

基于分布式系统的海量 DEM 金字塔切片并行构建方法

技术领域

本发明涉及地理空间数据并行处理技术领域，具体涉及一种基于分布式系统的海量 DEM 金字塔切片并行构建方法。

背景技术

目前，随着 DEM（数字高程模型 Digital Elevation Model）的需求日益增大，高精度、高分辨率 DEM 数据量呈几何级的增长，DEM 金字塔切片的理论与方法也趋于成熟，但如何对海量的 DEM 数据高效切片处理，是 DEM 广泛应用亟待解决的关键问题。随着并行计算、分布式计算平台的快速发展和趋于成熟，为海量 DEM 金字塔切片的并行计算提供高效的算力。结合并行计算和成熟的 DEM 采样算法、基于四叉树的分割算法实现 DEM 金字塔切片的并行构建方法，可提高 DEM 数据计算的效率。

近年来，国内外学者开始重点研究海量空间数据的分布式计算和分析，在构建 DEM 金字塔方面，一般还是集中在线程并行策略，主要是结合高性能计算机下利用多线程来构建海量 DEM 金字塔切片，这种方式能够充分发挥机器的性能，但是如果 DEM 数据量较大，一般的计算机内存不能够承受，达不到计算的效果。关于 DEM 分析计算方面也有一些专利。中国专利文献 089098709949.5 公开了一种面向并行数字地形分析海量 DEM 部署与调度方法，解决了海量 DEM 的地形分析的效率，此专利未涉及 DEM 金字塔切片的内容。中国专利文献 CN201710107308.X 公开了一种基于 DEM 的内流湖集水区划定方法，其包括获取设定区域内高程数字模型的 DEM 数据，并对其进行切割，得到基础 DEM 数据，根据基础 DEM 数据确立设定区域内的山脊线，并对空间单元进行分割计算空间栅格数据。本专利没有涉及并行和 DEM 金字塔的算法。中国专利文献 201410687961.4 公布了一种适于规则网格 DEM 的结构化多尺度重构方法，同样没有涉及金字塔建立和并行方法。中国专利文献 201410788404.1 公开了一种面向关系数据库扩展的自适应影像金字塔切片方法，提出了自适应影像切片模型，根据自适应影像切片模型自适应的选择影像切片的大小。影像切片涉及影像元的空值，不涉及 DEM 切片的拼接方面的方法。在软件方面，主要是商业 GIS 软件具有 DEM 金字塔切片的功能，这些方法的

主要采用不同的采样算法来切分 DEM，未实现分布式的 DEM 金字塔切片等内容。

因此，有必要开发一种基于分布式系统的海量 DEM 金字塔切片并行构建方法，能够对大规模或海量 DEM 的金字塔切片进行并行化切分，能够在高性能计算机集群环境下进行分布式计算，减少金字塔构建的时间，提高计算效率，同时可以结合临近平均值方法计算 DEM 切片边缘的值，以解决相邻切片间的平滑过渡。

发明内容

本发明要解决的技术问题是提供一种基于分布式系统的海量 DEM 金字塔切片并行构建方法，能够对大规模或海量 DEM 的金字塔切片进行并行化切分，能够在高性能计算机集群环境下进行分布式计算，减少金字塔构建的时间，提高计算效率，同时可以结合临近平均值方法计算 DEM 切片边缘的值，以解决相邻切片间的平滑过渡。

为了解决上述技术问题，本发明采用的技术方案是：该基于分布式系统的海量 DEM 金字塔切片并行构建方法，具体包括以下步骤：

S1 获取源 DEM 的元数据 Metadata；

S2 建立所述源 DEM 的局部坐标系，对所述源 DEM 进行切分形成若干个可并行计算的切片 Tile（切片，就是将一幅 DEM 切成很多大小一致的小块）；并根据所述源 DEM 的参数计算每一个所述切片 Tile 的索引值记为 Key，以<Key, Tile>键值对为计算单元组成一个保存源 DEM 的元数据的链表 List<Key, Tile>，将<Key, Tile>按照索引值 Key 分布到不同的计算节点进行计算；

