

直接表达区域的四叉树链式编码

邓朝晖 贾 华

(武汉测绘科技大学遥感与信息工程学院, 武汉市珞喻路39号, 430070)

摘 要 提出一种用于直接表达区域的数据结构——四叉树链式编码。与常规四叉树结构比较, 该方法具有压缩率高、表达直接等优点。本文还用 C 语言给出具体表达式和基本算法、效率分析。

关键词 区域表达; 四叉树; 链式编码

分类号 TP311.12

1 四叉树链式编码原理

四叉树是把一幅图像或栅格地图递归地等分为4个子区, 即四分区。如果一个四分区的的所有格网具有相同的值, 则这个子区就不再往下分割, 称为黑结点; 如果一个四分区的的所有格网值完全和目标不同, 则称为白结点, 否则称为灰结点。对灰结点要继续划分, 直到四分区为黑结点或白结点。所有黑结点的集合构成目标。

常规四叉树用指针来联系结点, 每个结点至少需6个量, 即父结点指针(前趋)、4个子结点指针(后继)和本结点的灰度或属性值表达。经线性化后, 随着分解层次的增加, 记录的黑结点代码长度也依次递增, 其代码长度 CL(Code Length)可表示为分解最深的层次码, 层次码实际上反映了结点的大小。

在每一个四分区, 其细分格网编码具有继承性, 即定位码还具有可压缩性。本文把线性四叉树定位码 LC 中的每一位用一个结点表示, 此结点仅由两个区段构成, 即一位四进制数字表示的数字区和指向下一层次数字的指针。用此种结点结构, 可建立表达四叉树的链式编码。

图1a 为结点的结构, 图1b 为链式编码结构图。从图1可看出四叉树链式编码的特点:

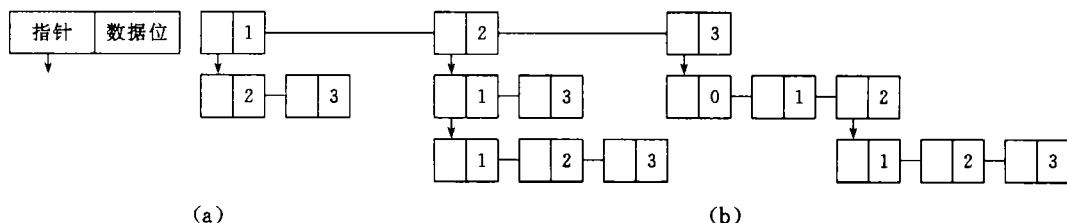


图1 链式编码结构图

- 1) 只表示四叉树的黑结点;
- 2) 同一分支各结点, 其定位码的前缀相同, 这些前缀数字在链式编码结构中是公有的, 可使四叉树总代码得到压缩;
- 3) 结点结构中指针为空者对应于一个黑结点;
- 4) 搜索过程在层与层之间时, 用指针传递; 在同层结点之间, 以线性方式进行;
- 5) 由图1b 中第1层出发到达某个指针为空结点所走过的各层数字形成的序列, 即为相应

线性二叉树结点的定位码。

2 二叉树链式编码结构的建立及基本算法

2.1 链式编码结构的建立

在数据索引和图幅分幅、数据压缩及 GIS 数据结构方面,二叉树结构都得到应用。二叉树应用效率的高低,关键在于表达形式的选取。本文对二叉树链式编码结构用 C 语言进行定义及表达。首先定义结点结构如下:

```
typedef struct quad __struct {
    char digi;
    void * data;
} quad;
```

在上述定义中,digi 为数据变量,data 为指向大儿子结点的指针。上下层结点之间的调用利用 C 语言的宏定义功能给出:

```
#define rquad1(v) (quad *) (quad.data)
```

利用循环嵌套定义,可得出任意层之间的关系。如两层之间关系:

```
#define rquad2(v) (i,j) (rquad1)(rquad1(v)[i])[j])
```

三层之间关系:

```
#define rquad3(v) (i,j,k) (rquad1(rquad1(rquad1(v)[i])[j])[k])
```

上述式中 i, j, k 分别为 1、2、3 层中兄弟结点位置。

从上述式子可看出,任一黑结点均可用一个表达式表达,等同于用二维数组表达区域中某个像素,表明动态通用串结构组织的链式编码可直接表达区域像素。而双指针表达的二叉树方法,像素的存取和表达是通过间接方式进行的,无法直接表达区域。

2.2 基本算法

2.2.1 删除算法

- (1) 首先判断该串 v 是否为空,为空则退出;
- (2) 如果需删除结点不在该串之内,退出, t 表示该串起始地址;
- (3) 删除结点在该串之内,确定删除结点位置 $k = p + n$, p 为起始位置, n 为结点个数;
- (4) 利用块移动函数,将位置 k 处 n 个数据移动到位置 p 处。

