

顾及时空语义的多主题瓦片数据优化检索方法*

仇林遥^{1,3}, 王萌^{1,3}, 朱庆^{2,3}, 杜志强^{1,3}

(1. 武汉大学 测绘遥感信息工程国家重点实验室, 湖北 武汉 430079;

2. 西南交通大学 地球科学与环境工程学院, 四川 成都 611756;

3. 地球空间信息技术协同创新中心, 湖北 武汉 430079)

摘要:针对虚拟地球可视化中多个逻辑图层叠加产生大量同名瓦片索引重叠引起的数据无效访问和内存冗余等问题,提出顾及时空语义的瓦片数据优化检索方法。在客户端实现顾及时空语义的自适应瓦片优选,将视点信息与数据集的时空范围、分辨率和优先级等语义信息自动匹配与自适应筛选过滤,显著提高目标瓦片数据的命中率;在服务器端实现面向主题的内存数据库瓦片数据缓存,进一步提高瓦片数据二次访问的响应速度。实验表明,该方法不受逻辑图层增量影响,保证目标瓦片较高的命中率,缓存方法进一步提升瓦片二次访问效率,显著提高面向多数据集的海量数据实时可视化性能。

关键词:瓦片金字塔;高分辨率遥感影像;自适应匹配;瓦片优选;内存数据库

中图分类号:TN95 **文献标志码:**A **文章编号:**1001-2486(2015)05-015-06

An optimal retrieval method of multi-theme image tiles considering the spatio-temporal semantics

QIU Linyao^{1,3}, WANG Meng^{1,3}, ZHU Qing^{2,3}, DU Zhiqiang^{1,3}

(1. State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China;

2. Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 611756, China;

3. Collaborative Innovation Center of Geospatial Technology, Wuhan 430079, China)

Abstract: Aiming at the invalid accesses and memory redundancies caused by data overlap of more than one logic map tile in the visualization process of virtual earth, an optimal retrieval method of multi-theme tiles considering the spatio-temporal semantics was proposed. On the client side, the self-adoptive retrieval of tiles considering spatio-temporal semantics automatically matches the viewpoint information with semantics of datasets like space-time range, resolution and priority, then the invalid selection was filtered and the efficiency of target tiles data hit ratio was improved. On the server side, the theme oriented data cache based on memory database was achieved to prompt the response speed of server in the second access of tiles. Finally, the experiments prove that this method can prevent stable and high hit-rate of target tile affected by amount of datasets and the cache method can further improve the efficiency of tiles access. The real-time visualization performance of massive image data is significantly improved.

Key words: tile pyramid; high resolution remote sensing images; self-adoption; optimal retrieval of tiles; memory database

通过航空航天遥感技术获取全球高时空分辨率的遥感影像并建立覆盖全球的数字地球已成为当代地理信息技术的重要标志^[1-4]。基于全球离散格网系统的空间数据组织方法实现了海量高分辨率遥感影像的网络可视化服务^[5],并广泛应用于灾害管理、环境监测、地籍管理、战场环境仿真等领域。现有面向公众服务的虚拟地球平台大多管理一个逻辑图层。逻辑图层指包含一个完整或者局部连续的影像金字塔的数据集合,以 Google

Earth 为例,金字塔层级上的影像来自 QuickBird, LANDSAT, IKONOS 等不同传感器^[6-7],但所有影像经过建库后属于同一个逻辑图层。在相同的层级和位置仅存在唯一瓦片,因此根据层级与行列号即可快速检索到目标瓦片。由于高分辨率影像数据快速且容易获取,仅一个逻辑图层难以满足多元遥感数据可视化应用需求。例如应急救援过程,需要应对多源数据记录灾区不同时相不同尺度的受灾情况与灾区场景。数据临时整合至同

* 收稿日期:2015-07-06

基金项目:国家自然科学基金资助项目(41171311,41471320);国家高分辨率对地观测系统应用系统(民用部分)建设资助项目(03-Y30B06-9001-13/15);四川省科技计划资助项目(2014SZ0106)

