数字地球系统中海量数据存储与传输方案研究

王 柏,胡谷雨,罗健欣
(解放军理工大学 指挥自动化学院,江苏 南京 210007)

摘 要:文中目的是设计一种基于四叉树索引的海量数据存储和传输方案,并将其应用在自主开发的数字地球系统中。使用四叉树索引完成海量影像与高程数据的存储方案设计,并在此基础上设计了一种基于海量数据的传输方案。该方案在数字地球系统的展示效果表明,这种基于四叉树索引技术的海量数据存储技术以及网络传输方案,能够满足影像、高程数据实时可视化的需要,并且利于确定视野可见范围内的地图块位置,利用截锥裁剪技术减少系统渲染的数据量,提高系统的执行效率。该方案在数字地球系统中发挥了重要作用,有一定的实用价值。

关键词:数字地球;四叉树;海量数据;网络传输
中图分类号:TP311 **文献标识码**:A **文章编号**:1673-629X(2012)03-0081-04

Research of Massive Data Storage and Transmission for Digital Earth System

WANG Bai, HU Gu-yu, LUO Jian-xin
(Institute of Command Automation, PLA University of Science & Technology, Nanjing 210007, China)

Abstract: A method used in massive data storage and transmission is needed, which is based on quadtree, also digital earth developed independently takes advantage of the method timely. Mostly, designed the database for image and altitude datas, and proposed the network transport based on massive data. The experiment indicates that this method can meet the needs of real-time visualization, also, it brings advantage to ascertain the map areas in sight, which can greatly decrease the number of datas to rend, and improve the performance of the system. At last, the methods played an important role in digital earth system, and it is practical.

Key words: digital earth; quadtree; massive data; network transport

0 引 言

美国前副总统戈尔的数字地球(Digital Earth)构想[1]一经提出即引起了各方面的广泛关注,许多国家都在积极开展关于数字地球的研究与探索工作。Google Earth是由Google于2005年6月推出的系列软件,它以三维地球的形式把大量卫星图片、航拍照片和模拟三维图像组织在一起,使用户从不同角度浏览地球;Skyline Globe Enterprise Solution(Skyline globe企业解决方案)是美国Skyline公司为网络运营三维地理信息提供的企业级解决方案,该方案包括了Skyline整套软件工具,让用户能够根据自己的需求建立个性化的三维地理信息系统;World Wind是由美国国家航空暨太空总署(NASA)阿莫斯研究中心的科研人员开发的开放源代码,它可以利用Landsat 7,SRTM,MODIS,GLOBE,Landmark Set等多颗卫星的数据,让用户体验三维地球遨游的感觉。

国内成熟的数字地球系统则相对较少,而数字地球在军事上的应用更是少见。但是数字地球在国防建设中却具有十分重要的意义,主要应用包括建立服务于战略、战术和战役的各种军事地理信息系统,并运用虚拟现实技术建立数字化战场等。此外,数字地球是一个典型的平战结合、军民结合的系统工程,建设中国的数字地球工程符合我国国防建设的发展方向。

因此,设计一个自主开发的数字地球系统很有必要。其关键技术之一是海量数据的存储与快速处理,连绵不断的地图影像数据以及地形数据信息,涉及到海量数据的可视化计算、存储处理等,现有的硬件和软件不可能支持整个大规模的图形数据进行一次性处理[2,3],要实现大规模影像高程数据的高速实时处理,海量数据的存储和调度是需要解决的首要问题。文中主要讨论数字地球关键技术中基于四叉树索引的海量

收稿日期:2011-07-26;**修回日期**:2011-10-29
基金项目:国家自然科学基金资助项目(60303023)
作者简介:王 柏(1987-),男,硕士研究生,研究方向为计算机应用技术;胡谷雨,教授,博士生导师,研究方向为网络管理。

数据存储和传输技术的研究。

1 海量数据存储设计

地图块的影像数据基于四叉树结构:第 1 层仅一个影像数据项,即整个地球的正向墨卡托投影,第 2 层有四个数据项,拼接即为第 1 层数据项,但分辨率提高了四倍,是第 1 层数据项的四个子节点,以此类推,构成一个金字塔式的数据集合。目前所使用影像数据最高分辨率已达到 20 层。