S3 根据是否定义了最大层次 MaxZoom，进行设置或计算最大层次 MaxZoom，再设置需要的输出切片 Tile 的参数；再根据设定的坐标系和切分规则计算输出切片 Tile 的参数，所述输出切片 Tile 的参数包括切片的长、宽、分辨率等；

S4 计算节点并行构建所分配的键值对<Key, Tile>，根据索引值 Key 合并同一计算节点上的高程值 Value；分别根据索引值 Key 按照所述输出切片 Tile 的长、宽、分辨率参数计算所对应的切片 Tile 的外包矩形 Extend，计算输入切片 Tile 中每个网格单元 cell 的坐标，再根据设定的输出坐标系，计算每个所述网格单元 cell 在输出坐标系的坐标；根据每个所述网格单元 cell 的坐标计算相应的所述网格单元 cell 所在的输出切片 Tile 的索引值 Key，并将其加入一个映射 Map

中，组成一个映射 $\text{Map}\langle \text{Key}, \text{List}\langle \text{cell} \rangle \rangle$ ，其中 Key 表示切片的索引值， List 表示链表， cell 表示网格单元；

S5 将计算节点上相同索引值 Key 的链表 $\text{List}\langle \text{cell} \rangle$ 合并；根据索引值 Key 合并所有计算节点上的链表 $\text{List}\langle \text{cell} \rangle$ ，形成一个键值对 $\langle \text{Key}, \text{List}\langle \text{cell} \rangle \rangle$ ；

S6 根据输出切片 Tile 的参数和坐标系，计算输出切片 Tile 的网格单元 cell 的坐标，接着根据采样方式从合并的键值对 $\langle \text{Key}, \text{List}\langle \text{cell} \rangle \rangle$ 中计算网格单元 cell 的高程值 Value ；

S7 利用 8 邻域算法，修正输出切片 Tile 边缘的高程值；根据生成的所述链表 $\text{List}\langle \text{Key}, \text{Tile} \rangle$ ，循环获取每个切片 Tile 的四边网格单元 cell 和与之相邻的 8 个切片 Tile ，取相邻切片 Tile 的网格单元高程值的平均值作为相应网格单元 cell 的高程值 Value ，达到切片 Tile 拼接的平滑；

S8 判断当前层次 Zoom 是否满足终止条件，若满足终止条件，则结束；若不满足终止条件则根据上一级层次计算下一级层次和下一级层次的切片 Tile 的外包矩形 Extend 的宽度和高度，计算方法： $\text{NZ} = \text{CZ} - 1$ ， $\text{W} = \text{EW}/2^{(\text{NZ}+1)}$ ， $\text{H} = \text{EH}/2^{(\text{NZ}+1)}$ ，其中 NZ 表示下一级层次， CZ 表示当前层次， EW 、 EH 表示原始 DEM 的宽度和高度， W 、 H 表示切片 Tile 的宽度和高度；

S9 根据上一级层次切片 Tile 和采样算法，并行计算生成下一级层次切片 Tile 和索引值 Key ；根据索引值 Key 、下一级的层次和切片 Tile 的参数计算当前层次切片 Tile 所对应的下一级层次的索引值 Key ，利用所述采样算法从当前层次切片 Tile 采样下一级层次切片的高程值，构成下一级层次切片 Tile 的子切片和索引值 Key ，在每个节点上合并相同的索引值 Key ，构成键值对 $\langle \text{Key}, \text{List}\langle \text{Tile} \rangle \rangle$ ；接着根据索引值 Key ，合并所有节点上相同的索引值 Key 的切片 Tile ，组成下一级层次的切片 Tile ，再执行步骤 S7；

S10 循环执行步骤 S8~S9，直到判断层次 Zoom 满足终止条件，所述终止条件为：若设定了最小层次 MinZoom ，则判断当前层次 Zoom 是否达到设定最小层次 MinZoom ，若达到，则终止；若未设定最小层次 MinZoom ，若当前层次 zoom 小于 0，则终止。

其中 Tile 表示切片， Key 表示切片的索引值， Value 表示高程值， List 表示链表， cell 表示网格单元， Zoom 表示层次， Map 表示映射， Extend 表示源 DEM 的外包矩形；采用上述技术方案，获取待切分源 DEM 的元数据 Metadata ，

