对地观测大数据处理: 挑战与思考

何国金*, 王力哲, 马艳, 张兆明, 王桂周, 彭燕, 龙腾飞, 张晓美

中国科学院遥感与数字地球研究所, 北京 100094

* 联系人, E-mail: hegj@radi.ac.cn

2014-08-28 收稿, 2014-10-29 接受, 2015-01-20 网络版发表

国家自然科学基金(60972142)和中国科学院遥感与数字地球研究所空间数据快速加工与综合服务技术研究规划项目资助

摘要 通过航天航空飞行器搭载的传感器对地球进行观测可以获得地球全面而系统的信息. 随着空间信息技术的高速发展, 对地观测领域步入了大数据时代. 在分析对地观测数据处理与服务的全流程及其算法的基础上, 总结出“大数据时代”对地观测数据处理面临的挑战——“数据密集型计算”问题, 并指出解决这一问题应该在系统平台、处理算法和服务模式3个方面开展创新性的研究工作.

关键词

对地观测大数据
数据密集型计算
遥感
对地观测

人类借助航天、航空对地观测平台实施对地球不间断地观测, 通过信息处理快速再现和客观反映地球表层的状况、现象、过程及其空间分布和定位, 服务于经济建设和社会发展. 21世纪以来, 以对地观测技术为核心的空间地球信息科技已经成为一个国家科技水平、经济实力和国家安全保障能力的综合体现.

遥感技术、通讯技术及计算机技术的快速发展, 带来了地观测数据的爆炸性增长和广泛应用. 尤其随着高分辨率对地观测时代的到来, 通过机载和卫星传感器等不同途径获取的遥感数据正以每日太比特(TB, 1 TB=1024 GB)级的速度增长, 其中单个遥感图像数据集的数据量就可达几十吉比特(GB). 据统计^[1], 单个数据应用中心将可提供数百个TB的遥感图像, 国家遥感数据存档量将达拍比特(PB, 1 PB=1024 TB)级, 而全球的遥感数据量将达到艾比特(EB, 1 EB=1024 PB)级. 海量遥感数据有待及时、有效地处理与分析.

与此同时, 遥感数据处理算法的“数据密集型计算”特性也日渐凸显. 在遥感数据处理的全流程中, 遥感数据处理速率一般远低于遥感数据获取与记录速率, 两者之间存在较大的数据吞吐性能差距. 卫星数据中心通常需要在一天内完成几百GB遥感图像

的各级数据处理以及各类遥感信息的提取与分析. 全球变化等复杂遥感应用则有必要建立高质量、连续、均一和综合的对地观测系统^[2]. 为实现全球变化关键参数和过程的多变量联合观(监)测, 往往需要对大区域甚至覆盖全球的多时相、多平台、多波段和多空间分辨率的遥感图像进行处理. 而及时准确的海上溢油轨迹预测、泥石流滑坡监测等突发性应急遥感应用通常对处理时效性要求很高, 需在几小时或几十分钟内完成大量的多源遥感图像处理运算.

因此, 有必要深入分析对地观测数据的特点以及大数据背景下对地观测处理所面临的问题与挑战, 并探讨可能的对策与解决方案.

1 对地观测领域进入大数据时代

21世纪人类对地球进行多尺度、全方位实时动态监测的能力进一步增强, 获取全球对地观测信息的遥感和定位卫星系统迅速发展, 遥感数据获取的技术和能力全面提高, 可以说对地观测领域进入了以高精度、全天候信息获取和自动化快速处理为特征的新时代.

自1964年8月28日发射第一颗对地观测卫星Nimbus以来, 到2011年12月全球共发射了514颗对地

引用格式: 何国金, 王力哲, 马艳, 等. 对地观测大数据处理: 挑战与思考. 科学通报, 2015, 60: 470-478

He G J, Wang L Z, Ma Y, et al. Processing of earth observation big data: Challenges and countermeasures (in Chinese). Chin Sci Bull, 2015, 60: 470-478, doi: 10.1360/N972014-00907

观测卫星。美国地球观测系统(Earth Observation System, EOS)计划的提出和实施带动了新一轮对地观测技术发展的浪潮,而地球科学事业(ESE)战略计划是对EOS的提升与延续,将地球系统科学的概念引入到计划中,把对地观测技术与面临的科学问题紧密结合起来(<http://science.nasa.gov/about-us/science-strategy/past-strategy-documents/earth-science-enterprise-plans/>);欧洲太空局以遥感卫星1号、2号以及环境卫星等而立足于世界对地观测技术前列;法国的高分辨率SPOT卫星系列在世界对地观测领域占据一席之