类似的,地图块的高程数据也基于四叉树结构,第 1 层有 9 * 18 个高程数据项,各个数据项的四个子节点构成第 2 层的数据项,即第 2 层有 9 * 18 * 4 个高程数据项,之后依此类推。高程数据和影像数据都是基于四叉树结构的,故数据库的设计也是大同小异,下面以影像数据为例进行两方面的介绍:数据库结构的设计和表结构的设计。

1.1 数据库划分算法

第 1 层至第 Lev 层所包含的影像数据个数分别为 $1,4,16\cdots\cdots4^{Lev-1}$。理论上第 20 层应该有数据 4^{19},约为 2749 亿。当然并不是所有地方都有 20 层的数据,相反,只有一些大城市才能够到达如此高的分辨率。为方便管理,将 10 亿定为一个数据库的数据上限即 DBMax = 1000000000,从而将海量影像数据存储到多个数据库中,提高了处理效率[4]。

数据库划分算法[5]:

1. 计算当前层数据总数。Sum = 4^{Lev-1}。

2. 判定 Sum>DBMax 是否成立,如果不成立,进入下一层并将 Sum 累加。即:Lev++,Sum+= 4^{Lev-1}。如果成立,转至步骤 3。

3. 计算当前数据总数应包含数据库个数。N = (int)(Sum/DBMax)+1。

4. 根据当前最高层号以及数据库的个数命名该层包含的数据库名称。

例如前 15 层共有数据项 Sum1 ~ 15 = $\frac{1*(1-4^{15})}{1-4}$ ≈3.6 亿,为方便数据的管理,避免因数据量过大导致遍历数据效率低下,将 1 ~ 15 层数据设计为第一个数据库,命名为 Map15,而第 16 层共有数据 Sum16 = 4^{16-1} ≈10.7 亿,故分为 2 个数据库,命名为 Map16_01 和 Map16_02。

1.2 表划分算法

影像数据项的个数随层数递增呈几何增长,因此前几层影像数据比较少,为方便管理,经过实践检验,现将 5 千万定为一个表内数据项个数的上限,这样可以使得效率较高,即:TableMax = 50000000。

表划分算法:

1. 计算当前层数据总数。Sum = 4^{Lev-1}。

2. 计算当前数据总数应包含表个数。N = (int)(Sum/ TableMax) +1。

3. 根据当前层号以及表的个数命名该层包含的表名称。

例如第 14 层共有数据项 Sum14 = 4^{14-1} ≈6.7 千万,故第 14 层数据分为 2 个表,命名为 Level14_01,Level14_02。

以上数据均为理论值,仅在有数据时才创建该数据所在表并将数据添加进表内。因此实际数据库没有理论上那么庞大。

1.3 数据具体分配流程

现有数据项所在层行列为 image(Lev,Row,Col)。如何获取该数据项所在的数据库和表名称呢?

分配算法:

1. 计算数据项在当前层的标号:Id = (Row-1) * $2^{(Lev-1)}$ +Col。

2. 确定数据项所在数据库名称:如果 Lev≤15,数据库名为 Map15,如果 Lev>15,则计算 n = (int)(Id/DBMax)+1,根据层号以及 n 即可确定数据库名(例如 Lev=17, n =2,数据库名称即为 Map17_02)。

3. 确定数据项所在表名称:计算 m = (int)(Id/TableMax)+1,如果 m =1,则直接用层号确定表名(例如 Lev=9,表名称即为 Level09),如果 m >1,则根据层号以及 m 值确定表名(例如 Lev=14, m =3,表名称即为 Level14_03)。

1.4 表结构设计

影像数据在数据库中的存储方式是二进制形式的,并且每个二进制影像数据对应着该数据在数字地球场景中的位置,所以,在数据库的表结构中应当包含影像数据的位置信息、二进制数据信息,以及数据信息的长度。表结构设计如表 1 所示[6]:

表 1 数据库中的表结构

列名	数据类型	说明
ID(主键)	bigint	数据在本层内的唯一标号
Row	int	数据项所在行号(上-下)
Col	int	数据项所在列号(左-右)
Map	image	数据(二进制形式)
Size	int	数据的长度

