

# 基于无人机平台的地表环境实时稠密点云生成与数字模型构建

胡博妮<sup>1</sup>, 陈霖<sup>1</sup>, 徐丙立<sup>2</sup>, 布树辉<sup>1</sup>, 韩鹏程<sup>1</sup>, 李坤<sup>1</sup>, 夏震宇<sup>1</sup>,  
李霓<sup>1</sup>, 李科<sup>3</sup>, 曹雪峰<sup>3</sup>, 万刚<sup>4</sup>

1. 西北工业大学, 西安 710072;

2. 陆军装甲兵学院, 北京 100072;

3. 信息工程大学, 郑州 450000;

4. 航天工程大学, 北京 101400

**摘要:**实时构建地表环境高逼真数字模型一直是遥感航测领域的研究热点,也是实现地理环境的虚拟映射并进而形成数字孪生地理环境的关键基础。针对当前地表环境三维数字模型构建中存在的速度慢、时效性低和大场景应用受限等问题,本文在综述现有国内外研究的基础上,结合团队研究基础,提出一个基于无人机平台的地表环境实时稠密点云生成与数字模型构建方法。研究通用实时定位与地图构建框架,打通了从数据获取到三维点云生成、数字模型重建、结果分析等技术链条,设计动态内存管理模块和联合GNSS优化的PI-SLAM,突破了在线获取数据与位姿解算、实时稠密点云生成、实时数字表面模型(DSM)和数字正射影像图(DOM)构建等关键技术。经实验验证,方法在稠密点云生成和数字模型构建精度与现有离线建模算法相当的同时,速度提升了30—50倍,达到在线采集数据并实时建模的程度。应用案例表明,本方法可应用于灾害预警救援、应急管理、作战模拟等高时效地形三维重建中,同时能够为数字孪生地理环境数字底座构建提供技术支撑。

**关键词:** 遥感, 实时重建, 稠密点云, 数字孪生, 地表环境, 数字模型

**中图分类号:** TP751/P2

**引用格式:** 胡博妮,陈霖,徐丙立,布树辉,韩鹏程,李坤,夏震宇,李霓,李科,曹雪峰,万刚.2024.基于无人机平台的地表环境实时稠密点云生成与数字模型构建.遥感学报,28(5):1206-1221

Hu B N, Chen L, Xu B L, Bu S H, Han P C, Li K, Xia Z Y, Li N, Li K, Cao X F and Wan G. 2024. Real-time dense point cloud generation and digital model construction of surface environment based on UAV platform. National Remote Sensing Bulletin, 28(5):1206-1221[DOI:10.11834/jrs.20232597]

## 1 引言

自美国密歇根大学 Michael Grieves 教授于2003年提出数字孪生概念以来,数字孪生从工业制造领域迅速扩展到各个领域(陶飞等,2018),并出现“万物皆可数字孪生”的发展态势。当前,数字孪生已从技术手段上升到学科研究方法(李浩等,2019;郭具涛等,2020)。在地理学领域,数字孪生方法也已经得到广泛关注与研究,比如数字孪生城市(李德仁,2021)、数字孪生流域

(刘昌军等,2022;夏润亮等,2021)、数字孪生铁路(朱庆等,2020)等已经从理论方法研究走向实践应用,并已产生良好的示范效果。在构建数字孪生地理系统的过程中,对地表环境的表面模型快速构建是实现从现实地理环境到虚拟地理环境映射的关键技术。

大范围地表环境数字模型构建主要依托卫星遥感和飞机航测方法完成。基于卫星图像的地表环境模型构建技术(潘雪琛等,2019)存在生成周期长、受天气影响大和数据时效性差问题,致

收稿日期:2022-11-15;预印本:2023-07-20

基金项目:国家自然科学基金(编号:41871322,62003272);陕西省重点研发计划项目(编号:2022GY-253);特色学科基础研究项目(编号:G2022WD);装备综合研究项目(编号:LJ2022C010133)

第一作者简介:胡博妮,研究方向为深度学习、复杂环境下的无人机定位与感知。E-mail:huboni@mail.nwpu.edu.cn

通信作者简介:布树辉,研究方向为实时建图与智能感知。E-mail:bushuhui@nwpu.edu.cn

使基于历史卫星影像的地表模型与实际地表环境相差甚大。无人机航空摄影测量作为一种机动灵活、高效快速、操作简便、作业成本低、适用范围广的遥感影像获取方式,在灾害救援、智慧农业、地球观测、协同作战等民用和军用领域得到了广泛应用(杨胜天等, 2021; 张继贤等, 2021)。然而, 现有飞机航测方法普遍基于离线照片导入重建, 存在计算复杂度高、处理速度慢的问题, 从派飞手执行飞行任务到最终产出模型往往需要数天之久, 无法满足下游任务和地理环境数字孪生由实到虚映射的实时性要求。因此, 如何在无人机有限机载算力和存储资源下实现在线数据获取和实时算法处理是当前地表环境数字模型构建所面临的核心问题。

本文结合自身研究基础提出一个基于通用实时定位与地图构建框架的无人机平台地表环境实时稠密点云生成与数字模型构建算法。在包含城市、农田、山区、戈壁等不同地表环境数据集的实验结果表明, 本文算法能在获得高质量地表环境稠密点云和数字模型的同时处理速度比现有算法(如: Pix4DMapper)快30—50倍, 平均在不到一秒钟的时间能内处理一幅高分辨率图像, 而一般航测无人机拍摄间隔也为1—2 s。将本文提出的算法应用于山洪灾害的应急救援工作——可实时重建救援人员无法到达的现场, 并勘测冲毁区、淤积区面积及定位, 辅助应急救援工作展开。

综上所述, 本文算法可在有限算力下实时重建地表环境稠密点云与数字模型。突破了现有航测算法的低时效瓶颈, 为基于无人机平台泛在测绘的大量普及和快速构建地表环境并实现现实地理环境向虚拟地理环境的实时映射提供了有效的技术支撑。

## 2 研究现状

地表环境模型构建过程中由于目标距离远、飞机载荷有限、环境干扰较大等因素, 通常采用基于视觉的被动式重建技术。相比主动式重建技术, 其不需要激光、声波、电磁波等能量源, 操作简单、易于实现, 成本低廉, 适用性强。本节将针对基于视觉的三维稠密点云生成与数字模型构建研究现状进行综述。

目前视觉环境重建最广泛使用的算法主要基于运动恢复结构 SfM (Structure from Motion) 和视

觉实时定位与地图重建 VSLAM (Visual Simultaneous Localization and Mapping) 技术, 分别应对离线精密重建和实时快速重建的应用需求。

### 2.1 基于 SfM 的地表环境重建算法

基于 SfM 重建的处理流程一般由特征点匹配、SfM 稀疏重建和稠密重建3大关键技术组成。特征点匹配的目的在于获得可靠的几何观测信息, 带有尺度信息的 SIFT 特征 (Lowe, 2004) 通过不同高斯标准差实现在尺度空间上的极值搜索, 能够提取对尺度、方向和噪声鲁棒的特征点被广泛应用图像拼接和三维重建等领域。Bay 等 (2006) 基于 Hessian 矩阵并结合箱式滤波实现 SIFT 特征快速计算, 提出了 SURF 算子, 将 SIFT 特征提取速度提升了约百倍。Rublee 等 (2011) 将 FAST (Rosten 和 Drummond, 2006) 特征点检测法和 BRIEF (Calonder 等, 2010) 特征描述符结合提出 ORB 特征提取算法, 在 SURF 的基础上又将特征提取速度提升了接近十倍。

在实践中, 由于利用特征描述子进行匹配往往包含错误匹配, 常常利用几何约束结合随机抽样一致性算法 RANSAC (Random sample consensus) 剔除。随着深度学习技术发展, 其对于深层语义特征强大的学习与表达能力提高了特征表达算法的在极端外观变化下的鲁棒性, 基于深度学习的特征匹配技术近几年得到了广泛关注。其中 SuperPoint (DeTone 等, 2018) 和 SuperGlue (Sarlin 等, 2020) 分别实现了基于深度学习的特征点提取和图匹配, 在光度变化、大视角差异方面表现优异, 但提取和匹配速度相对较慢, 无法满足实时性应用需求。Barroso-Laguna 和 Mikolajczyk (2023) 提出了一个结合手工制作和卷积神经网络学习的深度特征的多尺度关键点检测算法, 在保证匹配鲁棒性的同时大幅提高了其匹配效率。Wang 等 (2023) 将旋转群特征及局部旋转引入关键点监测和特征描述, 显著提高了点云配准的可靠性和效率。

在获取到图像之间的关联信息后, 基于 SfM 的稀疏重建部分会负责估计每个图像的拍摄位置, 特征点的三维位置和相机的内参信息。现有算法大体可以分为3类: 增量式、全局式和阶层式。其中增量式是从无序图像集合进行三维重建的一种流行策略, 已被多个影响广泛的开源重建算法采用

如 Bundler (Snavely 等, 2008), VisualSFM (Wu, 2011), COLMAP (Schönberger 和 Frahm, 2016) 等。增量式算法通过逐步添加新的图像以完成重建, 随之带来的明显缺陷是累积误差, 常常需要引入回环检测来减少偏移。为了避免重复执行耗时的后端优化同时克服漂移问题, 全局 SfM 重建算法应运而生, 与增量式不同, 它同时考虑整个视图, 而不是增量地向重建添加越来越多的图像, 因此仅需要一次后端优化即可完成重建 (Jiang 等, 2013)。但是在实际使用中全局式算法由于过分依赖正确的匹配结果, 其稳定性、重建精度和完整性比增量式差。与增量 SfM 重建和全局 SfM 重建算法相比, 分层 SfM 具有更好地并行处理能力, 一般被用于大规模的地图重建中。

稠密点云的生成主要基于多视图立体匹配理论, 一般可以分为点云法、深度法和体素法。早期的稠密点云估计常常在稀疏点云的基础上进行稠密化 (Lhuillier 和 Quan, 2005; Goesele 等, 2007), 其中比较有代表性的是 PMVS (Furukawa 和 Ponce, 2010), 容易产生大量缺漏。相比而言, 基于深度图的方法则更简洁灵活 (Galliani 等, 2015), 其中比较经典的算法是 PatchMatch (Tola 等, 2012; Barnes 等, 2009), 其重建效果好且稳定, 被广泛应用在各类三维重建软件中。而体素法则被深度学习方法广泛采用, 包括 MVSNet (Yao 等, 2018)、Fast-MVSNet (Yu 和 Gao, 2020) 等, 其特点是重建效果好, 但内存、显存开销较大, 难以适应面向无人机航拍的高分辨率图像重建的需求。NeRF (Mildenhall 等, 2020) 基于合成的角度实现了简单模型的三维重建, 将其与基于深度学习的体素法相结合, 可以降低算法对全监督深度信息的依赖并加速训练过程, 已成为三维重建领域一个新的研究热点, 如: NerfingMVS (Wei 等, 2021)、RC-MVSNet (Chang 等, 2022) 等。