地;加拿大则以雷达卫星系列为其对地观测技术的特色发展战略;日本制定了未来对地观测基本策略,并给出了未来卫星研制和发射计划日程;发展中国家印度也非常重视对地观测技术的发展,其成就已经引起世界广泛关注^[3]。从图1可以发现,进入21世纪后,全球对地观测进入了高速发展阶段,人类的生存、发展越来越依赖于对地观测技术。如图2所示,在未来的20年中,全球对地观测卫星发射计划将持续发展,我国所占发射份额快速增加^[4]。

政府间对地观测协调组织(Group on Earth Observ-

图1 1962~2012 年全球发射对地观测卫星情况^[4]

Figure 1 Earth observation missions in categorizing countries from 1962 to 2010^[4]

图2 2013~2035 年全球对地观测卫星发射计划^[4]

Figure 2 Launch plan of the Earth observation satellites in the world from 2013 to 2035^[4]

ations, GEO)已于2005年2月16日正式成立. 2014年1月15~17日, GEO在瑞士日内瓦召开了第10次全会和第三次部长级峰会. 会议一致认为要继续推动《GEOSS十年执行计划》的实施. 此计划致力于提高全球人类的生活质量. 目标是建立一个综合、协调和持续的全球综合地球观测系统, 同时在灾害、健康、能源、气候、天气、水、生态系统、农业和生物多样性等9个社会发展领域开展应用和信息服务, 为各国决策者提供从初始观测数据到专门产品的信息服务(<http://www.earthobservations.org/>).

1986年中国遥感卫星地面站的建立标志着中国的遥感应用进入了新的纪元. 在过去的30年间, 中国遥感卫星地面站先后接收了包括Landsat, SPOT, JERS, Radarsat, ERS, Envisat, CBERS, HJ, ZY和GF等国内外系列卫星数据, 截止2013年存档各类对地观测卫星数据资料达330余万景, 是我国最大的陆地观测卫星数据历史档案库. 其中, 仅美国陆地卫星Landsat TM和ETM影像就有63万景左右, 时间跨度为1986~2011年. Landsat 8也于2013年发射升空. 这些卫星数据以合适的空间分辨率记录着人类活动和自然变化, 成为最长时间系列的星载陆地观测数据集. 特别是我国陆地观测卫星数据全国接收站网建成以后, 密云、喀什和三亚3个接收站实现了覆盖我国全部领土和亚洲70%陆地区域卫星数据的接收(图3). 正在建设的极地站将进一步扩展我国卫星数据的接收范围. 另外, 气象和海洋等卫星系列也为人类认知地球提供了时空动态数据. 近年来, 随着城市地理信息系统发展和新一代高分辨率卫星系统相继投入应用, 数据的年增量加速提高, 每年约递增30~50 TB. “十二五”期间, 我国计划发射5~6颗地球观测卫星, 建成高空间分辨率、高时间分辨率和高光谱分辨率的对地观测系统, 数据量将大幅增加. 作为高分辨率对地观测系统的首发星, 高分一号卫星突破了高空间分辨率、多光谱与宽覆盖相结合的光学遥感等关键技术, 分辨率可达2 m, 经过相机多角度视场拼接, 优于16 m分辨率的视场可达800 km以上, 4天即可完成一次重访, 在分辨率和幅宽的综合指标上达到了目前国内外民用光学遥感卫星的领先水平(<http://www.cnsa.gov.cn/n1081/n7634/n516721/n516736/611179.html>). 而2012年升空的ZY-3卫星, 每天获取的数据量就在10 TB以上. 到2020年, 高分系统与其他观测手段相结合, 将形成具有时空协调、全天

时、全天候和全球范围观测能力的稳定运行系统. 可以说对地观测领域已经正式步入了大数据时代.

