

DOI: 10.11830/ISSN.1000-5013.201811005



运用 MongoDB 的三维时序遥感影像存储及 Web 应用

李嘉欣¹, 余劲松弟¹, 佟瑞菊²

(1. 福州大学 福建省空间信息工程研究中心, 福建 福州 350003;
2. 福建工程学院 交通学院, 福建 福州 350118)

摘要: 为更好地存储和管理时序遥感影像数据,将传统 tif 等格式的时序遥感影像数据集转换为 (x, y, t) 三维的 netCDF 格式,并对其进行八叉树瓦片划分后存储于 MongoDB 中;同时,使用三维 Morton 码建立索引机制.为验证该存储和管理系统,设计基于 Web 覆盖服务(WCS)的时序遥感影像 Web 服务原型,并在其基础上结合 Web 覆盖处理服务(WCPS)和 R 语言对影像作进一步的分析和处理.实验表明:文中所设计的时序遥感影像存储模型在时序检索和获取方面具有较好的性能,并可支持 WCPS, R 语言等互操作计算的扩展.

关键词: 时序遥感影像; netCDF; 八叉树; Morton 码; Web 覆盖服务; R 语言

中图分类号: TP 79 **文献标志码:** A **文章编号:** 1000-5013(2019)04-0535-07

3D Time-Series Remote Sensing Image Storage Using MongoDB and Web Application

LI Jiixin¹, YU Jinsongdi¹, TONG Ruiju²

(1. Spatial Information Research Center of Fujian, Fuzhou University, Fuzhou 350003, China;
2. Department of Transportation, Fujian University of Technology, Fuzhou 350118, China)

Abstract: In order to better store and manage the time-series remote sensing image data, we convert the data-sets of time-series remote sensing image in traditional tif format to (x, y, t) 3D netCDF format, then the data-sets of time-series remote sensing image in netCDF format were divided into octree tree tiles and stored in MongoDB, and use the 3D Morton code to build the indexing mechanism. In order to verify the storage and management system, we design a web coverage service (WCS)-based time-series remote sensing image service prototype, and further analyzes and processes of the image based on web coverage processing service (WCPS) and R language. Experiments show that the time-series remote sensing image storage model designed in this paper has better performance in time series retrieval and acquisition, and it can support the expansion of interoperability calculations such as WCPS and R language.

Keywords: time-series remote sensing image; netCDF; octree; Morton code; Web coverage service; R language

遥感影像数据作为一种超高容量的信息载体,在军事、资源、环境和防灾等领域起着不可替代的作

收稿日期: 2018-11-25

通信作者: 余劲松弟(1984-),男,副教授,博士,主要从事地理信息标准化、数字化对地观测、空间数据库和长期科学数据保存的研究. E-mail: yyx350@126.com.

基金项目: 国家自然科学基金资助项目(41401454);中国博士后科学基金资助项目(2015M582029);福建省教育厅科技项目(JAT160323)

用^[1]. 随着地理信息技术及数据挖掘技术的高速发展,从遥感影像提取信息的能力大幅度提高,多源多时相的遥感影像应用越来越广泛^[2]. 由于仅对单个时相的遥感数据进行分析难以达到时下对信息获取和遥感数据分析的要求,以及现有的遥感影像存储管理较多着重于其空间属性,在管理时序遥感影像时,仅增加一条时间标签或时间语义的记录^[3],难以高效地应对多时相或时序的遥感影像的获取和检索. 现有的二维空间遥感影像存储管理大多是将遥感影像划分为瓦片,并建立相应的索引后,存储于文件系统或关系数据库^[4-5]. 基于文件系统存储遥感影像能够实现影像的便捷管理,但不易实现对影像数据的并发访问及快速检索^[2]. 而基于关系数据库存储遥感影像,一方面,随着遥感影像数据量的增大,入库时间越来越长,难以支撑 PB 级的海量遥感影像管理;另一方面,关系数据库的关系表结构难以应对较为复杂的非结构化或半结构化数据. 相较于关系数据库,非关系数据库能够更灵活地存储和管理非结构化或半结构化数据^[6]. 在遥感影像网络共享方面,开放地理空间信息联盟(OGC)推出一系列面向 Web 服务的地理信息共享和互操作规范^[7]. 其中,Web 覆盖服务^[8](WCS)以 Coverage 的形式提供数字式的地理信息,不仅可以保持遥感影像数据的原始语义,还可以作为遥感影像数据共享规范. 为了更高效地组织管理时序遥感影像,便于时空查询,本文将传统的时序遥感影像数据集转换为 (x,y,t) 三维的 netCDF 格式,并对其进行瓦片划分,在原有四叉树瓦片划分的基础上,增加时间维度 t 的瓦片划分,形成三维八叉树瓦片划分方法,并研究其瓦片的存储和 Web 共享应用的问题.