地表环境数字模型的构建主要有基于二维和三维两种途径, 其中二维算法仅在地表环境较平坦时效果较好。基于等值面提取的 Marching Cubes (Lorensen 和 Cline, 1987) 方法和基于隐函数的泊松表面重建 (Kazhdan 和 Hoppe, 2013) 是最常见的基于点云求解模型表面的方法, 其中泊松表面重建方法能够根据采样比例和冗余自动调整插值, 无需显式参数设置, 自 2013 年 Kazhdan 开源其加速实现代码后被广泛使用。在网格生成后, 可以

进一步使用网格优化方法进行处理, 从而使表面更平滑及合理, 目前比较常用的网格优化方法包括拉普拉斯网格优化 (Nealen 等, 2006) 和最小二乘网格模型光顺算法 (Sorkine 和 Cohen-Or, 2004)。最后, 基于生成的网格模型和图像信息进行纹理贴图, 比较经典的实现是 Waechter 等 (2014) 年 ECCV 大会上提出的方法, 它采用图割 (Boykov 等, 2001) 的方法来选择视图, 并通过泊松编辑等技术让纹理自然过渡, 其纹理贴图方案至今仍被大部分三维重建系统参考或采用。

依托 SfM 技术发展了很多可用于地表环境稠密点云生成和数字模型重建的开源项目和商业软件如 OpenDroneMap、Pix4DMapper、DJI Terra 等。具有精度高的优点, 但普遍存在运算复杂度高、处理速度慢的问题, 无法满足大范围实时重建的需求。

## 2.2 基于 VSLAM 的地表环境重建算法

SLAM (Simultaneous Localization and Mapping) 技术的最早提出是为了解决无人设备在未知环境下的定位导航, 是无人机在未知环境中执行任务的首要 and 必需功能。基于应用层面可分为激光 SLAM 和视觉 SLAM (VSLAM), 由于无人机载荷限制和图像传感器成本低廉, 隐蔽性好等因素, 本文算法基于视觉源数据展开, 因此本节只对基于 VSLAM 的重建算法进行综述。

随着 VSLAM 技术不断发展, 越来越多研究人员尝试基于 VSLAM 完成快速乃至实时的地图生成, 其也被认为是 SfM 技术的实时版本。早期的 VSLAM 大多使用的是基于滤波的方法, 主要分卡尔曼滤波 (Bailey 等, 2006) 和粒子滤波 (Beevers 和 Huang, 2007) 两类, 难以有效利用大量观测信息, 研究人员探索了许多新的方法。PTAM (Klein 和 Murray, 2007) 是首个被熟知的基于图优化的 SLAM 系统, 由此开启了基于图优化的 SLAM 的大门。Strasdat 等 (2012) 从理论和实践角度讨论了滤波方法和图优化方法的优劣势, 并从此确立了图优化 SLAM 的主流地位。

Mur-Artal 等 (2015) 提出的 ORB-SLAM 基于 ORB 特征对传统基于 SLAM 的三维重建架构进行了大量优化, 并将单目 SLAM 带入了可实用的阶段。Bu 等 (2016) 在 IROS 会议上提出了一种基于单目 SLAM 技术的增量式自适应权重多波段方法—

Map2DFusion, 通过融合 GPS 信息辅以快速的图像融合算法, 成功展示了大场景下通过输入视频流, 系统可实时完成正射影像拼接。Hinzmann 等 (2018)、Wang 等 (2019) 实现了一个能增量生成稠密点云、DOM 和 DSM 的框架。然而, Hinzmann 的方法直接将稠密的点云投影到网格上生成高程图, 可能会导致缝合线周围出现轻微的颜色偏差。Wang 等 (2019) 的方法使用稀疏点云作为输入数据, 这导致在复杂地形上的效果较差。此外, OpenREALM (Kern 等, 2020) 尝试了利用 SLAM 技术在线生成数字高程模型和正射影像, 但所生成的结果精度和分辨率较差, 无法满足精细测量中准确度的需求。NICE-SLAM (Zhu 等, 2021) 通过引入层次式场景表示来整合多层局部信息, 实现了室内场景的稠密重建, 但由于不执行闭环检测, 不适用于长时间及大范围场景。

除上述基于多视图几何的 SLAM 重建系统外, 深度学习也在尝试改进和优化基于 SLAM 的三维重建算法。如, SfMLearner (Zhou 等, 2017) 尝试使用无监督神经网络解决单目深度估计同时也利用网络来估计图像的位置变化; CodeSLAM (Bloesch 等, 2018) 不仅使用深度学习来估计单目深度, 同时利用 SLAM 的位姿进一步优化深度估计;

Block-NeRF (Tancik 等, 2022) 将场景分解为单独训练的 NeRF, 使渲染能够扩展到任意大的环境; Mega-NeRF (Turki 等, 2022) 实现了无人机视角下的大规模三维建模, 但基于深度学习的方法普遍存在训练费时、需要大量数据以及无法实时重建和渲染的问题。

综上所述, 现有基于图像的三维重建算法大都基于 SfM 技术, 但其实时性无法满足许多下游任务的需求。基于 VSLAM 的重建算法又存在精度低无法满足测绘需求的问题。如何在保证精度的同时提高算法的实时性是现今视觉三维重建领域需要解决的主要问题。

### 3 实时地表环境稠密点云生成与数字模型构建

本文所设计实时地表环境稠密点云生成与数字模型构建算法与现有算法如: OpenMVS、COLMAP 相比, 能够在保证精度的情况下, 在速度上快 1—2 个数量级, 实现在无人机机载设备 (如 Intel NUC, DJI Manifold) 下实时运行的目的。总体算法流程如图 1 所示, 主要包括通用实时定位与地图重建框架、基于 PI-SLAM 的位姿解算、三维稠密点云生成、数字模型构建四大关键技术。

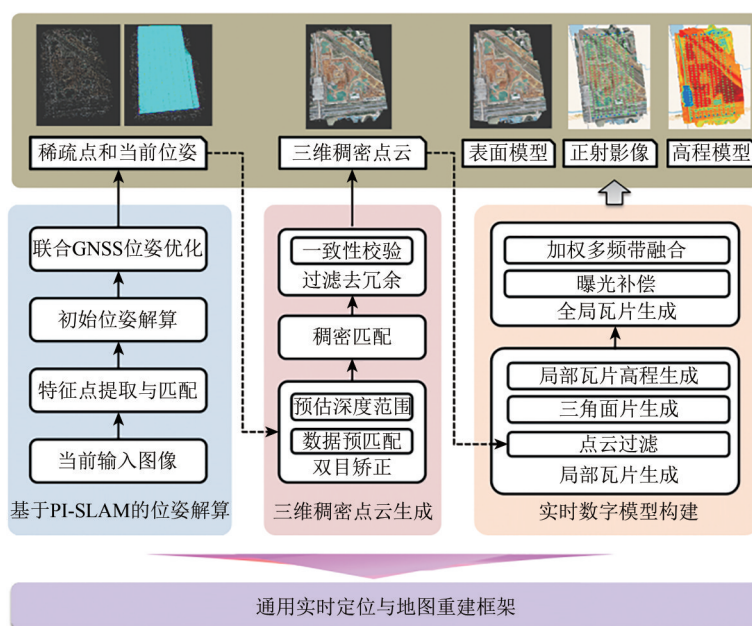


图1 总体算法流程图

Fig. 1 Real-time reconstruction algorithm framework

其中, 通用实时定位与地图构建框架为算法各个模块统一设计接口, 使得数据传输和模块间

消息传递效率大大提高, 算法性能得到最大程度发挥; 实时位姿解算首先解算初始化位姿, 通过

设计多项优化策略可加速获取鲁棒匹配,然后联合GNSS优化位姿,在解决漂移的同时使得建图过程具有真实的尺度信息;三维稠密点云生成包含数据预匹配、预估深度范围、稠密匹配、点云过滤和去冗余4个步骤,借助CUDA加速运算实时生成场景稠密点云;数字模型构建首先生成局部瓦片,这一步骤只依赖当前帧,采用多线程加速计算可达到实时的目的,在得到局部瓦片(即一个个高程图、正射图和权重图)后增量式处理局部瓦片来不断更新全局瓦片,并结合曝光补偿和加权多频带融合算法实现实时构建高质量拼接和平滑过渡的地表环境数字模型。

本文将逐个对算法细节进行介绍,最后基于多个数据集从稠密点云和模型还原度以及实时性角度对算法进行评估。

### 3.1 通用实时定位与地图重建框架

实时性需求不仅需要算法设计突破精度和效率难以双全的平衡问题,还对流程规划和基础框架提出了更高要求。传统的三维点云生成和数字模型构建算法大都基于离线批处理的方式,算法模块较多且接口不一致,使得数据传输和模块间消息传递耗时,限制了算法性能的发挥。同时,几乎所有的研究者都将注意力集中在自身理论的实现上,各个算法模块耦合度较高而弹性不够,不容易将其移植在其他系统上,这也在一定程度上阻碍了领域发展。

本文基于团队研究基础设计了一个通用的实时定位与地图构建框架,支持除C++外,Python和JavaScript等多种编程语言跨语言调用,有助于实时定位与地图重建前沿算法的集成,同时提高了模块测试代码的编写效率。解耦了从数据处理到算法开发、方案评估整个应用流程,可服务于无人机实时建图应用的整个生命周期,契合数字孪生地理环境构建中全生命周期平行仿真与决策推演的需要。具有可处理海量图像、计算高效,可在手机等嵌入式设备上移植,适用于多种复杂场景构建需求等特点。其整体框架流程如图2所示。

主要用于处理3个部分的交互:传感器数据集,实时定位与地图构建实现和应用程序。连接各个部分主要依靠两个关键组件:通用多编程交互接口Svar(Zhao等,2021)和高性能消息发布订阅组件Messenger。其中,Svar是最核心的组件,

其组织了所有算法和应用程序插件的实现,提供了多语言的交互能力,以及命令行参数的解析和传递,还是Messenger中消息的具体呈现形式。可以将其理解为一个高性能跨语言调用的桥梁,对于C++开发者,Svar可以提供类似于Python等语言运行时库导入的特性,让程序的依赖和编译变得简单,也使其更新和发布变得弹性,函数包内数据结构的改变不再影响程序的运行,通过Svar封装的动态链接库可以方便地被其他语言导入使用;对于Python、JavaScript等其他编程语言,也可通过Svar加载程序包并暴露给其他语言调用。Svar的多编程语言交互特性有助于解决基于Python等其他编程语言开发的算法模块,难以在基于SLAM算法的系统中部署的难题。Messenger模块的设计受ROS2消息传递体系结构的启发,进一步简化了用户操作,能够跨域的函数接口可以实现多点到多点的直接通信,算法各个模块之间的大部分通信都依赖该模块进行,在不给用户带来负担同时也可达到兼容现有的生态的目的。

