

文章编号: 1007-4619(2004)05-0465-06

实现栅格图形和图像数据矢量化提取的游程轮廓追踪法

谢顺平, 都金康, 王结臣

(南京大学 城市与资源学系, 江苏 南京 210093)

摘 要: 提出一种直接利用栅格图形游程编码进行图斑边界多边形矢量化提取的有效方法。对遥感图像专题信息和栅格型地理空间分析结果的矢量化提取是 GIS 的重要手段, 栅格数据结构可简化解决叠置分析等一类空间分析问题, 其中需要高效矢量化技术的支撑。本方法充分利用栅格游程编码所具有的可操作性强、检索和转换便捷、隐含横向差异信息、可处理高精度和大规格图形和图像等特征, 定义了“上邻游程”和“下邻游程”, 并采用相应有效查找算法和基于游程轮廓追踪的矢量化技术, 使得计算机处理栅格数据矢量化的规模、精度和效率明显提高, 可广泛应用于栅格型地理空间分析和遥感图像处理等方面。

关键词: 游程编码; 矢量化; 上邻游程; 下邻游程

中图分类号: P208/TP751.1 **文献标识码:** A

1 引 言

栅格数据模型结构是地理信息系统主要空间数据结构之一。大量的栅格结构图像可由遥感资料直接获得, 遥感数据的专题信息矢量化提取是 GIS 的重要内容。栅格数据结构非常易于解决叠置分析、缓冲区分析、衰减扩散分析等一类空间分析问题, 在解决地貌和水文问题的研究中也大量应用了栅格数据结构及其空间分析和模拟的方法, 这类分析结果通常需要矢量化转换。由于计算机性能和内存容量的提升以及各种栅格压缩格式的采用, 使得其可以处理的栅格图形规模和精度明显提高, 实现对不同尺度和复杂度栅格数据的快速、高效矢量化转换对栅格型地理空间分析具有重要作用。现有的栅格数据压缩编码, 如链码、各种形式的游程编码、块码、二叉树编码、线性四叉树编码^[1]等, 其中部分是面向存储的, 缺乏数据间的可操作性, 而行扫描游程编码具有形式简单、易于检索和可操作性强, 行压缩与解压方便, 非常适合于叠置分析等多种 GIS 空间分析和操作。因此, 直接利用游程编码进行栅格图斑边界的多边形矢量化具有较大的应用潜力。

传统的栅格矢量化技术和算法一般基于非压缩的

直接编码栅格、二值化处理和八邻域搜索等, 其缺点反映在对内存要求高、转化效率低、处理栅格的规模、精度和复杂度均受限制等。近年来一些研究者提出若干提高转换效率和处理规模的新方法, 拓展了解决此类问题的思路, 如双边界搜索法^[2]、散列线段聚合法^[3]、结点搜索法^[4]、无边界游程编码及其矢栅直接相互转换算法^[5]、基于线性四叉树的矢量化方法^[6]、“栅格技术”实现栅格数据矢量化的方法^[7]等。其中方法^[2]是基于栅格阵列的处理, 处理的栅格规模受到内存限制。方法^[3,4,7]对栅格处理基于外存, 附有大量中间数据的产生、存取和查找, 影响处理效率提高的幅度。方法^[5]需用额外的预处理弥补编码对栅格二维整体性的破坏, 占用了过多的处理时间。方法^[6]额外的标记存储开销和基于像元轮廓的追踪等影响算法效率的进一步改善。笔者结合多年的实践并借鉴已有的方法, 提出了一种利用行扫描游程长度编码(RLC)直接追踪图斑游程及其外围轮廓, 提取含有拓扑关系图斑边界的矢量化算法, 其运行效率、处理规模和实用性较以往有一定幅度提高。

2 追踪游程轮廓的矢量化方法

2.1 上邻游程和下邻游程的定义及其动态定位提取

为方便随机访问各行的游程编码, 本方法首先

收稿日期: 2003-03-04; 修订日期: 2003-04-21

基金项目: 国家自然科学基金资助项目(编号: 40170115)。

作者简介: 谢顺平(1957—), 男, 江苏南京人, 高级工程师, 南京大学计算机科学与技术系计算机及应用专业毕业, 主要从事地理信息系统、空间分析和模拟等方面的研究, 已发表论文20余篇、专著2本。