2 对地观测大数据处理与服务面临的挑战

2.1 对地观测大数据特点分析

从所涉及和应用的数据来看, 对地观测大数据具有鲜明的大数据“4V”特征. 对地观测不仅记录地球现状, 而且可以展现地球演变历史和预测未来发展, 数据体量十分巨大, 规模已达到EB级. 在数据类型上, 对地观测大数据包括文档、视频、图片和地理位置信息等, 涉及对地观测、科学模型、社会和经济等多种数据, 类型繁多. 对地观测大数据的数据来源多样, 不仅有大量的高分辨率遥感卫星数据, 基于有线或无线传感器的地基观测技术也迅速发展, 来自各种地基观测系统的数据量呈指数递增. 此外, 对地观测大数据处理系统具有对海量数据进行快速处理、实现数据到信息快速转化的能力, 能够为人类可持续发展面临的环境、灾害和生态等问题提供第一时间的信息服务支持^[5,6].

对地观测数据已成为空间地球信息科学研究的基础资料. 但是, 对地观测大数据的机理模型及其在科学发现中的理论与方法仍有待深入研究. 不仅要关注对地观测大数据所具有的大数据“4V”特征, 也要重视对地观测大数据的其他本质特性.

对地观测数据的“4V”特征: (1) 海量. 对地观测大数据具有海量数据特点, 高分辨率、高动态的新型卫星传感器不仅波段数量多、光谱和空间分辨率高、数据速率高、周期短, 而且数据量特别大, 仅EOS-AM和PM每日获取的遥感数据量就达TB级, 全球对地观测数据已经达到EB级. (2) 多源. 对地观测大数据的多源性, 一方面表现在数据来源和获取手段多样, 既有来自于分布全球的观测网络实时接收的大量遥感数据, 也有通过航空拍摄获得的遥感数据, 还包括大众用户通过互联网和带有地理信息的手持终端设备提供的个性化信息. 而且, 主、被动遥感在成像机理和成像模型等方面也存在巨大的差异. (3) 多时相. 遥感图像是某一时刻传感器对地观测的记录, 卫星通常按固定的轨道周期对地球进行重复观测. 单颗卫星时间分辨率的提高和在轨卫星数量的不断增加使得对地观测的采样间隔在缩短, 数据获取的频率大幅度增加. 通过地面传感网等手段获

图3 陆地观测卫星数据全国接收站网覆盖示意图

Figure 3 The coverage of China remote sensing satellite receiving station networks

取数据的频率则更高。(4) 高价值。对地观测数据的价值体现在商品价格和应用价值两个方面。虽然中低分辨率卫星数据已经逐步实现共享,但国际上高分辨率卫星遥感数据的价格仍然不菲,按数据种类的不同,每平方千米的价格在几十元到几百元人民币不等;实际上,对地观测数据的应用价值更为可观。对地观测数据不仅在科学研究、生态环境、土地资源、自然灾害和重大工程的监测与评估等方面得到广泛应用,而且也在数字地球、智慧城市建设

中发挥着重要作用,并逐步深入到大众生活,产生了巨大的经济价值和社会价值。

其他特征:(1) 异构。对地观测大数据的异构性一方面表现为系统异构,即数据生产所依赖的业务应用系统存在差异,如数据来自不同的数据中心;另一方面表现为模式异构,数据的逻辑结构或组织方式不同。(2) 多尺度。多尺度是对地观测大数据的重要特征,这是由于对地观测系统是由不同级别的子系统组成,各个系统都有各自的时空尺度,因而对地

观测大数据也具有空间多尺度和时间多尺度的特点,在不同的观察层次上所遵循的规律和体现的特征不尽相同. (3) 非平稳. 对地观测大数据因为具有广泛的获取方式和物理意义,因此从信息理论来说是典型的非平稳信号,即分布参数或者分布规律随时间发生变化,非平稳性正是经典遥感数据挖掘与分析理论所忽视的.

2.2 对地观测大数据处理面临“数据密集型计算”问题

对地观测大数据处理是整个对地观测系统的重要组成部分. 一个全流程的对地观测数据处理过程涉及从数据接收和记录、数据传输、数据预处理(辐射校正、系统几何校正)、深加工处理(精校正、正射校正、图像融合、图像变换)、数据产品存档与发布、信息提取与参数反演、以及专题应用(资源环境调查、灾害监测、全球变化)等多个环节. 中国遥感卫星地面站目前接收13颗国内外卫星(表1),在数据接收、记录与传输环节,2012年单颗卫星的数据下传码速率高达320 MB/s,多颗卫星的总码速率可达1.5 GB/s;密云、喀什和三亚各卫星接收站每天接收的单颗卫星数据为10~187.5 GB. 目前,在数据接收、数据记录和数据传输环节具备实时的数据处理能力,基本能与卫星数据下传的码速率同步. 但是,据地面站运行处理系统的统计数据显示,在预处理环节的整体数据处理速度普遍小于2 MB/s,深加工处理环节的数据处理速度则小于1 MB/s,而信息提取及专题遥感应用的数据处理速率也小于1 MB/s. 因此,在遥感数据预处理及后续处理环节的数据处理速率远

远落后于数据接收、记录与传输的速率^[7].