这样,就可以方便地利用视点依赖技术,确定当前场景内的地图块,根据地图块在数字地球场景中的位置访问数据库,读取对应的影像数据信息。

2 高效数据传输技术设计

数据储存到了服务器的数据库中,怎样由客户端

根据当前视野范围实时地更新下载相应地图块的影像与高程数据? 这就需要设计一个数据传输方案,高效快速地获取需要的数据。

2.1　报文格式

请求报文的内容主要包括当前报文的版本号、所要执行的操作类型、需要传输数据的数据类型、对数据的位置描述以及该描述的长度信息等。

对应的服务器返回的处理结果,包括结果信息长度和结果信息。结果信息可能是客户端需要的数据信息;也可能是出错信息,如数据库访问失败、数据不存在等情况。

请求报文和服务器返回的处理结果均经过了加密处理,保证了数据在网络传输过程中的安全[7]。

2.2　客户端设计

系统主要分为四个模块:数据更新、数据下载、存储和读取数据的文件系统、数据渲染。这里主要讨论数据下载模块。其主要任务就是与服务器建立连接,并确定需要更新的数据项位置(层行列),封装请求报文,向服务器发送请求报文,并等待服务器返回处理结果。将数据封装到缓存文件中。

2.3　服务器端设计

服务器端是根据请求报文内容进行分支处理的过程,通过类图说明各部分之间的关系(见图 1)。

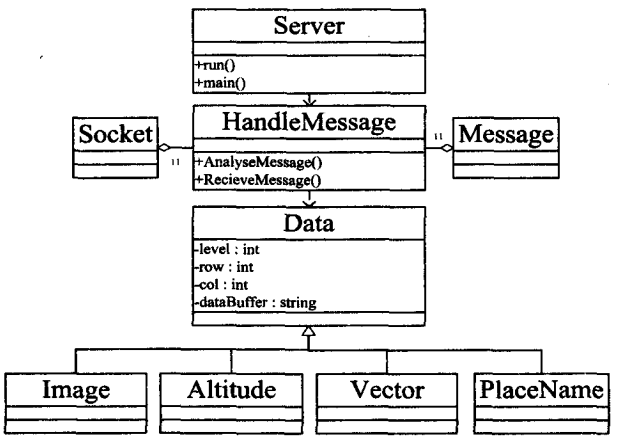

图 1　服务器端类图结构(简化)

服务器端大体分为三部分:服务器总线程部分(类 server)、数据接收部分、各种数据的分类处理部分。总线程部分主要功能就是监听服务端口,一旦有客户端连入,就开启子线程并返回给该客户端新的闲置端口为其服务。各种数据处理部分主要功能就是读取解析客户端的请求信息,根据层行列计算数据所在数据库以及表格位置,然后读取数据并封装返回给客户端。

