

文章编号: 1007-4619(2006)03-0326-06

遥感分类栅格图的快速矢量化方法

陈仁喜¹, 赵忠明¹, 潘 晶²

(1. 中国科学院 遥感应用研究所, 北京 100101; 2. 天津城市建设学院管理工程系, 天津 300384)

摘 要: 遥感数据成为 GIS 愈来愈重要的一种数据源。栅格数据与矢量数据之间的快速转换是 RS 与 GIS 集成中的关键技术之一。本文通过分析遥感分类图栅格像元之间的关系, 快速自动地提取矢量数据点及其连接信息, 进而利用连接信息快速生成弧段, 由弧段自动组建多边形并建立完整的拓扑关系。实验表明, 该方法是一种由遥感分类图自动转换成 GIS 通用数据的高效方案。

关键词: 弧段; 多边形; 矢量化; 数据转换

中图分类号: P208/TP751.1 **文献标识码:** A

A Fast Method of Vectorization for RS Classified Raster Map

CHEN Ren-xi¹, ZHAO Zhong-ming¹, PAN Jing²

(1. Institute of Remote Sensing Applications Chinese Academy of Sciences Beijing 100101, China;

2. Department of Management Tianjin Institute of Urban Construction Tianjin 300384, China)

Abstract: Remotely sensed data is becoming more and more important in GIS application. Data conversion between raster and vector is one of the key technologies for the integration of RS and GIS. In this paper, we analyzed the relations between pixel and pixel in the classified raster map. Based on the relations, all vector points and the connection information between them can be extracted. All arcs are traced rapidly by using the connection information between points. Polygons and the topological information between polygons are built automatically from the arcs. Finally, the experiment demonstrates that the method is a highly efficient scheme to automatically convert classified raster map to general format GIS data.

Key words: arc; polygon; vectorization; data conversion

1 引 言

遥感数据是地理信息系统非常重要的数据源, 随着地理信息系统的发展, 对遥感信息的依赖愈来愈强, 遥感数据专题信息矢量化提取是 GIS 的重要内容^[1]。虽然目前的 GIS 系统能够同时处理矢量数据和栅格数据, 但大量的分析和应用还是基于矢量数据的, 某些栅格数据分析的结果有时也需要进行矢量化表达, 因而, 栅格数据与矢量数据之间的转

换是不可避免的。在近年矢栅一体化的技术研究中, 高效的矢量栅格转换算法也占有重要地位。

矢量和栅格之间的转换在 GIS 中属于比较经典的问题, 因此积累了比较多的方法。如边缘跟踪法、散列线段聚合法^[2]、有向边界法^[3]、基于栅格技术的矢量化方法^[4]、基于拓扑关系的转换方法^[5]、基于游程编码轮廓追踪的方法^[1,6]等。方法^[2,3]源于边缘跟踪思想, 不能避免边缘跟踪法的缺陷: 处理速度较慢, 难于形成拓扑关系, 不能处理大型栅格图像。方法^[4,5]基于外存处理, 产生大量中间数据, 存

收稿日期: 2005-02-28; 修订日期: 2005-10-23

基金项目: 国家“863”高技术研究发展计划基金(2002AA133030)资助。

作者简介: 陈仁喜(1976—), 男, 武汉大学地图制图学与地理信息工程专业毕业, 现为中国科学院遥感应用研究所博士研究生。主要研究方向为 GIS、遥感图像处理。E-mail: sunex@163.com。

取查找速度不能得到提高,影响整个算法的效率。作者借鉴已有的方法,结合自己的实践,开发了一种栅格图矢量的有效方法。该方法能够处理大型图像,完全基于内存处理,便捷快速,采用拓扑结构的数据组织方式,能够处理复杂的岛状多边形,自动完成多边形属性的获取,是一种栅格数据到标准 GIS 数据的全自动化转换方案。

2 分类图的矢量化

遥感分类图是一种栅格数据,整个图像被分成几种地物类型,同类地物聚集成图斑。对分类图的矢量化就是要提取图斑之间的边界线,形成矢量弧段,并由弧段生成具有拓扑关系的矢量多边形(一个图斑的完整边界),每个多边形的属性由多边形边界所包围的图斑的属性确定。

栅格数据的每个像元(Pixel)都是有一定的大小,它并不是数学意义上的点(如 Landsat TM 图像的一个像元的大小是 30m×30m),因此栅格数据矢量化后的矢量数据都是面状信息,没有点和线。