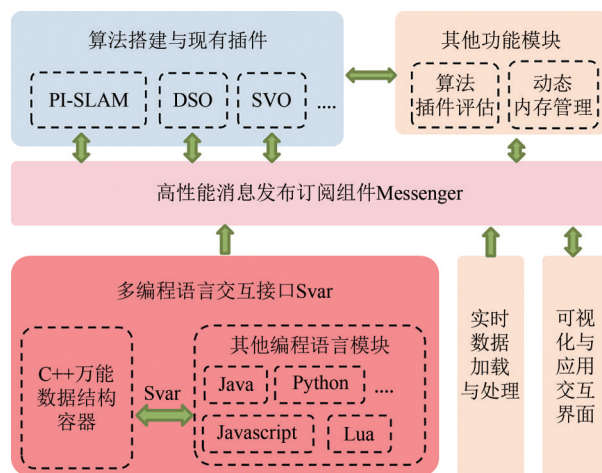


图2 通用实时定位与地图构建框架

Fig. 2 General framework of simultaneous localization and mapping

除此之外,面向大场景重建下机载内存不够的问题,我们仔细设计了程序内存使用,动态调整计算内容,避免出现内存爆炸的情况。动态内存管理采用一种基于LRU(Least Recent Use)策略的哈希分片缓存技术,避免了内存溢出。在处理高分辨率数据时,瓦片细节等级一般比较高,数百上千兆的数据很快就会将内存消耗殆尽。因此本文设计了一个基于LRU的哈希表,将长期不用以及未更新的瓦片,存储在硬盘中,缓解了算法

内存不足问题。此外显示模块借鉴 3D Tiles，采用多分片多层次（Multi-Layer, Multi-Details）的思想，根据显示设备分辨率和距离来调整显示内容，保证流畅显示。

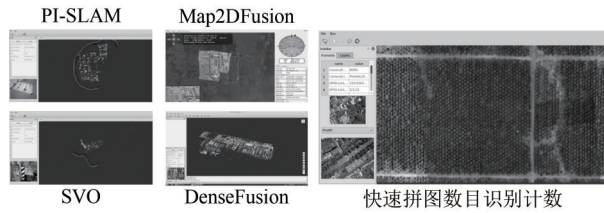


图3 基于框架实现的部分算法示例

Fig. 3 Part algorithms based this frame

所提出的框架已经在基于无人机平台的软件开发和实践中得到应用，实践表明，基于该框架能够有效集成团队自研算法：联合 GNSS 优化的大范围 SLAM 算法—PI-SLAM、实时数字正射影像图拼接算法—Map2DFusion、实时三维稠密重建算法—DenseFusion（Chen 等，2020）和基于实时正射影像图拼接的快速树木识别计数算法（Han 等，2022）以及其他研究 SVO（Forster 等，2014）等。目前该框架和 Svar 均已开源，研究人员可以直接使用框架中现有的功能或基于核心组件创建自己的功能模块来快速实现基于 SLAM 的智能算法，能够大幅缩短算法研发和项目实践周期，有望进一步推动无人机快速建图与相关智能算法的研发和应用落地。

### 3.2 基于 PI-SLAM 的实时位姿解算

PI-SLAM 实时位姿解算模块由初始化位姿解算和联合 GNSS 优化两部分组成，可有效解决现有 VSLAM 算法大场景下的定位漂移问题，在实时解算的同时有效保障定位精度。

(1) 初始化位姿解算。对于无人机拍摄的每一张图像，定位过程首先使用 SIFTGPU 在线提取特征点并计算描述子。SIFTGPU 是 SIFT 算法的 GPU 实现，能利用 GPU 并行处理图像像素，大大提高了计算速度，比 ORB 特征点更快。提取完特征点之后需要对特征点做匹配。匹配过程采用 RootSIFT（Arandjelović 和 Zisserman，2012）方法，通过对描述子归一化并计算平方根，来度量两个 SIFT 特征点的距离，该方法优于直接采用欧氏距离度量相似性的方法。另外，鉴于直接暴力匹配比较费时间而且效果不好，设计了如下策略快速获取鲁棒匹配：

1) 快速最近邻逼近搜索 FLANN（FastLibrary for Approximate Nearest Neighbors）（Muja 和 Lowe，2009）实现特征点快速匹配。一般的暴力匹配复杂度为  $O(n^2)$ ，基于 FLANN 将描述子转化到高维空间进行搜索，可以将复杂度降至  $O(n \log(n))$ 。

2) 基于 Lowe's 算法选取足够优秀的匹配点。如果搜索空间中最近点的距离除以次近点的距离小于某个值，则保留最近点，反之剔除。

3) 交叉匹配。交叉匹配本质上是保证特征点匹配的一致性。如果特征点 A 认为 B 是它的最佳匹配，同样的，B 也应该认为 A 是其最佳匹配。

4) 随机采样一致性 RANSAC（Random Sample Consensus）。RANSAC 算法能够通过不断循环迭代的方式从存在噪声的数据中找出符合模型的数据。

实验证明，通过以上策略过滤并计算本质矩阵，匹配过程耗时仅 7 ms，然后通过对本质矩阵奇异值分解即可得到相对位姿。

(2) 联合 GNSS 优化位姿。在无人机执行航拍任务时，默认 GNSS 信息始终可用，因此可以联合 GNSS 信息进行优化。GNSS 信息使得建图过程具有真实的尺度信息，这使得传统单目 VSLAM 中存在的尺度漂移问题变得容易解决。损失函数包括两部分：重投影误差  $l_r$  和 GNSS 误差  $l_g$ 。其中，重投影误差代表了视觉误差，表示地图点投影到图像平面后与匹配特征点的位置误差；GNSS 误差则约束了视觉所计算出的结果的尺度，不至于使视觉的累计漂移过大。具体损失函数如下：

$$l_r = \arg \min_{\xi} \frac{1}{2} \sum_{i=1} \left\| \mathbf{x}'_i - \frac{1}{s_i} \mathbf{K} \exp(\xi) \mathbf{p}_i \right\|_2^2 \quad (1)$$

式中， $\mathbf{x}'_i$  代表当前帧上的特征点， $\mathbf{K} \in \mathbb{R}^{3 \times 3}$  为相机内参， $\xi$  是定义在李代数下的相机位姿， $\mathbf{p}_i$  是三维点坐标。GNSS 误差  $l_g$  定义为

$$l_g = \left\| \mathbf{t}_{\text{GNSS}} - \mathbf{t}_{\text{SLAM}} \right\|_2 \quad (2)$$

总的误差函数定义为

$$\text{loss} = l_r + \beta l_g \quad (3)$$

为了避免  $l_r$  和  $l_g$  之间差距较大，影响优化，需要添加一个权系数  $\beta$ 。若 GNSS 的朝向不可靠，则通过调低  $\beta$  降低 GNSS 误差的影响。当重叠率较低时，匹配点都集中在局部区域，采用基础的 PnP 方法估计位姿误差会较大。此时 GNSS 信息就显得很重要，GNSS 误差项能够修正定位的尺度。在非线性优化中，GNSS 测量误差以及错误的匹配点对误

差函数将会带来更大的误差,在优化过程中会导致结果向错误的地方偏移。因此需要在优化中添加鲁棒核函数,限制错误数据对最终结果影响过大。鲁棒核函数对较小的误差基本保持原样,对于较大的误差会保持抑制。经过效果测试,Cauchy核函数限制异常值影响的效果比较好。另外,定位模块基于Ceres库的非线性优化算法来估计参数,可以得到包含位姿信息和稀疏点云信息的帧,并发送到稠密化模块生成稠密点云。

### 3.3 实时三维稠密点云生成

生成稠密点云主要包括4个步骤:数据预匹配、预估深度范围、稠密匹配、点云过滤和去冗余,其中数据预匹配和预估深度范围用于将相邻两帧矫正为双目,匹配完成后需要对深度信息进行过滤,保留具有一致性的点。

(1) 数据预匹配。在进行双目匹配时,帧与帧之间的基线距离会影响对深度值的估计。因此要选取合适的两帧进行稠密化,既不能太远导致视差过大,占用太多内存;也不能太近导致误差无法接受。对此,我们构造了一个定长队列,每当新完成一帧计算,就尝试在队列中搜索合适基线的帧作为匹配对。

(2) 预估深度范围。匹配过程中会生成一个 $W \times H \times D$ (分别代表图像宽、高和搜索范围)大小的代价矩阵,存储着聚合代价,这一步通常是最占用内存的环节。由于无人机硬件限制,其搜索范围 $D$ 也是有限制的,需要根据先验知识设置一个合理的搜索范围,缩小内存占用。由于PI-SLAM能够同步建图,我们通过PI-SLAM环节产生的地图点来估计场景的深度范围。需要注意的是,三角化出来的点也存在许多噪点,对此我们设计了一种基于密度的搜索方法来准确估计场景的深度范围。设当前帧中所有被观测到的地图点集为 $P_w$ ,当前帧位姿为 $(R, t)$ ,则当前帧中深度集为 $\{d|P_c = RP_w + t, d = P_c(2)\}$ 。具体如下:

- 1) 对 $d$ 进行排序,方便后续处理;
- 2) 确定距离阈值 $d_{thresh}$ ,设最大值为 $d_{max}$ 最小值为 $d_{min}$ ,密度阈值 $thresh = 0.1$ ,数量阈值 $n=10$ ;
- 3) 集合 $d$ 中的每一个点,都判断其左右 $d_{thresh}$ 距离内的点数量,如果数量超过 $n$ 则保留,若数量不足则判断为噪点,剔除;
- 4) 设新的深度集为 $d$ ,重新进行一遍以上流

程,最后取得最大最小值。

范围搜索算法是根据PI-SLAM产生的地图点的特点所设计的。其难点在于在城市场景中,高楼产生的少量有效的群聚点跟噪点的区分。而本文提出的基于密度的搜索算法,能够将密度过低的聚集点过滤掉,从而实现有效范围搜索。

(3) 稠密匹配。完成预匹配和深度范围预估后,还需要进行双目矫正。本文采用Bought's rectification方法用于最大化前后两帧的公共区域。采用SGM的CUDA加速版本来实现快速匹配。在KITTI数据集的测试结果显示,该CUDA加速的SGM算法在图像大小为 $1024 \times 440$ 像素,搜索视差范围为128,在采用4通道聚合的情况下,能够达到上百的帧率。