截止到2012年,3个接收站每天接收的总数据量为993.16 GB,每年接收的总数据量约为354 TB. 在传统的服务模式中,卫星地面数据处理系统常采用订单任务方式,只对少量用户请求的卫星数据进行处理,而大部分卫星数据则直接保存在数据存档系统中. 随着我国卫星接收站网布局的扩大以及数据中继卫星的发展,卫星下行数据量将大幅提高,将给对地观测数据处理的全流程带来巨大的数据吞吐压力,尤其是对于数据处理速率低且相当费时的深加工、信息提取以及应用处理等环节. 庞大的数据吞吐压力使得这些处理环节往往面临着“数据密集型计算”挑战性问题^[7].

表2给出了对地观测数据处理全流程中各个数据处理环节算法复杂度的分析结果. 从表中可以看出数据预处理算法的复杂度相对较低,而后续的信息提取与参数反演类算法以及深加工算法的复杂度相对较高^[7].

表 1 中国遥感卫星地面站接收国内外卫星情况
Table 1 The satellites being received by China remote sensing satellite ground station

	系列	编号	颗数
国产卫星	资源一号	02C	1
	资源三号	01	1
	环境减灾卫星系列	HJ-1A/1B/1C	3
	实践九号卫星	A/B	2
	GF	1	1
国际卫星	LANDSAT	8	1
	RADARSAT	2	1
	SPOT	5/6	2
	THEOS		1
合计			13

表 2 对地观测数据处理全流程中的算法复杂度分析
Table 2 The complexity analysis of the earth observation data processing algorithms

处理环节	算法	算法复杂度	并行性
数据预处理	辐射纠正	$O(n)$	好
	几何纠正	$O(n^2)$	好
	图像配准	$O(n^2 \log(n))$	好
深加工处理	K-T变换	$O(p^2bn^2)$ (p : 主成分数; b : 波段数)	较好
	图像卷积	$O(k^2n^2)$ (k : 卷积核阶数)	较好
	K-Mean分类	$O(rbm n^2)$ (r : 迭代次数; b : 波段数; m : 分类数)	一般
信息提取(分类)	Bayes分类	$O(bmn^2)$ (b : 波段数; m : 分类数)	较好
	BP神经网络分类	$O(bmn^2)$ (b : 波段数; m : 分类数)	较好

图4以CBERS-02和LANDSAT5为例,对整个数据处理流程的数据吞吐率进行对比分析。对地观测数据处理全流程包括从数据接收(RC)、数据记录(RD)、数据传输(TR),数据预处理-零级处理(L0)、辐射校正(L1)、几何校正(L2),深加工处理-精校正(L3)、正射校正(L4)、镶嵌(MO),信息提取-以图像分类(CL)、气溶胶反演(AQ)为例,专题应用-以海上溢油(OS)为例等^[7]。

从图4可以看出,在整个对地观测数据处理流程中,卫星数据在接收、记录和传输阶段基本具备实时处理能力,因为其处理速度与卫星下行码速率基本一致。但是,在数据预处理、深加工处理、信息提取以及专题应用等环节中,其数据吞吐速率直线下降,只有卫星下行码速率的1%~30%。因此,后续的数据处理环节无法满足实时的数据处理需求。而且,越往后的数据处理环节,其数据吞吐速率也越低。

2.3 对地观测数据增值服务需求日益旺盛

进入21世纪以来,各种商业卫星遥感数据不断涌现,高分辨率卫星影像的应用日益广泛,卫星遥感增值服务市场欣欣向荣。特别是美国国家航空航天局(NASA)和欧洲空间局(ESA)实行的卫星遥感数据共享以及低价商业分发政策,进一步推动了世界卫星遥感增值服务市场的快速发展。

由于原始数据销售的利润越来越低,世界卫星运营商已调整其发展战略以适应这种形势,即把增值服务作为获取新的利润增长点的一个最重要战略加以重视和发展。法国SPOTImage把“大力发展经加工处理的‘增值产品’和归档的、可多次销售的现成