1 时序遥感影像瓦片化及索引方法

时序遥感影像的瓦片化与传统二维遥感影像瓦片的划分不同,在地理空间二维 (x,y) 的基础上,增加时间 t 维度. 因此,对具有时间维度的时序遥感影像数据集进行瓦片化,需要在原有二维瓦片化的基础上向三维拓展,将传统的四叉树拓展为三维八叉树. 时序遥感影像的三维时空域可以基于三维笛卡尔坐标系定义,该时空域有 (x,y,t) 3 个相互正交的维度轴,每个三维实体对象的时空域 D 由每个维度轴的上界 l 和下界 u 定义,其中,时间轴 t 的下界默认为 1,上界为三维实体对象时间节点的个数,即

$$D = (l_x:u_x,l_y:u_y,1:u_t). \tag{1}$$

式(1)中: D 为时序遥感影像的三维时空域; l_x,l_y,u_x,u_y 分别为时序遥感影像在 x,y 轴的上下界;1 为时间轴 t 的默认下界; u_t 为三维实体对象时间节点的个数.

因此,八叉树瓦片划分方法可表示为对该三维时空域 D 中每个维度轴递归二分. 以第 1 次划分的第 1 个叶节点为例,八叉树划分的叶节点的时空域 D_1 可表示为

$$D_1 = \left[l_x:l_x + \left\lceil \frac{u_x - l_x}{2} \right\rceil, l_y:l_y + \left\lceil \frac{u_y - l_y}{2} \right\rceil, 1:\left\lceil \frac{u_t}{2} \right\rceil \right]. \tag{2}$$

式(2)中: D_1 为八叉树第 1 次划分的第 1 个叶节点. 在各维度进行递归二分,为保证瓦片大小为整数,中值数字取值为向上取整. 以 x 轴为例,其二分区间可表示为

$$\left[l_x:l_x + \left\lceil \frac{u_x - l_x}{2} \right\rceil \right], \quad \left[l_x + \left\lceil \frac{u_x - l_x}{2} \right\rceil:u_x \right].$$

在编码方法上,传统的 Morton 码是一种四叉树编码方式,二维数据可以通过 Morton 编码表示为一维数据^[9]. 当四叉树瓦片化方法向三维拓展为八叉树瓦片化方法时,相应地需要将 Morton 码向三维拓展. 三维 Morton 码可以由八叉树瓦片划分的叶节点的三维行列号 (X,Y,T) 计算得到^[10]. 首先,根据八叉树瓦片划分得到三维行列号,并将三维行列号转为二进制. 然后,以错位交叉的方法由二进制行列号得到二进制的三维 Morton 码. 最后,将其转换为十进制表示. 具体八叉树瓦片划分方法和三维 Morton 码错位交叉的计算,如图 1 所示.

在错位交叉方法中,错位交叉的先后顺序不同,会产生不同的 Morton 码序列,这将直接影响数据库内瓦片数据的存储顺序. 如按 x,y,t 的顺序进行错位交叉形成的 Morton 码,首先,注重于瓦片时间序列 t 轴的临近关系;其次,注重空间上 y 轴的临近关系,在形成整个 Morton 码序列时,可使临近时间序列的瓦片排列顺序尽可能连续,从而提高在数据库中按时间序列检索的性能.