作者简介:仇林遥(1988—),男,河南安阳人,博士研究生,E-mail:qiu_linyao@163.com;

杜志强(通信作者),男,副教授,博士,硕士生导师,E-mail:duzhiqiang@whu.edu.cn

一逻辑图层,一方面需要长时间的人机交互;另一方面,仅一个逻辑图层难以满足多主题图层的对比分析、多时相展示等可视化需求。当图层不唯一时,以金字塔层级和行列号为索引的离散格网方法难以确定当前视口所请求的瓦片属于哪个图层,传统方法通过瓦片编号在所有数据集获取目标瓦片,导致以下两个问题:①盲目遍历所有数据集容易产生大量无效查询,增加服务端的数据检索时间以及降低传输层中有效瓦片的占有率;②如果多个数据集在空间上存在交集(如图 1 所示),不仅同名瓦片产生内存冗余,而且绘制端难以准确辨识特定数据源的瓦片将导致绘制混乱,影响可视化的流畅度和精确性。针对以上问题,本文提出顾及时空语义的瓦片检索优化方法。

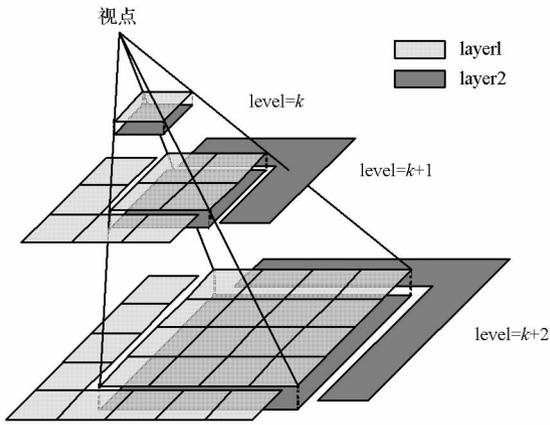


图 1 多图层叠加示意图

Fig. 1 Overlap of multiple logic layers

1 全球离散格网瓦片数组组织模式

全球离散格网的基本思想是采用倍率方法形成多分辨率层次,每层细分为大小相等的矩形瓦片,瓦片包含固定数据的采样点。通过 Plate Carree 等投影方式将地理坐标经纬度 $[-180^\circ, 180^\circ]$ 和 $[-90^\circ, 90^\circ]$ 范围内的地球表面投影成一个长宽比为 2 : 1 的规则矩形平面,以此为底面构建离散多分辨率影像瓦片金字塔,然后对各层进行均匀剖分^[8]。若影像瓦片像素大小为 $s \times t$, 函数 $f(l)$ 定义为第 l 层的空间分辨率,则瓦片的空间分辨率为(以 x 方向为例):

$$f_x(l) = 180.0 / (2^l \times s) \quad (1)$$

任意分辨率正射影像可以映射至金字塔中一段连续的层级。假设其空间范围为 A , 由左下角坐标 (x_l, y_l) 和右上角坐标 (x_h, y_h) 表示,像素大小为 $m \times n$, 定义 r 为影像的空间分辨率,则(以 x 方向分辨率为例):

$$r(x) = \frac{A \cdot x_h - A \cdot x_l}{m} \quad (2)$$

因此,可根据式(1)和式(2)可推算影像映射在金字塔中的最大层级^[8]:

$$l_{\max} = \max \left\{ \left\lceil \log_2 \frac{180.0}{r(x) \times s} \right\rceil, \left\lceil \log_2 \frac{180.0}{r(y) \times t} \right\rceil \right\} \quad (3)$$

由此可见,利用金字塔层级和行列号作为索引可检索唯一瓦片,即给定经纬度 (x, y) , 可在金字塔任意层定位到唯一的行号 X 和列号 Y :

$$X = \left\lfloor \frac{x + 180}{f_x(l)} \right\rfloor, Y = \left\lfloor \frac{y + 90}{f_y(l)} \right\rfloor \quad (4)$$