实际使用过程中,航拍数据动辄千万级像素数,搜索的视差范围也远不止128个像素。以 $4000 \times 3000$ 大小的图像为例,搜索256个像素匹配所构建的代价矩阵耗费 $4000 \times 3000 \times 256 \times 8 = 2457600000$  bytes = 22.8 GB。鲜有显卡有如此大的显存。因此本算法将根据显存的实际剩余量来调整图像切片,用时间换空间,减小内存的使用。

(4) 过滤去冗余。由于视差图不够精确,以及深度图中也存在一些错误,本文采用多个连续图像的一致性约束来重新确定每个点的深度值。

每当计算完成一个深度图后,都将其插入到一个队列中。当队列长度超过 $N$ (一般是奇数)即开始过滤。设队列中间帧为当前帧,将深度图通过相机投影公式转化三维点之后,再重新投影到其余的 $N-1$ 个帧上,得到了一个二维坐标点 $p_i$ 和深度值 $d'_i$ 。设该投影点在当前帧上的深度值为 $d_i$ ,对于任意一个被投影帧,设在该帧上的深度值为 $d'_i$ 。若满足则认为该帧与当前帧在该点处是一致的。如果该点在至少 $N_{con}$ 个帧上是一致的,那么该点就被保留,反之则被剔除。对于所有保留点,他们的深度值取为观测一致性点的均值,然后将其余帧的一致性置为0。本文中 $N$ 为7, $N_{con}$ 为4,阈值 $f_{eq}$ 为0.01。

经过一致性检验的深度图就可以输出为点云。输出过程会检查队列的最后一帧的一致性矩阵,如果某个点的一致性为0或是一致性小于 $N-1$ 则不输出该点,抑或者其深度值过小或者过大也不输出该点。对于其余点反投影到三维空间后发送给

地理信息融合部分进行处理。图4展示了岛屿数据集的表现,可以看出,一致性检查能够有效删除噪声点。

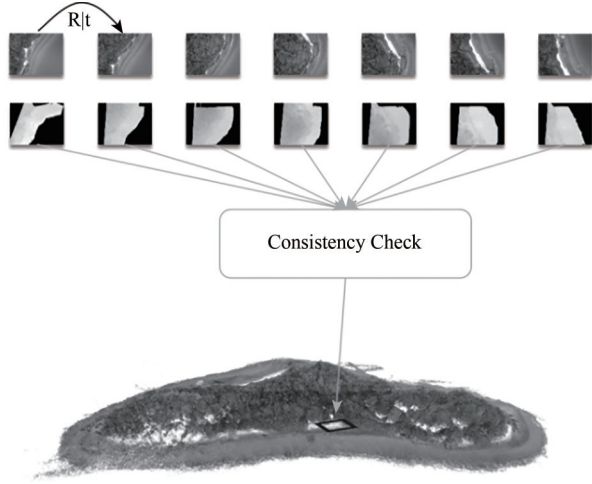


图4 生成稠密点云时基于一致性检查有效去除噪声点  
Fig. 4 Consistency checking can effectively remove noise spots when generating dense point clouds

### 3.4 实时数字表面模型构建

实时数字表面模型构建基于瓦片地图原理,包含局部瓦片生成和基于局部瓦片增量式更新融合全局瓦片两部分。

#### 3.4.1 局部瓦片生成

局部地图主要是为了求出当前点云覆盖区域的局部瓦片,后续将整合到全局地图中。最终生成的正射影像和高程图都是基于地图瓦片原理。为了提高计算速度,本文先针对每个计算完成的点云计算局部瓦片,再将瓦片融合到全局瓦片中。先局部后全局的处理方式有利于并行化处理,可以提高运行速度。目前本文产生的瓦片地图细节等级  $n=21$  (0.138 m/pixel)。

为了方便跟全局地图融合,局部地图中的瓦片不仅包含像素,还包含了一些融合的必要信息。一个瓦片包含3层:深度层用来生成高程图,权重图决定了不同瓦片的融合方式,图像层包含了影像。

(1) 点云过滤:点云过滤出于两个方面考虑,一方面21级的瓦片地图分辨率为0.138 m/s,而稠密点云的分辨率要高于这个水平,因此对于生成瓦片来说,稠密点云是有冗余的。另一方面本文最终生成的地图是2.5D的,不考虑需要全部的三维信息。因此在点云过滤环节需要先将点云降采

样,然后对于XY位置相同但是高度不同的点,只保留较高的点,这样就大大减小了后续的计算量。

(2) 三角面片生成:为了快速生成具有地理参考的3D网格,我们对当前点云对应的2D点采用Delaunay三角剖分,获取到一系列三角面片  $M_{\text{ing}}$ 。由于点云数据跟这些2D点是一一对应的,因此也就获得了三维空间中的一系列三角面片  $M_{\text{depth}}$ 。将这些三维点都转到ECEF坐标系下,然后再转到瓦片坐标系下,又获得了在瓦片坐标系下的一系列三角面片  $M'_{\text{depth}}$ 。

(3) 局部瓦片高程生成:高程是通过内插三角面片  $M'_{\text{depth}}$  生成的。由于稠密点云足够密,因此线性插值已经能满足不错的精度。设  $p_1, p_2, p_3, p_4$  已知点  $p_1, p_2, p_3$  和  $p_4$  的  $x_4, y_4$  值,求  $z_4$ 。由于可以定义一个面,求该面法向量  $N$ :

$$N = (p_2 - p_1) \times (p_3 - p_1) \quad (4)$$

由面上的点垂直于法线得:

$$(p_2 - p_1) \cdot N = 0 \quad (5)$$

整理得到式(6):

$$z_4 = z_1 - \frac{(x_4 - x_1)N \cdot z + (y_4 - y_1)N \cdot y}{N \cdot z} \quad (6)$$

由于点云十分密集,在插值的时候为了避免重复计算问题,设计Top-left规则来判定面片边缘部分插值规则。当瓦片中的像素中心在面片内时,该像素自然属于此面片。当像素中心在边缘部分时根据Top-left规则来判定归属。

(4) 局部瓦片正射生成:由于深度信息和图像中的位置都是已知的,那么正射影像的求解过程就变得简单了。由于  $M'_{\text{depth}}$  和  $M_{\text{ing}}$  中的三角面片是一一对应的,只需要将求解单应矩阵,将其中的像素映射过来即可。

(5) 权重系数生成:权重系数决定全局融合时哪些像素点被优先融合。权重系数由两部分决定。第一部分是正射系数  $w_o$ ,很显然,对于正射图而言,越靠近图像中间部分其正射性就越好。因此我们将距离图像中心的点置为1,把最边缘部分的点置为0。设图像上像素  $i$  距离中心点为  $d$ ,介于0到  $d_{\text{max}}$  之间。那么  $w_o^i = 1 - d_i/d_{\text{max}}$ ,第二部分是深度置信度系数,这与其距三角面片边缘距离有关,距离越远则置信度越低,同样的根据线性插值计算出  $w_d^i$ 。最终的权重  $w_i = w_o^i + w_d^i$ 。权重系数的设置优先保证靠近中心的像素点以及靠近点云的像素点被保留。可以看到,局部瓦片的生成操

作只依赖当前帧,通常也较为耗时,实际使用采用多线程加速计算。

### 3.4.2 全局瓦片生成

在得到局部瓦片之后,就得到了一个个高程图,正射图和权重图。由于本算法能够增量式的处理数据,因此全局瓦片是不断更新的。为了实现高质量的拼接以及平滑过渡,算法采用结合曝光补偿的加权多频带融合算法。曝光补偿是为了解决局部瓦片融合过程中发生的曝光不一致问题;而加权多频带融合则能让局部瓦片与全局瓦片重叠部分平滑过渡。

(1) 曝光补偿:目前常见的曝光补偿算法大多更注重效果,没有考虑到处理的效率问题。常规的曝光补偿算法为每幅图像都提供一个增益系数 $\rho$ ,它计算重叠部分全像素的强度差:

$$\tau = \frac{1}{2} \sum_{u_i \in R(i,j)} (\rho_i I_i(u_i) - \rho_j I_j(u_j))^2 \quad (7)$$

$\rho_i$ 和 $\rho_j$ 分别为图像 $I_i$ 和 $I_j$ 的增益系数。 $R$ 表示两图重叠区域,但是这种方法并不适合于此使用。一方面因为计算较为耗时,不满足实时要求;另一方面航拍图一般重叠率在70%—90%,重叠区域全像素处理会耗费过多计算资源。

因此借助于PI-SLAM提供的稀疏匹配关系对局部瓦片帧临接的帧求增益系数,如下式(8):

$$\tau = \frac{1}{2} \sum_{j=1}^n \sum_{u_i \in r(i,j)} (\rho_i I_i(u_i) - I_j(u_j))^2 \quad (8)$$

式中, $r(i,j)$ 代表图像 $I_i$ 和 $I_j$ 中具有共视关系的特征点。 $i$ 代表当前帧, $j$ 代表相邻帧。遍历与当前帧具有共视关系的所有帧,将共同观测的特征点计算其光强差后用最优方法求解参数 $\rho_i$ 。

(2) 加权多频带融合:解决了相邻帧曝光补偿问题,还需要将重合区域平滑融合。加权多波段图像融合方法将图像在高维空间融合后再逐层恢复到原尺寸,融合较为平滑。

对于局部瓦片中的每个瓦片来说,在往全局瓦片融合时,如果全局瓦片中不存在该瓦片,则直接复制过去即可。如果全局瓦片中已存在该瓦片。我们会对高程图和正射影像建立五层的拉普拉斯金字塔,对权重图建立五层的高斯金字塔。拉普拉斯金字塔第 $L_n$ 层的计算方式为1) 计算五层的高斯金字塔 $G$ 。2)  $L_n$ 定义为高斯金字塔中 $G_n$ 与 $G_{n+1}$ 的上采样的差。如下式(9):

$$G_{n+1} = \text{DOWN}(G_n \otimes \zeta_{5 \times 5}) \quad (9)$$

式中, $\otimes$ 是卷积符号, $\zeta_{5 \times 5}$ 代表高斯卷积核。由于没有计算 $G_6$ ,因此令拉普拉斯金字塔的最高层 $L_5$ 等于高斯金字塔的最高层 $G_5$ 。至此,这里一共建立了4个拉普拉斯金字塔: $L(\text{DEMI})$ ,  $L(\text{DEMG})$ ,  $L(\text{orthol})$ ,  $L(\text{orthog})$ ,以及两个高斯金字塔 $G(\text{weightl})$ 和 $G(\text{weightg})$ 。权重图组成的高斯金字塔是为了给拉普拉斯金字塔融合提供遮罩。