对调入内存的游程编码进行一次扫描处理,生成指向各行游程编码起始位置的索引表(图1),为便于标识游程的起始位置,本文采用的游程编码形式为{起始列,属性},一个游程的终止位置即为下一游程的起始位置,行上未游程的终止位置即为栅格图形的列数,显然,每行上的游程呈关键字为起始列的有序表形式。各游程编码的位置与所在行首游程起始位置之差即为该游程的行上序数。

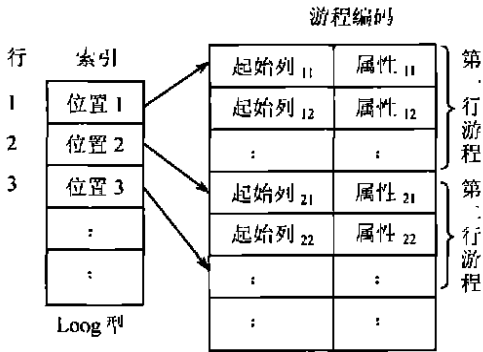


图 1 游程表及索引

Fig 1 Run-length code and index

由于矢量化时对图斑边界的追踪总是在相邻行或列上的相邻游程之间进行,一行上相邻游程可通过当前游程的前驱或后继访问获得,而相邻列上相邻游程虽然可通过预先产生的纵向静态指针链接,但需占用额外存贮空间。本文针对矢量化时沿图斑周边游程轮廓追踪的特点,对上邻游程和下邻游程进行定义,并采用对它们的实时动态查找方法。所谓上邻游程是指相对于某行的某个游程来说上一行游程中其头部位置小于本游程尾部位置而其尾部位置大于或等于本游程尾部位置的游程(图2(b)箭头指向的游程);所谓下邻游程是指相对于某行的某个游程来说下一行游程中其尾部位置大于本游程头部位置而其头部位置小于或等于本游程头部位置的游程(图2(b)箭头指向的游程)。为便于处理,在游程队列的首尾各设置一个属性为0的游程,代表图形上下边沿的外侧附加行,并作为首行游程的上邻游程和末行游程的下邻游程。

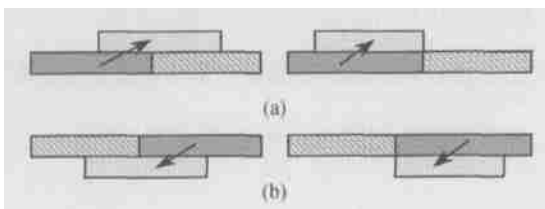


图 2 上邻游程和下邻游程

Fig 2 Up neighbor run-length and down neighbor run-length

由于地学图形随栅格行逐渐变化,相邻行上的游程数以及上下相邻游程的序数差异一般不是很大。笔者对十多幅不同栅格规模和图斑密度的栅格图形的游程编码进行计算统计,发现图形中相对于所有游程来说其上下相邻游程序数的平均位移量在1.3—1.8个游程之间,总的平均位移量约为1.5个游程。基于这一特点,本文采用一种“就近动态定向顺序查找”方法,即根据当前游程所在行及位置得到其在行上的序数,从相邻行的该序数游程开始查找,并从最初位置判别中确定是向前查找还是向后查找,从而根据定义条件就近迅速定位检索到上邻游程或下邻游程。

2.2 设置追踪方向和追踪标记

为标识追踪方向,设置方向数1、2、3、4,分别表示垂直向上、水平向右、垂直向下、水平向左。规定追踪弧段时其左侧为提取边界的图斑,右侧为其它图斑,这样对图斑外环边界的提取就自然地按逆时针方向进行,对图斑内环(洞)边界的提取则自然地按顺时针方向进行。

