

基于专题分类图的栅格矢量数据转换

谭炳香

关键词 栅格数据、矢量数据、地理信息系统

地理信息系统(Geographic Information System,简称GIS)中数据结构主要有两种,矢量数据和栅格数据。二者在原理上彼此独立,应用于不同的处理过程。例如美国环境系统研究所研制的ARC/INFO地理信息系统是一个典型的矢量数据结构的GIS,美国陆军工程兵团研制的GRASS是一个基于栅格数据的GIS。

矢量数据结构和栅格数据结构是GIS中不可缺少的两种数据结构。二者彼此独立,又各有优缺点。不同的GIS之间数据不能兼容,在数据处理中遇到许多麻烦。如果矢量数据和栅格数据能够互换,不同系统中的数据就可以传输,兼容共享,共同管理和处理,同时满足数据输入输出的不同要求^[1]。

另外,在不同系统中专题数据的自动更新问题还未完全解决。近几年国内外研制的GIS中,缺乏这种数据的自动更新能力。大部分系统仍采用手工数字化输入数据的方法。这样有几个缺点:①手工输入数据费时费力,工作量大,效率低精度有限。对卫星形像或航空的分析、解释最后绘制成图,都要花费大量的时间和人力。②系统中信息的实时性不好。手工数字化输出的地图是多种来源的空间信息、非空间信息经过综合处理然后绘制成图的,在时间上难以保证信息的稳定性,不能做到及时更新数据。③影响微机GIS的推广应用。因为这要求用户使用数字化仪等外设。近几年来,迅速发展的遥感技术,为解决数据自动更新创造了条件。数字图像空间分辨力和光谱分辨力的提高,计算机自动分类技术的进步,使得遥感数据的价值得到了广泛的承认。图像处理系统对卫星影像进行处理得到的一系列专题图像,不仅是分析和决策管理人员解决实际问题的基础,而且也是GIS的重要数据来源。目前,遥感图像解译时,专题信息的提取仍以目视解译为主。但随着人工智能和专家系统研究工作的发展和完善,图像解译工作会逐步走向自动化,遥感信息制图的自动化也将成为现实。在此情况下,如果能完成栅格数据到矢量数据的自动转换,对于解决数据更新问题,遥感信息的自动制图,以及图像系统和图形系统之间的接口,促使遥感和GIS的结合,都无疑有重要的意义。

1 国内外研究简介

在空间数据的转换中,从矢量数据到栅格数据的转换问题解决已久。但反转换法即从栅格数据到矢量数据的转换方法仍是国内外研究的对象^[2]。下面就国内外在这方面所做的工作

1991-08-07收稿。

谭炳香研究实习员(中国林业科学研究院资源信息研究所 北京 100091)。

作一下介绍。

已有的栅格矢量化转换算法大多是针对二值图象的。大量的算法假定整个二值图象已在内存, 或者对图象进行轴化(又称细化), 把灰度图变成二值图, 从而以某种检索规则来跟踪轮廓, 如8-邻域跟踪法^[3]; Pavlids 通过产生和检索行相邻图(LAG)来提取轮廓点; Parker^[4]利用了离散条件下的弦线定理等, 这些算法在存储和速度上很难令人满意。Cederberg^[5], Batchelor^[6]分别提出的通过栅格扫描来获得轮廓的表示, Lichtner 开发的处理栅格和矢量转换的综合软件RAVET, 这些算法在存储和速度上日趋完善, 但这些算法获得的结果是非结构化的矢量数据, 因而只能用于显示和编辑, 而不能进行图形操作和处理, 不能满足制图要求。