品”作为提高竞争力和获取利润的手段之一。Digital Global在不同的场合宣传自己“不仅仅是卫星数据提供商,同时也是信息的服务商”。增值服务已成为相关遥感机构关注的焦点,也将成为其获取利润的有效举措。实际上,行业应用也对遥感数据增值产品有广泛的需求。

Forecast International在《2013~2022年全球民用&商业遥感卫星市场》报告中预测,2013~2022年全球将发射113颗遥感卫星,市场规模将达到196亿美元。报告认为,尽管军队和政府仍然是遥感卫星数据的主要用户,但商业市场发挥着越来越重要的作用。对于非专业用户,卫星运营商将更多地提供经过深加工处理后的信息,而不是原始数据,以进一步扩大市场^[8]。

正如Frost & Sullivan公司2011年发布的“亚太地区卫星对地观测市场”研究报告中指出的那样:遥感市场的未来依赖于增值服务,该领域的竞争非常激烈。到2020年,卫星图像作为一种商品,从卫星图像提取有用信息的增值服务将决定商品的价值。对于相关企业来说,未来发展的方向将从卫星图像供应商转变为信息服务商(http://www.researchandmarkets.com/reports/1803560/asia_pacific_satellitebased_earth_observation)。

旺盛的对地观测数据增值服务需要“数据密集型计算”的技术支撑。举例来说,利用Landsat数据生成一幅30 m分辨率的除我国南海、东海以外的陆地部分的卫星影像,需要537景数据(图5(a)),其处理过程包括数据的正射校正、投影转换、色彩均衡、图像镶嵌以及图像增强等步骤,往往需要花费几天甚至十几天的时间来完成(图5(b))。从对地观测数据处理全流程分析可知,这些环节恰恰面临对地观测数据的“数据密集型计算”问题。

3 对地观测大数据处理与服务的几点思考

前面阐述了对地观测大数据的特点,系统分析了对地观测大数据处理全流程,指出了对地观测数据增值服务的旺盛需求以及对地观测大数据处理面临着“数据密集型计算”的挑战性问题。针对这一问题,对地观测大数据处理应该在系统平台、处理算法和服务模式3个方面开展创新性的研究工作。

3.1 系统平台

针对对地观测数据体量大、算法复杂度高、实

图4 对地观测数据处理全流程的数据吞吐率

Figure 4 The throughput of the earth observation data processing chain

图5 537景L5数据分布图(a)和镶嵌影像图(b)

Figure 5 The distribution (a) and the mosaic result (b) of 537 scenes of Landsat-5 images

时处理需求的特点,需要开展高性能对地观测数据处理软、硬件平台的研究.海量对地观测数据处理通常在卫星数据中心的高性能计算集群上开展.高性能计算集群的计算硬件平台一般包括通用刀片服务器、高性能互联网络以及海量存储设施.

借助计算机技术,尤其是高性能计算学科的最新成果,从高性能计算机软件方法角度提高对地观测大数据的处理性能^[9,10].需要重点开展高性能并行遥感文件系统的体系结构、高性能IO中间件^[11,12]、分布式内存体系结构的运行系统^[13]、并行对地观测数据处理的编程模型^[14]、复杂数据处理流程的优化调度算法^[15,16]等方面的工作.

基于上文讨论的高性能计算软、硬件平台,还需进一步开展针对对地观测数据的存储、计算、共

享和分发的算法软件以及数据结构的研究,包括时空元数据组织、多分辨率影像存储以及共享体系结构等等.

3.2 处理算法

对地观测大数据处理算法复杂、流程众多.同时必须注意到,对地观测大数据来源不仅仅是单一的空间传感器,地面观测、互联网和社交媒体等成为对地观测数据的重要来源.对于多源、高维、异质和复杂关联的对地观测数据,需要开展新型智能处理算法与处理模式的研究:(1)借助智能信息处理技术,尤其是信号处理、图像处理、模式识别和人工智能的最新前沿成果,开发新型处理算法,提高处理算法的智能化程度和效率,降低处理算法的复杂度,从根本上提高对地观测大数据处理效率.需要重点研究对地观测大数据的表征、学习、挖掘和知识发现的理论与方法^[17],以及适应对地观测大数据的认知模型、智能解译、目标提取与识别和信息融合技术等^[18];(2)新型计算模式,如群智感知、众包计算等,为对地观测大数据的协同处理与高效组织提供了全新的机遇.同时,借助这些新型对地观测大数据的计算模式,有望更好地促进对地观测大数据处理中“数据密集型计算”问题的解决,满足人类社会生活对空间信息的需求.

