问题转化为图像块的像素级配准问题.

对于地图中某一的视觉地图点, 记其世界坐标为  ${}^G p_i$ , 该地图点关联参考帧  $I_r$ . 当该地图点在当前帧  $I_k$  中被再次观测时, 若多模态数据间存在空间对齐误差, 则该点在当前图像上的重投影位置将发生错位, 导致投影处的观测灰度与参考纹理灰度产生显著的光度残差. 因此, 视觉前端的优化本质是寻找最佳的空间变换参数, 使得参考纹理块与当前观测图像块之间的光度残差  $r_v$  最小化, 从而消除视差伪影, 实现 RGB 图像数据中的环境纹理与点云数据构建的环境几何结构精确配准, 为后续的多模态岩石区域分割提供可靠的特征输入.

首先, 利用相机投影函数  $\pi(\cdot)$ , 将世界坐标系下的地图点  ${}^G p_i$  重投影到当前帧图像  $I_k$  的像素平面上, 得到像素坐标  $u_{i,k}$ :

$$u_{i,k} = \pi [ {}^C T_I ({}^G T_{I_k})^{-1} {}^G p_i ], \quad (2.3)$$

其中:  ${}^G T_{I_k}$  表示  $k$  时刻从机器人坐标系到世界坐标系的变换矩阵,  ${}^C T_I$  表示 IMU 与相机的外参变换矩阵.

考虑到野外环境光照条件会随云层遮挡或机器人姿态变化而剧烈波动, 模型引入光照补偿因子  $\tau$  (相对曝光时间比) 和仿射变换矩阵  $A_r$ . 则修正后的光度误差  $r_v$  可定义为:

$$r_v(x_k) = \tau \cdot I_k(u_{i,k}) - I_r(u_{i,r} + A_r \Delta u), \quad (2.4)$$

其中: 光照补偿因子  $\tau$  用于补偿野外动态光照引起

的亮度变化; 仿射变换矩阵  $A_r$  通过激光雷达提供的平面法向量修正因视角变化引起的图像块透视变形.

在此基础上, 多传感器融合的环境建图模块通过维护一个全局统一的哈希体素地图来实现异构传感器数据在物理空间中的配准. 对于地图中的任意体素  $V$ , 其状态空间定义为机器人周围环境的几何属性与纹理属性的集合:

$$V = \{n, q, \Sigma, S_{\text{geo}}, S_{\text{vis}}\}, \quad (2.5)$$

其中:  $n \in \mathbb{R}^3$  与  $q \in \mathbb{R}^3$  分别为该体素内局部拟合平面的法向量与中心点,  $\Sigma$  为平面参数的不确定度协方差矩阵;  $S_{\text{geo}}$  落入该体素的激光雷达点云集合, 用于持续修正三维地图的几何结构;  $S_{\text{vis}}$  为附着在几何点上的视觉图像块集合, 用于表征环境的纹理信息.

这种统一表达结构确保了点云数据提供的几何信息与 RGB 数据提供的视觉颜色信息在物理空间上的精确配准, 有效避免了传统基于松耦合的多传感器融合建图方法中常见的视差伪影. 系统采用 ESIKF 进行状态估计, 先利用  $r_{\mathcal{L}}$  更新几何状态, 再在几何约束的基础上利用光度残差  $r_v$  更新光度状态. 经过紧耦合处理, 多传感器融合的环境建图模块实时输出全局一致的彩色点云集合  $\mathcal{P}_{\text{color}}$ :

$$\mathcal{P}_{\text{color}} = \{p_m | p_m = [x, y, z, \kappa, n, r, g, b]^T\}, \quad (2.6)$$

集合中的每个点  $p_m$  均包含三维坐标  $(x, y, z)$ , 法向量  $n = [n_x, n_y, n_z]$ , 以及用于识别岩石材质差异的视觉特征  $(r, g, b)$  为后续的多模态语义分割任务奠定了坚实的特征基础.

## 2.2 多模态协同约束的岩石区域语义分割

尽管环境建图模块构建了野外环境的高精度彩色点云地图, 输出了包含空间坐标、法向量及光学纹理的全局点云集合. 但对于执行接触式作业的岩石取样机器人而言, 点云本质上仍是离散的非结构化集合, 缺乏指导动作的明确边界信息, 机器人还面临着语义感知的挑战, 这是机器人实现智能环境感知的关键. 野外环境具有高度的非结构化特征, 岩石、硬质土壤、植被与碎石混合分布, 若仅依赖传统的几何特征 (如平面拟合), 机器人难以区分平坦的岩石表面与硬质地面; 若仅依赖视觉特征, 一方面, 野外光照条件多变且不规则的岩石表面极易产生自遮挡阴影, 影响图像识别的精度, 另一方面, 野外存在大量诸如树干、土坡等相似颜色的干扰物体, 导致识别准确率下降.