基于 Morton 码的八叉树瓦片索引能较好地管理规则的时序遥感影像瓦片,检索时,用户只需输入目标影像区域的三维时空域 $D_m=(l_x:u_x,l_y:u_y,l_t:u_t)$,即可根据瓦片在三维时空的大小及整个时序遥感影像的三维时空域 D 计算出目标影像区域所包含瓦片的三维行列号,从而可以根据三维行列号计算出所包含瓦片的 Morton 码,以此实现根据三维时空坐标查询和检索时序遥感影像瓦片.瓦片划分的流程图,如图 2 所示.

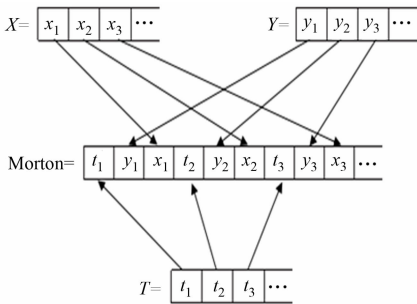


图 1 二进制三维 Morton 码错位交叉
Fig. 1 Binary 3D Morton code misalignment cross

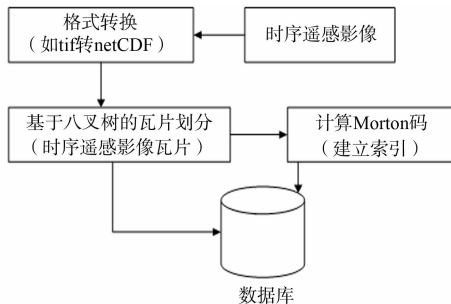


图 2 时序遥感影像瓦片划分流程图
Fig. 2 Time series remote sensing image tile division flow diagram

2 基于 MongoDB 的影像瓦片及元数据存储

将时序遥感影像的二进制瓦片文件和元数据分别存储.瓦片文件以 nc 格式存储于 GridFS,元数据以 BSON 格式存储.每个瓦片对应一个 fs.files 和多个 fs.chunks,其文件存储在 fs.chunks 中,其描述信息存储在 fs.files 中.fs.files 记录文件的序号 n 为瓦片 Morton 码,多个 fs.chunks 中的“file_id”与 fs.files 中的“_id”相对应,瓦片描述通过 Coverageid 与 BSON 元数据相关联.存储框架,如图 3 所示.

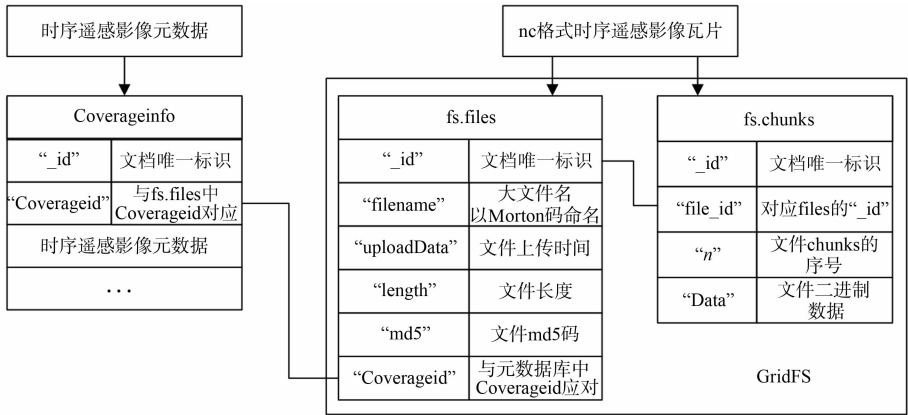


图 3 时序遥感影像存储结构
Fig. 3 Time series remote sensing image storage structure

时序遥感影像元数据相对于一般遥感影像元数据,除了需要具备基础元数据项外,还需要在索引域及时间信息上进行扩展. OGC Web 覆盖服务定义了 GetCapabilities, DescribeCoverage 和 GetCoverage 3 个核心操作,同时继承覆盖模型标准^[11].文献[11]将 Coverage 模型的地理要素以结构化树的形式组织起来,主要包括索引域(DomainSet)、值域(RangeSet)、值域类型(RangeType)和元数据(MetaData)4 个模块.因此,参考 OGC 制定的 Coverage 模型^[12],使用三维索引域描述时序遥感影像 (x,y,t) 三维时空范围.