后来开发的栅格矢量数据转换软件, 多数是灰度图像直接向矢量式数据转换。武汉测绘大学林开愚等为实现ARIES图象处理系统专题文件与RAMS图形处理系统的多边形文件的相互转换而设计的算法, 其基本思想仍是基于二值图像的邻域跟踪。北京大学张文中等设计了双边界定向搜索算法(DBDF), 其边界线搜索采用 2×2 栅格矩阵作为扫描窗口, 在每一个窗口内的四个栅格数据模式可以唯一地确定下一个窗口应移动的定向和该弧段的左右多边形编号, 大大提高了搜索速度, 拓扑关系也容易建立, 但未输出完整的矢量数据结构。北京大学李钦敏等在Geo-union 微机地理信息系统中设计了用一个 2×2 的子矩阵漫游整幅图象的算法, 需要对线段和结点进行一系列的处理才能获得较完整的数据结构和拓扑信息。中科院卫星地面站贾安等设计了一种直接从栅格式遥感专题图象到其多边形表示的算法, 基于游程编码栅格图象, 通过动态数据结构上指针操作, 在每一行上跟踪区域轮廓和拓扑关系的变化, 一次遍历原图来完成栅格数据的矢量转换, 从而使存储空间显著改善, 大大提高了速度。

本文设计的算法重点考虑以下几个因素: 一是在存储空间上有显著改善; 二是建立较完整的矢量数据结构和拓扑信息关系, 以满足不同输出设备的需要, 实现机助制图; 三是着眼于卫星遥感图像, 希望用遥感数据更新地理信息系统。本文以专题分类图为基础, 对栅格数据矢量化作了初步探讨。

2 栅格数据到矢量数据转换的原理与方法

栅格数据到矢量数据的转换实质上是图形的分割问题。其原理在于提取重要意义的点线面以及建立相应的数据文件(包括属性和拓扑信息), 把栅格图像矢量化。可以是二值图像向矢量式数据链的转换, 或者是灰度图像直接向矢量式数据转换。

2.1 一些约定

一幅专题分类图, 其分类数目及相应的类属性值是已知的, 因此建有一个图像类属性文件, 记录每类的编号和类属性值。

在专题分类图象中具有相同类属性值的相邻的点构成的集合称为一个栅格区域或栅格多边形(以下简称多边形), 这与矢量结构多边形是同名异构。

一个或几个的多边形完全包围在某个多边形内, 则这个或几个多边形称为岛多边形。

图像中属性发生变化处的象元构成边界点。相邻的具有相同左右或上下多边形属性的边界点构成多边形边(亦称线段)。如果某一边界点为3条或4条多边形边的公共点, 则称其为

结点, 结点同时又是多边形边的端点。相邻的多边形具有公共的多边形边。

图像中相邻或不相邻的具有相同左右或上下多边形属性的边界点构成的集合称为一线段类。相邻的或不相邻的属某类多边形的线段构成的集合称为一多边形类。

2.2 边界点和结点的提取

2.2.1 特殊处理 ①图像的四个顶点作为4个结点。②顶点之间的边界点, 也作为结点, 并与前一个顶点或结点或后一个顶点构成一条线段。

结点顺序编码, 记录每个结点的编码和X、Y坐标, 建立图轮廓点文件。

2.2.2 边界点和结点的跟踪 采用2*2栅格阵列作为窗口顺序沿行, 列方向对栅格图象全图扫描。①如果窗口内的四个栅格有且仅有两个不同的编号(灰度值), 则该四个栅格点构成一个边界点, 如图1。②如果窗口内的四个栅格有三个以上不同编号(灰度值), 则该四个栅格点构成一个结点, 如图2。③对于对角线上栅格两两相同的情况, 由于造成多边形的不连通, 也当作结点处理, 如图2。

边界点和结点的情况共有14种, 如下:

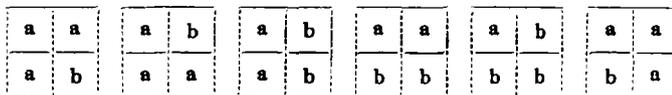


图1 边界点的六种结构

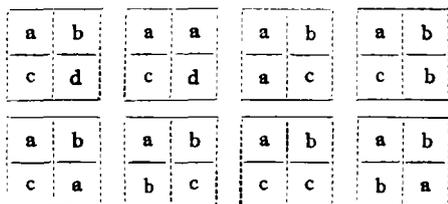


图2 结点的八种结构

记录每个边界点或结点的X、Y坐标, 并保留其左区或右区多边形的属性信息, 且标出它是否为结点。

2.3 边界线的跟踪

2.3.1 相同属性线段点的跟踪 相同属性线段点的跟踪就是将所属左右多边形的属性

信息相同的边界点或结点归为一个线段类。选定任意一边界点或结点为起点, 根据边界点或结点所属的左右多边形的属性信息, 对所有边界点和结点扫描。左右多边形的属性信息相同的归为一起, 构成一线段类的点集合。构成线段类的点仍然保留其结点标示符。

2.3.2 线段点的排序 图像中, 同一类别的多边形可能有一个以上, 故某两类的公共边就可能有一条以上。因此构成某一线段类的点集合可能是几条线段的点的集合, 必须将每一条线段的点区分开来, 同时按一定顺序排列。

线段点的排序逐线段类进行。从一个线段类的点集合中选择一个结点作为线段的起点, 并记为当前点。按最小距离原则寻找后继点, 并转移当前点。跟踪过的点作标记。如此判断跟踪至到无后继点为止, 这样构成一条线段的点就由左到右或由上到下完成排序。给每一条线段编码, 记录其左右区多边形的属性信息, X、Y坐标, 建立点线面的拓扑关系文件, 同时给结点编码, 建立结点与线段的拓扑关系文件。

检查一线段类的点是否跟踪完毕, 如果无剩余点, 则此线段类点的排序结束。否则说明此线段类至少由两条线段构成。从剩余点中寻找一个结点作为另一条线段的起点, 如同上述方法寻找后继点直到无后继点为止。

所有线段类的点排序完毕时, 线段跟踪结束。

2.4 多边形的跟踪

2.4.1 同类多边形线段的跟踪 同类多边形线段的跟踪是按多边形类逐个进行的。若线段所属的左或右多边形属性与某一类别 I 的属性值相同, 则线段是 I 类的一条边。对所有线段进行扫描, 判断其归属。这样可求出构成某类多边形的线段号及数目。

2.4.2 多边形线段的排序 由于某类别可能有两块以上, 所以构成某类别多边形的线段可能分属于几个多边形, 故必须将它们分开来, 且按一定方向排列起来。原理同线段点的排序一样。

选某一类多边形的第一条线段作为起始线段, 记为当前线段。寻找与当前线段的终点相等的起点或终点所属的线段, 则该线段为当前线段的后继线段, 并将其转为当前线段。如此判断跟踪, 直到无后继线段为止。这样一个多边形的线段排序完毕, 或一多边形类的线段排序完毕。给多边形编码, 记录多边形的编码和组成它的线段的有向编码, 建立多边形和线段的拓扑关系文件。

所有多边形类的线段排序完毕时, 多边形的跟踪结束。

2.4.3 岛多边形的处理 岛多边形是由一条封闭的线段构成的, 根据多边形的线段数确定其是否为岛多边形, 线段数为一的多边形就是岛多边形。

岛多边形边的起点和终点重合, 因此当作一个结点处理。

2.5 矢量数据的优化

多边形边的优化包括多余点的删除和多边形边的平滑。

在一条线段上的 3 个连续点, 如果在一条直线上, 则认为三个点中的中间一点为多余点, 予以删除, 这个过程称多余点的删除。这样做的目的是压缩数据, 节省内存, 提高矢量操作和存储效率。对下列实例进行处理后数据存储空间减少 39.16%。