针对上述野外环境中特有的感知挑战, 现有的

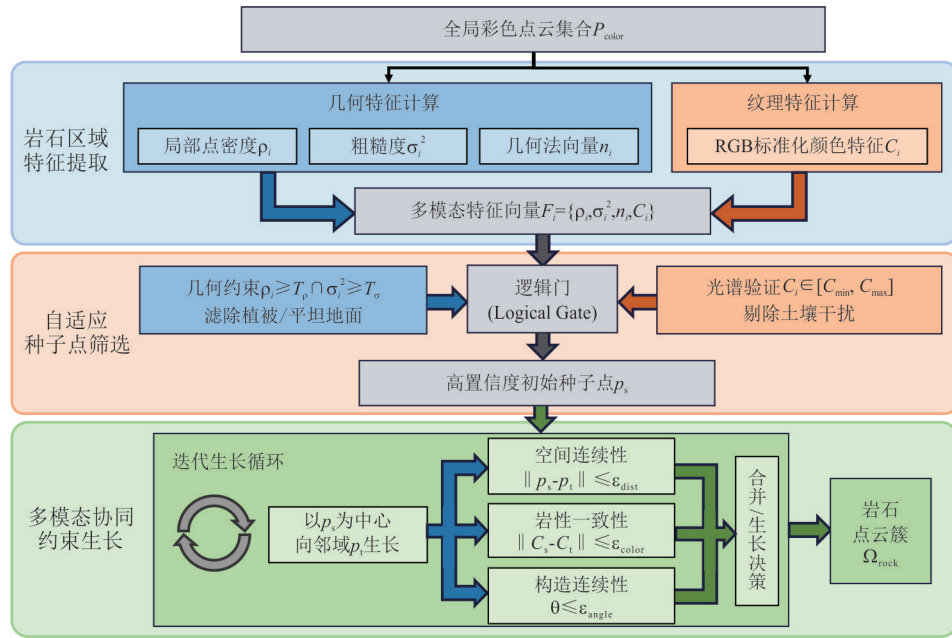


图 3 多模态协同约束的岩石区域语义分割算法

Fig. 3 Multimodal cooperative constraint-based semantic segmentation algorithm for rock regions

有监督深度学习方法虽然在特定数据集上表现优异,但高度依赖大规模标注样本.考虑到野外岩石的形态、纹理及分布具有极大的随机性,构建完备的训练集成本极高且模型在不同野外场景下的泛化能力受限.相比之下,区域生长法(region growing, RG)作为一种经典的无监督分割策略,利用点云局部的

物理属性相似性(如法向连续性、曲率一致性)将相邻点聚类为独立区域,具有无需预训练、物理意义明确且能自然保留物体三维拓扑结构的优势.然而,传统区域生长算法通常依赖人工选取种子点,且仅利用单一几何特征,在野外噪声干扰大、环境复杂的条件下,极易出现因种子点选取不当导致的分割失败或因为点云特征失效引起的过分割现象.

为了克服传统方法的局限性并适应野外自动化作业需求,本文提出一种基于几何—纹理协同约束的岩石区域分割方法.该方法首先根据岩石区域点云密度、表面非平滑等物理先验特征,自动筛选出最具代表性的岩石核心区域作为生长起点,以解决人工选点的不确定性问题;然后以优选种子点为起点,在其周围邻域范围内根据几何结构与光学纹理的协同约束搜索,满足阈值的与种子点合并,并设定为新的区域生长起点,以此进行区域生长迭代,将点集中所有的相似特征区域聚合实现野外环

境中岩石区域与非岩石区域的分割,为后续机器人取样作业提供精确的空间基础,如图3所示.

**2.2.1 基于物理先验的自适应种子点筛选** 为了兼顾环境感知的全局一致性与取样作业的实时性,考虑到单帧数据在空间分布上呈现显著的稀疏性与离散性.这种稀疏性不仅会导致局部点密度 $\rho$ 与表面粗糙度 $\sigma^2$ 等关键统计特征的计算严重失真,且单一视角观测不可避免地存在自遮挡现象,导致岩石背部或侧面形态缺失.因此,本文结合建图模块的时序累积优势,基于机械臂的作业半径 $R_{work}$ ,从多源融合建图模块输出的体素地图中实时截取位于机器人前方的局部点云子集 $\mathcal{P}_{local}$ :

$$\mathcal{P}_{local} = \{p_i \in \mathcal{P}_{color} \mid \|p_i\|_2 \leq R_{work}\}. \quad (2.7)$$

该子集 $\mathcal{P}_{local}$ 集合了机器人历史运动过程中对当前区域的多角度高频观测数据,有效填补单帧扫描的点云稀疏以及盲区.在此基础上,为了满足区域生长算法对邻域搜索的高频需求,本文针对离散点集 $\mathcal{P}_{local}$ 构建KD-Tree空间拓扑索引,并对任意点 $p_i$ 搜索其半径 $r$ 范围内的 $\mathcal{N}_i$ .

在非结构化的野外场景中,岩石区域与非岩石区域的土壤、植被等耦合严重,单一维度的几何特征或视觉难以实现对岩石区域的精确锁定.通过对野外非结构化环境的原始点云进行特征空间采样分析,如图4所示.

