

遥感目视解译图的扫描数字化*

潘 贤 章

(中国科学院南京土壤研究所 210008)

SCANNING DIGITIZING OF VISUAL INTERPRETATION MAP OF REMOTELY SENSED IMAGES

Pan Xianzhang

(*Institute of Soil Science, Academia Sinica, Nanjing 210008*)

关键词 扫描数字化, 细化, 矢量化, ARC/INFO 系统

遥感目视解译图等线画图输入地理信息系统, 传统的一般都采用手工数字化方法, 由于这种方法手工量大, 且精度和速度等与操作者工作熟练程度和耐性等有很大关系, 所以人们逐渐青睐扫描数字化这种自动化程度高, 更为快速精确的数字化方法。

这里的扫描数字化是指首先将线画图扫描成栅格式的二值图象, 再通过细化和矢量化处理生成矢量式数据的过程。由于这种数字化过程统统由计算机完成, 既提高了效率, 也提高了精度。我们开发了一套扫描矢量化软件, 配以 A₄ 平板扫描仪和 Photostyler 扫描软件即可完成扫描数字化过程。现将方法介绍如下。

一、线画图的扫描

扫描输入对图件的质量要求较高。原始判读图要清绘整洁, 着墨均匀且连续不断, 一般要求线宽 0.2 mm 左右, 不填代码。

扫描使用 A₄ 幅面或更大幅面的平板式扫描仪, 扫描软件用的是窗口下运行的 Photostyler 1.2 版。主要扫描参数如下: 采样点大小为 0.1 mm × 0.1 mm, 进行黑白二值扫描, 阈值为 100, 可根据图件着墨的浓淡进行适当调整。

先启动扫描仪, 待其检测完毕后, 启动窗口软件, 运行 Photostyler; 然后将图正面朝下覆盖于扫描仪上, 预扫图象, 框出图象范围, 开始正式扫描。扫描出的图象上若线条有断点, 可重新设置阈值后再扫描, 直至取得最佳的二值图象。扫描后的图象要求以非压缩 TIFF 格式存贮。

* 本文曾经曾志远教授斧正, 特此致谢。

二、线条的细化处理

由于扫描图象上线宽度往往大于 2 个采样点,而矢量化要求的只是单线宽度,因此必须对原始扫描图进行细化处理,过程如下。

(一) TIFF 格式文件的数据转换

非压缩二值 TIFF 文件分为文件头和数据块两大部分。文件头中第 30 和 31 字节存放图象宽度信息,第 42 和 43 字节则存放图象高度信息,文件头长 194 字节(不同扫描图象文件头长度可能不一样)。数据块部分按扫描先后依次记录了各采样点的值,每 8 个采样点的 0 值或 1 值用一个字节表示,因此在细化时先要将这些数据还原成单个采样点的形式。这个过程并非单独进行,以免全部展开后数据要占用相当大的硬盘空间。原始扫描图的数据转换还包括图象的黑白反视,即使得图上有线条的地方值为 1,背景值为 0。以后的一切处理都在此基础上进行。

(二) 扫描图象的预处理

扫描过程中由于各种各样原因会使得线条不均匀,出现孔洞和短枝,这些必须先消除掉,否则会影响后面的细化和矢量化效果。孔洞和短枝的消除有两种方法,一是在扫描后直接对图象进行必要的修改,另一种是在细化软件中加入这个处理功能^[1]。我们实际采用第一种方法,因为该方法直观简便。

(三) 扫描图线条的细化

细化采用经典细化法^[2]。该算法以连通性准则为基础,凡符合图 1 的两个模板及模板 a 的一次 90° 旋转和模板 b 的 3 次 90° 旋转模板的 P 象素就是骨架象素,不能删除。模板 a 和 b 中标记为 A 或 B 的区域中,至少有一个象素必须为非零,标记为 2 的位置也可以是 1。

