

土地利用扫描图件的注记自动识别与入库研究

李小娟 尹连旺 官辉力 赵文吉 孙永华

(首都师范大学资源环境与GIS北京市重点实验室 三维信息获取与应用教育部重点实验室,北京 100037)

摘要 土地利用基础图件建库是国土资源信息化和“数字国土”的重要组成部分。在传统建库过程中,图件上的大量注记主要是人工识别、手工录入到数据库中,影响了数据库建库的进度,并使数据库质量难以保障。为此提出了扫描图件数字注记自动识别与属性自动入库技术流程和方法,设计开发了较为实用的软件。在包头市土地利用详查数据库建库中的应用表明,上述技术方法合理,注记自动识别精度在87%以上,属性的自动入库准确率超过90%,具有较高的实用价值。

关键词 土地利用图 数字注记 自动识别 属性入库

中图分类号: TP79 P208 文献标识码: A 文章编号: 1006-8961(2006)10-1439-04

Identification of Annotation in Scanned Landuse Map and It's Importation to Database

LI Xiao-juan, YIN Lian-wang, GONG Hui-li, ZHAO Wen-ji, SUN Yong-hua

(Key Laboratory of Resources Environment and GIS of Beijing, Key Laboratory of 3D Information Acquisition and Application of Ministry of Education Capital Normal University, Beijing 100037)

Abstracts Spatial database construction of landuse maps plays an important role in our national land resources information management. Identifying the annotation and importing the attributes to database is a manual work in the traditional landuse database constructing, which costs a lot of time and brings low quality. In this paper, the approach of identifying the numeric annotation in scanned landuse map and importing the attributes to database automatically is discussed, and the practical software is developed. Applied in the landuse database constructing in Baotou, this methods is proved available and efficient.

Keywords landuse map, numeric annotation, automatic identification, attributes importation to database

1 引言

我国土地利用基础图件的数据库建设是国土资源大调查和国土资源信息化的基础与重要组成部分。土地利用基础图件的建库工作主要包括两部分:首先是进行基础图件的扫描矢量化,提取面状地类图斑、线状地物和行政界线等;第二,根据原始图件中的注记和手簿进行矢量图层的各种属性录入工作,包括地类码、图斑号、长度、宽度和行政权属等。

在空间要素矢量化方面,随着地图扫描矢量化技术的快速发展,市场已经出现了一些比较优秀的扫描矢量化软件,如 VpStudio, R2V, WinTopo 等,这些软件在提高地图扫描矢量化效率和质量方面都做出了重要贡献。在注记识别方面,由于地图注记类型多样(如包括符号注记、文字注记和数字注记等),注记的自动识别是一个比较复杂的问题。我国从20世纪80年代开始开展了地形图注记的识别研究^[1-5],取得了一定的进展,但迄今为止还没有实用的软件工具。

基金项目:国家自然科学基金资助项目(40471090);北京市教委基金项目(KM200510028013)

收稿日期:2005-04-28;改回日期:2005-08-19

第一作者简介:李小娟(1965~),女,副教授。1999年于中国科学院遥感应用研究所获地图学与地理信息系统博士学位。主要研究方向为地理信息技术在资源环境动态监测中的应用。E-mail: xiaojuanli@vip.sina.com

我国国土广袤,土地利用基础图件数量庞大(农区1:10 000比例尺),图上注记以数字类注记占优势,每个面状图斑都有以数字进行编码的地类码、图斑编号,每个线状地物都有长度、宽度等注记,数字注记数量是非常大的。采用手工数字注记录入,一方面限制了工作进度,同时也易产生录入错误。为此,结合包头市土地利用数据库建库工作,开展了土地利用图中数字注记自动识别和入库研究,开发了工具软件,在建库中的应用表明这些工具软件具有较高的实用性。

2 土地利用基础图件数据库建库技术流程

图 1 是土地利用基础图件数据库建库的技术流程。土地利用基础图件数据库建库主要包括两项内容:一是对各种点、线、面要素的矢量化,二是对地图上的文字和数字注记进行识别和属性的输入,主要包括注记自动识别、注记编辑(连接独立字符、删除错误字符)、注记规范化处理、空间一致性匹配、属性数据传递与检查纠正等。