岩石、植被与土壤在几何结构上存在本质差

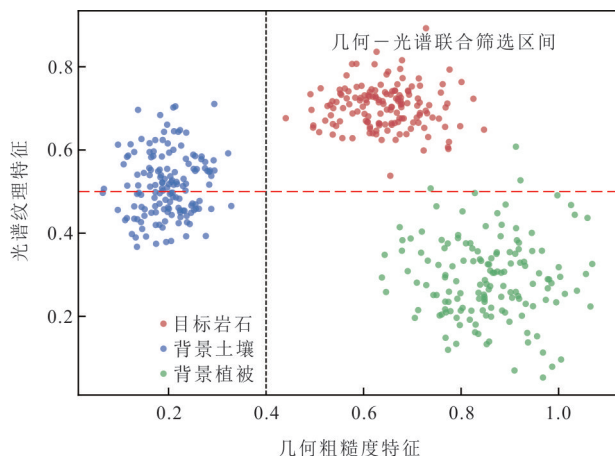


图 4 野外岩石区域与非岩石区域的特征分布统计

Fig. 4 Statistical distribution of feature characteristics between rock and non-rock areas in the field

异. 岩石作为致密介质, 其表面反射具有连续性, 局部点密度  $\rho_i$  高于背景. 同时, 由于风化作用形成的微观起伏, 其粗糙度  $\sigma_i^2$  呈现特定的分布区间, 而植被受体积散射影响表现为高离散、低密度特征. 针对几何特征相似的场景 (如干硬土块与岩石), 引入光谱颜色特征  $C_i$  作为辅助验证, 目标岩石在特定的色调区间内具有显著的反射一致性, 这为区分异物同形干扰提供了关键判据. 基于上述分析, 利用  $p_i$  邻域点集解算各点的多模态特征向量  $F_i = \{\rho_i, \sigma_i^2, \mathbf{n}_i, C_i\}$ , 其计算模型如下,  $\rho_i$  为局部点密度, 定义为邻域球体内的有效点数:

$$\rho_i = \text{Count}(p_j) \text{ s.t. } p_j \in \mathcal{N}_i, \quad \|p_j - p_i\|_2 \leq r, \quad (2.8)$$

式中:  $\text{Count}(\cdot)$  为计数函数,  $r$  为邻域搜索半径.

基于主成分分析解算粗糙度  $\sigma_i^2$  与几何法向量  $\mathbf{n}_i$ , 计算邻域点集  $\mathcal{N}_i$  的协方差矩阵  $M_i$ :

$$M_i = \frac{1}{k} \sum_{p_j \in \mathcal{N}_i} (p_j - \bar{p}_i)(p_j - \bar{p}_i)^T \quad (2.9)$$

式中:  $k$  为邻域点数,  $\bar{p}_i$  为邻域几何中心. 对  $M_i$  进行特征值分解, 得到特征值  $\lambda_1 \geq \lambda_2 \geq \lambda_3$  及其对应的特征向量  $v_1, v_2, v_3$ .

几何法向量  $\mathbf{n}_i$  取最小特征值对应的特征向量, 即  $\mathbf{n}_i = v_3$ , 其表征了岩石表面的局部形状. 粗糙度  $\sigma_i^2$  为最小特征值在总方差中的占比:

$$\sigma_i^2 = \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3}. \quad (2.10)$$

光谱颜色特征  $C_i$  则是通过提取彩色点云中的 RGB 特征, 并进行归一化处理, 得到颜色向量  $C_i = [r_i, g_i, b_i]^T$ , 可用于区分不同材质.

区域生长法通过选取种子点, 依据生长准则对

相邻点进行判断, 当没有相邻点满足生长准则时即停止生长. 其中, 选取的初始种子点的优劣将直接影响到算法计算效率以及最终的分割效果. 基于此, 本文结合岩石区域区别于背景环境的物理特性, 设计了基于致密性—粗糙度—光谱多维物理先验的自适应种子点筛选策略.

首先, 根据野外岩石的几何形态特征, 构建疑似岩石点集合  $S_{\text{rock}}$ . 基于统计规律的密度阈值  $T_\rho$  与粗糙度阈值  $T_\sigma$ , 构建联合约束模型:

$$S_{\text{rock}} = \{p_i \in \mathcal{P}_{\text{local}} \mid (\rho_i \geq T_\rho) \cap (\sigma_i^2 \geq T_\sigma)\}. \quad (2.11)$$

利用高密度来滤除了植被与噪声, 利用高粗糙度剔除了平坦的硬质地面, 可以有效锁定了潜在的岩石目标区域.

在此基础上, 为更加精确的区分具有相似粗糙度的硬质土块与岩石, 排除了异物同形的干扰, 本文利用了岩石与土壤在光谱反射率上的显著差异, 引入光学光谱特征作为材质验证约束. 在几何初筛集合  $S_{\text{rock}}$  中, 选取颜色向量  $C_i$  落在岩石典型光谱区间  $[C_{\min}, C_{\max}]$  且局部密度最大的  $K$  个点 (取  $K=5$ ) 作为初始生长种子. 同时, 优先选择高密度中心点作为种子, 确保了生长过程始于数据质量最高、特征最显著的岩石核心部位, 避免了从岩石与背景交界处的模糊区域开始生长导致的误差扩散.

此外, 若当前视场内因遮挡未搜索到满足条件的种子点, 系统将根据里程计信息, 在机器人位移超过一定距离后触发重取样, 确保对连续分布岩石的持续检测.

### 2.2.2 基于几何—纹理协同约束的区域生长模型

在确定高置信度的岩石种子点  $p_s$  ( $p_s \in S_{\text{rock}}$ ) 后, 算法向邻域点  $p_i$  执行迭代扩张. 为了在复杂的野外环境中精确划定岩石区域与非岩石区域的物理边界, 本文, 构建了几何—纹理—空间三维协同约束模型, 只有当邻域点  $p_i$  同时满足以下 3 个一致性准则时, 才将其判定为岩石区域的一部分.