3.3 服务模式

长期以来,遥感数据的分发都是按照“接收-处理-存档-分发”的流程来进行的.通常,数据提供商首先需要完成遥感数据的接收与处理,并对数据进行归档入库,然后用户才能通过检索、下订单的方式来订购或下载数据.在这种分发模式下,从数据提供商接收遥感数据到用户拿到数据产品通常需要几天甚至更长的时间.随着对地观测领域传感器技术的发展、海量多源遥感数据获取能力的提高,多源数据处理呈现出精细化分工与协同式综合并存的发展局面;同时,各种遥感应用又需要得到不同卫星、不同区域或国家数据中心的数据支持,对遥感数据处理提出了复杂多样的需求.应重点研究遥感数据产品标准化、系列化的指标体系与分类体系,主动服务、智能服务、一站式服务的理论、方法与技术,有针对性地提高对地观测大数据处理与产品生产的能力,提高遥感数据产品服务的效率.同时,还应重

视对地观测技术与其他地理信息技术、互联网技术、物联网以及云计算技术等高新技术的深度融合,综合利用卫星、航空、地面乃至大众观测数据,形成新的信息服务模式和产业链。面向未来发展和用户需求,依托中国遥感卫星地面站基础设施和数据资源,中国科学院遥感与数字地球研究所开展了新一代遥感数据服务模式研究,该服务模式集数据、计算及服务于一体,一方面可提供陆地观测卫星下行数据的实时主动推送^[19],同时对下行数据中的异常变化信息进行快速分析与发布,提高数据服务的响应速度,提升对自然灾害和突发事件的快速监测与预警服务能力;另一方面,通过建立“RTU-Ready To Use”产品库,向用户提供标准化、系列化和多样化的数据产品服务,从而逐步从形式和内容上改变

中国遥感卫星地面站传统的卫星数据服务模式,适应对地观测大数据时代的应用需求。

4 结束语

在阐述大数据时代遥感数据特点的基础上,对遥感数据处理流程进行了分析,认为大数据时代的遥感数据处理与分析面临的一个重大挑战是“数据密集型计算”问题,它是制约遥感数据规模应用的关键因素之一。如何快速、自动地进行遥感大数据的处理和分析,进而完成空间数据产品的主动、智能的信息服务,是大数据时代对地观测领域面临的一个严峻课题。本文认为,解决这一问题应该在系统平台、处理算法和服务模式3个方面开展创新性的研究工作。