融合过程实际上并不直接对原图进行融合,而是先将拉普拉斯金字塔融合。对于拉普拉斯金字塔的每一层,都参考局部权重 $G(\text{weightl})$ 和全局权重 $G(\text{weightg})$ ,保留权重较大的部分。最后新的全局瓦片就按如下式(10)所示,从拉普拉斯金字塔中恢复出来。

$$G_n = L_n + \text{UP}(G_{n+1}) \quad (10)$$

经过以上步骤,无人机在拍摄时所造成的曝光不同和重叠区域的高程差都会被平滑。

### 3.5 实验结果与方法对比

本节将对本文提出的实时稠密点云生成与数字模型构建方法进行评估。评估所用数据集为DroneMap2数据集,大多数来自大疆无人机拍摄,包含城市、农田、岛屿、山区、戈壁等不同地表环境,图像分辨率均为4 K,有效像素达到千万量级。

(1) 稠密点云评估。相比于Pix4DMapper中使用的传统多视角立体视觉MVS(Multi View Stereo)重建三维场景,本文方法基于双目匹配融合策略也能够实时输出高质量的点云结果。由于基于MVS的方法在空三阶段就能够获取到场景的稀疏信息,这在一定程度上给后续处理带来了方便,因此从理论上来说,增量式的方法精度相比于MVS方法要差。对此,本文采取了一致性检查过滤、加权融合等措施来提高点云的质量。

本节将算法生成的稠密点云与Pix4DMapper的稠密重建结果进行交叉对比来评估算法的相对精度。在运行环境上,由于算法整个在Linux系统开发完成,Pix4DMapper只有Windows版,因此两者运行时的平台不同,但是硬件相同。在参数设定上,Pix4DMapper采用高质量模式来获取稠密点云,而本算法仅使用默认参数。在评估策略上,计划将双方输出的稠密点云导入CloudCompare2后,通过迭代最近点算法ICP(Iterative Closest Point)来将点云对齐,并通过点到最近点的距离来衡量

点云之间的差异。

具体数据选择 DroneMap2 中的 mavic-library、mavic-warriors、mavic-garden、mavic-yangxian 和 pr4-village，包含了村庄、山脉、城市建筑物以及街道等场景。图 5 显示了本文方法和 Pix4DMapper 之间

的错误分布图像和统计结果。从统计图表中发现，在图 5(a) 中 90% 的稠密点距离 Pix4DMapper 结果在 1.31 m 以内，50% 在 0.63 m 以内。在图 5(b) 中 90% 在 1.42 m 以内，50% 在 0.95 m 以内。

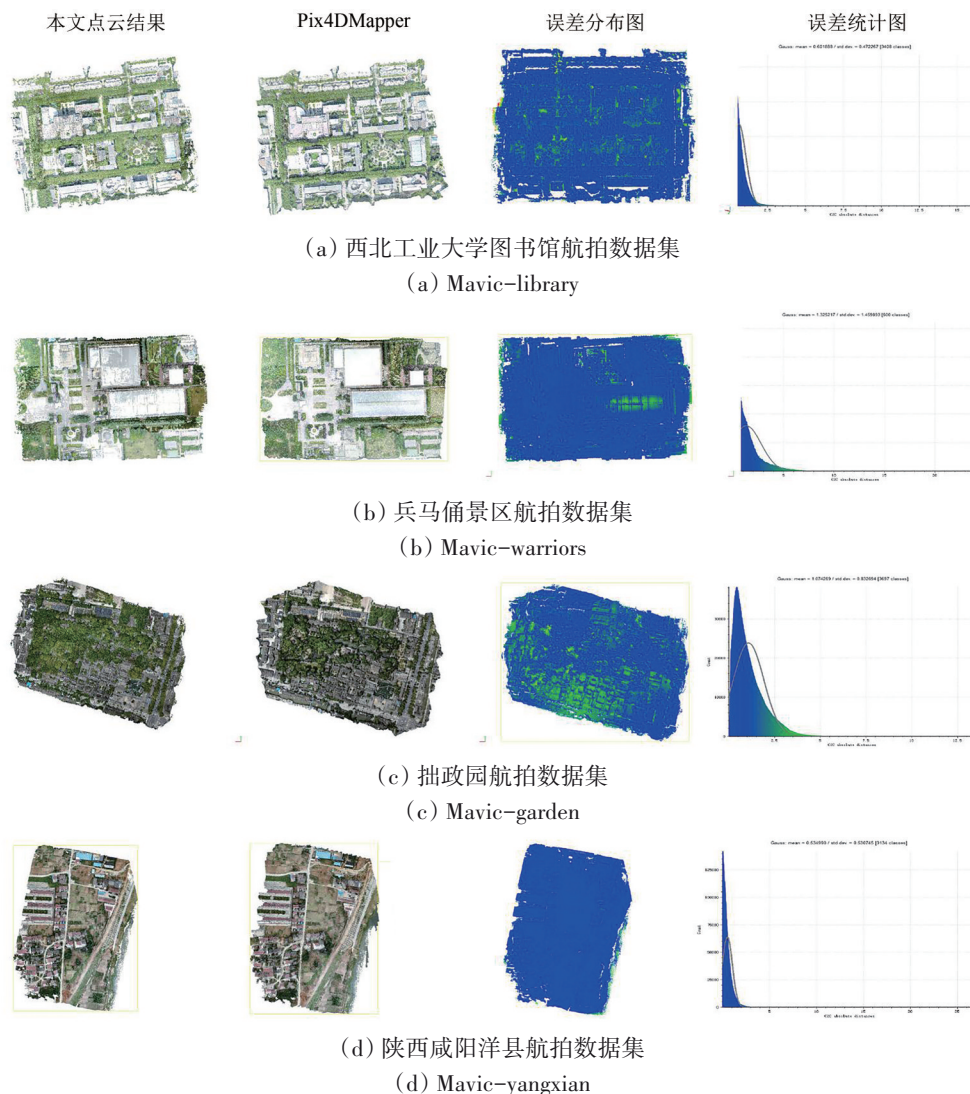


图 5 本文算法与 Pix4DMapper 的稠密点云评估

Fig. 5 The dense point cloud evaluation comparison with Pix4DMapper

对于 mavic-campus 数据集来说，293 张图一共产生了 1.4 亿个点融合到最终的点云文件中。由于我们的框架是实时生成点云的，因此不包含全局的光束平差。然后通过加入 GNSS 约束，以及融合模块选取了一系列可靠的支撑点，保证了全局尺度的准确性，同时在流程上，一致性检验剔除了异常值，保证了点云质量。实验表明，本文所提的方法能够获得高质量的稠密点云，而且还可以

清楚地看到建筑、树木、沟壑等三维结构。

(2) 数字模型评估。算法基于稠密点云实时构建数字正射影像、数字高程模型和数字表面模型等地表环境数字模型，本节将对比另一个实时重建框架 TerrainFusion (Wang 等, 2019)，评估算法模型重建的效果。如图 6 所示，左侧的是本框架，右侧的是 TerrainFusion。可以看出 TerrainFusion 在平原，丘陵地带能够产生不错的结果，但是在

有建筑物的场景下, 其结果十分扭曲。而本文算法结果出自稠密点云, 因此细节相比于TerrainFusion要好得多, 在有建筑物的场景下重建精度更好。

实验证明本文依托稠密点云实时构建的正射影像和高程图, 质量都显著提升, 尤其是在地表起伏变化明显的复杂地形环境中。

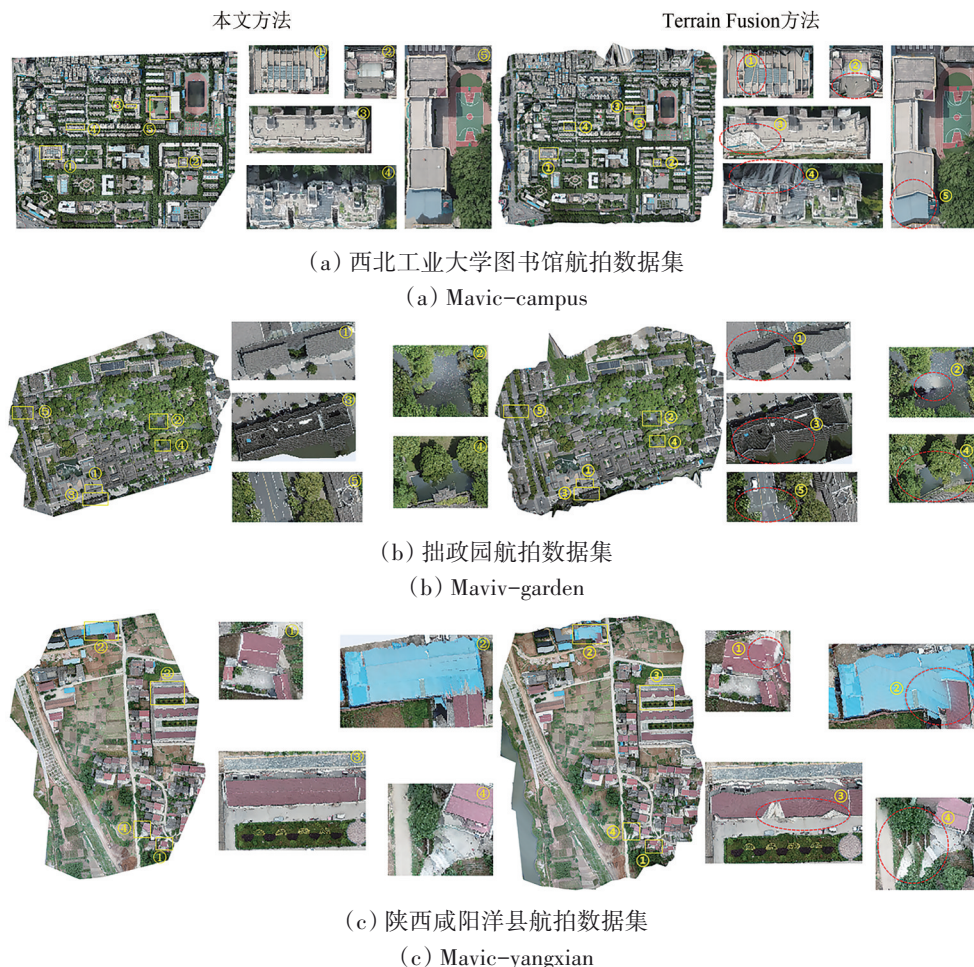


图6 本文算法与TerrainFusion数字正射影像模型评估

Fig. 6 The Digital Orthophoto Model (DOM) comparison with TerrainFusion

(3) 效率评估。本文的目的是实现一个实时的地表环境重建系统, 计算效率是我们的首要指标。传统软件例如 Smart3D, Pixi4DMapper 的速度非常慢, 而且使用起来专业性较强, 对计算设备的 CPU, RAM 和显存要求比较高。有时候如果配置不合适, 还有可能出现耗费几个小时计算失败的情况。本节比较了本框架和 TerrainFusion、Pix4DMapper 在同样配置电脑上运行相同数据集所花费的时间。