在分类图的矢量化中,有如下一些约定和定义:

- (1) 栅格数据矢量化时只有面信息;
- (2) 栅格数据中的每个像元都是有大小的,矢量的最小图斑是一个像元;
- (3) 弧段:图斑与图斑之间的连续的分界线;
- (4) 结点:弧段两端的端点;
- (5) 坐标点:弧段上的顶点;
- (6) 在结点处,至少在三个方向上存在边界,而对于坐标点只有两个方向存在边界(图 1);

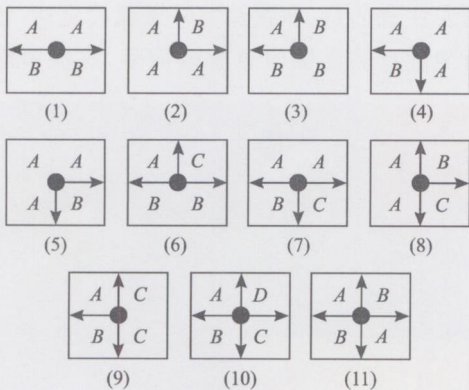


图 1 结点和坐标点

Fig 1 Nodes and vertex points

(7) 每个矢量多边形由一个外环和多个内环组成(或没有内环)。

矢量化既要考虑大图像的处理,又要顾及处理算法的速度。本文采取的方法是先从栅格图像中提取出所有可能的结点和坐标点,以及结点、坐标点的连接信息,然后完全在内存中对结点和坐标点进行处理,完成弧段的生成和多边形的组建及其拓扑关系的建立。

3 矢量化实现方法

3.1 结点、坐标点的提取

矢量化要提取的矢量点并不对应于图像中的像素点,而是四个像元之间的点。提取采用 2×2 大小的窗口进行,根据窗内四个像元的属性确定是否存在结点或坐标点,可以分为图 1 的 11 种情况,其中 1-5 属于坐标点,6-11 属于结点。其中第 1 种坐标点属于冗余坐标点,剔除该种类型坐标点不但不会影响数据精度,还能减少数据量,所以在提取结点和坐标点的过程中,忽略该种坐标点。

对于每个结点或坐标点,采用如下的结构体记录其信息。

```

struct PointNode
{
    int x;           //x坐标(列号)
    int y;           //y坐标(行号)
    int adj[4];
    int type;
}

```

其中, x, y 为该点的行列坐标; adj[4] 分别记录该点 0, 1, 2, 3 四个方向上(依次对应为右、上、左、下)的连接信息,这些连接信息在矢量化各个步骤中起着不同的重要作用:

- (1) 在提取结点、坐标点的过程中,记录其四个方向上是否存在边界线标志;
- (2) 在建立结点、坐标点连接信息的过程中记录该方向上与该点具有连接关系的结点或坐标点的序号,若值为负表示没有连接点;
- (3) 在弧段跟踪完毕之后,记录其四个方向上与其连接的弧段的标号。

另外,类型 type=2, 3, ..., 11, 对应于上面的 10 种结点和坐标点情况(第 1 种情况已忽略),表明该矢量点的类型。

根据图 1, 逐行分析栅格单元之间的情况,就能快捷地提取出图像中的结点和坐标点,并记录各点的坐标信息、坐标点类型及其边界标志信息。在此

过程中,无需将整个图像载入内存,而每次只需两行图像数据即可,处理完之后,下移一行,因而内存需求较小,可以适应大图像的处理。对于图像第 1 行、第 1 列及最后一行、最后一列,在图像的上下各增加一行,左右各增加一列单元值为零的栅格,避免特殊处理。

3.2 结点、坐标点连接信息的建立

从栅格图中提取出的结点和坐标点,都是孤立的点(图 2)。将孤立点连接成弧段的必要信息是结点或坐标点之间的连接信息。在上面提取结点和坐标点的过程中,提取出的各数据点记录了四个方向是否具有边界(有边界记为 true 否则记为 false),根据这些标志信息可以正确地找到该点具体和哪些点具有上下左右的连接关系。该过程通过对所有矢量数据点进行一次扫描完成。

