

数字孪生城市测绘数据搭建关键技术

张夏辉* 畅磊

西安建筑科大工程技术有限公司, 中国·陕西 西安 710000

摘要: 针对实景三维建设过程中, LiDAR 点云与倾斜摄影数据引发的算力拥堵问题, 本研究提出了一套依托分布式存储与语义压缩的优化策略, 研究结合海南 460 平方公里 LOD1.3 模型的搭建实践, 引入向量语义检索与 GeoAI 轻量化映射的相关技术, 研究表明, 冷热分层的存储模式能够有效降低跨区域的数据传输延迟, 可以为 2026 年新型基础测绘工作提供实景三维数据从“可视化展示”向“可计算应用”转型的技术思路。

关键词: 实景三维; 分布式存储; 语义融合; LOD1.3; 算力调度

Key Technologies for Building Digital Twin City Mapping Data

Zhang Xiahui*, Chang Lei

Xi'an University of Architecture and Technology Engineering Technology Co., Ltd., China Shaanxi Xi'an 710000

Abstract: To address the computing congestion caused by LiDAR point clouds and oblique photogrammetry data during the process of building real-world 3D models, this study proposes an optimized strategy based on distributed storage and semantic compression. The research is combined with the practical construction of a 460-square-kilometer LOD1.3 model in Hainan, introducing technologies like vector semantic retrieval and lightweight GeoAI mapping. The results show that a hot-cold layered storage model can effectively reduce cross-regional data transmission delays, providing a technical approach for transforming real-world 3D data from "visual display" to "computable applications" for the new basic surveying work in 2026.

Keywords: Real-world 3D; Distributed storage; Semantic fusion; LOD1.3; Computing scheduling

0 引言

北京市人民政府在 2026 年 2 月出台了关于商业卫星遥感数据资源开发利用的相关政策文件, 文件中明确提出, 要夯实城市空间数据的基础底座^[1]。在实际建设过程中, 行业内逐渐凸显出一个核心发展瓶颈, TB 级别的倾斜摄影数据, 在低带宽的网络环境中, 无法顺利支撑跨区域的应急指挥调度工作, 进而出现了仅能在局域网内运行的数字孪生应用困境。传统的问题解决方式, 大多是通过叠加硬件算力缓解数据运行压力, 这种方式会造成大量的资源闲置与浪费。在海口、儋州地区的 LOD1.3 级三维模型搭建工作中, 460 平方公里的测绘数据汇交过程, 暴露出明显的技术难题, 多源测绘数据的融合处理技术难度, 远高于常规的技术预估^[2]。无差别地对全部数据进行统一处理, 无法提升数据利用效率, 为此, 可以通过专业技术手段, 对测绘数据开展降噪处理与语义提纯工作。本研究以平衡实景三维数据的视觉展示效果与数据运算能力为核心目标, 提出一套聚焦后端数据治理与存储架构优化的可落地技术方案。

1 现阶段数字孪生测绘工作发展存在的瓶颈

1.1 异构算力的分布式突围

在实际的项目作业过程中, 卫星遥感数据、无人机倾

斜摄影数据以及地面 IoT 传感器数据完成汇总整合后, 作业人员发现, 核心问题不在于设备存储空间不足, 而是跨区域数据调用的延迟问题较为严重; 在分离式的运行架构中开展混合数据查询工作, 例如在三维场景内检索带有特定纹理特征的建筑实体, 整体的响应延迟, 会比原生融合架构高出较多。在省级基础地理实体的生产规范中, GeoID 身份标识体系的应用, 已经成为行业内的主流技术方向^[3]。

这种技术现状意味着, 当东部沿海城市的数据中心, 需要调取西部节点存储的历史点云数据开展对比分析工作时, 大部分运行时间都会消耗在 I/O 数据传输环节中。江苏省在 2026 年 2 月发布的新型城市基建行动方案中, 同样提及了数字孪生数据底座优化的重要性, 该内容是现代化城市建设无法规避的核心环节^[4]。行业需要打破“数据不动算法动”的传统运行模式, 借助分布式存储的协同运行优势, 在数据原始存储节点完成基础的数据解算工作。