对点、线、面要素的矢量化而言,一般的矢量化软件可进行自动识别,精度也较高;对于数字注记的

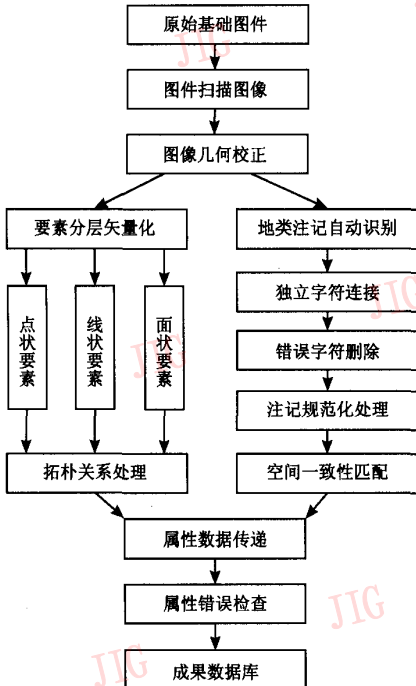


图 1 土地利用基础图件数据库建库技术流程

Fig. 1 The technologic flow of building landuse data base

自动识别环节,经过对比研究,选用 Ablesoft 公司 R2V 软件进行数字注记的自动识别,该软件操作相对简单,自动识别效果较好。

3 数字注记自动识别与编辑处理

对数字注记进行自动识别,首先是要进行注记图层剪裁,根据需要对注记层进行区域(ROI)裁剪,具体方法是:利用一个面状图层(包括:图廓、矩形区域或多边形边界)与注记层进行叠加分析,去掉边界外的注记要素。

使用 R2V 软件对数字注记进行识别,识别后的数字注记需要计算机进行自动分析、编辑处理,形成有意义和正确的数字注记。其中主要是组注记自动连接和无用注记删除处理。

3.1 组注记自动连接

数字注记自动识别后,出现最多的问题是注记分离,即将一个完整的数字注记分解成几组注记。例如,图 2 是经过 R2V 软件对数字注记进行自动识别的初步结果,圆圈标示中,地类图斑编号 2015 被识别为 20 和 15 两组数字注记,因此要进行组注记的自动连接。

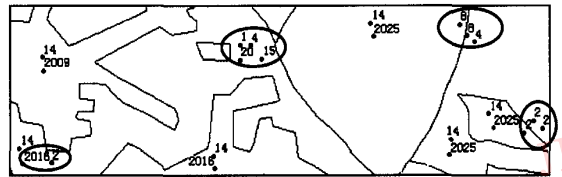


图 2 数字注记自动识别初步结果

Fig. 2 The preliminary result of identifying the numeric annotation

组注记自动连接的方法是:根据某个字符的位置,结合注记的行高和宽度生成一个矩形搜索工具,检索落在该矩形内的相关字符,按照从左到右的顺序依次连接各字符形成注记。具体技术流程如图 3 所示。

3.2 无用注记的删除

经过组注记连接处理后,还存在一些无用的注记,它们一般是图中的个别点位、线状地物交叉口等被错误识别为数字注记,这些注记需要删除(如上图右侧圆圈处)。由于这些注记一般为独立注记,或注记为单一字符,根据这个特点,计算机可以自动判断并逐一删除,最终形成正确的注记。图 4 是

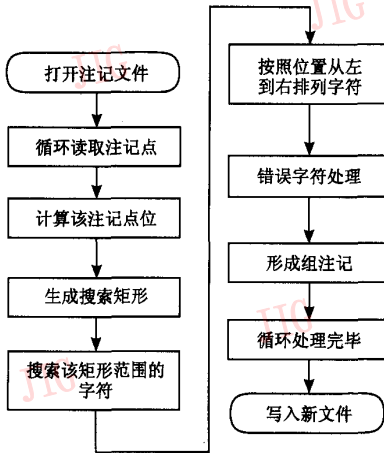


图 3 组注记自动连接技术流程

Fig. 3 The technologic flow of automatic connection between the group annotations