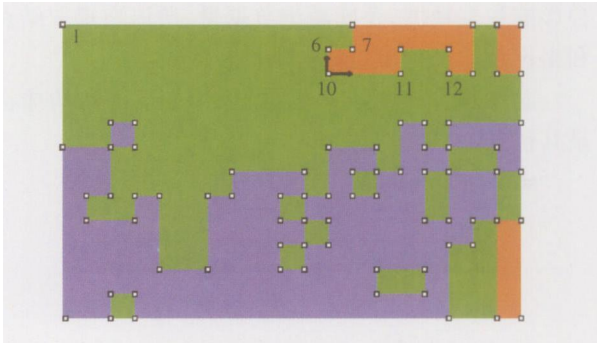


图 2 由栅格图提取的结点和坐标点

Fig 2 Nodes and vertex points extracted from the raster map

注意,在提取结点和坐标点的过程中,是按行从左到右依次处理的,所以提取出的数据点也具有相同的排列规则,且这些点都存储在一个一维数组中(图 2 中标出了某些点的序号)。这样便可以非常容易并快速地建立数据点之间的连接关系。考查第 n 号点(n 为该点在一维数组中的下标号),具有下列一些连接规则:

- (1)若在“0 方向”具有邻接点,该邻接点必为 $n+1$ 号点;
- (2)若在“2 方向”具有邻接点,该邻接点必为 $n-1$ 号点;
- (3)若在“1 方向”具有邻接点,该邻接点必为在 n 号点前面、与 n 号点具有相同 x 坐标且最靠近 n 号的坐标点;
- (4)若在“3 方向”具有邻接点,该邻接点必为在 n 号点后面、与 n 号点具有相同 x 坐标且最靠近 n 号的坐标点。

若 n 号点某个方向上没有邻接点,则将相应的 $adj[k]$ ($k=0, 1, 2, 3$) 赋一个负值,如 $adj[3] = -1$,表示 3 方向上没有邻接点。

根据此规则,可以快速找到每个点的四个方向邻接点。例如,图 2 中 10 号点的右邻接点为 11 号点、上邻接点为 6 号点。通过该步骤处理,所有数据点建立了上、下、左、右四个方向上与其他数据点的连接信息,以便用于后面弧段的跟踪和生成(图 2)。

实际上该步骤可以融入到矢量点提取过程中,和前一步骤同时完成。

3.3 弧段的生成

一般存在两种类型的弧段:开放弧段和封闭弧段。必须按照先跟踪开放弧段后跟踪封闭弧段的顺序来进行处理。

3.3.1 跟踪开放弧段

开放弧段的跟踪从“结点”开始。从全图中找出一个未跟踪过的结点,先从该结点的 0 方向开始,若 $adj[0]$ 的值为负数,表示该方向没有弧段,则转向下一个方向 1;若 $adj[0]$ 的值不为负数,则表示 0 方向上有新的弧段, $adj[0]$ 中的数值就是新弧段下一个坐标点的序号,根据跟踪到的该坐标点的连接信息进而可以跟踪得到另一个新的坐标点,如此下去,直到跟踪到某个结点为止,说明该弧段跟踪完毕。按照同样的方法,对其他三个方向进行处理,则该结点处的弧段跟踪完成。如图 3,从点 P_1 开始,在方向 0 上,跟踪得到弧 $P_1 \rightarrow P_2$;在方向 1 上,跟踪得到弧 $P_1 \rightarrow V_1 \rightarrow V_2 \rightarrow \dots \rightarrow V_9 \rightarrow P_2$;在方向 2 上跟踪得到弧 $P_1 \rightarrow P_3$ (图 3)。

需要注意的是,在跟踪完一条弧段之后,该弧段上所有坐标点的连接信息应该清除,这些清除了连

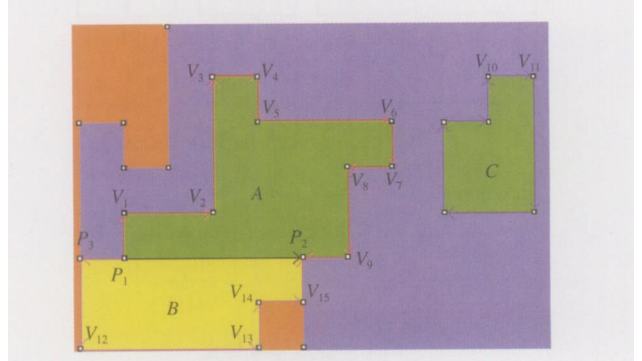


图 3 弧段的跟踪与多边形组建

P_1, P_2, P_3 为结点, V_1, V_2, \dots, V_{15} 为坐标点

Fig 3 Tracing arcs and building polygons

P_1, P_2, P_3 are nodes and V_1, V_2, \dots, V_{15} are vertex points

接信息的坐标点成为无效点,表示不再有弧段与它们有关。而对于弧段两端的结点,当前搜索方向上的连接信息也应该清除,防止后面对同一条弧段进行重复跟踪。当结点四个方向上的弧段都跟踪完毕之后,该结点也将成为无效点。