一是空间连续性准则, 该准则用于约束地质实体的物理连通性:

$$\|p_s - p_i\|_2 \leq \epsilon_{\text{dist}}, \quad (2.12)$$

式中: 阈值  $\epsilon_{\text{dist}}$  依据激光雷达在机器人有效作业距离内的平均取样密度设定.

对于接触式取样作业, 目标表面的连续性是保障钻进稳定性的前提. 野外环境中常存在悬浮尘埃、飞虫或植被缝隙产生的离散噪点, 其空间间距通常显著大于实体表面的点云间距. 通过该约束,

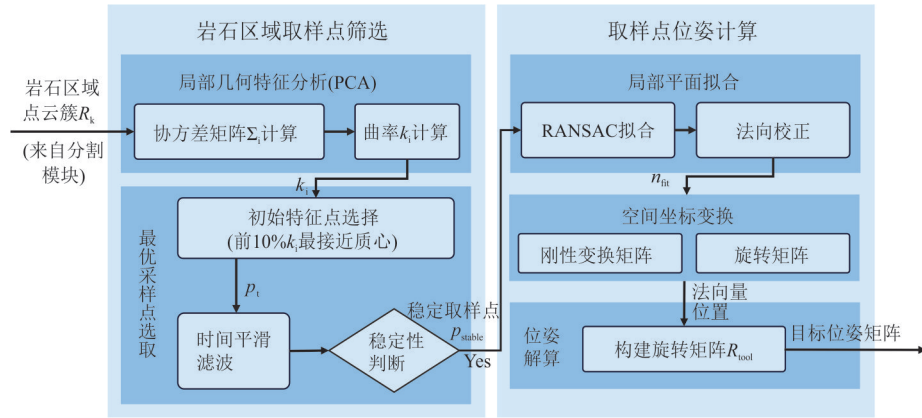


图5 取样点位姿自动生成策略

Fig. 5 An automated sampling-pose generation strategy

算法能够有效滤除离群噪声,防止机器人将非实体的虚假点误判为可作业区域,避免机械臂执行空操作或发生碰撞风险。

二是岩性一致性准则,该准则基于岩石与背景(如土壤、风化碎屑)在光谱反射特性上的显著差异,构建材质约束:

$$\|C_s - C_t\|_1 = |r_s - r_t| + |g_s - g_t| + |b_s - b_t| \leq \epsilon_{\text{color}}, \quad (2.13)$$

在野外,经常出现“异物同形”现象,即紧贴岩石根部的硬质土壤在几何平整度上与岩石表面高度相似.若仅依赖几何特征,生长过程极易误判到土壤区域,导致机器人误将土壤识别为岩石而执行无效取样.引入光学特征约束,能够利用光谱差异在几何过渡区强制截断生长,精确锁定岩石的岩性边界,确保取样点落在坚硬的岩石本体上。

三是构造连续性准则,该准则利用法向量夹角约束表面的几何平滑度.尽管岩石表面微观粗糙,但在宏观尺度上其构造变化具有连续性:

$$\theta = \arccos\left(\frac{\mathbf{n}_s \cdot \mathbf{n}_t}{\|\mathbf{n}_s\| \|\mathbf{n}_t\|}\right) \leq \epsilon_{\text{angle}}, \quad (2.14)$$

式中: $\mathbf{n}_s$ 代表当前种子点的法向量, $\mathbf{n}_t$ 代表邻域点的法向量。

在野外自然光照条件下,不规则的岩石表面极易产生自遮挡阴影,导致同一块岩石在明暗交界处出现剧烈的颜色突变,时视觉特征失效,若无几何约束,岩石将被错误地分割为多个碎片(欠分割),影响机器人对岩石整体形态的评估.构造连续性准则利用法向量的一致性,迫使生长过程跨越阴影边界,保证了岩石实体的完整提取,为机械臂规划无碰撞的接近路径提供了完整的几何包络信息。

当区域生长迭代终止后,为了进一步提升分割

结果的可用性,需对生成的初级点云簇进行形态学优化.针对传感器噪声产生的非地质微小斑块,本文设定阈值进行滤波;同时,针对视线遮挡导致的目标断裂,计算簇间欧氏距离,将小于10 cm的邻近簇进行同源合并.最终,输出的集合 $\Omega_{\text{rock}}$ 具备清晰的物理边界,可作为后续取样点选取与位姿计算模块的输入对象。

### 2.3 取样点位姿自动生成

岩石取样机器人通过语义分割获得岩石点云簇 $\mathcal{R}_k \in \Omega_{\text{rock}}$ 后,需在不规则的岩石表面推理出最佳作业位姿,以提高其在敲击作业时的破岩效率.为了弥补环境感知与机械臂执行之间的空间差异,结合机器人取样作业的执行约束,本文提出一种取样点位姿自动生成策略,如图5所示。

为了确保机械臂在野外非结构化环境下实现高效且安全的物理交互,取样点的选择需同时满足易破碎性与可达稳定性两个关键指标.基于点云数据的几何特征统计分析如图6所示,岩石表面的几何形态与其物理采样难度存在强相关性.平坦区域(低曲率区)虽然便于接近,但在冲击作业时难以形成有效的应力集中,导致破岩失败率较高;相比之下,岩石表面的棱角与凸起部位(高曲率区)在冲击载荷下更易发生脆性断裂。