有关基础元数据部分,根据文献[13]制定的影像和格网数据内容模型标准,该标准针对当前广泛使用的影像和格网数据进行分类,并针对不同数据类别制定基础元数据,以及设定可选项和必选项,可支持扩展.其中,对于光学影像,文献[13]定义了获取时间、影像描述、镶嵌情况、波段信息、卫星平台和传感器等基础元数据项.

针对三维时序遥感影像元数据,不仅需要在索引域扩展出时间维度信息,并记录三维遥感影像时间节点,还需要修改获取时间的元数据,按照三维遥感影像时间节点的排列顺序存储时序遥感影像所有时相的影像获取时间.其余基础元数据项因为同属同一数据源,在影像描述、镶嵌情况、波段信息等基础元数据部分均保持一致.因此,时序遥感影像元数据总体可分为 2 部分:1) 基于文献[13]定义的基础元数据项;2) 基于文献[11]定义的三维索引域和值域信息,并在此基础上增加 TimeSeries 记录来描述时序遥感影像的时间信息.具体时序遥感影像元数据表结构,如图 4 所示.

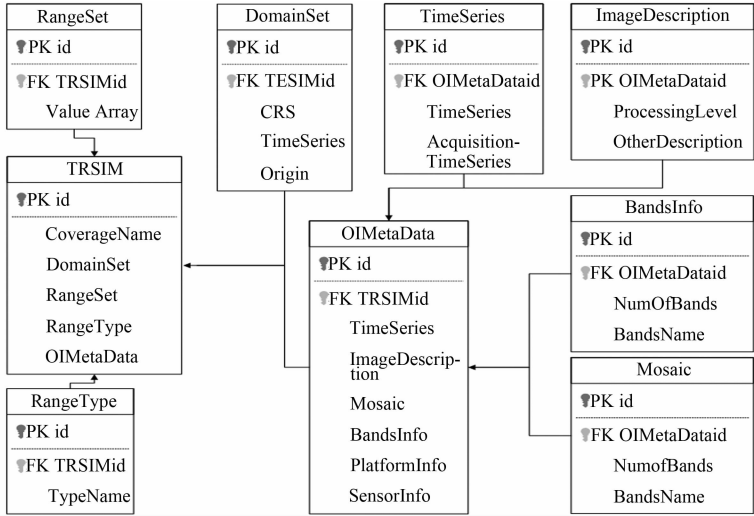


图 4 时序遥感影像元数据表结构

Fig. 4 Time series remote sensing image metadata table structure

在时序遥感影像元数据表结构中,DomainSet,RangeSet,RangeType,OIMetaData 分别存储时序遥感影像的索引域、值域、值域类型及光学影像元数据,在遥感影像元数据中,OIMetaData 部分主要记录光学遥感影像基础元数据.DomainSet 除了需要记录目标时空域的二维索引域,还需要记录时空域的时间序列,构成三维时空坐标系的三维时空索引域.其中,DomainSet/TimeSeries 仅记录时序遥感影像的时间序列号,其序列号对应的遥感影像获取时间记录在 OIMetaData/TimeSeries 中.除了 TimeSeries 外,在光学影像的基础元数据中,ImageDescription 记录影像处理级别及其他描述信息;Mosaic 记录镶嵌信息;BandsInfo 记录波段信息;PlatformInfo 记录卫星平台信息;SensorInfo 记录传感器信息.

3 性能测试比较与分析

为验证所提的时序遥感影像三维瓦片化存储模型的性能,将所提的三维格网化存储与不划分瓦片存储,以及传统的遥感影像二维瓦片化存储的影像获取性能进行比较.将自变量设置为获取目标影像瓦片时间节点的个数(N),因变量设置为检索并获取目标影像瓦片所需响应时间(t),为排除第 3 方格式转换和影像融合所造成的影响,该性能比较试验不进行格式转换.性能测试结果,如图 5 所示.