作为数据切分和投影转换的基础；建立源 DEM 左上角为原点，向下向右为正方向的坐标系，根据默认的切片 Tile 的宽度和高度对源 DEM 图幅进行切分，根据初步切分的 Tile 的行列号组成索引值 Key，构建一个以索引值 Key 为键，切片 Tile 为值的键值对；将键值对分布在不同的计算节点进行以下的计算；根据设定的 MaxZoom 或计算的 MaxZoom，以及设定的输出切片 Tile 的宽高、坐标系和采样算法，每个计算节点对所分配的输入切片 Tile 进行切分计算，输出索引值 Key 和一个包含网格单元 cell 的链表；接着在每个计算节点上合并相同索引值的网格单元；最后合并不同计算节点上相同的索引值 Key 的 cell，输出多个 Key/Value 对；根据索引值计算相邻的 8 个索引值 Key，获取相邻切片 Tile 与当前切片 Tile 的边缘高程值，将相应的边缘值的平均值作为当前切片 Tile 的边缘值，达到切片 Tile 拼接时的平滑；接着循环计算下一级的层次 Zoom、切片 Tile 的宽高等数据，以上一级的输出链表 List<Key, Tile>作为输入，计算下一级的链表 List<Key, Tile>；循环终止条件是当前层次 Zoom 小于设定的 MinZoom 或 0；其中采样方式为利用采样算法进行采样，采样算法可以是最邻近像元法，双线性内插法，三次立方卷积法等，均为现有技术。

作为本发明的优选技术方案，所述步骤 S1 中所述元数据 Metadata 包括原始分辨率、坐标系，以及在源坐标系下源 DEM 的外包矩形 Extent 数据。获取源 DEM 的 Metadata 的目的是为了后续计算的需要，这些计算包括将源数据进行初步的切分，以便于后续计算的并行化。

作为本发明的优选技术方案，所述步骤 S2 中建立源 DEM 的局部坐标系具体为：建立以源 DEM 左上角为原点，向下向右为正方向的坐标系；根据设定切片 Tile 的宽和高，将源 DEM 数据切分为多个切片 Tile，为了将源 DEM 进行初步的切分，以构成链表 List<Key, Tile>，其中索引值 Key 的计算方法：

$$\text{Key} = C + "_" + R, C = (x - XMin) / TW, R = (y - YMin) / TH$$

其中，C、R 表示切片 Tile 的行列号，x、y 表示切片 Tile 任意一点的切片坐标，XMin 表示源 DEM 的外包矩形 Extent 的 X 最小值，YMin 表示源 DEM 的外包矩形 Extent 的 Y 最小值，TH 是初始切片 Tile 的高，TW 是初始切片 Tile 的宽。

作为本发明的优选技术方案，所述步骤 S3 若未设定层次 MaxZoom，则需要计算层次 MaxZoom，计算过程是根据源 DEM 的外包矩形 Extent 的左下点坐

标和右上点的坐标，并将左下点坐标和右上点的坐标进行投影转换，转换为经纬度坐标，其中左下点坐标是（SLon，SLat），右上点的坐标是（ELon，ELat）；接着根据公式计算层次 MaxZoom，计算公式如下：

$$\begin{aligned}
 dLat &= ((ELat - SLat) / 180.0) \cdot \pi \\
 dLon &= ((ELon - SLon) / 180.0) \cdot \pi \\
 lat1 &= (SLat / 180.0) \cdot \pi \\
 lat2 &= (ELat / 180.0) \cdot \pi \\
 a &= (\sin(dLat / 2) \cdot \sin(dLat / 2)) + (\sin(dLon / 2) \cdot \sin(dLon / 2) \cdot \cos lat1 \cdot \cos lat2) \\
 dist &= R \cdot 2 \cdot \arctan\left(\sqrt[3]{a} / \sqrt[3]{1-a}\right) \\
 MaxZoom &= (\log(2\pi R / dist - tileSize) / \log(2))
 \end{aligned}$$

其中，MaxZoom 为最大层次，dLat 表示纬度差，dLon 表示经度差，lat1 表示 SLat 的弧度，lat2 表示 ELat 的弧度，a 表示计算的中间值，dist 表示两点的距离，tilesize 表示 Tile 设置的最大边长，R 表示 WGS84 椭球的长半轴， $R=637817$ 。计算输出切片 Tile 的 Extend，需要适当的扩大，目的是为了避免输出切片 Tile 拼接不会出现裂缝等问题。