栅格图形中可能离散分布着多个同属性的图斑—为了标识已提取其边界的图斑,追踪时对垂直向上通过其右侧的游程属性取负,这样就确保了该图斑的边界在后续扫描中不会再次被选中作为入口,追踪时游程属性的提取因取其绝对值而不受影响。对于图斑中含有内环边界的情形,可用图3说明,图中实线箭头表示垂直向上按逆时针方向追踪通过图斑外环右侧边界时对左侧相邻游程标记,虚线箭头表示垂直向上按顺时针方向追踪通过图斑内环左侧边界时对左侧相邻游程标记,不难看出这种游程属性标记对复杂图斑的处理也是正确的。所有属性的图斑边界矢量化提取完成以后,栅格游程编码的属性恢复正值。

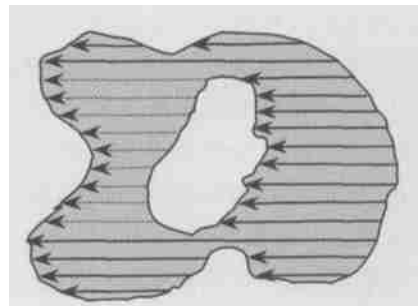


图 3 含有内环图斑的游程标记

Fig 3 Mark of run-length code

2.3 图斑弧段的提取方式

栅格图斑边界弧段的提取有两种形式,一种是边界完整闭合弧段的提取,完成一条弧段提取以回到起点为标志,另一种是无冗余的边界弧段提取,它要求在结点处划分弧段且共同边界弧段只提取存储一次,完成一条弧段提取以右码改变或回到起点为标志。为实现后一种形式的弧段提取,可根据弧段两侧属性值的大小关系决定是否存储该弧段,对其左侧属性(提取图斑的属性)大于右侧属性的弧段予以存储,否则不予存储。用提取标志 0, 1, 2 分别表示不提取、提取弧段、提取闭合弧段。因为弧段追踪开始时其提取标志就可确定,故对无需提取的弧段,只追踪路径不生成弧段坐标。矢量库由弧段坐标文件和索引文件组成,弧段存储时保存的拓扑和索引信息包括弧段左码、右码、头结点的位置、坐标点数等。

2.4 弧段提取算法

本方法完成一类图斑边界的提取需扫描一遍游程编码队列,寻找该类型各图斑边界的追踪入口,而弧段的追踪主要围绕边界处的相邻游程进行。为确保弧段完整性,对首条非闭合弧段作缓存处理,待追踪闭合时再和最后弧段拼合存储。下面是提取一类属性若干图斑边界弧段的详细算法步骤。

(1) 寻找入口

为适应任何图斑(含孤岛)边界入口定位,按由下往上的行序从当前位置扫描后续各行游程编码,若当前游程的属性和提取边界图斑的属性相同,其所在行该游程尾部下端即为弧段追踪入口,左侧游程指针指向该游程,记录起始位置和坐标并以右侧游程属性作为弧段初始和当前右码,若无右侧游程则取右码为 0,置初始和当前追踪方向数为 1,设置当前弧段的存储标志,转步骤 2 继续。若游程已扫描到尾部,则结束该属性图斑边界的提取。

(2) 推进和确定新的追踪方向

当前方向为 1 时:将左侧游程属性取负(标记),垂直向上推进一个游程高度,并判其上邻游程属性是否等于提取图斑属性,若同,再判两游程结束位置是否相同,若相同,新的追踪方向保持不变,当前左侧游程指针指向上邻游程,更新当前右码;若不相同,即上邻同属性游程结束位置较大,新的追踪方向转为水平向右(方向 2),右侧游程指针指向当前游程的下一游程,右码不变,新的左侧游程指针指向上邻游程。若上邻游程属性不等于当前图斑属性,则新的追踪方向转为水平向左(方向 4),新的右