由图 5 可知:相较于无瓦片划分,对遥感影像进行二维或三维瓦片化都可以明显地提高遥感影像子集获取的效率,瓦片化存储可以明显减小影像子集获取的数据量;对于二维和三维瓦片化存储,在获取单个时间的遥感影像瓦片时,由于三维瓦片文件大小大于二维瓦片,无法体现出其瓦片获取性能的优势,但随着获取遥感影像瓦片的时间节点个数的递增,三维瓦片可以极大地减小影像瓦片获取时的 IO 操作的次数,从而在获取多个时间节点的遥感影像瓦片时有较高的效率,可更高效地管理时序遥感影像.

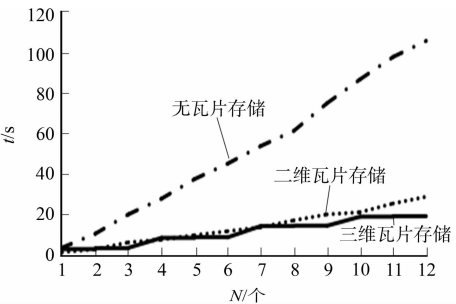


图 5 3 种存储方式瓦片获取性能比较
Fig. 5 Comparison of three storage mode tile acquisition performance

4 基于 WCS 的时序影像管理服务及 WCPS 和 R 语言计算

在基于 MongoDB 的时序遥感影像存储的基础上,为进一步完善时序遥感影像的管理,首先,以时序遥感影像数据集作为数据来源,使用 Gdal 作为空间数据转换工具,将时序的遥感影像集转换为具有时间维度的 netCDF 格式,并进行瓦片划分后存入 MongoDB;然后,构建基于 OGC WCS 的 Web 覆盖服务原型,根据 WCS 服务原型检索,获取目标时空域的时序遥感影像子集,并使用 Web 覆盖处理服务(WCPS)和 R 语言对目标时空域的时序遥感影像进行进一步分析和计算。

Gdal 是一种在 X/MIT 许可协议下的开源栅格空间数据转换工具,它利用抽象数据模型来表达所支持的各种文件格式^[14]. Gdal 不仅可以对各种栅格进行读写,还可以实现栅格空间数据的格式转换、几何校正、重投影、重采样、裁剪等。

Rasdaman 是由德国不莱梅雅各布斯大学和 Rasdaman 公司合作开发的多维栅格数据管理系统,支持 OGC WCS 2.0 核心及其拓展,包括 WCS-T 和 WCPS 等. WCPS 是一种独立于 Web Coverage 的查询语言,支持对 Coverage 进行代数运算、逻辑运算和聚合运算等. R 是一种用于统计分析和绘图的语言和操作环境,Shiny 是一个开源的 R 包,为使用 R 构建 Web 应用所提供的一种 Web 开发框架,可以进行快速数据分析和可视化。

文中设计的基于 MongoDB 的 WCS 服务原型将原始遥感影像增加时间维度,使用 Gdal 将时序的遥感影像数据集转换为 netCDF 格式后进行瓦片划分,并将 netCDF 格式的三维 (x,y,t) 瓦片存储于 GridFS 中;在用户发送获取影像请求时,根据用户发送的时空域请求,快速索引瓦片,最终调用 Gdal 进行瓦片融合,并转换为用户需要的格式返回给用户. 数据流程图,如图 6 所示。

该服务原型可以结合 Rasdaman 的 WCPS 和 R 语言进行进一步的数据挖掘,形成时序遥感影像获取和处理服务链。首先,在该 WCS 快速原型及影像计算服务链中,WCS 服务原型可直接获取用户请求的目标遥感影像数据. 然后,将用户请求的目标遥感影像数据上传至 Rasdaman,并发送 WCPS 计算请求,对遥感影像进行初步计算后,将 WCPS 计算结果传递给 Shiny,并使用 R 语言对计算结果进行进一步统计分析. 最后,获取的遥感影像经过 WCPS 计算和 R 语言分析得到可视化结果. WCS 快速原型及影像计算服务链架构,如图 7 所示. 图 7 中:IBI 为基于遥感指数的新型城市建筑用地指数。