作为本发明的优选技术方案，所述步骤 S4 中网格单元 $cell=\langle x,y,Value \rangle$ ，其中 x，y 是在切片 Tile 的坐标，Value 是高程值。

作为本发明的优选技术方案，所述步骤 S4 并行计算输入切片 Tile，根据输出切片 Tile 的参数将输入切片 Tile 的 DEM 值划分到不同的输出切片 Tile 中；其中首先根据输入切片 Tile 的索引值 Key 计算 Tile 的外包矩形 Extend，计算公式：

$$\begin{aligned}
 MinX &= ExMinX + col \cdot W \\
 MinY &= ExMinY + row \cdot H \\
 MaxX &= ExMinX + (col + 1) \cdot W \\
 MaxY &= ExMinY + (row + 1) \cdot H
 \end{aligned}$$

其中，MinX、MinY、MaxX、MaxY 分别表示外包矩形 Extend 的最小最大 XY 坐标值，col、row 是输入切片 Tile 的行列号，ExMinX、ExMinY 表示源 DEM 的外包矩形 Extend 的最小的 X、Y 坐标值，H 表示切片 Tile 的高，W 表示切片 Tile 的宽。

作为本发明的优选技术方案，所述步骤 S7 利用 8 邻域算法，修正切片 Tile 边缘的高程值 Value 时，对输出的链表 List<Key, Tile>并行计算相邻的切片

Tile, 取四周除交点外的相邻网格单元 cell 值, 并对上下或左右的 cell 取值, 将取到的 cell 值取平均值作为指定网格单元 cell 的高程值 Value; 四个角点的高程值的计算是取切片 Tile 的相邻的另外的三个切片 Tile 与相应角点的值, 并取平均值作为相应网格单元 cell 的高程值 Value。

与现有技术相比, 本发明具有的有益效果为: 该基于分布式系统的海量 DEM 金字塔切片并行构建方法具有以下优点: 利用高性能计算集群, 实现了对大规模 DEM 数据构建 DEM 金字塔切片的并行化, 可减少 DEM 金字塔构建的时间, 提高计算效率; 对切片的边缘高程值进行 8 邻域平滑处理, 以避免 DEM 可视化和分析计算的平滑。

附图说明

图 1 为本发明的基于分布式系统的海量 DEM 金字塔切片并行构建方法的工作流程图;

图 2 为本发明的基于分布式系统的海量 DEM 金字塔切片并行构建方法的局部坐标系切分 DEM 示意图;

图 3 为本发明的基于分布式系统的海量 DEM 金字塔切片并行构建方法的单个计算节点计算示意图;

图 4 为本发明的基于分布式系统的海量 DEM 金字塔切片并行构建方法的多个计算节点合并示意图;

图 5 为本发明的基于分布式系统的海量 DEM 金字塔切片并行构建方法边缘平滑示意图;

图 6 为本发明的基于分布式系统的海量 DEM 金字塔切片并行构建方法可视化图;

其中: 图 (a) 为 17 级图; 图 (b) 为 17、16 级镶嵌显示图; 图 (c) 为 17 级的 Wireframe 图; 图 (d) 为 17、16 级镶嵌 Wireframe 图。

具体实施方式

下面将结合本发明的实施例图中的附图, 对本发明实施例中的技术方案进行清楚、完整的描述。

实施例: 如图 1~6 所示, 该基于分布式系统的海量 DEM 金字塔切片并行构建方法, 具体包括以下步骤:

其中 Tile 表示切片, Key 表示切片的索引值, Value 表示高程值, List 表示链表, cell 表示网格单元, zoom 表示层次, Map 表示映射, Extend 表示源 DEM 的外包矩形, <Key, Tile>表示键值对;

S1 使用分辨率 30 米的源 DEM 数据和一部分高精度源 DEM 数据, 数据格式是 tif, 设定的坐标系为 WGS84; 分别研发基于 Flink 分布式平台的金字塔切片程序;