侧游程指针指向上邻游程,更新当前右码,左侧游程指针不变(图 4)。转步骤 3 继续。

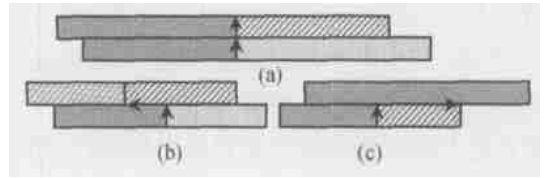


图 4 当前方向为 1 的下一追踪方向
Fig 4 The next racing direction of direction 1

当前方向为 2 时:若方向左侧游程结束位置小于等于右侧游程结束位置,水平向右推进到左侧游程结束位置,新的追踪方向转为垂直向上(方向 1),左侧游程指针不变,更新右码。否则,水平向右推进到右侧游程结束位置,右侧游程指针指向下一个游程,若新的右侧游程属性等于提取图斑属性,则下一追踪方向转为垂直向下(方向 3),左侧游程指针更新为指向当前右侧游程,右码保持不变,否则右码为新的右侧游程属性且追踪方向不变(图 5)。转步骤 3 继续。

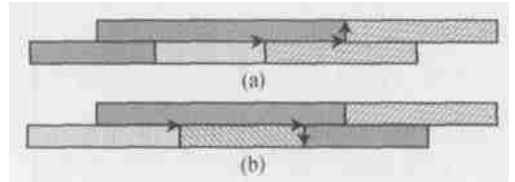


图 5 当前方向为 2 的下一追踪方向
Fig 5 The next racing direction of direction 2

当前方向为 3 时:垂直向下推进一个游程高度。若方向左侧游程的下邻游程属性等于提取图斑属性,则再判别两游程的起始位置,若相同,追踪方向不变,左侧游程指针指向下邻游程,更新当前右码;若不相同,即下邻游程起始位置在前,新的追踪方向转为水平向左(方向 4),右侧游程指针指向当前左侧游程的前一游程,右码保持不变,新的左侧游程指针指向其下邻游程。若下邻游程属性不等于当前图斑属性,则新的追踪方向转为水平向右(方向 2),右侧游程指针指向下邻游程,更新当前右码,左侧游程指针不变(图 6),转步骤 3。

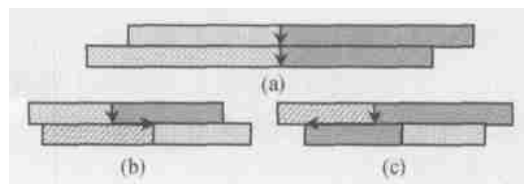


图 6 当前方向为 3 的下一追踪方向
Fig 6 The next racing direction of direction 3

当前方向为 4 时:若左侧游程起始位置大于等于右侧游程起始位置,水平向左推进到左侧游程起始位置,下一追踪方向转为垂直向下(方向 3),更新右码。否则,左侧游程起始位置在前,水平向左推进到右侧游程起始位置,右侧游程指针指向前一个游程,若新的右侧游程属性等于提取图斑属性,则下一追踪方向为垂直向上(方向 1),左侧游程指针更新为指向当前右侧游程,右码保持不变,否则右码为新的右侧游程属性且追踪方向不变(图 7),转步骤 3。

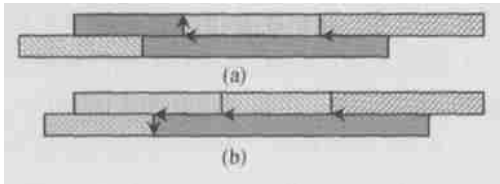


图 7 当前方向为 4 的下一追踪方向

Fig 7 The next racing direction of direction 4

(3) 计算坐标

当提取标志为 1,且弧段右码发生改变或追踪方向发生变化时,根据当前位置计算其结点或拐点坐标,转步骤 4 继续;当提取标志为 2,且追踪方向发生变化时,根据当前位置计算其拐点坐标,转步骤 5 继续;当提取标志为 0,且弧段右码发生改变,则一条弧段追踪结束,若此时已回到起点,将栈中独立弧段退出并存储,转步骤 1 继续下一图斑的提取,否则设置新弧段的起点坐标、更新初始右码和方向及提取标志。转步骤 2 继续。

