

新型基础测绘中多源异构数据融合与集成方法探

王永 胡广志

湖北省国土测绘院, 湖北省武汉市, 430010;

摘要: 随着新型基础测绘向智能化、立体化发展, 多源异构数据的高效利用愈发重要。本文主要针对新型基础测绘中多源异构数据的融合与集成加以分析。先阐述遥感影像、激光点云等数据类型, 及结构、时空、精度等异构性表现。随后分析数据标准化、时空基准统一等技术挑战, 基于此提出相应的技术方法: 预处理阶段进行元数据标准化、时空基准统一和数据轻量化; 融合算法涵盖几何、特征、语义三级; 集成架构采用分布式存储、服务化平台及动态更新机制。希望通过这些方法可以有效解决多源异构数据问题。

关键词: 新型基础测绘; 多源异构数据; 融合; 集成; 方法

DOI: 10. 64216/3080-1508. 25. 06. 038

新型基础测绘向智能化、立体化发展, 带来多源异构数据。其类型多样, 有遥感影像、激光点云等, 且存在结构、时空、精度等异构性, 导致数据孤岛、处理低效等问题。传统方法难以应对, 影响测绘成果质量与应用。基于此, 探究多源异构数据融合与集成方法十分必要, 对提升地理信息服务能力、推动新型基础测绘发展意义重大。

1 新型基础测绘中多源异构数据的类型与特征

新型基础测绘技术的发展使得数据来源日益多元化, 这些数据在支撑测绘工作的同时, 也因自身特性呈现出复杂状况。明确其具体类型划分, 以及不同数据间的异构性表现, 是开展融合与集成工作的基础。

1.1 数据类型划分

遥感影像数据涵盖高分七号、World View 系列等高分辨率卫星影像以及航空光学影像、无人机影像, 以 TIFF、JPEG2000 格式为主, 包含光谱与空间几何信息, 覆盖范围广、现势性强, 常用于地表覆盖分类和变化检测。激光点云数据分机载 LiDAR 与地面三维激光扫描两类, LAS、LAZ 格式存储, 能够直接获取厘米级三维坐标与反射强度, 单项目数据量可达 TB 级, 高密度且细节丰富。

倾斜摄影数据需借助无人机多镜头相机获取, 生成含纹理与几何信息的 OSGB 格式三维模型, 适合城市三维建模, 但存在模型接缝、纹理拉伸问题。矢量与专题数据包括道路、水系等矢量要素及土地利用等专题数据, SHP、GDB 格式为主, 语义丰富但现势性滞后。传感器网络数据如 CORS 数据、物联网监测数据, 实时性强, 时间维度信息丰富, 用于动态监测与精度校正^[1]。

1.2 数据异构性表现

结构异构体现在结构化、半结构化与非结构化数据并存。像矢量属性表这类结构化数据, XML 元数据等半结构化数据, 以及影像、点云等非结构化数据, 组织方式差异显著。时空异构涉及时间与空间基准的不一致, 时间上卫星影像与地面观测数据可能存在采集时差, 动态监测数据时间戳格式混乱, UTC 与北京时间混用; 空间上存在 2000 国家大地坐标系与地方独立坐标系、1985 国家高程基准的转换问题。

精度异构表现为不同数据源精度跨度大, 卫星影像平面精度为米级, 地面激光点云则达厘米级, 此种差异会引发数据融合时的精度冲突。此外, 语义冲突与信息冗余也很突出, 同一地理实体在不同数据源中描述不同, 比如, 建筑物在矢量、点云、影像中分别表现为封闭多边形、密集点簇、特定光谱区域, 极易导致实体匹配错误, 且多源数据重叠区存在信息冗余, 增加存储与处理成本。

2 多源异构数据融合与集成的技术挑战

多源异构数据的特性给融合与集成带来不少阻碍。这些挑战涉及数据标准、基准统一、处理效率以及语义信息等多个方面, 展开对这些问题的深入了解是找到解决办法的前提。

2.1 数据标准化难题

多源数据采集设备、生产流程没有统一标准, 坐标系、精度指标等元信息常不完整或表述混乱。比如部分无人机倾斜摄影数据缺失相机内参校准参数, 直接影响与激光点云的配准精度。元数据必填项执行不严格,

同一类数据的格式、字段定义存在差异,导致数据共享时需要反复人工校对,无疑增加了预处理成本^[2]。

2.2 时空基准统一问题

空间上,新型测绘数据需要统一到 2000 国家大地坐标系,但实际仍有地方坐标系数据,转换时易产生误差。时间上,动态监测数据时间戳格式混乱(UTC+8 与北京时间混用),卫星影像与地面观测数据存在采集时差,导致时序数据融合出现时间错位,影响动态分析准确性。