获取全国 30 米源 DEM 的元数据 Metadata, 包括原始的分辨率、坐标系 WGS84 以及 Extend;

S2 对所述输入源 DEM 数据建立局部坐标系, 设定初始划分切片的高和宽均为 126 (网格), 利用滑动窗口方法对源数据进行切分, 接着根据 Tile 在局部坐标系的行列号计算 Key, 构建一系列的<Key, Tile>键值对, 按照 Flink 的算法将键值对数据分布到不同的计算节点上进行计算; 建立以源 DEM 左上角为原点, 向下向右为正方向的坐标系; 其中索引值 Key 的计算方法:

$$\text{Key} = C + "_" + R, C = (x - XMin) / TW, R = (y - YMin) / TH$$

其中, C、R 表示切片 Tile 的行列号, x、y 表示切片 Tile 任意一点的切片坐标, XMin 表示源 DEM 的外包矩形 Extent 的 X 最小值, YMin 表示源 DEM 的外包矩形 Extent 的 Y 最小值, TH 是初始切片 Tile 的高, TW 是初始切片 Tile 的宽;

S3 根据是否定义了最大层次 MaxZoom, 进行设置或计算最大层次 MaxZoom, 再根据设定的坐标系和切分规则计算输出切片 Tile 的参数; 所述输出切片 Tile 的参数包括切片的长、宽、分辨率等; 若未设定最大层次 MaxZoom, 则需要计算最大层次 MaxZoom, 计算过程是根据源 DEM 的外包矩形 Extend 的左下点坐标和右上点的坐标, 并将左下点坐标和右上点的坐标进行投影转换, 转换为经纬度坐标, 其中左下点坐标是 (SLon, SLat), 右上点的坐标是 (ELon, ELat); 接着根据公式计算最大层次 MaxZoom, 计算公式如下:

$$\begin{aligned}
dLat &= ((ELat - SLat) / 180.0) \cdot \pi \\
dLon &= ((ELon - SLon) / 180.0) \cdot \pi \\
lat1 &= (SLat / 180.0) \cdot \pi \\
lat2 &= (ELat / 180.0) \cdot \pi \\
a &= (\sin(dLat / 2) \cdot \sin(dLat / 2)) + (\sin(dLon / 2) \cdot \sin(dLon / 2) \cdot \cos lat1 \cdot \cos lat2) \\
dist &= R \cdot 2 \cdot \arctan\left(\sqrt[3]{a} / \sqrt[3]{1-a}\right) \\
MaxZoom &= (\log(2\pi R / dist - tileSize) / \log(2))
\end{aligned}$$

其中, MaxZoom 为最大层次, dLat 表示纬度差, dLon 表示经度差, lat1 表示 SLat 的弧度, lat2 表示 ELat 的弧度, a 表示计算的中间值, dist 表示两点的距离, tileSize 表示 Tile 设置的最大边长, R 表示 WGS84 椭球的长半轴, $R=637817$;

计算输出切片 Tile 的 Extend, 需要适当的扩大, 目的是为了避免输出切片 Tile 拼接不会出现裂缝等问题; 所述步骤 S3 中设置输出切片 Tile 的参数为: 宽高均为 65 (网格)、坐标系为 WGS84、索引值 Key 的命名遵循 TMS 标准;

S4 计算节点并行构建所分配的键值对 <Key, Tile>, 根据索引值 Key 合并同一计算节点上的高程值 Value; 分别根据索引值 Key 按照所述输出切片 Tile 的长、宽、分辨率参数计算所对应的切片 Tile 的外包矩形 Extend, 计算输入切片 Tile 中每个网格单元 cell 的坐标, 对初步切分的 Tile 分别计算; 首先根据索引值 Key 计算 Tile 的 Extend; 计算输入切片 Tile 中每个网格单元 cell 的坐标, 其中 cell=<x,y,Value>, 其中 x, y 是在 Tile 的坐标, Value 是高程值的坐标, 根据设定的输出坐标系 WGS84, 计算 cell 在输出坐标系的坐标, 根据计算结果计算此 cell 所在的输出切片 Tile 的索引值 Key, 并将其加入 Map 中, 组成一个 Map<Key, List<cell>>; 将计算节点上相同索引值 Key 的 List<cell>合并, 原理见图 3 所示; 所述步骤 S4 并行计算输入切片 Tile, 根据输出切片 Tile 的参数将输入切片 Tile 的 DEM 值划分到不同的输出切片 Tile 中; 其中首先根据输入切片 Tile 的索引值 Key 计算 Tile 的外包矩形 Extend, 计算公式:

$$\begin{aligned}
MinX &= ExMinX + col \cdot W \\
MinY &= ExMinY + row \cdot H \\
MaxX &= ExMinX + (col + 1) \cdot W \\
MaxY &= ExMinY + (row + 1) \cdot H
\end{aligned}$$