2 顾及时空语义的自适应瓦片优选算法

面向唯一图层,根据视点位置能够检索和定位任意瓦片。然而,随着数据集的增加,当前方法缺乏对数据集的有效甄选手段,针对这一问题,提出了时空语义标注方法,从多个维度对数据集的可视范围进行约束,实现瓦片自适应匹配算法。

2.1 时空语义标注方法

2.1.1 语义描述

根据《ISO/TC - 211 19115 元数据规范》对遥感数据语义的描述^[9],建立针对影像瓦片数据的语义描述集合,作为瓦片过滤的基本依据。语义描述包括以下内容:

1) 主题语义。针对数据来源、传感器类型、关联事件等方面的差异,建立主题约束。例如在灾区影像数据可视化任务中,不同灾害类型对数据分辨率、传感器偏好不同。主题语义有效增加不同数据集的关联约束,支持用户从应用层面对数据进行批量取舍。

2) 分辨率语义。描述数据集在三维空间垂直维度(垂直于 XY 平面)的“可视范围”,指导视点在瓦片选择过程中自动剔除可视深度以外的数据集。例如在视点接近地表时,用户关心的是高分辨率影像数据集表达的地物细节特点,而非全球背景数据集的低分辨率纹理。

3) 优先级语义。描述数据集调度流程的逻辑顺序,对数据集的深度可视范围做进一步约束。通过建立优先等级与金字塔的局部连续层级的映射关系,实现视点位置与数据集的实时关联分析,分辨率语义和优先级语义是在垂直维度实现瓦片自适应匹配的关键。

4) 时间语义。描述数据集在时间维度的排列顺序和生命周期,以时间戳形式定义数据集的时间有效范围,为数据集的序列分析和多时相展示提供支持。

5) 空间语义。描述数据集在 XY 平面的空间

范围,利用最小外接包围盒表达。结合分辨率语义信息,将数据集的空间范围映射至分辨率可见的各个金字塔层级,将经纬度信息转换为金字塔各级的行列 ID,有助于瓦片的快速确定与剔除。

2.1.2 描述方法

采用资源描述框架 (Resource Description Framework, RDF) 对数据集的语义标注进行表达。RDF 文件对每个数据集的时空语义进行有效的结构化组织,并转变成计算机易于识别和解析的资源信息,实现数据集的自适应匹配与瓦片自动检索。

表1展示了RDF文件的描述片段:在RDF文件中瓦片数据集作为一个类进行描述,时空语义作为类的属性进行表达,包括名称 (Name)、主题 (Theme)、最小外接包围盒 (Box)、影像分辨率 (Resolution)、优先级 (Priority)、起始日期 (Start Time)、终止日期 (End Time) 和数据源 (Data Source) 等,每个数据集属于类的一个实例。

2.2 瓦片自适应匹配过程

传统基于视点位置的瓦片检索方法无法辨别多个逻辑图层中同名瓦片的差异,采用遍历数据

集的方法查询。本文的改进之处在于通过RDF文件描述数据集的时空语义特点,在瓦片检索过程中结合当前视点位置信息对瓦片进行自适应匹配和筛选。图2展示了瓦片自适应匹配流程。

表1 RDF文件描述片段

Tab.1 Snippets of a RDF file

```

<!-- http://www.semanticweb.org/TileData#Individual2 -->
< owl: NamedIndividual RDF: about = "&TileData; Individual2" >
< RDF:type rdf:resource = "&TileData; Tile_data_set"/ >
< RDFs:label >鲁甸县灾前影像 </RDFs:label >
< Name >鲁甸县灾前影像 </Name >
< Theme >鲁甸地震 </Theme >
< Box > 103. 144195, 26. 978907, 103. 674071, 27. 538004
</Box >
< Resolution > 2.5 </Resolution >
< Priority > 3 </Priority >
< Start_Time > 2012-07-14-21-21-13 </Start_Time >
< End_Time > 2014-08-03-16-30-00 </End_Time >
< Data_Source > Ludian\DOM_20140803163000\tms. xml
</Data_Source >
</owl:NamedIndividual >