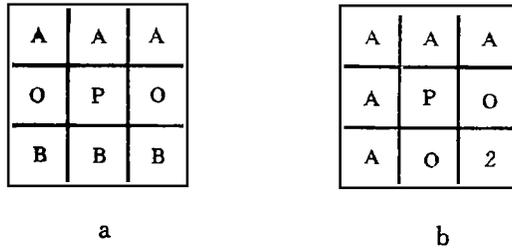


图 1 连通性模板 a 和模板 b

经典细化法首先处理宽度为 2 的线段,以免它们被删除。具体细化过程是:对每个黑象素点进行判断,符合连通条件的象素首先被标记为 2,非骨架象素被标记为 3,每轮处理后非骨架象素被置为 0,如此循环处理直至所有非骨架象素全部变为背景象素,这样最后保留的就是骨架象素,即形成了宽度为 1 的线条。

该算法结果黑象素之间是 8 连接的,非常适合于矢量化的要求,但运算速度较慢,可以进行改进。我们曾试过其它多种算法(包括 T.Y. Zhang 的快速细化法等^[4]),但细化效果没有此法好。

三、矢量化过程

矢量化处理指依次找出各线段上所有点行列位置,并用一串数据表示该线段位置的过程。数据存贮采用以下格式:

标记值 1——曲线 1 的标记

行坐标、列坐标——曲线 1 起始点行、列位置

行坐标、列坐标——曲线 1 上其它点位置

行坐标、列坐标——曲线 1 最后一点行、列位置

END——曲线 1 结束

标记值 2——曲线 2 的标记

行坐标、列坐标——曲线 2 各点位置

END——曲线 2 结束

……——各曲线各点位置

END——矢量文件结束

这种格式数据用 GENERATE 命令可以很容易读入 ARC/INFO 信息系统。

矢量化主要过程是这样的^[2]:

(1) 标记节点和端点为 -1。

(2) 对每一节点,依次搜索与它相连的各曲线段的各个象素位置,并给不同曲线段分配一个不同的标记值,将此值和曲线上各点行列坐标一并记入数据文件,并用 END 标记各曲线段结束。同时将已被搜索过的曲线点标记为 2。如此跟踪完所有的节点。

(3) 进行闭合曲线的跟踪。闭合曲线可以认为是首尾节点相同的线段,跟踪方法同上,即在完成以上(1)和(2)步骤后,再从左至右,从上到下搜索象素为 1 的点,并将最先出现的点定为节点,由此跟踪完整条曲线。并如此直至完成所有闭合曲线的跟踪。

(4) 在数据文件末置 END,以标记该矢量数据文件的结束。这样就完成了曲线的矢量跟踪,实现了从扫描到矢量化的主要过程,生成的矢量式数据文件输入 ARC/INFO 系统之后,即可进行一些必要的修改和编辑,建立拓扑关系,输入图斑属性等。

四、实例与讨论

为检验软件质量及扫描数字化的可行性,我们选取一小块遥感判读图进行扫描、细化及矢量跟踪,结果见图 2、3、4。

比较这三个图,可以看出细化后曲线正好位于扫描黑线的中间,这说明细化方法能够完成曲线特征的提取,效果良好。矢量跟踪回放图(图略)与细化结果是一致的,说明矢量化也可行。

以上的例图很小,实际由于遥感解译图幅面很大,给数据处理带来很多新的问题。下面着重对几个有关于大幅面图处理的问题作一讨论。

1. 大图象的分块细化问题 一幅 1:25 万遥感解译图按 0.1 mm 采样点扫描成图后达 6000×5000 (列×行)点,一次装入内存处理在普通 PC 机上不易实现,一般都分块细化。分块处理会造成接缝处线段变形,并带来很多新的接线头。解决这个问题一般采用重叠处理的方法,即上下两块各重叠一行处理,而左右两块则重叠 8 列(这是因为我们没有预先进行 TIFF 文件数据的展开)。由于保证了后面矢量化数据来源于非重叠区域,也就保证了穿过相邻两块的曲线的完整性。