由于上述方法生成的线段因栅格效应成阶梯状, 显然应加以平滑处理。采用各种分段多项式和样条函数等方法均可。

3 实验与结论

本文设计的算法在 386 计算机上得以成功实现。对中国水系专题分布图的一部分 300*300 的窗口(图 3 略)进行了操作, 取得满意的结果。所开图像窗口有 18 个类别, 经转换处理得到 65 条线段, 22 个多边形(包括一个岛多边形)。原图的存储空间为 90 K, 转换后的矢量数据存储空间为 25 K, 数据存储空间减少 72.3%。获得的矢量数据结构具有良好的拓扑关系, 利于图形编辑和查询, 后处理方便, 可以从多种输出设备上输出结果, 如在屏幕上显示[见图 4, 5(图 5 略)]或从绘图机上输出结果(图 6), 从打印机上输出

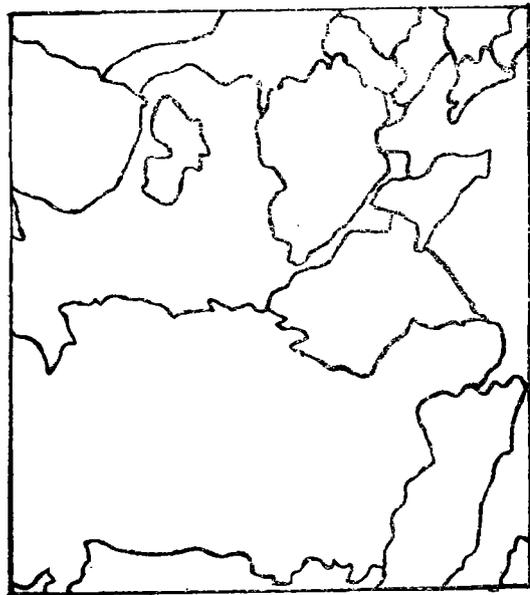


图 4 矢量轮廓图

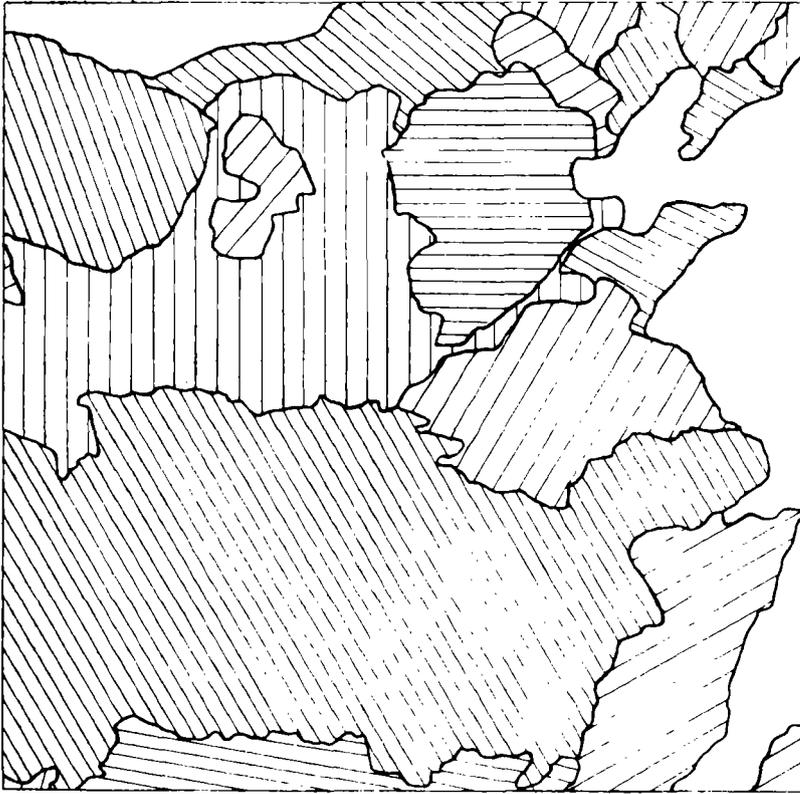


图6 DMP-52MP绘图机输出的矢量图

拓扑信息关系表。

本算法综合了栅格数据和矢量数据的长处,使不同数据达到统一,在不同的技术系统中兼容共享,提高了数据利用率。本算法不仅对专题分类图进行格式转换是实用的,而且形成的矢量数据能直接满足绘图机等绘制专题图件的需要。对于遥感卫星影象,可以在机助分类识别后,或在进行图像处理及目视判读后,得到专题分类图。经过栅格矢量转换,从绘图设备上输出结果是可行的。从而为遥感信息的自动制图奠定了基础。