故而,为了提升机械臂末端取样器在敲击作业时的破岩效率,优先在岩石区域内寻找几何特征突出且位置稳定的作业点.设任意岩石区域的点云簇 $\mathcal{R}_k = \{p_1, p_2, \dots, p_n\}$ ,对于点 $p_i$ ,利用主成分分析法计算其邻域协方差矩阵 $\Sigma_i$ :

$$\Sigma_i = \frac{1}{|N_i|} \sum_{p_j \in N_i} (p_j - \bar{p}_i)(p_j - \bar{p}_i)^T. \quad (2.15)$$

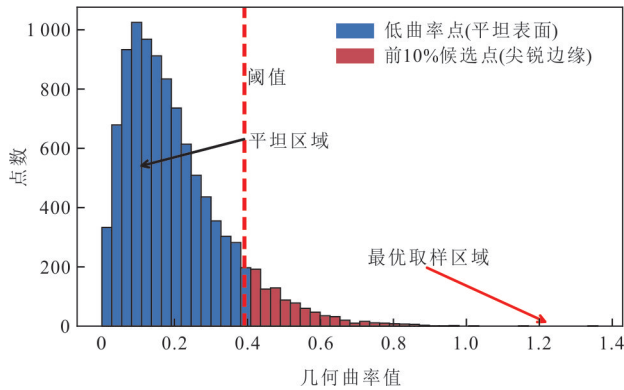


图 6 岩石表面曲率统计分布

Fig. 6 Statistical distribution of rock surface curvature

对协方差矩阵  $\Sigma_i$  进行特征值分解, 得到 3 个特征值  $\alpha_1, \alpha_2, \alpha_3$ , 描述点在各主方向上的离散程度. 并基于特征值解算岩石区域的局部曲率  $\kappa_i$ :

$$\kappa_i = \frac{\alpha_{\min}}{\alpha_1 + \alpha_2 + \alpha_3}. \quad (2.16)$$

该指标能够有效量化表面的凹凸程度. 策略选取曲率排名前 10% 的特征点作为候选集  $p_{\text{top}}$ , 并从中优选距离区域质心最近的点作为初始取样点  $p_i$ . 此外, 针对相机抖动或传感器噪声引起的位置跳变问题, 引入时序滑动平均滤波器. 构建历史点序列  $H_i = \{p_{i-N+1}, \dots, p_i\}$ , 通过计算序列均值  $\bar{p}_i$  并与上一时刻稳定位置  $p_{\text{stable}}$  进行欧氏距离校验 (阈值  $\delta = 0.05 \text{ m}$ ), 实现取样点坐标的平滑更新:

$$p_{\text{stable}} \leftarrow \beta \cdot p_{\text{stable}} + (1 - \beta) \cdot \bar{p}_i, \quad (2.17)$$

该机制确保了机器人能够锁定岩石表面的稳定几何特征, 避免因感知噪声导致的机械臂动作震荡.

确立稳定取样点后, 需结合机械臂的执行约束将其转化为机械臂运动控制器可执行的 6-D 目标位姿. 在其局部邻域内提取有效深度值  $z$ , 针对已知的岩石取样点  $P = (x, y, z)$ , 通过 random sample consensus (RANSAC) 方法拟合取样点平面, 机械臂末端执行器沿着该平面法向量方向进行敲击取样, RANSAC 方法鲁棒性强、可有效去除离群点.

在取样点  $p$  所在点云簇附近, 构建固定大小的感兴趣区域 ROI, 提取该区域的稠密三维点云  $P$ , 对点集  $P$  进行 RANSAC 平面拟合, 获取最优的平面参数  $(a, b, c, d)$ , 最小化点  $p_i$  到平面的距离:

$$D_i = \frac{|ax_i + by_i + cz_i + d|}{\sqrt{a^2 + b^2 + c^2}} < \tau, \quad (2.18)$$

其中,  $\tau$  为设定的内点距离阈值. 最终单位法向量为:

$$\mathbf{n}_{\text{fit}} = \frac{(a, b, c)^T}{\sqrt{a^2 + b^2 + c^2}}. \quad (2.19)$$

为了确保取样器能够沿着垂直于岩石表面的方向进行有效敲击, 规定法向量方向需指向相机一侧:

$$\mathbf{n} = \begin{cases} \mathbf{n}_{\text{fit}}, & \mathbf{n}_{\text{fit}} \cdot \mathbf{Z}_{\text{cam}} > 0 \\ -\mathbf{n}_{\text{fit}}, & \text{otherwise} \end{cases}, \quad (2.20)$$

其中:  $\mathbf{Z}_{\text{cam}} = (0, 0, 1)^T$  表示  $\{C\}$  下的前方方向. 利用预先标定的手眼变换矩阵  $T_{\text{base}}^{\text{cam}}$  和旋转矩阵  $R_{\text{base}}^{\text{cam}}$ , 将感知空间下的取样点位置  $P$  与法向量  $\mathbf{n}$  统一映射至执行空间 (机械臂基座坐标系), 得到  $P_{\text{base}}$  和  $\mathbf{n}_{\text{base}}$ . 最终, 为了构造完整的机械臂末端姿态, 令末端执行器的 Z 轴对齐于法向量, 并通过引入固定参考向量  $\mathbf{v}_{\text{ref}} = (1, 0, 0)^T$  利用正交约束补全 X 轴与 Y 轴, 生成目标旋转矩阵  $R_{\text{tool}}$ . 组合位置与姿态信息, 即可得到机械臂末端的目标位姿矩阵:

$$T_{\text{target}} = \begin{bmatrix} R_{\text{tool}} & P_{\text{base}} \\ 0_{1 \times 3} & 1 \end{bmatrix}. \quad (2.21)$$

该目标位姿可用于后续的机械臂控制与轨迹规划, 实现岩石敲击取样.