```

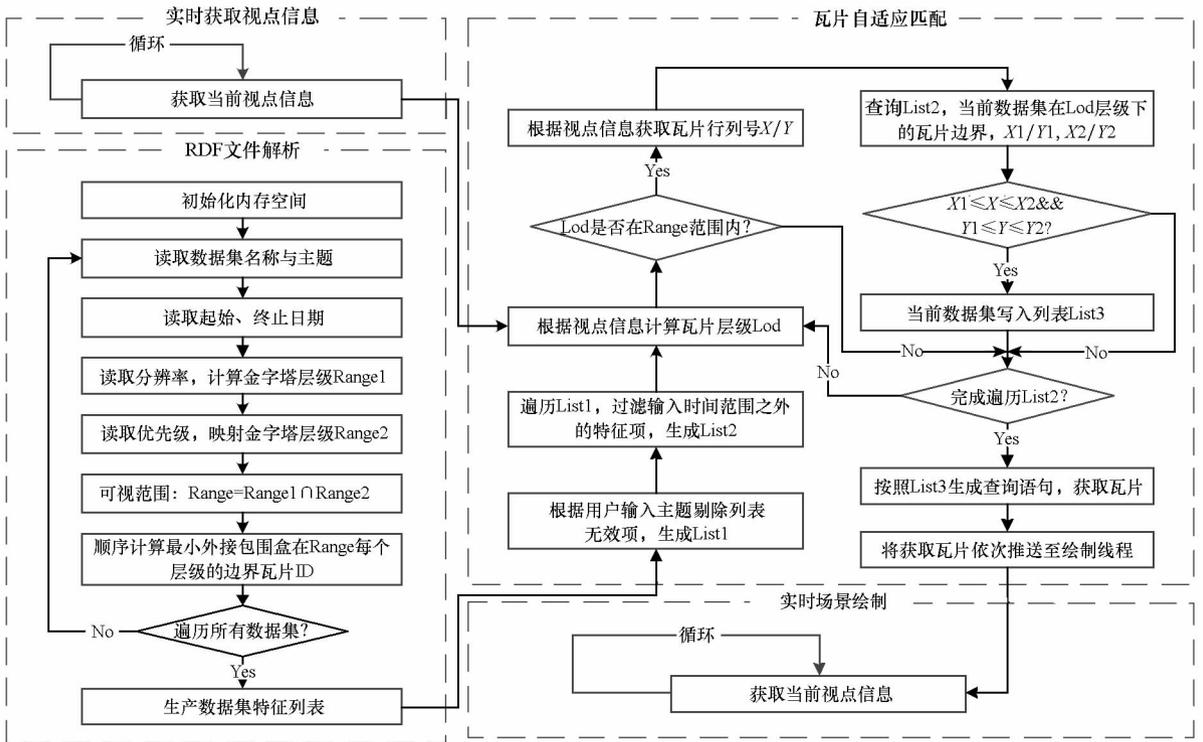


图2 瓦片自适应匹配流程

Fig. 2 Self-adaptive match process

首先遍历并结构化存储所有数据集的时空语义信息,从多个层次对其语义特点进行量化描述:数据集的名称、主题和数据源作为粗粒度的标识

信息进行存储。起始、终止日期记录时间维度的可视范围。影像分辨率与优先级控制垂直维度的可视范围。根据式(3)推算当前数据集分辨率对

应的最大金字塔层级 Level, 确定 0 至 Level 的可视范围 Range1; 根据优先级 Priority 与局部连续金字塔层级建立的映射关系得到可视范围 Range2, 进一步对 Range1 和 Range2 求交集, 确定数据集在垂直维度的范围内。最小外接包围盒控制数据集在水平二维平面的可视域。根据式(4)计算数据集在 Range1 \cap Range2 的各金字塔层级上处于包围盒左下和右上位置的瓦片 ID, 作为范围边界标识。由此, 语义信息量化转换后存储, 准备自适应匹配。