2.4　客户端与服务器交互设计

客户端与服务器的交互是多对一的关系,服务器通过多线程技术支持多用户的同时访问。具体交互流程如图 2 所示。

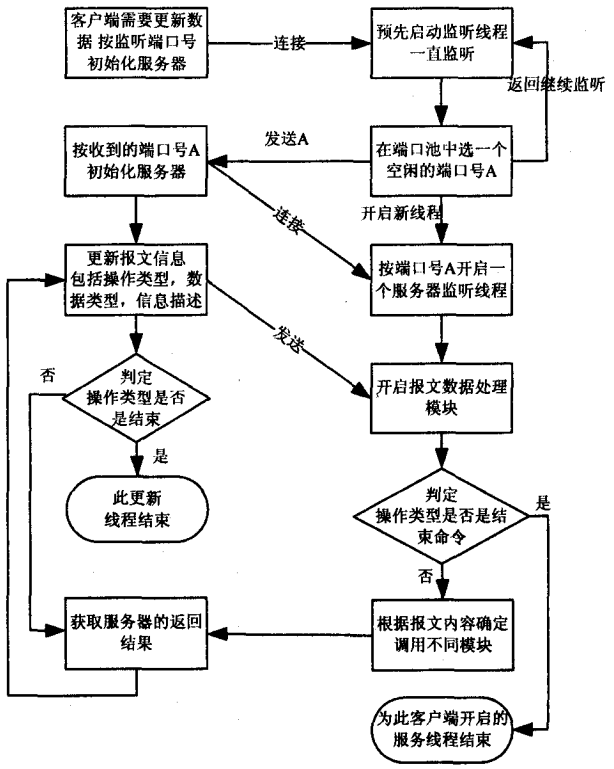

图 2　客户端与服务器交互流程

1. 服务器端端口 5050 一直处于监听状态,客户端在访问服务器时首先连接此端口。

2. 当客户端需要更新数据时,则连接服务器监听端口。

3. 服务器与客户端建立连接,之后在备用传输数据端口池中选择空闲的端口号发送给客户端,返回并重新监听连入的客户端,以支持多用户同时访问。

4. 服务器端利用多线程技术在服务器端按发送的传输端口创建服务器监听线程,为该客户端服务。

5. 客户端收到服务器发来的空闲传输端口号后就按此端口号再次连接服务器。此后的客户端与服务器的交互都是通过该连接端口完成。

6. 客户端根据数据位置以及相应的操作封装请求报文,发送给服务器。

7. 服务器端收到客户端的请求,解析报文内容,根据数据类型执行不同处理模块。最后将从数据库获取的数据信息返回给客户端。

8. 客户端收到服务器发来的数据内容,将其封装在本地磁盘缓存中,以供程序使用。然后循环执行步骤 6。

如上所述,客户端开启后确定需要下载的影像数据所在图层块位置,封装请求报文并发送,服务器对接收的报文进行解析后,根据图层块位置访问数据库并读出对应的影像数据,返回给客户端。当客户端结束时,同时发送结束报文给服务器,服务器端将分配给该

客户端的端口关闭,结束此服务线程[8]。

3　应用实例及结论

文中的研究成果已成功应用于我们自主开发的数字地球系统中。目前,服务器端将400G的影像数据与高程数据存储在SQL Server 2005中,并使用Java部署服务器端接收客户端的请求并为其服务。客户端利用视点依赖技术确定当前视野范围内的地图块位置(层行列),向服务器端发起连接请求,下载需要的影像与高程数据,服务器端根据地图块位置访问数据库,获取数据信息,并返回给客户端。客户端即可利用这些数据完成可视化场景的渲染[9,10]。系统效果如图3~图5所示。

图3　500万米高空鸟瞰中国

图4　1000米高空鸟瞰天安门

图5　峰峦叠嶂的喜马拉雅山脉

实验结果表明,文中提出的基于四叉树索引技术的海量影像与高程数据存储技术以及网络传输方案,能够满足数据实时可视化的需要;通过对视野可见范围区域外地图块的裁剪,大大减少了系统渲染的数据量;通过四叉树索引提供的快速搜索算法迅速得出可见地图块位置,进一步获取影像与高程数据进行实时渲染[11,12],提高了系统的执行效率。

4　结束语

该数字地球系统可以作为军事应用的平台,通过加载各种对应模型实现虚拟数字化战场;可以用于通讯设备的性能监控、环境的实时显示、线路铺设场景模拟等,可以实时反映某设备线路状况,包括导线温度值、微风震动波形等。

参考文献:

[1] Gore A. The Digital Earth-Understanding Our Planet in the 21st Century[R]. Los Angeles California: California Science Center, 1998.

[2] 唐泽圣. 三维数据场可视化[M]. 北京:清华大学出版社, 1999:55-70.

[3] 谭　兵,徐　青,周　杨. 大区域地形可视化技术研究[J]. 中国图象图形学报,2003,8(5):578-584.

[4] 周　鹏,鱼　滨. 基于P2P分布式数据库实时更新[J]. 计算机技术与发展,2007,17(5):144-147.

[5] Date C J. An Introduction to Database Systems[M]. Beijing: China Machine Press, 2002:327-440.

[6] 李晓黎. ASP+SQL Server网络应用系统开发与实例[M]. 北京:人民邮电出版社,2004:117-119.

[7] Wallach D S. A Survey of Peer-to-Peer Security Issues[C]// Proceedings of International Symposium Security. Berlin, Germany: Springer-Verlag, 2002:42-57.

[8] Richter J. Windows 核心编程[M]. 北京:机械工业出版社, 2000.

[9] 普建涛,查红斌. 大规模复杂场景的可见性问题的研究[J]. 计算机研究与发展,2005,42(2):236-246.

[10] 罗健欣,倪桂强. 高性能快速地形漫游算法[J]. 计算机工程,2004,34(8):21-24.

[11] Hoppe H. Progressive Meshes[C]//Computer Graphics (SIGGRPAH'96 Proceedings). [s. l.]:[s. n.], 1996:99-108.

[12] Bradley D. Evaluation of Real-Time Continuous Terrain Level of Detail Algorithms[EB/OL]. 2003-08. http://www.derekbradley.ca/Publications/Papers/carleton03_TerrainLOD.pdf.