表 1 中展示了本文算法和 TerrainFusion、Pix4DMapper 在 DroneMap2 中不同数据集上的计算时间, 涵盖山地、城市、郊区和平原等不同地表环境。结果表明, 本文算法在定位、稠密重建和模型融合方面比 Pix4DMapper 快 30—50 倍, 平均

在不到 1 s 的时间能内处理一幅高分辨率图像。一般来说, 航测时无人机拍摄间隔也为 1—2 s。因此, 本文所提方法的速度能够满足基于高分辨率无人机航拍图像实时三维重建的应用需求。

## 4 应用案例

湖北恩施屯堡乡马者村沙子坝滑坡是恩施地质灾害重点监测对象之一, 属特大型土质滑坡。2020年6月8日入梅以来, 连续的强降雨诱发马者村沙子坝滑坡体滑移, 产生滑坡泥石流并堵塞清江形成堰塞湖 (程怀蒙等, 2021)。常规监测技术很难及时发现受灾地区, 更难获取受灾地区全面信息, 对抢险救援工作开展造成极大不便。

表 1 Dronemap2不同数据集处理时间对比

Table 1 The comparison of processing times of different data sets of Dronemap2 in seconds

| 数据集               | 图像数量 | 分辨率       | 飞行高度/m | 区域面积/km <sup>2</sup> | 处理时间/s        |        |        |
|-------------------|------|-----------|--------|----------------------|---------------|--------|--------|
|                   |      |           |        |                      | TerrainFusion | Pix4D  | 本文方法   |
| mavic-factory     | 359  | 4000×3000 | 153    | 0.545                | 1003.89       | 4811.0 | 106.23 |
| mavic-fengniao    | 109  | 4000×3000 | 60     | 0.073                | 377.302       | 3303.0 | 44.31  |
| Plantom3-olathe   | 160  | 4000×3000 | 122    | 0.257                | 161.295       | 2514.0 | 67.26  |
| mavic-yangxian    | 165  | 4000×3000 | 100    | 0.113                | 171.719       | 2935.0 | 40.89  |
| P4r-village       | 136  | 4864×3648 | 250    | 1.076                | 288.272       | 2762.0 | 61.50  |
| mavic-campus      | 293  | 4000×3000 | 213    | 0.974                | 368.182       | 4707.0 | 92.13  |
| mavic-warriors    | 95   | 4000×3000 | 143    | 0.157                | 218.324       | 1623.0 | 27.11  |
| Phantom4-mountain | 81   | 4864×3648 | 300    | 0.929                | 173.598       | 1752.0 | 39.90  |
| mavic-huangqi     | 229  | 4000×3000 | 188    | 0.478                | 809.678       | 4351.0 | 74.94  |

而现有遥感测绘技术主要基于卫星影像和航空遥感影像提供的地形信息，通过绘制山洪灾害评估本底影像图，预测灾害可能的冲毁区和淤积区，无法实时获取灾害发生现场真实的地表环境资料，可能造成救援工作延误。对此，我们基于本文提出的算法对沙子坝两段山体滑坡变形区域进行实时稠密点云和数字模型构建，实验时间为2020年7月。实时采集图像分辨率为4864×3648像素、数据量分别为131和191张、飞行高度为150 m和200 m、整

体面积为0.4 km<sup>2</sup>和0.65 km<sup>2</sup>，最终实时重建总耗时分别为62.56 s和91.89 s，重建结果如图7，由左至右分别为稠密点云、数字表面模型、数字高程模型和数字正射影像图。基于实时构建的数字正射影像图，通过QGIS软件可以快速得出滑坡变形区域的面积及地理坐标，其中，第一段变形区域的总面积为222329.479 m<sup>2</sup>，第二段为112436.588 m<sup>2</sup>，部分关键点地理坐标标注如图8所示。

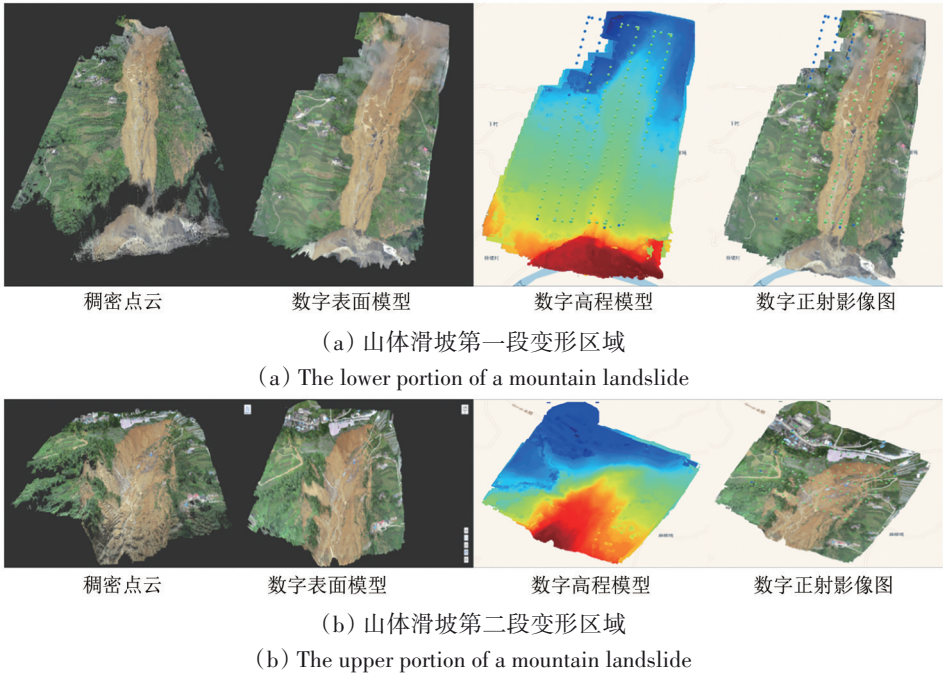


图 7 山洪灾害地表环境实时稠密点云与数字模型构建结果

Fig. 7 Real-time dense point cloud and digital model construction results of flash flood disaster

结果表明，使用本文算法能够在平均在半秒钟的时间内处理一幅高分辨率灾害发生地航拍图像，实时生成山体滑坡区域地表环境稠密点云和数字模型，还原灾区真实三维地貌；基于实时生

成的灾区地表环境数字模型还可实现快速分析测绘，得到灾害发生现场准确的变形区、冲毁区面积及位置，对抢时抢险指导应急救援工作，减小人员伤亡与经济损失都具有十分重要的意义。

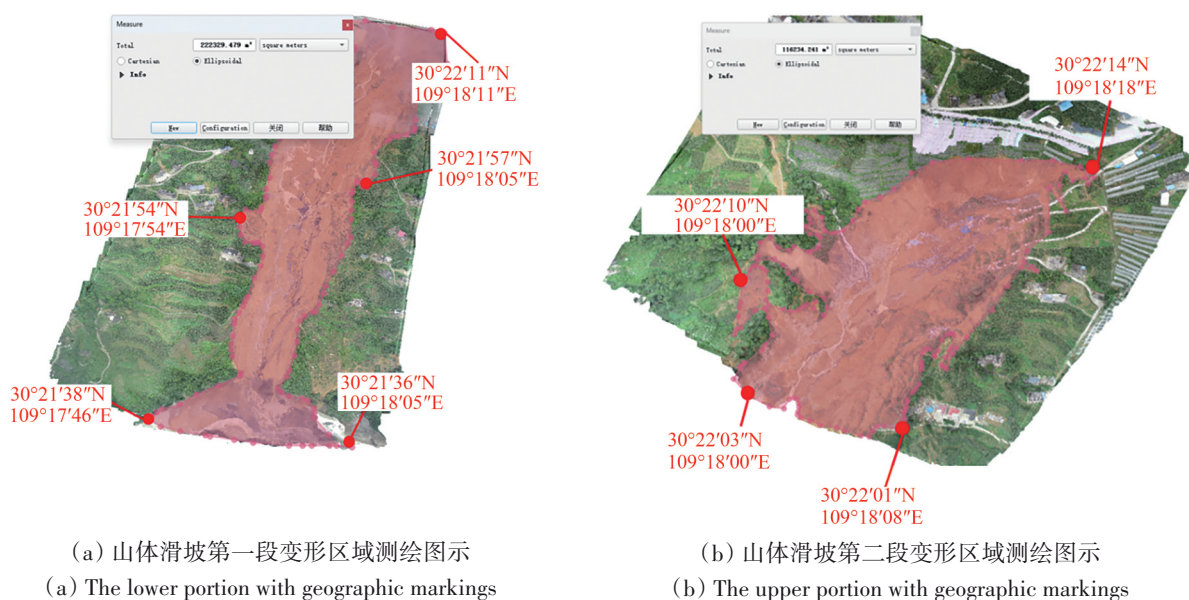


图8 基于实时构建的数字模型测绘分析结果  
Fig. 8 Mapping and analysis results based on digital model

## 5 结 论

针对当前地表环境三维数字模型构建中存在的速度慢、时效性低和大场景应用受限等问题, 本文提出一个基于无人机平台的地表环境实时稠密点云生成与数字模型构建方法, 通过设计通用实时与定位框架和动态内存管理模块提高算法时效性, 并研究联合 GNSS 优化的 VSLAM 算法—PI-SLAM, 在提高位姿解算时效性的基础上保障其精度。突破了大范围航测应用场景中地表环境稠密点云生成与数字模型构建精度和效率难以兼顾的难题。实验评估和应用案例表明, 研究形成的算法能够在地表模型构建精度与现有离线建模算法相当的同时, 速度提升几十倍, 达到在线采集数据并实时建模的程度。可为灾害预警、应急管理、协同作战等高时效需求任务提供全面的、第一手的地表环境信息。

另外, 由于算法成果突破了地理环境数字孪生构建中由实到虚实时映射的核心难题。因此将其赋能数字孪生, 为数字孪生地理环境应用提供数字底座, 在实时映射地表环境三维信息的基础上, 进行平行仿真和决策推演。还有望在动态监测场景目标信息的基础上, 实现紧急预警、方案推演评估、决策优化等功能, 进一步提高所应用系统的自动化和智能化水平。