3.3.2 跟踪封闭弧段

将所有开放弧段跟踪完之后,全图中剩下的有效点都是坐标点,封闭弧段的跟踪可以从任意一个坐标点开始,最后必定回到起点,完成该条封闭弧段的跟踪。同样对跟踪完的坐标点的连接信息要进行清除,以防重复跟踪。如图 3 中,由坐标点 V_{10} 开始跟踪,得到封闭弧 $V_{10} \rightarrow V_{11} \rightarrow \dots \rightarrow V_{10}$ 。

3.3.3 弧段左右面域属性的获取

弧段跟踪完成之后,还应该获得每条弧段两侧面域的属性,其属性值需要从原始分类图中获得。取弧段起点和第二个顶点之间的连线为边界线,取得边界线左侧紧邻的像元值,即为左侧多边形属性;同理,取边界线右侧紧邻的像元为弧段右侧多边形属性。在这里记录每条弧段左右面域的属性,主要用于后面组建多边形时确定多边形的属性。

3.4 多边形的自动组建及其拓扑关系的建立

由弧段自动组建多边形是 GIS 中的经典算法。本文先建立弧段和结点的连接关系,然后采用“弧段双向搜索法”组建多边形。

3.4.1 结点、弧段连接信息建立

在开始进行多边形组建之前,先建立结点和弧段之间的连接关系,即显式的记录每个结点和哪些弧段相连接。一个结点最多和 4 条弧段具有连接关系,可用每个结点的 $adj[4]$ 分别记录四个方向与之相连接的弧段号,若在某个方向 $i(i=0, 1, 2, 3)$ 上没有连接弧段,则将相应的 $adj[i]$ 赋值为负数 -1 。

3.4.2 弧段双向搜索法组建多边形

弧段为两个多边形之间的公用边界,所以在建立多边形过程中每条弧段只能使用两次,反之,由每条弧段,按不同的方向跟踪,可以得到两个不同的多边形。这里从一条弧段开始,采用“双向搜索法”来组建多边形,即先正向、后逆向的搜索步骤(图 3)。

如图 3,假设以弧 $P_1 \rightarrow P_2$ 作为起始弧,先进行正向搜索,点 P_1 作为多边形第一个顶点,点 P_2 作为第二个顶点,在顶点 P_2 处采用右旋最小角法搜索得到下一条弧段 $V_9 \rightarrow P_2$,接着在顶点 V_9 处采用右旋最小角法搜索得弧段 $V_8 \rightarrow V_9, \dots$,最后回到第一个顶点 P_1 ,搜索一圈完毕,得到多边形 A。正向搜索完

毕后,然后进行逆向搜索,将点 P_2 作为新多边形第一个顶点,点 P_1 作为第二个顶点,在顶点 P_1 处采用右旋最小角法搜索得到下一条弧段 $P_1 \rightarrow P_3$,而后在顶点 P_3 处采用右旋最小角法搜索得弧段 $P_3 \rightarrow V_{12}, \dots$,最后回到第一个顶点 P_2 ,搜索完毕,得到多边形 B。这样,对每条弧段进行两个方向的搜索便能够得到公用该弧段的两个多边形。

在这里,弧段具有正向和逆向两个方向,对于每条弧段,采用两个标志(正向标志和逆向标志),在组建多边形的过程中,每生成一个新的多边形后,对于组成该多边形的弧段,更新正向标志或逆向标志,具体为:若弧的走向和多边形走向一致,则弧的正向标志置为 true;若弧的走向和多边形走向相反,则弧的逆向标志置为 true。

当一条弧的正向标志和逆向标志都为 true 时,不再对该弧段进行处理。这种设置双向标志的作用在于:(1)保证每条弧段只使用两次;(2)不会产生重复多边形,避免判断重复多边形的步骤,加快了处理速度。

注意,本文采用右旋最小角法搜索弧段时,无需计算每条弧段和当前弧段的夹角,因为每个结点处的弧段具有垂直关系,可以直接由当前弧段方向得到最小右旋方向上的弧段(图 3)。