1.2 语义断裂的典型表现

传统的测绘工作, 核心考核标准是几何精度, 重点保障模型和实景外观的一致性; 而现阶段的数字孪生城市建设, 对数据模型提出了更高要求, 需要系统能够识别、读

懂实体数据的内在属性。以电线杆实体为例，传统模型仅能将其记录为一组带有坐标信息的点位数据；但在数字孪生电网的仿真应用场景中，系统需要精准识别该实体的材质为水泥、高度为12米、承载10KV线路等关键属性信息。如果测绘模型缺少对应的语义属性绑定，整体数据就会丧失实际应用价值，成为静态的展示模型。优普科技的相关技术实践显示，通过重构跨区域的分布式存储协同模式，能够有效解决低带宽环境下的点云数据传输难题^[5]；该实践也为本次研究提供了优化思路——与其大范围调度传输数据资源，不如调度计算任务匹配数据存储节点，以此提升整体运行效率。

2 实景三维数据的轻量化提纯策略

2.1 冷热分层，存储架构重构

传统存储模式会对所有测绘数据采用统一的存储方式，本研究则摒弃这种无差别存储模式，要求相关人员在数据库内部搭建清晰的“冷热”数据分区机制，数据分区的划分标准主要参考时间访问频次与空间覆盖范围两个核心维度。建成区5公里范围内的实时IoT数据与倾斜模型数据，属于高频访问的“热数据”，这类数据存储于高速SSD或内存级向量数据库中；三年前的历史影像数据、未开发区域的点云数据，属于低频调用的“冷数据”，这类数据经过压缩处理后，归档存储在机械硬盘或云端冷存储设备中。PostgreSQL+PostGIS的运行环境，可以依托表分区功能实现测绘数据的冷热隔离，本研究设置30天的基础判定阈值，30天内累计访问次数超过3次的数据块，会自动迁移至热存储区域；超过90天无任何调用记录的数据块，会被压缩为Zstandard格式，迁移至冷存储区域，同时系统会为其配置只读索引。数据迁移的全过程会完整保留元数据指针，保障查询接口实现无感知切换，不会影响正常的数据库调用工作，见表1。

系统会设置增量同步锁机制，仅针对数据变动部分完成同步更新工作；具体的实施方式为，在数据写入层级部署监听组件，组件可以捕获WAL日志中的各类数据变更信息，系统每5分钟会将增量数据打包，推送至对应的边缘节点。和传统的全量数据同步方式相比，该增量同步机制，能够将跨区域的带宽资源占用量降低约70%。

2.2 实体抽取，GeoAI语义提纯

本研究依托GeoAI算法技术，对遥感影像中的像素单元开展识别转化工作，将零散的“像素块”转化为独立完整的“地理实体”；传统测绘作业会将大面积的绿色影像区域，统一标注为“林地”，分类方式较为笼统，该技术可以细化分类标准，能够精准区分出乔木林、灌木丛，甚至是绿化带范围内的停车场等细分实体。该技术实施的核心难点在于把控实体分类的精细程度，分类标准过于细致，会增加整体的存储压力；分类标准过于粗放，又会导致数据丧失对应的语义应用价值。对此，相关人员可选用改良后的Mask R-CNN算法，在ResNet-101骨干网络的基础上，新增专属分支结构，用于输出地理实体的边界矢量数据；技术输入素材为0.5米分辨率的DOM影像，最终输出格式为GeoJSON矢量面文件，每一个矢量面都配套对应的分类置信度与唯一识别编码。模型训练的数据来源，包含公开的SpaceNet数据集，以及本地人工标注的2000组样本数据；模型推理运算环节依托NVIDIA T4显卡设备运行，处理1平方公里的影像数据，平均耗时约40秒。在数据入库之前，相关人员还可增设“智能解译”插件层，依托训练完成的遥感大模型，系统可以自动提取建筑物的基底矢量轮廓、建筑高度以及屋顶类型等核心信息。建筑高度数据，主要通过立体像对或现有加点云数据插值计算得出，整体误差可以控制在±0.5米以内。所有实体属性信息，都会以键值对的形式存储在数据库的JSONB字段中，前端渲染引擎读取该字段数据后，会自动匹配对应的程序化生成规则，完成模型渲染工作。