其中, MinX、MinY、MaxX、MaxY 分别表示外包矩形 Extend 的最小最大 XY 坐标值, col、row 是输入切片 Tile 的行列号, ExMinX、ExMinY 表示源 DEM 的外包矩形 Extend 的最小的 X、Y 坐标值, H 表示切片 Tile 的高, W 表示切片 Tile 的宽;

S5 根据索引值 Key 合并所有计算节点上的 List<cell>; 由于输入切片 Tile 分布在不同的计算节点上, 而输出切片 Tile 可能包含不同的输入切片 Tile, 也就是说, 输出切片 Tile 的数据可能在不同的计算节点, 需要将这些数据拉到同一个节点上进行合并, 也就是 Flink 的 Reduce 操作, 原理见图 4 所示。

S6 根据输出切片 Tile 的参数和坐标系, 计算输出切片 Tile 的网格单元 cell 的坐标, 使用设置的采样方法, 从合并的<Key, List<cell>>中计算 cell 的高程值, 本实施例中使用的是最邻近采样算法;

S7 利用 8 邻域算法, 修正切片 Tile 边缘的高程值 Value; 根据生成的所述链表 List<Key, Tile>, 循环获取每个切片 Tile 的四边网格单元 cell 和与之相邻的 8 个切片 Tile, 取相邻切片 Tile 的网格单元高程值的平均值作为相应网格单元 cell 的高程值 Value, 达到切片 Tile 拼接的平滑; 根据输出的键值对<Key, Tile>, 根据索引值 Key 的 TMS 命名规则计算每个切片 Tile 相邻的 8 个切片 Tile, 取相邻 Tile 的 cell 值, 并作平均值作为相应 cell 的值, 达到 Tile 拼接的平滑; 原理示意见图 5 所示;

S8 判断当前层次 Zoom 是否满足终止条件, 若满足终止条件, 则结束; 若不满足终止条件则根据上一级层次计算下一级层次和下一级层次的切片 Tile 的外包矩形 Extend 的宽度和高度, 计算方法: $NZ = CZ - 1$, $W = EW/2^{(NZ+1)}$, $H = EH/2^{(NZ+1)}$, 其中 NZ 表示下一级层次, CZ 表示当前层次, EW、EH 表示原始 DEM 的宽度和高度, W、H 表示切片 Tile 的宽度和高度;

S9 根据上一级层次切片 Tile 和采样算法, 并行计算生成下一级层次切片 Tile 和索引值 Key; 根据索引值 Key、下一级的层次和切片 Tile 的参数计算当前层次切片 Tile 所对应的下一级层次的索引值 Key, 利用所述采样算法从当前层次切片 Tile 采样下一级层次切片的高程值, 构成下一级层次切片 Tile 的子切片和索引值 Key, 在每个节点上合并相同的索引值 Key, 构成键值对<Key, List<Tile>>; 接着根据索引值 Key, 合并所有节点上相同的索引值 Key 的切片

Tile，组成下一级层次的切片 Tile，再执行步骤 S7；本实施例中使用的是最邻近采样算法；

S10 循环执行步骤 S8~S9，计算下一级的 List<Key, Tile>，直到判断当前层次 Zoom 满足终止条件，所述终止条件为：若设定了最小层次 MinZoom，则判断当前层次 Zoom 是否达到设定最小层次 MinZoom，若达到，则终止；若未设定最小层次 MinZoom，则当前层次 Zoom 小于 0，则终止。

以上所述仅为本发明的较佳实施例而已，并不用于限制本发明，凡在本发明的精神和原则之内，所做的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。