瓦片自适应匹配流程随可视化流程启动, 首先根据用户输入信息获取可视化主题和时相。前者体现了操作者的兴趣对象, 通过信息主题与数据集的相关属性进行匹配, 过滤后形成相关数据集列表 List1。进一步与时间信息进行比对, 剔除列表内输入时间区间以外的数据集, 得到列表 List2, 缩小后续自动匹配和检索的范围。随着视点位置的移动, 实时获取视点信息, 包括视点的高程和经纬度位置, 根据视点高程与金字塔层级划分的映射关系计算视点当前所处的层级 Lod, 结合式(4)计算视点可见瓦片的行列号。然后遍历数据集列表 List2, 将 Lod 与数据集的可见深度进行匹配, 选择垂直维度可见范围包含 Lod 的数据集, 进一步查询该数据集的金字塔数组, 找到 Lod 层级的瓦片边界, 将待请求瓦片的行列号与之进行匹配, 包含瓦片的数据集构成列表 List3, 由此确定最终需要检索的数据集对象。

上述过程能够保证场景浏览时瓦片仅在唯一或极少数数据集中请求。但少量特殊情况下可能出现匹配错误。比如, 当数据集的实际边界与外接最小包围盒重叠度较低时, 少数不在数据集内的瓦片可能被请求。选择内存数据库作为服务器缓存用以提高客户端二次访问速度和正确率。

3 面向主题的内存数据库缓存方法

随着数据集的不断增长, 瓦片总量随之增加。面对海量瓦片, 调度过程中磁盘 I/O 负载较高, 使用将所有数据放在磁盘上进行管理的传统数据库应用方式很难满足高并发、高时效的访问需求, 在服务器端构建内存数据库作为瓦片缓存数据库, 可以有效提高瓦片调度效率和系统性能。

3.1 内存数据库特点

由于存储介质的特性不同, 相对于常规的磁盘数据库, 内存数据库具有更高的访问速度和更低的系统延迟, 并且不受磁盘 I/O 瓶颈限制^[10]。近十几年来, 内存的发展一直遵循摩尔定律, 成本

不断降低, 数据库管理系统将工作数据集放入内存变得可行。另外, 采用分布式内存数据架构容易突破单台普通服务器内存容量低的限制, 最大程度发挥缓存的作用。

3.2 面向主题的内存数据库结构

在瓦片调度过程中, 较多瓦片会不止一次地被访问, 形成“热点数据”, 而每次从磁盘调度瓦片容易增加检索时耗和磁盘 I/O 负载。将数据缓存至内存数据库, 可以明显提高瓦片二次访问的效率。

结合瓦片调度过程, 海量瓦片结构单一、文件大小均衡、操作频率较高的特点, 适合采用 Key-Value 模型的内存数据库, 常见的此类内存数据库包括 Redis, Memcached 和 Riak 等。它支持基于键值对的操作和读取, 数据的写入和读取效率较高^[11-12]。

在数据密集的服务器端构建面向主题的内存数据库(如图 3 所示), 系统架构主要包括以下层次: 代理服务器、逻辑控制脚本、内存数据库集群和磁盘数据库。其中代理服务器接收客户端请求, 发挥负载均衡和请求分发的作用; 逻辑控制脚本在服务器端负责实时生成 RDF 文件、解析代理服务器请求并向内存数据库写入标识符、控制请求响应的优先级。内存数据库集群接到请求时, 优先请求集群中的数据, 有效地降低二次请求的时间消耗, 提升数据传输速度; 瓦片数据库负责海量影像的磁盘存储管理。