(4) 存储弧段

若弧段右码未变且未回到起点,转步骤 2 继续,否则,对自闭合弧段将其连同拓扑和索引信息一起存储到矢量库中,一个图斑边界提取结束,转步骤 1 继续;对图斑的首条非自闭合弧段,将其坐标连同右码一起存入栈中待处理;对图斑的中间弧段,将其连同拓扑和索引信息一起存储到矢量库中;对图斑的末条弧段,若其右码与栈中保存的首条弧段右码一致,则将当前弧段与栈中退出的弧段合并后连同拓扑和索引信息一起存储到矢量库中,若两弧段右码不同,则将当前弧段和栈中退出弧段分别连同各自的拓扑和索引信息存储到矢量库中,一个图斑边界提取结束,转步骤 1 继续。设置新弧段的起点坐标、更新初始右码和方向及提取标志,转步骤 2 继续。

(5) 判闭合和存储

若追踪回到入口起点,则将该弧段连同索引信息一起存储到矢量库中,一个图斑完整边界提取结

束,转步骤 1 继续。否则,转步骤 2 继续搜索。

3 光滑处理

对矢量化提取的弧段在存储前应根据需要进行光滑处理,一般以保持图斑的主体轮廓特征和弧段所围面积基本不变为原则。为此本文采用一种自适应的线性迭代法(抹角法)^[8],为避免因光滑插入过多节点,线性迭代的次数控制在 3 次以内,对需抹去的角点而言,其两边上参与线性迭代的初始点并不是最近的两侧节点,而是取两边的中点和距离角点 n 倍像元边长的点中较近的点, n 取值越大,抹角范围和矢量变形也越大,光滑效果好,图形面积失真大, n 取值小,抹角范围小、图形保持原图斑主体轮廓越显著,但光滑效果差,所围面积失真小,需根据像元精度权衡折中。

4 试验结果

为验证本方法的正确性和效率,以便于对比,我们用实现本方法的 Visual C++ 程序对不同规模和图斑数的多幅土壤类型栅格图形分别在 P II 266 128M Windows98, P III750 128M Windows xp 和 P IV 1.8G 256M Windows xp3 种机型环境上进行提取图斑边界的试验,其性能统计参数由表 1 列出,由于试验用图较为复杂,为和传统方法对比,表中附有直接编码法的试验结果。从表中可以看出:当图形的图斑数(程序矢量化闭合图斑计数获得)不变时,本方法矢量化时间随栅格规模增加,但增幅远低于直接编码法,当图斑数急剧变大时,这种矢量化时间随栅格规模增加的速率在逐步降低,表现出对大规格、高精度复杂图形一定的适应性。对比本方法和[4][5][7]方法在 P II 机上的试验结果来看,本方法的效率较后者也有明显提高。图 8 和图 9 分别为用于试验 1 的南京江宁县土地利用类型栅格图和试验获取的矢量图。

5 结 论

由于本方法直接对经压缩获得的图形游程编码进行矢量化提取,无需生成、存取和处理中间结果,处理的数据量和算法的时间复杂度取决于图形游程编码的压缩比,压缩比愈高,算法的时间复杂度就愈低。由于地学图形的特点以及复杂图形常用高精度表示,采用游程编码的压缩比仍然较高,所以本方法较适合于大规模、高精度复杂图形的矢量化处理,可