3.4.3 多边形拓扑关系建立

考虑到可能出现岛状多边形的情况,还需要建立多边形之间的拓扑关系,形成复杂多边形。由前述方法得到的多边形自然地分成顺时针多边形(多边形边界为顺时针走向)和逆时针多边形(多边形边界为逆时针走向)两种。其中,只有顺时针多边形才可能成为岛状多边形的内环,而逆时针方向多边形只能是多边形的外环,因此,只需查找每个逆时针多边形包含哪些顺时针多边形即可。

由于采用了双向搜索法,在内岛多边形处,自然生成两个重叠的反向多边形环(图 4), P_2 和 P_3 实际上是重合的,为了显示需要,将它们离开一点空隙。 P_2 和 P_3 形成复杂的岛状多边形, P_1 形成另外一个内部多边形。

3.5 矢量数据的简化压缩

由前述方法得到的矢量数据,边界精度很高,完全和原始图像边界重合,这样使得矢量数据点很多,数据量很大,有必要对矢量数据进行一定的简化压缩。

本文采用道格拉斯-普克(Douglas-Peucker)算

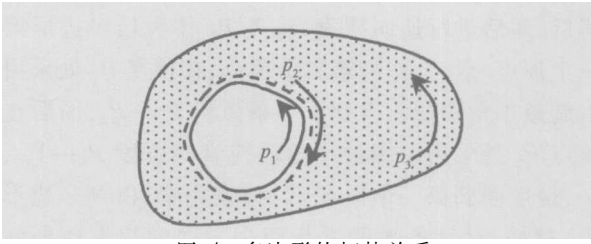


图 4 多边形的拓扑关系

Fig 4 Topo-relation between polygons

法(图 5),该算法按一个线目标为单元对数据进行压缩,是一个从整体到局部,即由粗到细的方法来确定曲线压缩后保留点的过程,能够保证各级特征点的保留。其优点是具有平移、旋转的不变性,给定曲线与限差后,抽样结果一致。

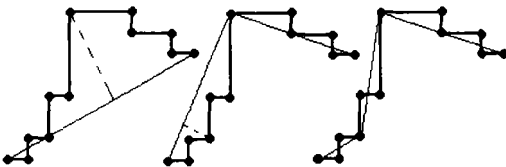


图 5 道格拉斯-普克压缩

Fig 5 Douglas-Peucker compression

对每一条弧段,首末两点虚连一条直线,求弧段上所有点到直线的距离,并找到最大距离值 d_{max} ,用 d_{max} 与限差 D 相比:

(1)若 $d_{max} < D$,则这条弧段上的中间点全部舍去;

(2)若 $d_{max} \geq D$,保留 d_{max} 对应的坐标点,并以该点为界,把曲线分为两部分,对这两部分重复使用这种方法。

通过限差 D 可以控制简化程度, D 的大小视具体需要而定。

本文采用拓扑关系数据组织方式,多边形公共边界只存储一次,对所有弧段目标进行简化处理后,不会出现多边形边界裂缝问题。

3.6 矢量数据的存储

本文在矢量化过程中,采用的是具有拓扑关系的数据组织方式,其优点在于:(1)公用边界只存储一次,大大减少数据量;(2)能够处理复杂的岛状多边形。

在具体操作时,弧段记录的是每个坐标点的序号,多边形边界记录的是组成该多边形的弧段的序号,这种结构完全和 E00 的数据格式相同,所以可以将矢量结果数据存为 E00 文件。另外,这种拓扑

结构的数据结构也很容易转换成以实体为单元的数据组织方式,如 ArcView 的 shape 文件。

4 试验与结论

为了验证本文算法的有效性,在 Visual C++ 环境下开发了相应的程序,进行了两组试验,第一组是针对不同复杂程度栅格图的试验,第二组是对同一幅图像分别放大到原图整倍情况下的试验,结果如表 1 和表 2。从试验结果来看,该算法的处理速度与图像的大小几乎没有什么关系,主要与图像的复杂程度有关,如结点和坐标点的数量、弧段的数量、多边形的个数。

表 1 不同栅格图的试验

Table 1 Experiment for different raster map

| 编号 | 图幅大小 | 矢量点数 | 弧段数 | 多边形数 | 处理时间 /s |
|----|-----------|-------|-------|-------|---------|
| 1 | 510×438 | 5918 | 471 | 332 | 0 |
| 2 | 1022×876 | 11319 | 1922 | 1279 | 1 |
| 3 | 2566×2187 | 44425 | 4991 | 3899 | 5 |
| 4 | 4203×3718 | 53456 | 24105 | 10480 | 12 |