客户端根据逻辑控制脚本生产的 RDF 文件, 组合包含名称、主题和影像分辨率信息的数据请求并发送至服务器; 服务器接到请求之后, 逻辑控制脚本将请求解析组合形成“名称: 主题: 影像分辨率”的唯一标识符作为进行瓦片检索的标识符 Key; 逻辑控制脚本利用标志符优先对内存数据库进行检索, 通过哈希函数 $H(x)$, 确定 Key 在内存中的位置 $Addr = H(Key)$; 如果数据存在, 则返回数据给服务器; 如果对无效记录标识列表进行检索, 确保目标瓦片号有效后对磁盘数据库进行检索。如果磁盘中数据存在, 将数据返回给客户端并将此条数据缓存到内存数据库中, 否则, 返回数据为空的消息给客户端, 同时向缓存发送无效记录标识。

4 实验与分析

4.1 实验环境与数据

实验设备为 6 台戴尔 Edge Power R710 服务

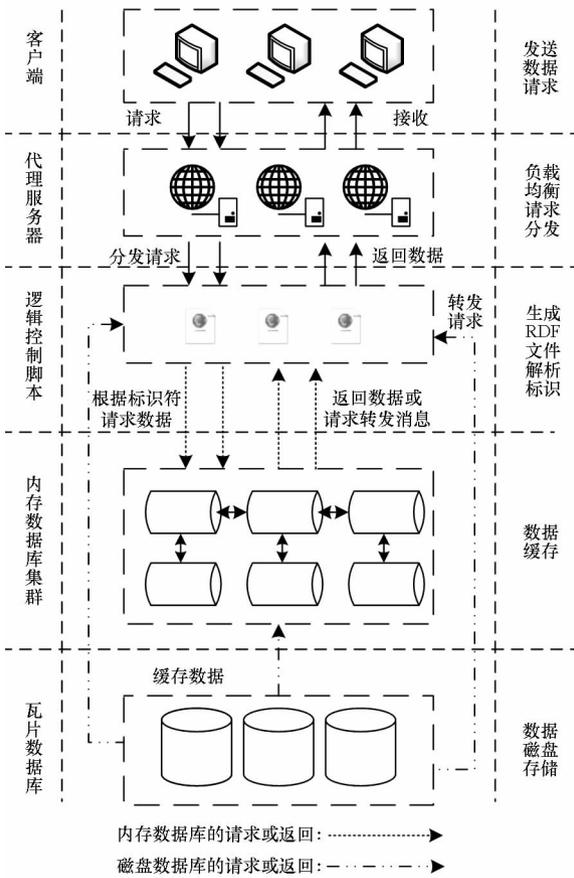


图 3 服务器端数据请求解析和响应流程

Fig. 3 Data request analysis and response on the servers

器,磁盘阵列共有 5TB 容量,服务器操作系统为 CentOS 6.4。实验数据包括全国各省 10m 分辨率遥感影像、鲁甸地震(2014)和尼泊尔地震(2015)等灾区乡镇高分辨率卫星、航空遥感影像,数据集超过 30 个,数据总量达到 3TB。瓦片数据库选择 MongoDB 作为磁盘数据库,内存数据库选择 Redis 作为服务器缓存。其中,6 台服务器分别部署 Redis 节点(每台分配内存 16GB)搭建分布式内存数据库(共 96GB)。

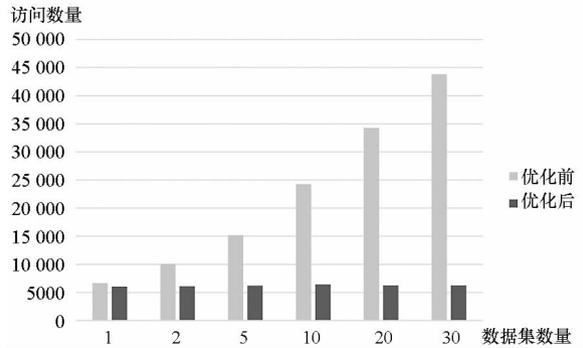
4.2 实验结果与分析

实验一:瓦片请求效率对比

实现瓦片检索优化前后的数据访问效率对比实验。为了保证不同数据环境下有效瓦片请求数量一致,设置统一的三维视点飞行路径。测试不同数据集条件下,优化前后有效瓦片数量的访问效率。测试结果显示:

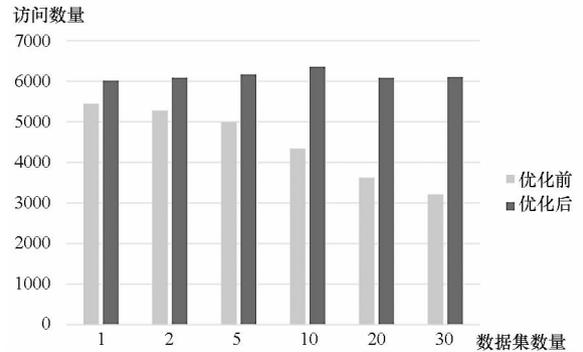
相同三维场景操作状态下,一般瓦片检索和访问总量与数据集数量呈正相关,然而有效瓦片的访问量与有效访问比率随数据集的增加而明显降低。相比之下,优化后的检索方法不受数据集数量的影响,始终保证瓦片总访问量和有效瓦片

访问量的稳定性(如图 4(a)和图 4(b)所示),同时保证有效访问百分比维持在 98% 以上(如图 4(c)所示)。由于传统方法缺乏无效瓦片的自动辨识与过滤方法,导致网络带宽大部分被此类瓦片占据,严重影响客户端实时获取目标瓦片,场景刷新速率随数据集的增加明显下降。经测试,优化方法能够将局域网下平均有效访问速度从 600KB/s 提升至 20MB/s,保证较高的带宽占用率,在多数数据集环境下显著提高海量数据访问与可视化性能。



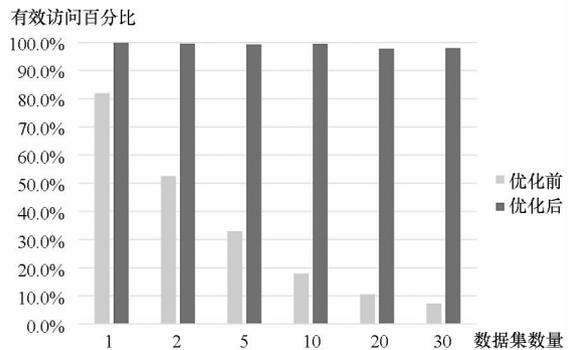
(a) 瓦片访问总量对比

(a) Contrast of total tile query



(b) 有效瓦片访问量对比

(b) Contrast of valid tile query



(c) 有效瓦片访问率对比

(c) Contrast of valid query rate

图 4 自适应瓦片匹配效率对比

Fig. 4 Contrast of self-adaptive tile match efficiency

实验二:缓存访问效率对比

设置 10 台客户端同时对覆盖多数数据集的同

一地区进行场景浏览,包括平移、缩放、漫游和路径飞行,并发访问服务器。测试增加内存数据库前后,代理服务器接收来自数据库响应的时间差异。如图 5 所示,随着操作时间增加,直接从磁盘数据库中获取瓦片的请求响应速率无明显变化,而增加缓存设计的数据库响应时间逐渐降低。此外,缓存容量和响应速率的变化趋势相反,说明本文面向主题的“键-值”设计充分发挥分布式内存数据库的检索能力,能够满足多用户并发条件下海量影像可视化的性能需求。