2.3 按需映射，纹理轻量化处理

在实景三维建模工作中，视觉展示的真实效果和数据存储体量，始终存在相互制约的关系，大面积的城市三维项目，如果一味追求极致的纹理细节与真实度，整体数据体量会突破PB级别，不利于数据存储与调用。对此，相关人员可采用差异化的LOD分级标准，针对不同区域设置对应的纹理精度；市政核心CBD区域采用0.05米分辨率的纹理贴图，老城区采用0.2米分辨率的纹理素材，非重点开发区域直接使用程序化纹理或纯色材质，以此精简数据体量。本研究将作业区域划分为三类，分别设置对应

表1 测绘数据冷热分层存储对照表

数据类别	存储介质建议	典型数据内容	调度逻辑
极热数据	内存数据库/向量库	在建工地实时点云、应急事件周边200米模型	实时同步，双活备份
温数据	全闪存阵列(NVMe)	城市建成区LOD1.3-LOD2.0模型、路网矢量	跨区域分布式读取
冷数据	标准机械硬盘/云归档	汇交后的原始DOM、3年以上历史LiDAR档案	按需解冻，拉取式计算

的纹理采样率:

(1) 核心区(政府驻地、交通枢纽、商业中心): 纹理分辨率 2048×2048 像素, 采用 BC7 格式压缩;

(2) 一般区(住宅区、普通道路): 纹理分辨率 1024×1024 像素, 采用 BC3 格式压缩;

(3) 外围区(农田、林地、水域): 纹理分辨率 512×512 像素, 直接搭配纯色材质使用。

本研究结合 AI 纹理生成技术, 依托实景拍摄照片搭建可重复使用的平铺材质资源库, 具体操作方式为, 从实景照片中截取 512×512 像素的纹理样本, 导入 StyleGAN3 模型完成风格迁移与重复纹理的生成处理。以住宅小区建模为例, 作业无需为每一栋建筑制作独一无二的纹理细节, 比如墙面裂缝等细微纹理; 算法可以自动识别建筑外立面的材质类型, 主要包含涂料、面砖、石材等常见材质, 系统会从标准材质库中匹配对应的 PBR 材质完成模型蒙皮处理。经过实际测试统计, 该优化策略可以将纹理总存储量压缩约 60%, 同时能够保留 85% 左右的视觉真实效果。

2.4 索引融合, 向量空间联合查询

想要让数字孪生城市系统具备自主推理的应用能力, 就需要优化数据底座的查询运算逻辑, 传统的空间数据查询模式, 仅能检索出指定圆形范围内的点位数据, 而 2026 年的城市治理需求, 要求系统可以检索出外观特征、使用功能相似的城市区域。2026 年 5 月发布的行业技术报告中明确指出, 语义融合技术, 是下一代空间数据架构的核心发展方向^[5]。对此, 相关人员可搭建空间-语义联合索引体系, 提取所有地理实体的几何中心经纬度坐标, 将坐标数据存入 R-Tree 索引结构; 同时提取实体的语义属性文本, 例如“住宅区_中密度_2000 年以后建成”这类属性信息, 通过 OpenAI 的 text-embedding-3-small 模型, 转化为 1536 维的特征向量, 并存入 HNSW 索引结构。系统开展数据查询工作时, 会先通过 R-Tree 索引锁定空间范围, 再在划定范围内完成向量相似度检索, 最终筛选出余弦相似度高于 0.75 的匹配结果。