2.3 海量数据处理效率瓶颈

激光点云、倾斜摄影模型等单项目数据量达数十 TB,传统单机处理模式很难应对。比如对城市级点云进行去噪分类,常规算法至少需要数天,无法满足高效生产需求。数据读取、传输耗时占比较高,硬件存储与计算能力的不匹配,加剧了处理延迟。

2.4 语义冲突与信息冗余

不同数据源对同一地理实体描述有语义差异。建筑物在矢量中是封闭多边形,在点云中是密集点簇,在影像中是特定光谱区,导致实体匹配错误。多源数据重叠区易出现信息冗余,比如同一区域的影像和点云数据冗余,既浪费存储资源,又增加后续处理的计算负荷^[3]。

3 多源异构数据融合与集成的技术方法

针对多源异构数据融合与集成面临的各类挑战,需要从技术层面构建系统解决方案。这些方法涵盖数据预处理的标准化与归一化,到融合算法的多级递进,再到集成架构的分布式与服务化设计,多种技术方法汇聚起来形成一套完整的处理体系,以破解数据融合集成中的难题。

3.1 数据预处理:标准化与归一化

3.1.1 元数据标准化

元数据标准化需要先建立新型基础测绘数据元标准,强制要求所有数据源包含五要素:空间基准(坐标系、投影方式)、时间基准(采集时间、数据版本)、精度指标(平面/高程中误差)、数据格式、生产设备参数。针对元信息缺失的数据,需开发自动提取工具补全,比如从相机 EXIF 信息中恢复无人机影像的拍摄时间与位置参数;对表述混乱的元数据,如“精度”字段同时出现“米”“厘米”单位,通过规则引擎统一转换

为“毫米级”“厘米级”“米级”三级分类。同时,采用 XML 格式封装元数据,确保不同系统间可直接解析,减少人工校对成本。

3.1.2 时空基准统一

空间基准统一以 2000 国家大地坐标系为目标,地方坐标系数据采用七参数法转换,当转换区域 $<500\text{km}^2$ 时,用分区四参数法控制残差在 0.1m 以内;高程数据统一至 1985 国家高程基准,通过似大地水准面模型修正不同区域的高程异常。时间基准统一为北京时间,对无时间戳的数据,结合采集日志和太阳方位角反推精确时间;对格式混乱的时间戳,如 2023-10-01T08:00Z 与 2023-10-01 16:00,可通过正则表达式批量转换为 YYYY-MM-DD HH:MM:SS 格式,消除时序数据融合的时间错位^[4]。

3.1.3 数据轻量化处理

针对激光点云可以采用体素格网+非均匀采样组合策略:广场、农田等平坦区域按 1 点/ m^2 保留点密度,陡坎、墙角等建筑物边缘、地形突变处按 10 点/ m^2 保留,在保证细节的前提下减少 60% 以上数据量,同时用 LAZ 格式压缩存储,与 LAS 格式相比,LAZ 格式可以节省 70% 空间。倾斜摄影模型通过 LOD 技术生成多分辨率瓦片,其中 Level 0 (最低精度)用于大范围浏览,Level 5 (最高精度)用于局部细节查看,单模型数据量可以从 GB 级降至 MB 级。针对影像数据则采用波段联合压缩的方式处理,将多光谱影像的红、绿、蓝波段与全色波段融合,随后用 JPEG2000 无损压缩,以平衡精度与存储需求。

3.2 数据融合算法:从几何配准到语义融合

3.2.1 几何级融合

影像与点云配准分两步:先从高分辨率影像中提取 SIFT 特征点,与激光点云生成的强度影像进行粗配准,保留匹配误差小于 2 个像素的特征对;随后用 ICP 算法迭代优化,计算点到平面的距离误差,根据计算数值逐步调整变换矩阵,使最终配准误差控制在 0.5 个像素以内。矢量与影像融合采用边缘检测+缓冲区分析:对矢量道路数据做 1m 缓冲区,在影像中用 Canny 算子提取道路边缘,将边缘线与缓冲区叠加,修正矢量数据的位置偏差,使平面精度提升 30%^[5]。

3.2.2 特征级融合

采用多源特征组合分类地物:从遥感影像中提取 N

DVI 指数、波段反射率均值等光谱特征,从激光点云中提取点云高度分位数、坡度等高程特征,从矢量数据中提取建筑物邻接关系、道路节点度等拓扑特征。将这些特征归一化至 $[0, 1]$ 区间后,输入随机森林算法模型,通过 10 折交叉验证优化参数,使建筑物、道路等主要地物的分类精度提升至 95%以上,较单一数据源分类结果的混淆矩阵对角线数值提高 15%-20%。