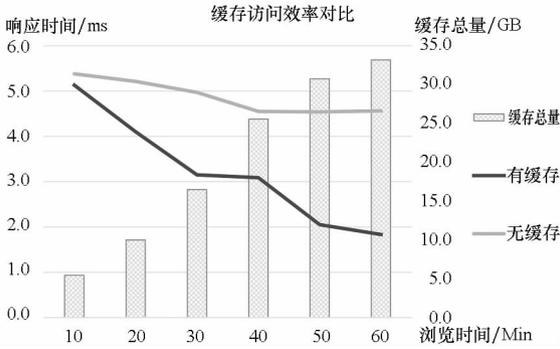


图 5 缓存访问效率对比

Fig. 5 Contrast of cache access efficiency

5 结论

本文针对虚拟地球可视化中多个逻辑图层叠加产生同名瓦片数据重叠引起的数据无效访问和内存冗余等问题,提出顾及时空语义的瓦片数据优化检索方法。在客户端设计顾及时空语义的自适应瓦片优选算法,实现瓦片的自动匹配与过滤,显著提高目标瓦片的命中率,有效解决传统方法瓦片检索低效的问题;在服务器端构建面向主题的分布式内存数据库缓存,利用多字段混合键实现内存数据的高效检索,有效提升瓦片二次数据的访问效率。方法不受逻辑图层个数和数据量影响,满足海量影像数据三维可视化的高性能需求。

参考文献 (References)

[1] 龚健雅. 3 维虚拟地球技术发展与应用[J]. 地理信息世界, 2011,9(2): 15 - 17.
GONG Jianya. The development and application of 3-D virtual

Earth technology[J]. Geomatics World, 2011,9(2): 15 - 17. (in Chinese)

[2] 郭华东,王力哲,陈方,等. 科学大数据与数字地球[J]. 科学通报, 2014,12: 1047 - 1054.
GUO Huadong, WANG Lizhe, CHEN Fang, et al. Scientific big data and digital Earth [J]. Chinese Science Bulletin, 2014,12: 1047 - 1054. (in Chinese)

[3] Grossner K E, Goodchild M F, Clarke K C. Defining a digital Earth system[J]. Transactions in GIS,2008, 12(1): 145 - 160.

[4] Guo H D. China's Earth observing satellites for building a digital Earth [J]. International Journal of Digital Earth, 2012, 5(3): 185 - 188.

[5] 童晓冲. 空间信息剖分组织的全球离散格网理论与方法[J]. 测绘学报, 2011, 40(4): 536.
TONG Xiaochong. The principles and methods of discrete global grid systems for geospatial information subdivision organization[J]. Acta Geodaetica et Cartographica Sinica, 2011, 40(4): 536. (in Chinese)

[6] Potere D. Horizontal positional accuracy of Google Earth's high-resolution imagery archive[J]. Sensors, 2008, 8(12): 7973 - 7981.

[7] Yu L, Gong P. Google Earth as a virtual globe tool for Earth science applications at the global scale: progress and perspectives[J]. International Journal of Remote Sensing, 2012, 33(12): 3966 - 3986.

[8] 刘义,陈萃,景宁,等. 利用 MapReduce 进行批量遥感影像瓦片金字塔构建[J]. 武汉大学学报(信息科学版), 2013, 38(3): 278 - 282.
LIU Yi, CHEN Luo, JING Ning, et al. Parallel batch-building remote sensing images tile pyramid with MapReduce [J]. Geomatics and Information Science of Wuhan University, 2013, 38(3): 278 - 282. (in Chinese)

[9] 朱庆,李海峰,杨晓霞. 遥感信息聚焦服务的多层次语义约束模型[J]. 武汉大学学报(信息科学版), 2009, 34(12): 1454 - 1457.
ZHU Qing, LI Haifeng, YANG Xiaoxia. Hierarchical semantic constraint model for focused remote sensing information services[J]. Geomatics and Information Science of Wuhan University, 2009, 34(12): 1454 - 1457. (in Chinese)

[10] Yue P, Jiang L C. BigGIS: how big data can shape next-generation GIS [C]//Proceedings of the 3rd International Conference on Agro-geoinformatics, IEEE, 2014: 1 - 6.

[11] Hecht R, Jablonski S. NoSQL evaluation: a use case oriented survey[C]//Proceedings of International Conference on Cloud and Service Computing,2011:336 - 341.

[12] Han J, Haihong E, Le G, et al. Survey on NoSQL database[C]//Proceedings of the 6th International Conference on Pervasive Computing and Applications (ICPCA), IEEE, 2011: 363 - 366.