2. 细化后图形短枝的处理 通常在细化后需去掉短枝,但分块处理后客观上要产

生很多短枝,尤其在边缘地方,若将所有短枝删除,势必造成真正骨架线条的流失,致使矢

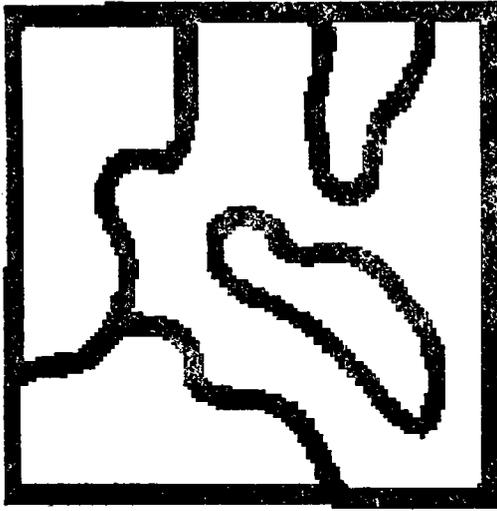


图 2 原始扫描图

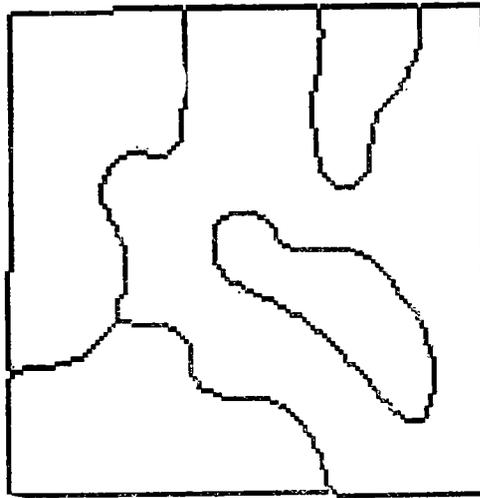


图 3 细化处理后的图

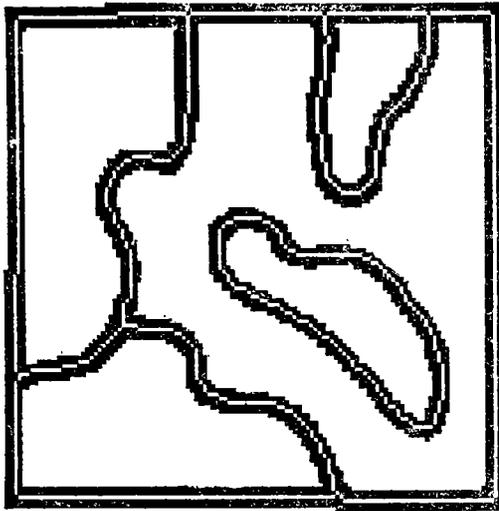


图 4 原始扫描图与细化图叠加显示

量化后形成很多断线的情况。当然加上一些严格条件,可以保证某些正常线条不致流失,但要给编程带来很多困难,考虑到 ARC/INFO 软件在做 Clean 处理时可以消除短枝,并且该软件具有较强的图形编辑功能,所以细化后我们就未做短枝消除的处理,而是把它交给 ARC/INFO 系统去做。

参 考 文 献

1. 林春蔚等, 1993: C 环境下地图图像矢量化及图形编辑技术与实例。61—62 页, 海洋出版社。
2. 朱振兴, 1993: 从栅格数据直接建立拓扑关系的算法研究。环境遥感, 第 8 卷 3 期, 232—240 页。
3. T. Pavlidis, 1987: 计算机图形显示和图象处理算法(中译本)。183—185 页。科学出版社。
4. T. Y. Zhang and C. Y. Suen, 1984: A Parallel Algorithm for Thinning Digital Patterns. CACM 27, No.3, pp. 236—239.