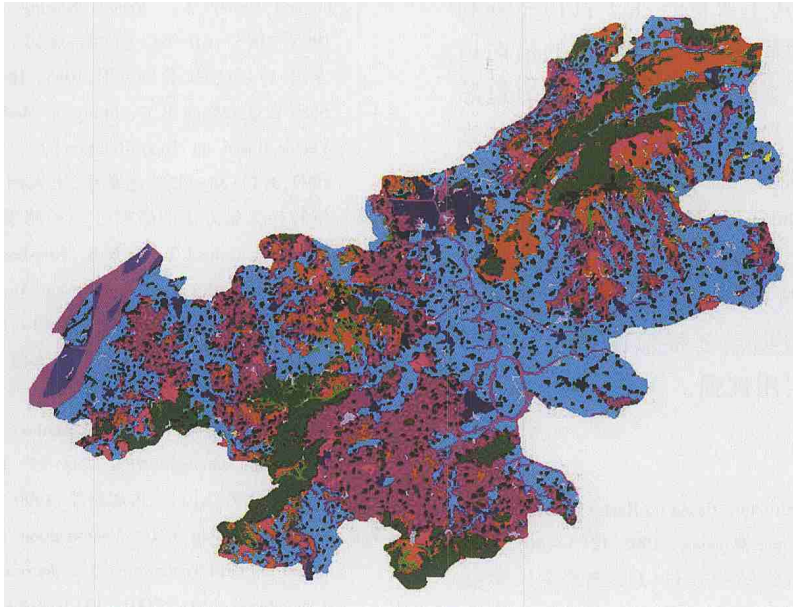


图 8 土地利用类型栅格图

Fig 8 Raster map for land use

表 1 不同规模和图斑数的栅格图形的矢量化处理时间统计

Table 1 Process time of vectorization for different rasters

试验 编号	图形 规格	类型 数量	图斑 数量	弧段 数量	直接编码法执行时间/s			游程轮廓追踪法执行时间/s		
					P II 时间	P III 时间	P IV 时间	P II 时间	P III 时间	P IV 时间
1	2764× 2798	19	4204	6991	46.320	3.870	1.750	0.550	0.260	0.120
2	5528× 5596	19	4204	6991	64.570	15.120	6.370	0.940	0.420	0.190
3	3048× 3474	23	12989	216354	22.840	6.710	2.640	1.920	0.850	0.400
4	6096× 6948	23	12989	21354	76.480	23.940	8.850	2.880	1.290	0.610
5	2476× 3512	22	51683	115762	25.920	9.260	3.680	4.350	2.240	1.120
6	4952× 7024	22	51683	115726	81.540	33.130	10.820	6.190	3.270	1.560

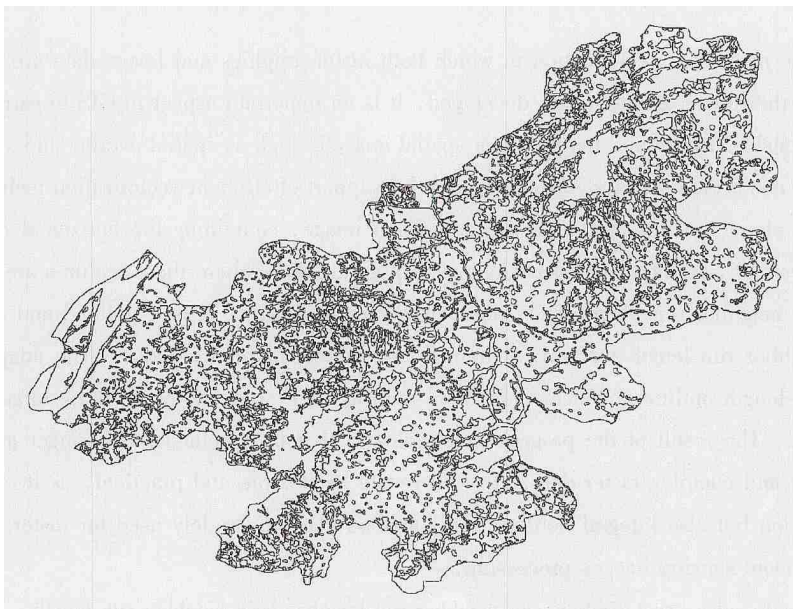


图 9 经矢量化获得的矢量图

Fig 9 Vector map after vectorization

获得较高的转换效率。其主要优点在于:(1)充分利用游程编码的特点,处理简捷,沿游程轮廓的横向追踪跨结点进行,使追踪平均步距较基于像元的追踪法明显提高;(2)采用适当形式的游程编码及其索引,上邻游程和下邻游程的定义及合理查找方法,提高了对游程的定位和纵向追踪速度;(3)追踪标记的设计既保持游程编码形式不变,又省去了用于标记的内存开销;(4)既可无冗余地生成含有拓扑关系的图斑边界弧段,还可以根据需要提取图斑边界的完整闭合弧段,灵活性和实用性强。