以下为 C 语言算法:

```
int vstrdel (quad __struct * v, int p, int n)
{
    int k, l;
    char * t;
    l = sizeof (struct quad __struct);
    if (vstr __inited(v) {
        if (p >= 0 && n > 0) {
            t = (char *) v->data;
            memmove (t + p * l, t + k * l, n * l);
        }
    }
```

```
    }  
    return 0;  
}
```

2.2.2 插入算法

- (1) 首先判断该串 v 是否为空, 为空则退出;
- (2) 确定从位置 p 处开始插入 n 个结点, t 表示该串起始地址;
- (3) 将从位置 p 处起 m 个结点往后挪 n 个位置;
- (4) 将从串 s 起始地址起的 n 个结点移入起始地址为 t 并往后偏移 p 个位置处。

以下为 C 语言算法:

```
int vstrins (quad __struct * v ,int p ,void * s ,int n )  
{  
    char * t ;  
    int l ,m ;  
    l =sizeof (struct quad __struct);  
    if(vstr __inited( v ) && p >=0) {  
        t =(char *) v ->data;  
        m =length( v ) - p ;  
        memmove( t + ( p + n ) * l ,t + p * l ,m * l );  
        memmove( t + p * l ,s ,n * l );  
        return l ;  
    }  
    else return 0;  
}
```

3 四叉树链式编码效率分析

3.1 空间效率分析

文献[2] 给出了四叉树空间效率分析的方法与结果。取一个 $2^m \times 2^m$ 黑色正方形区域, 此区域位于 $2^n \times 2^n$ 画面上($n \geq m$)。此正方形区域位于不同位置会形成不同的四叉树, 图 2 为最坏情况。对于外设存储效率来说, 本文所述编码方法的效率同[2]

中线性数字搜索树编码效率。四叉树黑结点表示数为:

$$B = (2^{m+1} - 1) + \sum_{i=0}^{m-1} (2^{m-i+1} - 3) = 3(2^{m+1} - m) - 5$$

链式编码黑结点表示数为:

$$\begin{aligned} B' &= (2 \times 2^{m+1} + m) + [\sum_{i=0}^{m-1} 4(2^i - 1) + 4(2^{m-1} - 1) + 2] \\ &= 5 \times 2^{m+1} - 3m - 6 \end{aligned}$$

[2]中指出, 当图像分辨率较高时(如大于 256×256), 链式编码效率优于四叉树及线性四叉树, 分析过程见[2]。

对于指针及动态通用串表达来说, 比较四叉树及链式编码存

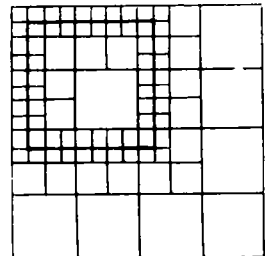


图2 正方区域图像的最坏情况

储方式可看出:二叉树表达每个结点至少需 3 个域——1 个数据域和 2 个指针域,而链式编码方法每个结点只需 2 个域——1 个数据域,1 个指针域。所以,假设共有 m 个结点,每个数据域占用 z 个数据位,则二叉树黑结点数据位数为 $C = 3mz$,而链式编码黑结点数据位数为 $C' = 2mz$,两者压缩比 $\alpha = 1.5$ 。

3.2 时间效率分析

如采用通常的双指针运算,假设有 m 个结点,则压缩与恢复时的时间效率为 $o(2m)$,因为涉及到双指针赋值与运算;而对于动态通用串方法,由于在同层只以线性方式追加,层与层之间只以单向指针联系。设层数为 L ,则时间复杂度为 $o(L + m)$,因通常情况下, L 远小于 m 。故一般来说,用本文方法时间效率为双指针二叉树方法的 1 倍。

参 考 文 献

- 1 邓朝晖,贾 华. 矢量数据的一种动态分层组织方法. 武汉测绘科技大学学报,1995,20(2):118~122
- 2 周洞汝. 姜海涛. 用于区域表达的线性数字搜索树编码. 计算机辅助设计与图形学学报,1992(3):1~7
- 3 龚健雅. GIS 矢量栅格一体化数据结构与面向目标数据模型的研究:[学位论文]. 武汉:武汉测绘科技大学,1993
- 4 Gahegan M. An Efficient Use of Quadtree in Geographical Information System. Geographics Information System,1989(3):303~307
- 5 Dyer C R. The Pace Efficiency of Quadrees CG &IP,1982(8):348~355
- 6 Shaffer C Q. A Geographic Information System Based on Quadrees. Geographics Information System, 1990(2):254~259

Lineal Quadtree Encoding for Direct Regional Representation

Deng Zhaohui Jia Hua

(School of Remote Sensing and Informatics, WTUSM, 39 Luoyu Road, Wuhan, China, 430070)

Abstract This paper presents a kind of data structure that can be used to directly represent region—lineal quadtree encoding. Compared with ordinary quadtree, it has some advantages including a high rate of compression and a direct representation. Furthermore, concrete represatative equations have been listed.

Key words regional representation; quadtree; lineal encoding