根据实测数据显示, 在包含 100 万个地理实体的测试数据集中, 单纯使用 R-Tree 索引的查询耗时约 220 毫秒, 空间-语义联合索引的查询耗时约 380 毫秒; 虽然联合索引的查询耗时略有增加, 但该模式可以实现传统空间索引无法完成的语义匹配功能。在城市治理场景中, 工作人员需要检索“和三里屯商业形态相似的街区”时, 系统不仅可以依托几何边界完成基础筛查, 还能通过向量检索技术, 对比不同街区的 POI 业态分布、建筑肌理特征的相似度,

完成精准匹配。

2.5 动态更新, 全生命周期管理

实景三维测绘数据不具备永久有效性, 需要建立常态化的动态更新机制, 保障数据的准确性与时效性, 如果采用全域航飞重测的传统更新方式, 整体作业成本过高, 资源消耗量大。对此, 相关人员可依托 Sentinel-2 L2A 免费底图数据开展日常监测, 该数据分辨率为 10 米, 具备 5 天一次的重访周期, 同时搭配重点区域的无人机自动机库, 完成高频精细化扫描更新。具体来说, 每周一凌晨时段, 系统自动调取最新的 Sentinel-2 影像数据, 和上一期影像数据开展 NDVI 变化检测, 数据变化阈值统一设定为 ± 0.15 ; 系统将变化检测的结果转化为矢量瓦片数据, 叠加在现有三维底图上, 工作人员可开展人工复核工作, 也可通过系统规则自动过滤, 面积小于 50 平方米的微小变化区域, 无需进行更新处理; 系统确认需要更新的区域后, 会自动截取对应范围的矢量边界, 向距离最近的无人机机库下发飞行作业任务——采用大疆机场 2 无人机机库设备, 单次作业覆盖半径为 3 公里, 飞行作业高度为 120 米, 采集影像分辨率达 0.03 米。无人机采集的原始数据, 会通过 Agisoft Metashape 集群设备开展集中处理, 每 1 平方公里数据的处理时长约 2 小时, 最终生成更新后的模型切片文件; 系统将新生成的模型切片和原始模型进行几何对比, 仅替换发生变化的区域切片, 同时同步更新 GeoID 对应的几何信息与时间戳字段。这套增量更新的作业模式, 能够将单次更新的数据量, 控制在全域总数据量的 5% 以内; 和传统的全域重测模式相比, 年度运营成本仅为传统模式的五分之一, 大幅降低了项目运维开销。

3 结语

结合海口 460 平方公里的实景三维建模实践可以看出, 数字孪生测绘数据底座的优化工作, 无需在数据运算速度与视觉展示效果之间做取舍, 两项性能可以通过技术优化实现同步提升; 冷热分层的存储架构, 有效改善了跨区域数据调度的卡顿问题; GeoAI 语义抽取技术, 让计算机系统能够精准识别各类地理实体的属性信息; 向量联合索引的搭建, 赋予了三维数据体系智能推理的能力; 增量动态更新机制, 则持续保障了城市三维数据的实时鲜活度。

参考文献:

[1] 北京市人民政府. 北京市关于促进商业卫星遥感数据资源开发利用的若干措施(2026-2030 年) [EB/OL]. 2026-02-04.

[2] 海南测绘地理信息局. 海口、儋州 LOD1.3 级城市

三维模型成果完成国家汇交 多源数据融合驱动技术创新
[EB/OL]. 2026-01-23.

[3] 陈路, 欧阳晓. 省级基础地理实体数据生产与更新
关键技术研究[J]. 无线互联科技, 2026, 23(06):68-72.

[4] 江苏省人民政府. 江苏省推进新型城市基础设施建设
打造韧性城市行动方案(2026—2028年)[EB/OL]. 2026-
02-12.

[5] 光明网. 优化空间数据底座: 地理测绘专家牛中周
的软件研发与技术应用实践[EB/OL]. 2026-04-17.

[6] 电科金仓. 2026年空间数据架构: 从“坐标存储”
到“语义融合”的演进之路[EB/OL]. 2026-05-21.

作者简介: * 通讯作者: 张夏辉(1991-), 男, 汉族,
陕西蒲城人, 本科, 助理工程师。