参 考 文 献(References)

- [1] Zhao F Discussion of Data Structure Based on Raster Data[J] . *Journal of Institute of Surveying and Mapping*, 1995, 12(3): 201—205. [赵夫来. 基于栅格的数据结构探讨[J] . 测绘学院学报, 1995, 12(3): 201—205.]
- [2] Wu L, Liu Y, Zhang J, et al, Geographical Information System—Principle, Method and application[M] . Beijing: Science Press, 2002, 150—155. [邬伦, 刘瑜, 张晶等. 地理信息系统——原理、方法和应用[M] . 北京: 科学出版社, 2002, 150—155.]
- [3] Huang B, Chen Y. New Approaches for Mutual Transferring of

Vector and Raster[J] . *Remote Sensing Technology and Application*, 1995, 10(3): 61—65. [黄波, 陈勇. 矢量、栅格相互转换的新方法[J] . 遥感技术与应用, 1995, 10(3): 61—65.]

- [4] Shen Z Q, Wang R C. Study on Method of trans-foming Grid To Vector Based on Topo Relation[J] . *Journal of Remote Sensing*, 1999, 3(1): 38—42. [沈掌泉, 王人潮. 基于拓扑关系原理的栅格转换矢量方法的研究[J] . 遥感学报, 1999, 3(1): 38—42.]
- [5] Wu H Y, Gong J Y, Li D R. Non-boundary Run-length, Encoding System for Faster and it's Relevant Algorithms[J] . *Acta Geodaetica et Cartographica Sinica*, 1998, 27(1): 63—68. [吴华意, 龚健雅, 李德仁. 无边界游程编码及其矢栅直接相互转换算法[J] . 测绘学报, 1998, 27(1): 63—68.]
- [6] Tang H, Sheng Y H. A New Method of Vectorization [J] . *Peijing Survey and Mapping*, 1999, (3): 13—15. [唐宏, 盛业华. 一种新的矢量化方法[J] . 北京测绘, 1999, (3): 13—15.]
- [7] Zhang X C, Pan Y H. Vectorization Technique for GIS Grid Data Based on“Grid Technique”[J] . *Journal of Computer-aided Design & Computer Graphics*, 2001, 13(10): 896—900. [章孝灿, 潘云鹤. GIS 中基于“栅格技术”的栅格数据矢量化技术[J] . 计算机辅助设计与图形学学报, 2001, 13(10): 896—900.]
- [8] Zhang W Z, Xie S P. Microcomputer Cartograph[M] . Beijing: High Education Press, 1990, 139—141. [张文忠, 谢顺平. 微机地理制图[M] . 北京: 高等教育出版社, 1990, 139—141.]

A Method for Tracing Run-length Outline to Implement Vectorization of Raster Graphics and Image Data

XIE Shun-ping, DU Jin-kang, WANG Jie-chen

(Department of Urban and resources, Nanjing University, Nanjing 210093, China)

Abstract: An effective method of vectorization in which both raster graphics and image data are converted into polygon features through tracing the run-length outline is developed. It is an important aspect in GIS to extract thematic vector information from remote sensing images and the resultant spatial analysis such as spatial overlay and spatial diffusion, which can be simplified in the raster data structures needs a powerful support of efficient vectorization technique. The run-length codes of raster have the characteristics of storing high precision image, containing the horizontal difference information, easy access, rapid conversion, operability and moderate compression ratio where these features are well shown in the paper. The concept of ‘up neighbor run-length’ and ‘down neighbor run-length’ is well defined and the efficient algorithms for seeking vertical neighbor run-length for vertical tracing are developed in this paper. The adaptation of vectorization technique based on run-length outline enhances computer’s ability and efficiency of handling large scale and high precision complex raster data. The result of the program test indicates that this method has stronger adaptability and weaker sensitivity to large scale and complex raster data. It is proved to be flexible and practical, as it can extract not only the arc links with topo-relation but also integral borders. This method could be widely used for raster-type geographical spatial analysis and also remote sensing images processing.

Key words: run-length codes; vectorization; up neighbor run-length; down neighbor run-length