5 应用案例分析

5.1 研究区及实验数据

开封市位于我国黄河中下游平原东部,与交通枢纽郑州相邻,其城市发展历程可以追溯到春秋战国时期,是我国历史文化古城发展的缩影,研究开封市的城市发展进程和趋势对建设我国古代历史名城有着重要的参考意义. 开封市主城区位于东经 $113^{\circ}52'15''\sim 115^{\circ}15'42''$,北纬 $34^{\circ}11'45''\sim 35^{\circ}01'20''$ 。

实验数据来源于美国地质勘探局 USGS 下载的 Landsat 5 TM 影像,产品级别为 L1T,地面分辨率为 30 m,从 2000—2011 年期间共有 12 幅影像,日期分别为 2000 年 7 月 27 日,2001 年 7 月 14 日,2002 年 8 月 2 日,2003 年 9 月 22 日,2004 年 7 月 22 日,2005 年 8 月 26 日,2006 年 6 月 26 日,2007 年 7 月 31 日,2008 年 8 月 2 日,2009 年 9 月 22 日,2010 年 8 月 8 日和 2011 年 7 月 10 日. 文中实验主要选择夏季或盛夏时期的遥感影像数据,该季节植被较为茂密,闲置耕地较少,减少荒地对 IBI 的计算和提取结果的影响。

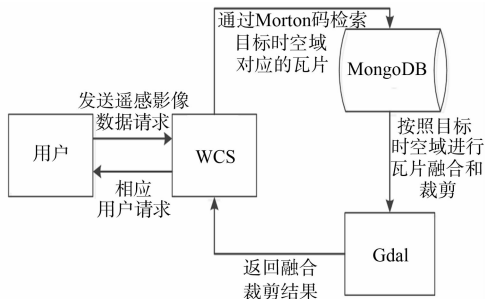


图 6 WCS 原型数据流程图

Fig. 6 WCS prototype data flow diagram

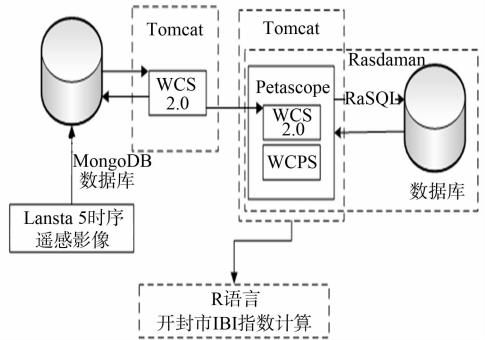


图 7 WCS 原型及影像计算服务链架构图

Fig. 7 WCS prototype and image calculation service chain architecture diagram

5.2 案例分析

通过所提的 WCS 原型服务可以检索并获取开封市主城区的时序遥感影像数据,并上传至 Rasda-man. 以徐涵秋^[15]提出的 IBI 计算为例,调用 WCPS 分别计算出每年开封市主城区的归一化建筑指数(NDBI)、修正归一化水体指数(MNDWI)和土壤调节植被指数(SAVI). 以 2000 年开封市 NDBI 指数计算为例,WCPS 语句为

```
for c in (kaifeng2000)
  return encode((c.4-c.3)/(c.4+c.3),"png")
```

上述语句中: c 为 2000 年开封市遥感影像的整体变量; $c.3, c.4$ 分别为用于 NDBI 计算的近红外波段和中外波段.

WCPS 计算 NDBI 的可视化结果,如图 8 所示. 使用 R 语言对 WCPS 计算出的 NDBI, MNDWI 和 SAVI 3 种指数进行 IBI 指数计算,最终获得开封市主城区 IBI 指数的计算结果,如图 9 所示. 图 8, 9 中: 仅展示变化较为明显的计算结果(2002, 2004, 2008, 2010 年).