参考文献

- 1 Lü X F, Cheng C Q, Gong J Y, et al. Review of data storage and management technologies for massive remote sensing data. *Sci China Tech Sci*, 2011, 54: 3220–3232 [吕雪锋, 程承旗, 龚健雅, 等. 海量遥感数据存储管理技术综述. *中国科学: 技术科学*, 2011, 41: 1561–1573]
- 2 Xu G H, Ge Q S, Gong P, et al. Societal response to challenges of global change and human sustainable development. *Chin Sci Bull*, 2013, 58: 3161–3168 [徐冠华, 葛全胜, 宫鹏, 等. 全球变化和人类可持续发展: 挑战与对策. *科学通报*, 2013, 58: 2100–2106]
- 3 Gao F, Feng Y, Hou C M, et al. Development strategy on earth observing technology of the main countries (in Chinese). *Remote Sens Technol Appl*, 2006, 6: 565–576 [高峰, 冯筠, 侯春梅, 等. 世界主要国家对地观测技术发展策略. *遥感技术与应用*, 2006, 6: 565–576]
- 4 Guo H D. *Scientific Satellite for Global Change Research* (in Chinese). Beijing: Science Press, 2014 [郭华东. 全球变化科学卫星. 北京: 科学出版社, 2014]
- 5 Li D R, Tong Q X, Li R X, et al. Current issues in high-resolution Earth observation technology. *Sci China Earth Sci*, 2012, 55: 1043–1051 [李德仁, 童庆禧, 李荣兴, 等. 高分辨率对地观测的若干前沿科学问题. *中国科学: 地球科学*, 2012, 42: 805–813]
- 6 Guo H D, Wang L Z, Chen F, et al. Scientific big data and digital Earth (in Chinese). *Chin Sci Bull*, 2014, 59: 1047–1054 [郭华东, 王力哲, 陈方, 等. 科学大数据与数字地球. *科学通报*, 2014, 59: 1047–1054]
- 7 Ma Y. A research on the key techniques of parallel computing platform for data-intensive remote sensing image processing (in Chinese). Dissertation for the Doctoral Degree. Beijing: Institute of Electronics Chinese Academy of Sciences, 2013 [马艳. 数据密集型遥感图像并行处理平台关键技术研究. 博士学位论文. 北京: 中国科学院电子学研究所, 2013]
- 8 Forecast International. *The Market for Civil & Commercial Remote Sensing Satellites. Analysis Report*. Newtown: Forecast International, 2013. 2013–2022
- 9 Wei J B, Liu D S, Wang L Z. A general metric and parallel framework for adaptive image fusion in clusters. *Concurr Comp-Pract E*, 2014, 26: 1375–1387
- 10 Zhang W F, Wang L Z, Liu D S, et al. Towards building a multi-datacenter infrastructure for massive remote sensing image processing. *Concurr Comp-Pract E*, 2013, 25: 1798–1812
- 11 Ma Y, Zhao L J, Liu D S. An asynchronous parallelized and scalable image resampling algorithm with parallel I/O. In: Gabrielle A, Jaroslaw N, Edward S, et al., eds. *Proceedings of ICCS 2009, Part II*, LNCS 5545. Heidelberg: Springer, 2009. 357–366
- 12 Li G, Ma Y, Wang J, et al. Preliminary through-out research on parallel-based remote sensing image processing. In: Vassil N A, Geert D V A, Peter M A S, et al., eds. *Proceedings of ICCS 2006*, LNCS 3991. Heidelberg: Springer, 2006. 880–883
- 13 Ma Y, Wang L Z, Liu D S, et al. Distributed data structure templates for data-intensive remote sensing applications. *Concurr Comp-Pract E*, 2013, 25: 1784–1797

- 14 Ma Y, Wang L Z, Liu D S, et al. Generic parallel programming for massive remote sensing data processing. In: Cluster Computing (CLUSTER), 2012 IEEE International Conference. Beijing: IEEE, 2012. 420–428
 - 15 Zhang W, Wang L Z, Ma Y, et al. Design and implementation of task scheduling strategies for massive remote sensing data processing across multiple data centers. Software Pract Exper, 2014, 44: 873–886
 - 16 Ma Y, Wang L Z, Albert Y, et al. Task-Tree Based Large-Scale Mosaicking for massive remote sensed imageries with dynamic DAG scheduling. IEEE T Parall Distr, 2014, 25: 2126–2137
 - 17 Wang L Z, Lu K, Liu P, et al. IK-SVD: Dictionary learning for spatial big data via incremental atom update. Comput Sci Eng, 2014, 16: 41–52
 - 18 He G J, Zhang X M, Jiao W L, et al. A study on the data mining strategy based intelligent information processing technologies for satellite remote sensing (in Chinese). Sci Tech Engng, 2005, 24: 1911–1915 [何国金, 张晓美, 焦伟利, 等. 基于数据挖掘机制的卫星遥感信息智能处理方法研究. 科学技术与工程, 2005, 24: 1911–1915]
 - 19 Liu J B, Yang J, Chen F, et al. Location-based instant satellite image service: Concept and system design. Int J Digit Earth, 2014, doi: 10.1080/17538947.2014.942395
-

Processing of earth observation big data: Challenges and countermeasures

HE GuoJin, WANG LiZhe, MA Yan, ZHANG ZhaoMing, WANG GuiZhou, PENG Yan, LONG TengFei & ZHANG XiaoMei

Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China

Earth observation uses spaceborne sensors to obtain complete and systematic information about the earth. With the help of advances in spatial information technologies, the field of earth observation has entered the era of “big data”. By analyzing the entire process of earth observation data processing and associated algorithms, this paper concludes that the challenges facing earth observation “big data” processing are those of data-intensive computing. We suggest that to solve this problem, there should be research and innovation in the areas of system platforms, processing algorithms, and services models.

earth observation big data, data-intensive computing, remote sensing, earth observation

doi: 10.1360/N972014-00907