### 3 岩石取样机器人实验与应用

为了验证所提方法在真实地质勘查场景中的有效性和可行性, 本节对岩石取样机器人进行案例研究, 并在野外实地场景进行测试.

#### 3.1 岩石区域分割实验

本实验选取了具有代表性的户外岩石与碎石混合场景作为测试区域. 该场景包含裸露岩块、植被干扰、大小不均的碎石堆积以及不平整地形, 能够较全面地反映地质踏勘任务中常见的自然岩石环境, 如图 7 所示.

在该场景中, 利用搭载激光雷达与工业相机的取样机器人进行连续感知, 并运行多传感器融合的环境建图算法获取 3 组带颜色信息的三维点云数据, 3 组数据分别为 S1、S2、S3. 这 3 组点云覆盖不同角度、不同距离与不同光照条件下的真实岩石表面, 且本文数据的正确标签是采用点云数据处理软件对上述数据进行人工分割获取的.

采用上述 3 组数据进行实验分析, 并与传统区域生长法进行比较. 针对每一组彩色点云数据, 采用本文提出的岩石区域分割方法进行处理, 得到对应的岩石区域分割结果, 俯视图如图 7 所示. 基于下指标进行分割精度对比, 定义公式为:



图7 野外环境与分割结果  
Fig. 7 Field environment and segmentation results

表 1 分割精度对比

Table 1 Comparison of segmentation accuracy

| 序号 | 内容        | 分割方法    |         |
|----|-----------|---------|---------|
|    |           | 区域生长法   | 本文方法    |
| S1 | 正确分割点数(个) | 35 627  | 86 693  |
|    | 总点数(个)    | 96 788  | 96 788  |
|    | 正确率(%)    | 36.81   | 89.57   |
| S2 | 正确分割点数(个) | 32 996  | 79 618  |
|    | 总点数(个)    | 92 376  | 92 376  |
|    | 正确率(%)    | 35.72   | 86.19   |
| S3 | 正确分割点数(个) | 32 127  | 88 776  |
|    | 总点数(个)    | 101 540 | 101 540 |
|    | 正确率(%)    | 31.64   | 87.43   |

$$A = \frac{\sum_{i=1}^k TP_i}{\sum_{i=1}^k (TP_i + FN_i)}, \quad (3.1)$$

式中： $A$  表示分割精度； $k$  表示类别个数（本文  $k=2$ ）； $TP_i$  表示第  $i$  类被正确分割的点云数量； $FN_i$  表示被错分割为其他区域的点云数量。

具体的分割结果如表 1 所示。本文提出的基于几何—纹理协同约束的岩石区域分割方法 3 组点云中均能有效识别主要岩石区域，3 组数据的分割精度分为 89.57%、86.19% 和 87.43%，远优于传统的区域生长法。

上述实验结果表明，本文方法能够在复杂自然环境下保持稳定性能。分割结果清晰地分离了岩石主体与背景区域，误分割比例较低，验证了所提方法在野外真实场景中的可靠性和泛化能力。

### 3.2 取样点位姿计算实验

由于机器人在实际冲击取样作业中的最佳机械臂末端取样器操作距离约为 0.7 m，因此实验中将工业相机与目标岩石样本固定在约 0.7 m 的相对距离，并在岩石表面手动选取多个典型取样点。控制机械臂末端工具直接触碰取样点，记录取样点的真实三维位姿，此值与本文算法计算的三维位姿进行对比并计算误差。

表 2 取样点位姿实验误差结果

Table 2 Results of sampling point pose experiment

| 实验序号 | 位置误差(m) | 姿态误差(°) |
|------|---------|---------|
| 1    | 0.006 9 | 1.43    |
| 2    | 0.006 4 | 1.14    |
| 3    | 0.007 1 | 1.66    |
| 4    | 0.007 0 | 1.25    |
| 5    | 0.007 4 | 1.72    |

取样点位置误差采用欧氏距离定义如下：

$$\epsilon_{\text{pos}} = \sqrt{(x_p - x_t)^2 + (y_p - y_t)^2 + (z_p - z_t)^2}, \quad (3.2)$$

其中： $(x_t, y_t, z_t)$  为真实位置， $(x_p, y_p, z_p)$  为预测位置。

取样点姿态误差采用夹角误差定义如下：

$$\epsilon_{\text{angle}} = \arccos(\mathbf{n}_{\text{true}} \cdot \mathbf{n}_{\text{pred}}) \times \frac{180}{\pi}, \quad (3.3)$$

其中： $\mathbf{n}_{\text{true}}$  和  $\mathbf{n}_{\text{pred}}$  分别为真实与预测的单位法向量。实验结果如表 2 所示。

分析表 2 中的实验数据可知，不同采样点间的误差存在微小波动，平均位置误差为 0.696 cm，平均姿态误差为 1.44°。这些波动的产生主要受以下物理与环境因素的综合影响，一方面，野外岩石表面普遍存在由于风化作用形成的微观起伏与不规则纹理。由于 RANSAC 算法是在 ROI 内寻找最优拟合平面，岩石表面的微观凹凸会导致点云分布不完全符合理想平面假设，导致在随机采样迭代过程中不同子集产生的法向量会存在小范围波动，造成姿态误差。另一方面，激光雷达在复杂野外光照环境下存在固有的测距噪声，且岩石表面的反射特性不一，会引起局部点云坐标的细微跳变，且测量真实位姿时也存在毫米级的测量误差。