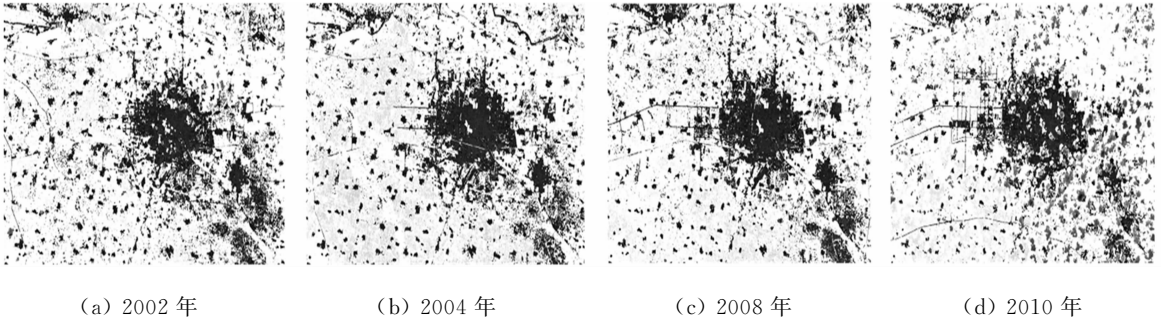


图 8 WCPS NDBI 指数计算结果

Fig. 8 NDBI index calculation result by WCPS

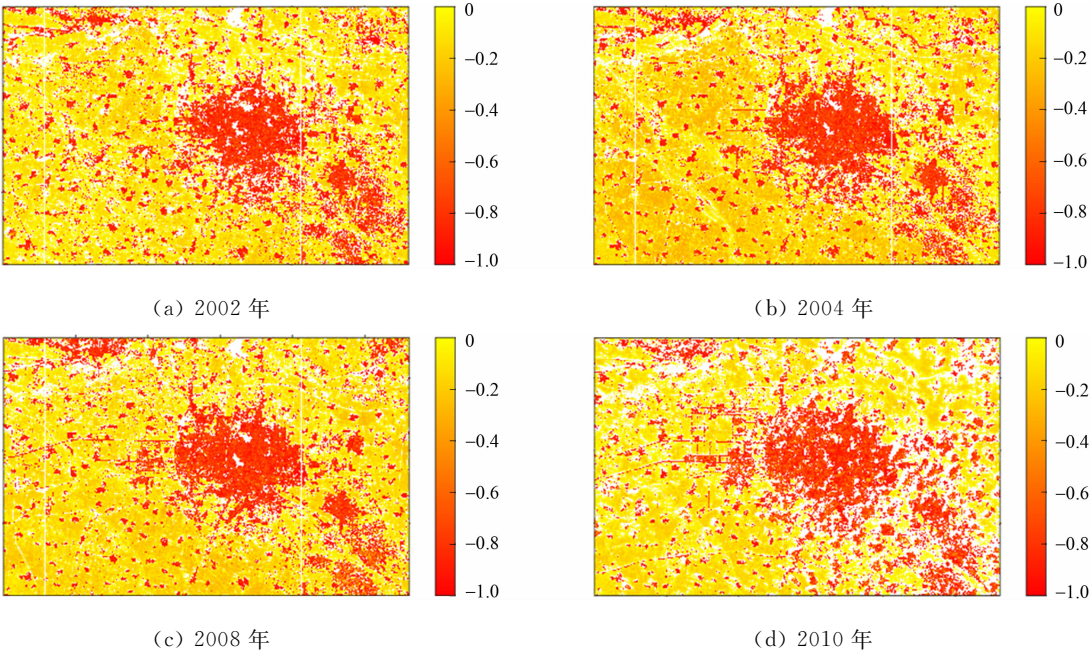


图 9 R 语言 IBI 指数计算结果

Fig. 9 IBI index calculation result by R language

由图 8, 9 可知: NDBI 指数可视化结果和 IBI 指数可视化结果可以较为直观地看出开封市主城区主要城市建筑的空间分布. 相较于 NDBI 指数, IBI 指数能够更加有效地排除土壤、水体及植被对建筑物提取的影响, 且新的指数可以进一步影响城市生态环境的因子研究, 对研究城市发展历程和建筑物规划分布有指导性意义.