志 谢 感谢赵勇博士和王伟对本文实时重建

算法所做的贡献以及对本文工作所提出的宝贵建议。

## 参考文献 (References)

- Arandjelović R and Zisserman A. 2012. Three things everyone should know to improve object retrieval//2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence: IEEE: 2911-2918 [DOI: 10.1109/CVPR.2012.6248018]
- Bailey T, Nieto J, Guivant J, Stevens M and Nebot E. 2006. Consistency of the EKF-SLAM algorithm//2006 IEEE/RSJ International Conference on Intelligent Robots and Systems. Beijing: IEEE: 3562-3568 [DOI: 10.1109/IROS.2006.281644]
- Barnes C, Shechtman E, Finkelstein A and Goldman D B. 2009. Patch-Match: a randomized correspondence algorithm for structural image editing. ACM Transactions on Graphics, 28(3): 24 [DOI: 10.1145/1531326.1531330]
- Barroso-Laguna A and Mikolajczyk K. 2023. Key.Net: keypoint detection by handcrafted and learned CNN filters revisited. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(1): 698-711 [DOI: 10.1109/TPAMI.2022.3145820]
- Bay H, Tuytelaars T and Van Gool L. 2006. SURF: speeded up robust features//9th European Conference on Computer Vision. Graz: Springer: 404-417 [DOI: 10.1007/11744023\_32]
- Beevers K R and Huang W H. 2007. Fixed-lag sampling strategies for particle filtering SLAM//Proceedings 2007 IEEE International Conference on Robotics and Automation. Rome: IEEE: 2433-2438 [DOI: 10.1109/ROBOT.2007.363684]
- Bloesch M, Czarnowski J, Clark R, Leutenegger S and Davison A J. 2018. CodeSLAM - learning a compact, optimisable representation for dense visual SLAM//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE: 2560-2568 [DOI: 10.1109/CVPR.2018.00271]
- Boykov Y, Veksler O and Zabih R. 2001. Fast approximate energy min-

- imization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11): 1222-1239 [DOI: 10.1109/34.969114]
- Bu S H, Zhao Y, Wan G and Liu Z B. 2016. Map2DFusion: real-time incremental UAV image mosaicing based on monocular SLAM//2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Daejeon: IEEE: 4564-4571 [DOI: 10.1109/IROS.2016.7759672]
- Calonder M, Lepetit V, Strecha C and Fua P. 2010. BRIEF: binary robust independent elementary features//11th European Conference on Computer Vision. Heraklion: Springer: 778-792 [DOI: 10.1007/978-3-642-15561-1\_56]
- Chang D, Božić A, Zhang T, Yan Q S, Chen Y C, Süssstrunk S and Nießner M. 2022. RC-MVSNet: unsupervised multi-view stereo with neural rendering//17th European Conference on Computer Vision. Tel Aviv: Springer: 665-680 [DOI: 10.1007/978-3-031-19821-2\_38]
- Chen L, Zhao Y, Xu S B, Bu S H, Han P C and Wan G. 2020. DenseFusion: large-scale online dense pointcloud and DSM mapping for UAVs//2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Las Vegas: IEEE: 4766-4773 [DOI: 10.1109/IROS45743.2020.9341413]
- Cheng H M, Xu D L and Wang S T. 2021. Application of geophysical prospecting in emergency responseto landslide disaster--A case study of Shaziba Landslide in Enshi, Huabei. *Chinese Journal of Engineering Geophysics*, 18(2): 273-281 (程怀蒙, 徐玳笠, 王思桐. 2021. 滑坡灾害应急处置中物探技术的作用——以湖北恩施州屯堡乡沙子坝滑坡为例. *工程地球物理学报*, 18(2): 273-281)
- DeTone D, Malisiewicz T and Rabinovich A. 2018. SuperPoint: self-supervised interest point detection and description//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Salt Lake City: IEEE: 337-349 [DOI: 10.1109/CVPRW.2018.00060]
- Forster C, Pizzoli M and Scaramuzza D. 2014. SVO: fast semi-direct monocular visual odometry//2014 IEEE International Conference on Robotics and Automation (ICRA). Hong Kong, China: IEEE: 15-22 [DOI: 10.1109/ICRA.2014.6906584]
- Furukawa Y and Ponce J. 2010. Accurate, dense, and robust multiview stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8): 1362-1376 [DOI: 10.1109/TPAMI.2009.161]
- Galliani S, Lasinger K and Schindler K. 2015. Massively parallel multiview stereopsis by surface normal diffusion//Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago: IEEE: 873-881 [DOI: 10.1109/ICCV.2015.106]
- Goessele M, Snavely N, Curless B, Hoppe H and Seitz S M. 2007. Multi-view stereo for community photo collections//2007 IEEE 11th International Conference on Computer Vision. Rio de Janeiro: IEEE: 1-8 [DOI: 10.1109/ICCV.2007.4408933]
- Guo J T, Hong H B, Zhong K K, Liu X J and Guo Y. 2020. Production management and control method of aerospace manufacturing workshops based on digital twin. *China Mechanical Engineering*, 31(7): 808-814 (郭具涛, 洪海波, 钟珂珂, 刘骁佳, 郭宇. 2020. 基于数字孪生的航天制造车间生产管控方法. *中国机械工程*, 31(7): 808-814) [DOI: 10.3969/j.issn.1004-132X.2020.07.006]
- Han P C, Ma C B, Chen J, Chen L, Bu S H, Xu S B, Zhao Y, Zhang C H and Hagino T. 2022. Fast tree detection and counting on UAVs for sequential aerial images with generating orthophoto mosaicing. *Remote Sensing*, 14(16): 4113 [DOI: 10.3390/rs14164113]
- Hinzmann T, Schönberger J L, Pollefeys M and Siegwart R. 2018. Mapping on the fly: real-time 3D dense reconstruction, digital surface map and incremental orthomosaic generation for unmanned aerial vehicles//Hutter M and Siegwart R, eds. *Field and Service Robotics*. Cham: Springer: 383-396 [DOI: 10.1007/978-3-319-67361-5\_25]
- Jiang N J, Cui Z P and Tan P. 2013. A global linear method for camera pose registration//Proceedings of the 2013 IEEE International Conference on Computer Vision. Sydney: IEEE: 481-488 [DOI: 10.1109/ICCV.2013.66]
- Kazhdan M and Hoppe H. 2013. Screened poisson surface reconstruction. *ACM Transactions on Graphics*, 32(3): 29 [DOI: 10.1145/2487228.2487237]
- Kern A, Bobbe M, Khedar Y and Bestmann U. 2020. OpenREALM: real-time mapping for unmanned aerial vehicles//2020 International Conference on Unmanned Aircraft Systems (ICUAS). Athens: IEEE: 902-911 [DOI: 10.1109/ICUAS48674.2020.9213960]
- Klein G and Murray D. 2007. Parallel tracking and mapping for small AR workspaces//2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality. Nara: IEEE: 225-234 [DOI: 10.1109/ISMAR.2007.4538852]
- Lhuillier M and Quan L. 2005. A quasi-dense approach to surface reconstruction from uncalibrated images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3): 418-433 [DOI: 10.1109/TPAMI.2005.44]
- Li D R. 2021. Smart cities based on digital twins. *Internet World*, (7): 12 (李德仁. 2021. 基于数字孪生的智慧城市. *互联网天地*, (7): 12)
- Li H, Tao F, Wang H Q, Song W Y, Zhang Z F, Fan B B, Wu C L, Li Y P, Li L L, Wen X Y, Zhang X S and Luo G F. 2019. Integration framework and key technologies of complex product design-manufacturing based on digital twin. *Computer Integrated Manufacturing Systems*, 25(6): 1320-1336 (李浩, 陶飞, 王昊琪, 宋文燕, 张在房, 樊蓓蓓, 武春龙, 李玉鹏, 李琳利, 文笑雨, 张新生, 罗国富. 2019. 基于数字孪生的复杂产品设计制造一体化开发框架与关键技术. *计算机集成制造系统*, 25(6): 1320-1336) [DOI: 10.13196/j.cims.2019.06.002]
- Liu C J, Lü J, Ren M L, Chen S, Zhang X L, Song W L and Zhang D W. 2022. Research and application of digital twin intelligent flood prevention system in Huaihe River Basin. *China Flood and Drought Management*, 32(1): 47-53 (刘昌军, 吕娟, 任明磊, 陈胜, 张晓蕾, 宋文龙, 张大伟. 2022. 数字孪生淮河流域智慧防洪体系研究与实践. *中国防汛抗旱*, 32(1): 47-53) [DOI: 10.16867/j.issn.1673-9264.2021375]
- Lorensen W E and Cline H E. 1987. Marching cubes: a high resolution 3D surface construction algorithm. *ACM SIGGRAPH Computer Graphics*, 21(4): 163-169 [DOI: 10.1145/37402.37422]
- Lowe D G. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2): 91-110 [DOI: 10.1023/B:VISI.0000029664.99615.94]
- Mildenhall B, Srinivasan P P, Tancik M, Barron J T, Ramamoorthi R and Ng R. 2020. NeRF: representing scenes as neural radiance fields for view synthesis//16th European Conference on Computer