参 考 文 献

- 1 Wilkinsim G G, Fisher P F. 地理信息系统的目前发展状况及将来趋势. 地理译丛, 1988, (4), 29~33.
- 2 张光宇, Lee Y C. 地理信息系统的回顾与发展、测绘通报, 1990, (4):31~36, (5):31~40.
- 3 胡友元, 黄杏元. 计算机地图制图. 北京测绘出版社, 1987.
- 4 Parker J R. Extracting vectors from raster images. Comput. and Graphics, 1988, 12, (1).
- 5 Cederberg R. Chain-link coding and segmentation for raster scan device. Comput. Graphics Image Process <CGIP>, 1979, (10), 224~234.
- 6 Batchelor B G, Mariow B K. Converting run code to chain code. Cybernetics Systems Int. J., 1981, (12), 237~246.

A Algorithm for the Extraction of Polygonal Representation from a Thematic Classification Map

Tan Bingxiang

Abstract A simple and improved algorithm is developed which extracts the polygonal contour representation from a raster format thematic classification map. The design adopts a 2×2 window which is used to scan and extract the boundary pixels of adjacent polygons along each pair of lines of the map. Contours are thus generated as a list of (X, Y) coordinate pairs, together with the attributes of the left and right regions at both sides of the boundary point. In the process of line segment tracking and polygon tracking, various attributes and topological relations are established. The structure, attribute and topological information of the vector data obtained are integrated and standardized, and they are easy to inquire and operate, and make the postprocessing flexible. The algorithm described in this paper has been implemented on a 386 microcomputer.

Key words raster data, vector data, geographic information system

Tan Bingxiang, Assistant engineer (The Research Institute of Forest Resource Information Techniques, CAF Beijing 100091).

“大青山实验局森林资源现代化经营管理技术的研究”成果 整体达到国际先进水平

本课题是林业部“七五”重点项目,由中国林科院组织,资源信息所、热林中心、林研所、热林所部分同志共同完成,1992年12月在中国林科院热林中心(广西凭祥)通过了林业部科技司主持的专家鉴定。鉴定意见认为:本项研究系统分析了森林资源与营林生产、经营决策三者的关系,提出森林资源经营管理由信息反馈、实施反馈和决策反馈三个反馈环构成。并根据这一理论设计了先进的经营管理模式,可不断进行资源数据更新,提出可靠的现状和动态统计值,机辅制定营林计划,指导编制作业设计,从而改变资源工作中的静态管理局面。按照上述理论模式,开发了科学的技术系统及其微机软件 FORMAN(微机森林管理系统)和 PCGIS(微机地理信息系统),利用信息技术完成资源信息和管理信息的交换,提高动态经营和资源动态管理水平。

研究创造性地提出了全林整体模型系统和营林措施效果的定量分析方法。两个专业软件具备林业资源信息管理的各项功能。地理信息系统可直接访问资源数据库,处理和输出多级区划的各种专题图。FORMAN 可以直接将营林生产信息自动反馈到数据更新系统。并可进行随机统计。上述成果与相应的立地营林技术研究配套,可大幅度提高生产和经营管理水平,保证资源档案的连续性,促进森林资源持续发展。研究提出的技术系统及其微机专业软件整体上达到世界先进水平。其中经营管理理论模式、全林整体生长模型系统及营林措施定量评价技术处于国际领先水平。

(中国林科院资源信息研究所 李希菲)