由图 9 可知: IBI 指数可视化结果可以明显看出开封市城区有较为明显的向东西方向发展的趋势,

开封市西边毗邻我国中原交通枢纽郑州市,一方面,这种向西发展的模式可以更好地和郑州东区发展相互结合,使郑汴一体化进程进一步加快;另一方面,城区建筑物分布逐渐变为松散,城市中掺杂较多空白区域,这表明在 2002—2010 年间,开封市的城市发展在向西扩张,加速周边区域城市化的同时,对老城区进行了进一步的生态改造,有效地优化了开封市主城区的生态环境,这种发展经验也对我国其他历史名城的现代化和城市化建设有指导意义。

6 结束语

基于 MongoDB 进行时序遥感影像的三维格网化存储研究,且以 WCS 的形式发布为 Web 遥感影像获取服务,可以具有针对性地解决区域时序遥感影像的存储和检索问题。同时,该影像获取服务可以进一步与 WCPS 服务结合,形成影像获取和影像计算 Web 服务链,并支持与其他影像处理服务协同对遥感影像数据进行深度挖掘或进一步计算研究。

参考文献:

[1] 李芳,邬群勇,汪小钦. 基于 OGC 规范的遥感影像数据服务研究[J]. 测绘信息与工程,2009,34(4):30-32. DOI:10.14188/j. 2095-6045. 2009. 04. 011.

[2] 刘桂红. 多源海量影像数据库关键技术研究[D]. 北京:北京建筑工程学院,2011.

[3] 仇林遥,王萌,朱庆,等. 顾及时空语义的多主题瓦片数据优化检索方法[J]. 国防科技大学学报,2015,37(5):15-20. DOI:10.11887/j. cn. 201505003.

[4] 谢毅,陈丹捷,魏国丽. 一种基于文件系统的海量遥感影像存储组织结构[J]. 福建电脑,2011,27(1):110-111. DOI:10.3969/j. issn. 1673-2782. 2011. 01. 056.

[5] 刘伟,刘露,陈莹,等. 海量遥感影像数据存储技术研究[J]. 计算机工程,2009,35(5):236-239. DOI:10.3969/j. issn. 1000-3428. 2009. 05. 081.

[6] 陈超,王亮,闫浩文,等. 一种基于 NoSQL 的地图瓦片数据存储技术[J]. 测绘科学,2013,38(1):142-143. DOI:10.16251/j. cnki. 1009-2307. 2013. 01. 030.

[7] 龚健雅,高文秀. 地理信息共享与互操作技术及标准[J]. 地理信息世界,2006(3):18-27. DOI:10.3969/j. issn. 1672-1586. 2006. 03. 003.

[8] PETER B. OGC[®] WCS 2.0 interface standard-core-corrigendum; OGC 09-110r4[S]. [S. l.]:[s. n.],2012.

[9] 张天蛟,严泰来,王海蛟,等. 基于 Morton 码的土地空间网格数据组织与检索[J]. 农业工程学报,2013(增刊):235-243. DOI:10.3969/j. issn. 1002-6819. 2013. z1. 037.

[10] 韩慧民,郑少开. 基于线性八叉树的点云数据存贮与索引[J]. 大众科技,2014(6):39-41. DOI:10.3969/j. issn. 1008-1151. 2014. 06. 014.

[11] Norwegian Technology Centre. Final text of CD 19123 geographic information; Schema for coverage geometry and functions; ISO/TC 211 N 1475[S]. Norwegian:ISO/TC 211/WG 2/19123 Editing Committee,2003.

[12] 谭玉敏,池天河,唐中实. GML 的空间信息映射模式[J]. 华侨大学学报(自然科学版),2004,25(1):87-90. DOI:10.3969/j. issn. 1000-5013. 2004. 01. 022.

[13] International Standards Organization. Geographic information; Content components and encoding rules for imagery and gridded data (part 1); Content model; ISO/TS 19163-1; 2015(E)[S]. Geneva:[s. n.],2015.

[14] 葛亮,何涛,王均辉,等. 基于 GDAL 的瓦片切割技术研究[J]. 测绘与空间地理信息,2014(7):130-132. DOI:10.3969/j. issn. 1672-5867. 2014. 07. 041.

[15] 徐涵秋. 一种基于指数的新型遥感建筑用地指数及其生态环境意义[J]. 遥感技术与应用,2007(3):301-308. DOI:10.3969/j. issn. 1004-0323. 2007. 03. 001.

(编辑: 李宝川

责任编辑: 钱筠

英文审校: 吴逢铁)