- Vision. Glasgow: Springer: 405-421 [DOI: 10.1007/978-3-030-58452-8\_24]
- Muja M and Lowe D G. 2009. Fast approximate nearest neighbors with automatic algorithm configuration//Proceedings of the Fourth International Conference on Computer Vision Theory and Applications. Lisboa: SciTePress: 331-340 [DOI: 10.5220/0001787803310340]
- Mur-Artal R, Montiel J M M and Tardós J D. 2015. ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics*, 31(5): 1147-1163 [DOI: 10.1109/TRO.2015.2463671]
- Nealen A, Igarashi T, Sorkine O and Alexa M. 2006. Laplacian mesh optimization//Proceedings of the 4th international Conference on Computer Graphics and Interactive Techniques in Australasia and Southeast Asia. Kuala Lumpur: ACM: 381-389 [DOI: 10.1145/1174429.1174494]
- Pan X C, Jiang T, Yu A Z, Wang X and Zhang Y. 2019. Geo-positioning of remote sensing images with reference image. *Journal of Remote Sensing (in Chinese)*, 23(4): 673 - 684 DOI: 10.11834/jrs.20197490. (潘雪琛, 姜挺, 余岸竹, 王鑫, 张一. 2019. 基准影像数据辅助遥感影像几何定位. *遥感学报*, 23(4): 673-684) [DOI: 10.11834/jrs.20197490]
- Rosten E and Drummond T. 2006 Machine learning for high-speed corner detection//9th European Conference on Computer Vision. Graz: Springer: 430-443 [DOI: 10.1007/11744023\_34]
- Rublee E, Rabaud V, Konolige K and Bradski G. 2011. ORB: an efficient alternative to SIFT or SURF//2011 International Conference on Computer Vision. Barcelona: IEEE: 2564-2571 [DOI: 10.1109/ICCV.2011.6126544]
- Sarlin P E, DeTone D, Malisiewicz T and Rabinovich A. 2020. Super-Glue: learning feature matching with graph neural networks//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE: 4937-4946 [DOI: 10.1109/CVPR42600.2020.00499]
- Schönberger J L and Frahm J M. 2016. Structure-from-motion revisited//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE: 4104-4113 [DOI: 10.1109/CVPR.2016.445]
- Snaveley N, Seitz S M and Szeliski R. 2008. Modeling the world from internet photo collections. *International Journal of Computer Vision*, 80(2): 189-210 [DOI: 10.1007/s11263-007-0107-3]
- Sorkine O and Cohen-Or D. 2004. Least-squares meshes//Proceedings Shape Modeling Applications, 2004. Genova: IEEE: 191-199 [DOI: 10.1109/SML.2004.1314506]
- Strasdat H, Montiel J M M and Davison A J. 2012. Visual SLAM: why filter?. *Image and Vision Computing*, 30(2): 65-77 [DOI: 10.1016/j.imavis.2012.02.009]
- Tancik M, Casser V, Yan X C, Pradhan S, Midenhall B P, Srinivasan P, Barron J T and Kretschmar H. 2022. Block-NeRF: scalable large scene neural view synthesis//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 8238-8248 [DOI: 10.1109/CVPR52688.2022.00807]
- Tao F, Liu W R, Liu J H, Liu X J, Liu Q, Qu T, Hu T L, Zhang Z N, Xiang F, Xu W J, Wang J Q, Zhang Y F, Liu Z Y, Li H, Cheng J F, Qi Q L, Zhang M, Zhang H, Sui F Y, He L R, Yi W M and Cheng H. 2018. Digital twin and its potential application exploration. *Computer Integrated Manufacturing Systems*, 24(1): 1-18 (陶飞, 刘蔚然, 刘检华, 刘晓军, 刘强, 屈挺, 胡天亮, 张执南, 向峰, 徐文君, 王军强, 张映锋, 刘振宇, 李浩, 程江峰, 戚庆林, 张萌, 张贺, 隋芳媛, 何立荣, 易旺民, 程辉. 2018. 数字孪生及其应用探索. *计算机集成制造系统*, 24(1): 1-18) [DOI: 10.13196/j.cims.2018.01.001]
- Tola E, Strecha C and Fua P. 2012. Efficient large-scale multi-view stereo for ultra high-resolution image sets. *Machine Vision and Applications*, 23(5): 903-920 [DOI: 10.1007/s00138-011-0346-8]
- Turki H, Ramanan D and Satyanarayanan M. 2022. Mega-NeRF: scalable construction of large-scale NeRFs for virtual fly-throughs//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 12912-12921 [DOI: 10.1109/CVPR52688.2022.01258]
- Wachter M, Moehrl N and Goesele M. 2014. Let there be color! Large-scale texturing of 3D reconstructions//13th European Conference on Computer Vision. Zurich: Springer: 836-850 [DOI: 10.1007/978-3-319-10602-1\_54]
- Wang H P, Liu Y, Hu Q Y, Wang B, Chen J G, Dong Z, Guo Y L, Wang W P and Yang B S. 2023. RoReg: pairwise point cloud registration with oriented descriptors and local rotations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8): 10376-10393 [DOI: 10.1109/TPAMI.2023.3244951]
- Wang W, Zhao Y, Han P C, Zhao P C and Bu S H. 2019. TerrainFusion: real-time digital surface model reconstruction based on monocular SLAM//2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Macau, China: IEEE: 7895-7902 [DOI: 10.1109/IROS40897.2019.8967663]
- Wei Y, Liu S H, Rao Y M, Zhao W, Lu J W and Zhou J. 2021. Nerfing-MVS: guided optimization of neural radiance fields for indoor multi-view stereo//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE: 5590-5599 [DOI: 10.1109/ICCV48922.2021.00556]
- Wu C. 2011. VisualSFM: a visual structure from motion system. [2023-03-13]. <http://ccwu.me/vsfm/doc.html>.
- Xia R L, Li T, Yu W and Li Y R. 2021. Digital twin theory of watershed and its application in flood control of the Yellow River. *China Water Conservancy*, (20): 11-13 (夏润亮, 李涛, 余伟, 李英睿. 2021. 流域数字孪生理论及其在黄河防汛中的实践. *中国水利*, (20): 11-13) [DOI: 10.3969/j.issn.1000-1123.2021.20.006]
- Yang S T, Wang P F, Wang J, Lou H Z and Gong T L. 2021. River flow estimation method based on UAV aerial photogrammetry. *National Remote Sensing Bulletin*, 25(6): 1284-1293 (杨胜天, 王鹏飞, 王娟, 娄和震, 巩同梁. 2021. 结合无人机航空摄影测量的河道流量估算. *遥感学报*, 25(6): 1284-1293) [DOI: 10.11834/jrs.20209082]
- Yao Y, Luo Z X, Li S W, Fang T and Quan L. 2018. MVSNet: depth inference for unstructured multi-view stereo//Proceedings of the 15th European Conference on Computer Vision. Munich: Springer: 785-801 [DOI: 10.1007/978-3-030-01237-3\_47]
- Yu Z and Gao S H. 2020. Fast-MVSNet: sparse-to-dense multi-view stereo with learned propagation and gauss-newton refinement//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE: 1946-1955 [DOI: 10.1109/CVPR42600.2020.00202]
- Zhang J X, Liu F and Wang J. 2021. Review of the light-weighted and small UAV system for aerial photography and remote sensing. *National Remote Sensing Bulletin*, 25(6): 1284-1293 (张俊霞, 刘飞, 王健. 2021. 轻量化小型无人机系统用于航空摄影和遥感. *国家遥感学报*, 25(6): 1284-1293)

- tional Remote Sensing Bulletin, 25(3): 708-724 (张继贤, 刘飞, 王坚. 2021. 轻小型无人机测绘遥感系统研究进展. 遥感学报, 25(3): 708-724) [DOI: 10.11834/jrs.20210052]
- Zhao Y, Zhao P C, Xu S B, Chen L, Han P C, Bu S H and Jiang H K. 2021. Svar: a tiny C++ header brings unified interface for multiple programming languages. arXiv: 2108.08464 [DOI: 10.48550/arXiv.2108.08464]
- Zhou T H, Brown M, Snavely N and Lowe D G. 2017. Unsupervised learning of depth and ego-motion from video//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE: 6612-6619 [DOI: 10.1109/CVPR.2017.700]
- Zhu Q, Zhu J, Huang H P, Wang W and Zhang L G. 2020. Real 3D spatial information platform and digital twin Sichuan-Tibet railway. High Speed Railway Technology, 11(2): 46-53 (朱庆, 朱军, 黄华平, 王玮, 张利国. 2020. 实景三维空间信息平台与数字孪生川藏铁路. 高速铁路技术, 11(2): 46-53) [DOI: 10.12098/j.issn.1674-8247.2020.02.008]
- Zhu Z H, Peng S Y, Larsson V, Xu W W, Bao H J, Cui Z P, Oswald M R and Pollefeys M. 2021. NICE-SLAM: neural implicit scalable encoding for SLAM//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE: 12776-12786 [DOI: 10.1109/CVPR52688.2022.01245]

## Real-time dense point cloud generation and digital model construction of surface environment based on UAV platform

HU Boni<sup>1</sup>, CHEN Lin<sup>1</sup>, XU Bingli<sup>2</sup>, BU Shuhui<sup>1</sup>, HAN Pengcheng<sup>1</sup>, LI Kun<sup>1</sup>, XIA Zhenyu<sup>1</sup>, LI Ni<sup>1</sup>, LI Ke<sup>3</sup>, CAO Xuefeng<sup>3</sup>, WAN Gang<sup>4</sup>

1. Northwestern Polytechnical University, Xi'an 710072, China;

2. Army Armored Force Academy, Beijing 100072, China;

3. University of Information Engineering, Zhengzhou 450000, China;

4. University of Aerospace Engineering, Beijing 101400, China

**Abstract:** The development of high-fidelity 3D digital model of the land surface environment in real time has become essential in many fields, including urban planning, agricultural surveying and mapping, disaster management, and military applications. Building a high-fidelity digital model of the land surface environment in real time is the key foundation for realizing the virtual mapping of the geographic environment and then forming a digital twin geographic environment. However, current methods for constructing these models suffer from several challenges such as slow speed, low timeliness, and limited application in large scenarios. To overcome these challenges, this paper proposes a new real-time dense point cloud generation and digital model construction method based on unmanned aerial vehicle (UAV) platforms. The proposed algorithm remarkably improves the speed and accuracy of land surface model construction compared with existing algorithms and improves the degree of online data collection and real-time modeling. The algorithm is based on a general simultaneous localization and mapping framework. The technical chain from data acquisition, data processing, feature extraction and matching to 3D point cloud generation, digital model reconstruction, and result analysis is unobstructed. The algorithm breaks through the key technologies in scene reconstruction such as online data acquisition and pose problem solving, real-time dense point cloud generation, digital surface model, and digital orthophoto map construction.

According to the experimental results of the dataset containing different land surface environments such as cities, farmland, mountains, and deserts, the algorithm proposed in this paper can process high-quality land surface environment dense point clouds and digital models while being 30—50 times faster than existing algorithms such as Pix4DMapper. On average, it can process a high-resolution image in less than 1 second, whereas the interval for general aerial survey drones is 1—2 seconds. Furthermore, the application of the algorithm proposed in this paper to emergency rescue work of mountain flood disasters can assist in the real-time reconstruction of inaccessible areas, survey the area and location of washed-out and sediment areas, and support emergency rescue work.

The proposed real-time dense point cloud generation and digital model construction method based on UAV platforms represents a considerable advancement in the field of geographic information systems. It can revolutionize various geographic fields, including disaster warning and management, emergency response, military applications, and agricultural and urban planning. Additionally, due to the breakthrough of the algorithm, it solves the core problem of real-time mapping from reality to virtual in the construction of digital twins in geographic environments. Therefore, it can empower digital twins and provide a digital foundation for digital twin applications in geographic environments. Based on the real-time mapping of 3D information of the ground environment, parallel simulation and decision making can be conducted. Therefore, it is also expected to achieve functions such as emergency warning, scheme evaluation, and decision optimization based on the dynamic monitoring of target information in the scene, further improving the automation and intelligence level of the applied system.

**Key words:** remote sensing, real-time reconstruction, dense point cloud, digital twin, terrain environment, digital model

**Supported by** National Natural Science Foundation of China (No. 41871322, 62003272); Key Research and Development Project of Shaanxi Province (No. 2022GY-253); ND Basic Research Funds (No. G2022WD)