表 2 同栅格图在不同分辨率下的试验

Table 2 Experiment for the same raster map at different resolution

| 图幅大小 | 1倍原图 | 2倍原图 | 4倍原图 | 8倍原图 |
|---------|------|------|------|------|
| 处理时间 /s | 12 | 13 | 14 | 16 |

本文方法相比于文献 [2, 3],不会严重受到图像大小的影响,且能形成拓扑关系,可以处理大型图像;与文献 [4, 5]相比,完全基于内存处理,速度得到提高。总结本文算法,具有如下一些特点:

(1)采取先分行提取全图结点、坐标点,而后进行处理的措施,使得算法能够处理大型栅格数据。

(2)在弧段跟踪、多边形建立的过程中,先显示地建立结点、坐标点、弧段之间的连接信息关系,避免搜索,加快了处理速度。

(3)完全采用拓扑关系数据结构,能够很好地处理复杂多边形,自动生成通用 GIS 数据文件。

本文的矢量化算法已开发成成熟的软件模块,经检验,能够完成大型遥感分类栅格图的快速矢量化(图 6 和图 7),具有较高的实用价值。

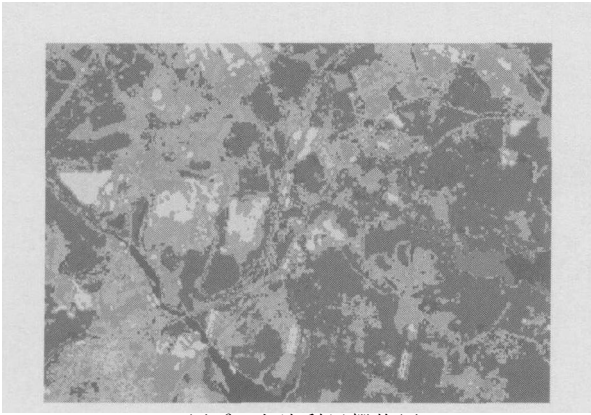


图 6 土地利用栅格图

Fig 6 Raster map for land use

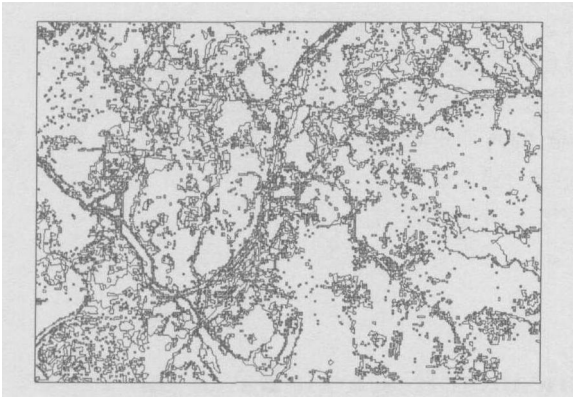


图 7 矢量图

Fig 7 Vector map

参考文献 (References)

[1] Xie S P, Du J K. Approach of Vectorization for GIS Raster Data

Based on Run-length Encoding System [J]. *Acta Geodaetica et Cartographica Sinica*, 2004, **33**(4): 323- 327. [谢顺平, 都金康等. 基于游程编码的 GIS 栅格数据矢量化方法 [J]. 测绘学报, 2004, **33**(4): 323- 327.]

[2] 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.]

[3] Li Z C, Liu C Y. An Efficient Directed Boundary Method for Vectorization of Dot Matrix Image [J]. *Computer Applications and Software*, 1997, **14**(3): 48- 51. [李占才, 刘春燕. 点阵图形矢量化的高效方法——有向边界法 [J]. 计算机应用与软件, 1997, **14**(3): 48- 51.]

[4] 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): 895- 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): 895- 900.]

[5] Shen Z Q, Wang R C. Study on a New Method for Transferring Grid to Vector Using the Topological Theory [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.]

[6] Xie S P, Du J K. A Method for Tracing Run-length Outline to Implement Vectorization of Raster Graphics and Image Data [J]. *Journal of Remote Sensing*, 2004, **8**(5): 465- 470. [谢顺平, 都金康等. 实现栅格图形和图像数据矢量化提取的游程轮廓追踪法 [J]. 遥感学报, 2004, **8**(5): 465- 470.]