综合上述分析，所提方法在 0.7 m 的实际工作距离下能够稳定地获得厘米级的三维定位精度（平均位置误差约 0.696 cm），同时具备较高的姿态方向估计精度（平均法向量夹角误差为 1.44°），满足岩石精准取样的实际应用需求。

## 4 结论

本文面向野外地质踏勘任务中岩石区域识别困难、无标注数据缺乏、以及从语义分割到机械臂执行缺乏有效映射等关键瓶颈，提出了一种野外岩石取样机器人的多模态融合感知方法。实验表明，基于几何—纹理协同约束的岩石分割方法在真实

野外环境中表现稳定,能够有效减少光照变化、阴影及背景混淆带来的误分割问题.此外,所设计的取样点位姿生成算法可在复杂岩石表面上自动筛选满足接触与作业约束的可操作域,可实现厘米级的定位误差和高精度的法向量估计.本文方法实现了从环境感知到机械臂取样动作的闭环,使四足机器人能够在野外场景执行全自动岩石取样作业.

## References

- Bar, A., Lohdefink, J., Kapoor, N., et al., 2020. The Vulnerability of Semantic Segmentation Networks to Adversarial Attacks in Autonomous Driving: Enhancing Extensive Environment Sensing. *IEEE Signal Processing Magazine*, 38(1): 42–52. <https://doi.org/10.1109/msp.2020.2983666>
- Cadena, C., Carlone, L., Carrillo, H., et al., 2016. Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age. *IEEE Transactions on Robotics*, 32(6): 1309–1332. <https://doi.org/10.1109/tro.2016.2624754>
- Campos, C., Elvira, R., Rodriguez, J. J. G., et al., 2021. ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial, and Multimap SLAM. *IEEE Transactions on Robotics*, 37(6): 1874–1890. <https://doi.org/10.1109/tro.2021.3075644>
- Engel, J., Koltun, V., Cremers, D., 2017. Direct Sparse Odometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(3): 611–625.
- El Madawi, K., Rashed, H., El Sallab, A., et al., 2019. RGB and Lidar Fusion Based 3D Semantic Segmentation for Autonomous Driving. *IEEE Intelligent Transportation Systems Conference (ITSC)*, Piscataway, N.J., Institute of Electrical and Electronics Engineers (IEEE), USA, 7–12.
- Forster, C., Zhang, Z. C., Gassner, M., et al., 2016. SVO: Semidirect Visual Odometry for Monocular and Multi-camera Systems. *IEEE Transactions on Robotics*, 33(2): 249–265. <https://doi.org/10.1109/tro.2016.2623335>
- Gorai, A. K., Raval, S., Patel, A. K., et al., 2021. Design and Development of a Machine Vision System Using Artificial Neural Network-Based Algorithm for Automated Coal Characterization. *International Journal of Coal Science & Technology*, 8(4): 737–755. <https://doi.org/10.1007/s40789-020-00370-9>
- Guo, Y. F., Xu, P., Li W., et al., 2025. 3D Perception and Reconstruction of Underground Coal Mine Roadways Based on Multi-Sensor Fusion. *Earth Science*. doi: 10.3799/dqkx.2025.274(in Chinese with English abstract).
- Hu, Q. Y., Yang, B., Xie, L. H., et al., 2021. Learning Semantic Segmentation of Large-Scale Point Clouds with Random Sampling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25: 1–1. <https://doi.org/10.1109/tpami.2021.3083288>
- Lin, J. R., Zheng, C. R., Xu, W., et al., 2021. R<sup>2</sup>-LIVE: A Robust, Real-Time, LiDAR-Inertial-Visual Tightly-Coupled State Estimator and Mapping. *IEEE Robotics and Automation Letters*, 6(4): 7469–7476. <https://doi.org/10.1109/lra.2021.3095515>
- Li, K. L., Li, M., Hanebeck, U. D., 2021. Towards High-Performance Solid-State-LiDAR-Inertial Odometry and Mapping. *IEEE Robotics and Automation Letters*, 6(3): 5167–5174. <https://doi.org/10.1109/lra.2021.3070251>
- Liu, X., Gong, G., Hu, X., et al., 2024. Cognitive Enhancement of Robot Path Planning and Environmental Perception Based on Gmapping Algorithm Optimization. *Electronics*, 13(5): 818(in Chinese with English abstract).
- Lin, F. F., Lin, T. L., Yao, Y., et al., 2025. VPA-Net: A Visual Perception Assistance Network for 3d Lidar Semantic Segmentation. *Pattern Recognition*, 158: 111014. <https://doi.org/10.1016/j.patcog.2024.111014>
- Liu, H. W., Xu, G. Q., Liu, B., et al., 2025. A Real Time LiDAR-Visual-Inertial Object Level Semantic SLAM for Forest Environments. *ISPRS Journal of Photogrammetry and Remote Sensing*, 219: 71–90. <https://doi.org/10.1016/j.isprs.2024.11.013>
- Mur-Artal, R., Montiel, J. M. M., Tardos, J. D., 2015. ORB-SLAM: A Versatile and Accurate Monocular SLAM System. *IEEE Transactions on Robotics*, 31(5): 1147–1163. <https://doi.org/10.1109/tro.2015.2463671>
- Pan, B. W., Sun, J. K., Leung, H. Y. T., et al., 2020. Cross-View Semantic Segmentation for Sensing Surroundings. *IEEE Robotics and Automation Letters*, 5(3): 4867–4873. <https://doi.org/10.1109/lra.2020.3004325>
- Pang, Y. E., Li, X., Dong, Z. K., et al., 2024. Convolutional Neural Network-Based Model for Recognizing TBM Rock Chip Gradation. *Automation in Construction*, 163(1): 105414. <https://doi.org/10.1016/j.autcon.2024.105414>
- Qin, T., Li, P. L., Shen, S. J., 2018. VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator. *IEEE Transactions on Robotics*, 34(4): 1004–1020. <https://doi.org/10.1109/tro.2018.2853729>
- Sun, K., Mohta, K., Pfrommer, B., et al., 2018. Robust Stereo Visual Inertial Odometry for Fast Autonomous Flight. *IEEE Robotics and Automation Letters*, 3(2): 965–972. <https://doi.org/10.1109/lra.2018.2793349>