流程与算法如图 5 所示。

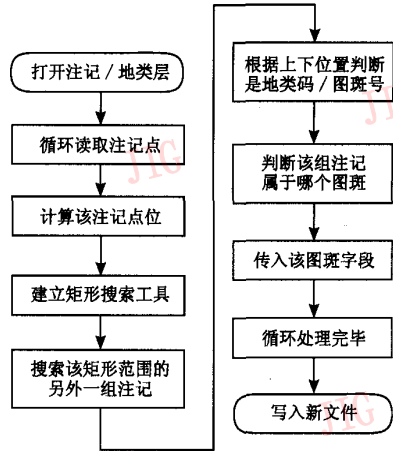


图 5 注记自动传入数据库流程

Fig. 5 The technologic flow of importing the annotations to the database

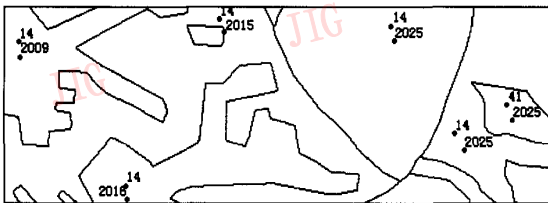


图 4 经过组注记连接和无用注记删除后的注记

Fig. 4 The final annotations generating by connecting the group annotations and deleting the useless annotations

过组注记连接和无用注记删除处理后形成的正确注记层。

4 数字注记的空间匹配与自动入库

上述处理后的数字注记是以 ArcView Shape 点状图层文件进行保存的,既每个点对应一个数字注记,数字注记存储在点状要素的属性表中。针对不同地类和点、线、面状要素类型,运用空间位置匹配的思想进行属性数据向土地利用属性数据库的传输。

对于面状图斑中的地类码和图斑编号,注记空间匹配和自动入库主要过程是:读取一个注记点,并搜索与其对应的另一个注记,按照空间位置判断它们属于哪个图斑;根据注记的上下位置,判断注记是地类码,还是图斑编号;然后计算机自动将其分别传入到该地类图斑的地类码和图斑编号字段中,处理

根据上述方法和技术流程,基于 VB6.0 和 Esri Mapobjects 2.0 开发了土地利用图数字注记自动识别与属性传递的处理软件。除了 R2V 在注记识别样本训练外,其余工作基本都由计算机自动处理。注记录入后,还要人工进行必要的属性检查,以确保数据库属性准确无误。该软件的功能主要包括扫描图像纠正,图幅裁剪,注记识别,组注记连接,错误注记删除,去除独立注记和注记属性自动入库。

扫描图像的注记自动识别与入库的关键问题在于注记识别的精度和预处理。在包头市土地详查数据库建设中应用上述技术方案进行地类码和图斑号注记的自动识别和入库,通过实验区 193 幅图件的统计试验,R2V 软件的注记自动识别精度在 87% 以上,属性的自动入库准确率超过 90%,极大地提高了工作的效率。同时,也减轻了后期属性数据库质量检查工作。

参考文献 (References)

- 1 GUO Bing-xuan, LI De-ren, LEI Zhen. Numeric annotation abstraction from map based on contour shape analysis[J]. Geomatics and Information Science of Wuhan University, 2001, 26(3): 275 ~ 278. [郭丙轩,李德仁,雷震. 基于轮廓形状分析的地形图数字注记提取[J]. 武汉大学学报(信息科学版), 2001, 26(3): 275 ~ 278.]
- 2 CHEN Rui, ZHANG Zu-xun, ZHANG Jian-qing. Identification of digital elevation annotation in scanned map [J]. Geomatics and Information Science of Wuhan University, 2002, 27(2):194 ~ 198.

- [陈睿,张祖勋,张剑清. 扫描地形图中数字高程注记的提取和识别[J]. 武汉大学学报(信息科学版), 2002, 27(2):194~198.]
- 3 Liu Shao-chuang, Lin Zhong-jian. Digits of map lettering recognition using back-propagation learning [J]. Journal of Wuhan Technical University of Surveying and Mapping, 1994, 19(3):195~198. [刘少创,林宗坚. 基于神经网络的地图数字注记识别[J]. 武汉测绘科技大学学报, 1994, 19(3):195~198.]
- 4 XU Zhan-wu, ZHANG Tao, LIU Xiao-lin. Automatic extraction and recognition of numbers in topographic maps[J]. Journal Of Chinese Information Processing, 2001, 15(2):51~56. [徐战武,张涛,刘肖琳. 地形图数字注记的自动提取与识别[J]. 中文信息学报, 2001, 15(2):51~56.]
- 5 Huang Wen-qian, Yang Qi-he. Recognition of explanation and numeral notes of map symbols [J]. Engineering of Surveying and Mapping, 1998, 7(1): 25~30. [黄文睿,杨启和. 地图符号的说明注记和数字注记识别[J]. 测绘工程, 1998, 7(1): 25~30.]