3.2.3 语义级融合

构建地理实体语义本体库,需定义建筑物、道路、植被等多类核心实体的语义属性(如建筑物的结构类型、楼层数;道路的车道数、路面材质)。对不同数据源的语义描述做映射:点云中的高反射率($>80\%$) + 高度 $>3\text{m}$ 区域、影像中的矩形轮廓、低植被覆盖区域、矢量中的封闭多边形、住宅属性,均关联至建筑物语义类别。通过语义相似度计算(如编辑距离算法),解决街巷与支路统一为道路子类等同名异义问题,实现跨数据源的实体统一描述^[6]。

3.3 数据集成架构:分布式与服务化

3.3.1 分布式存储架构

分布式存储架构需采用分布式文件系统(HDFS) + 关系型数据库+NoSQL 数据库混合方案:其中,HDFS 按块(128MB/块)存储海量影像、点云等非结构化数据,通过副本机制,一般情况下默认 3 副本,保证数据可靠性;PostgreSQL/PostGIS 存储矢量数据与元数据,利用空间索引(R 树)加速查询;MongoDB 存储倾斜摄影模型的轻量化瓦片,按地理范围分片,如 $10\text{km} \times 10\text{km}$ 为一个分片。对 10TB 级点云数据,分散存储于 10 个节点的 HDFS 集群,单节点存储压力降低 90%,并行读取速度提升 8 倍。

3.3.2 服务化集成平台

基于微服务架构分三层设计:接入层开发多源数据适配器,支持 SHP、LAS、OSGB 等数十种以上格式的自动解析,通过元数据比对过滤无效数据;处理层将点云配准、影像分类等融合算法封装为 Docker 容器,通过 Kubernetes 动态调度资源,当任务量增加时自动扩容计算节点;服务层提供标准化 API 接口,包括 WMTS(影像瓦片服务)、3D Tiles(三维模型服务)、WFS(矢量要素服务),支持数据的在线查询、裁剪、下载,响应时间控制在 1 秒以内^[7]。

3.3.3 动态更新机制

采用增量更新与全量更新结合的模式:对城市建成区等变化快的区域,每月用无人机倾斜摄影做增量更新,仅处理新增/拆除的建筑物、改扩建道路等变化要素,数据量仅为全量更新的 1/5;针对山区、农田等变化慢的区域,每年用卫星影像做全量更新,替换超过三年的历史数据。通过 Z-order 曲线等时空索引快速定位变化区域,结合版本控制(需注意每个要素保留 5 个历史版本),使更新效率提升 60%,同时保证数据回溯能力^[8]。

4 结语

综上所述,多源异构数据融合与集成是新型基础测绘的关键。本文所提出的预处理、融合算法及集成架构等技术方法,可以有效解决数据标准化、处理效率等问题。通过标准化、统一基准、轻量化等预处理,结合三级融合算法与合理集成架构,能够提升数据处理质量与效率。未来,该领域研究人员还需要持续优化方法,以适应技术发展,为新型基础测绘提供更有力的数据支撑。

参考文献

- [1] 车一鸣,史长斌,李强,等.海量多源异构基础地理实体数据组织管理研究[J].测绘科学,2023,48(3):49-56.
- [2] 张政,张江水,曹一冰,等.面向地理实体数据的多维组织方法和分布式异构数据库存储[J].时空信息学报,2025,32(1):40-51.
- [3] 职爽,范业稳,毛涛敏,等.新型城镇化多源时空数据可共享性快速识别研究[J].测绘地理信息,2023,48(6):116-120.
- [4] 曾政祥,秦卫祥,张婧,等.航空多源数据融合的细小物体建模技术[J].遥感信息,2024,39(6):48-55.
- [5] 董传胜,孙久虎,高希舰,等.基于多源数据融合的实景三维山东构建[J].时空信息学报,2024,31(6):745-753.
- [6] 周思凡,王履华,沈健,等.面向实景三维的"一码多态"数据管理与应用[J].测绘工程,2024,33(5):50-56.
- [7] 甘斌,方登茂,金鑫,等.基础地理实体数据库应用管理系统设计与实现[J].城市勘测,2024(2):17-22.
- [8] 张向军,李茜楠,马强,等.地理实体数据建设关键技术研究及应用探索[J].地理空间信息,2024,22(8):64-67.