- Shin, Y. S., Park, Y. S., Kim, A., 2020. DVL-SLAM: Sparse Depth Enhanced Direct Visual-LiDAR SLAM. *Autonomous Robots*, 44(2): 115—130. <https://doi.org/10.1007/s10514-019-09881-0>
- Wang, Y. Z., Sun, S. Y., 2021. Image-Based Rock Typing Using Grain Geometry Features. *Computers & Geosciences*, 149: 104703. <https://doi.org/10.1016/j.cageo.2021.104703>
- Wang, Y., Zhang, J., Chen, Y. H., et al., 2024. An Automated Learning Method of Semantic Segmentation for Train Autonomous Driving Environment Understanding. *IEEE Transactions on Industrial Informatics*, 20(4): 6913—6922. <https://doi.org/10.1109/tii.2024.3353874>
- Xie, T., Sun, Q. H., Sun, T., et al., 2025. DVDS: A Deep Visual Dynamic Slam System. *Expert Systems with Applications*, 260(10): 125438. <https://doi.org/10.1016/j.eswa.2024.125438>
- Xu, X. B., Zhang, L., Yang J., et al., 2022. A Review of Multi-Sensor Fusion SLAM Systems Based on 3D LIDAR. *Remote Sensing*, 14(12): 2835. <https://doi.org/10.3390/rs14122835>
- Xu, K., Hao, Y. F., Yuan, S. H., et al., 2025. AirSLAM: An Efficient and Illumination-Robust Point-Line Visual SLAM System. *IEEE Transactions on Robotics*, 41(5): 1673—1692. <https://doi.org/10.1109/tro.2025.3539171>
- Yan, X., Gao, J., Zheng, C., et al., 2022. 2DPASS: 2D Priors Assisted Semantic Segmentation on Lidar Point Clouds. *European Conference on Computer Vision (ECCV)*, Springer, Berlin, Germany, 677—695.
- Yang, J. T., Kang, Z. Z., Yang, Z., et al., 2022. A Laboratory Open-Set Martian Rock Classification Method Based on Spectral Signatures. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1—15. <https://doi.org/10.1109/tgrs.2022.3175996>
- Yao, X., Shi, J., Chen, J., et al., 2024. From Direct Exploration to AI Exploration: Evolution and Insight, *Development and Research Centre*, 15(4): 671—679 (in Chinese with English abstract).
- Zhang, J., Kaess, M., Singh, S., 2017. A Real-Time Method for Depth Enhanced Visual Odometry. *Autonomous Robots*, 41(1): 31—43. <https://doi.org/10.1007/s10514-015-9525-1>
- Zhang, J., Singh, S., 2018. Laser-Visual-Inertial Odometry and Mapping with High Robustness and Low Drift. *Journal of Field Robotics*, 35(8): 1242—1264. <https://doi.org/10.1002/rob.21809>
- Zhao, L., Zhou, H., Zhu, X. G., et al., 2023. LIF-Seg: LiDAR and Camera Image Fusion for 3D LiDAR Semantic Segmentation. *IEEE Transactions on Multimedia*, 26: 1158—1168. <https://doi.org/10.1109/tmm.2023.3277281>
- Zhu, H., Pan, B., Wu, W., et al., 2023. A Review of Rock Mass Discontinuity Information Collection and Identification Methods, *Journal of Basic Science and Engineering*, 31(6):1339—1360 (in Chinese with English abstract).
- Zhang, S., Shan, J. J., Liu, Y. B., 2025. Approximate Inference Particle Filtering for Mobile Robot SLAM. *IEEE Transactions on Automation Science and Engineering*, 22: 7967—7978. <https://doi.org/10.1109/tase.2024.3475735>

## 中文参考文献

- 郭云飞, 徐鹏, 李伟, 等, 2026. 基于多传感器融合的煤矿井下巷道三维感知与重建方法. 地球科学.
- 姚晓峰, 施俊法, 陈骥, 等. 2024. 从“直接找矿”到“智能找矿”: 脉络与启示, 矿产勘查, 15(4): 671—679.
- 朱合华, 潘柄屹, 武威, 等. 2023. 岩体结构面信息采集及识别方法研究进展, 应用基础与工程科学学报, 31(6):1339—1360.