

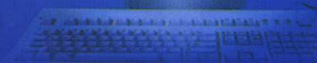


全国高技术重点图书·信息获取与处理领域

地图信息识别与提取技术

郝向阳 编著

010101010
1 1
0 0
1 1
010101010
010101010
010101010
1 1
0 0
1 1
010101010
010101010



测绘出版社

全国高技术重点图书·信息获取与处理领域

地图信息识别与提取技术

郝向阳 编著

测绘出版社

· 北 京 ·

内容简介

本书是一本系统介绍地图信息识别与提取原理和算法的学术专著。全书共分七章:第一章论述了地图信息识别与提取技术的发展现状;第二章介绍了现有的各种地图信息识别与提取方法;第三章重点介绍了基于结构特征的地图信息识别与提取的基本原理;第四章介绍了地图扫描图像的自动预处理方法;第五章介绍了各种主要地物符号的识别与提取算法;第六章介绍了等高线的识别与提取算法;第七章介绍了注记信息识别与提取的各种方法。

本书可作为测绘专业高年级学生和研究生的教学参考书,对从事测绘学、地理学、数字图像处理和模式识别领域工作的广大科技人员也有较大的参考价值。

图书在版编目(CIP)数据

地图信息识别与提取技术/郝向阳编著. —北京:测绘出版社,2001.6
ISBN 7-5030-0969-1

I. 地... II. 郝... III. 地图信息-信息处理
IV. P283.7

中国版本图书馆 CIP 数据核字(2000)第 15662 号

测绘出版社出版发行

社址:北京宣武区白纸坊西街3号 邮编:100054

三河市艺苑印刷厂印刷

2001年6月第1版·2001年6月第1次印刷

成品尺寸:148mm×210mm·印张:7

字数:188千字·印数:0001—2000

定价:16.80元

《全国高技术重点图书》 出版指导委员会

主 任：朱丽兰

副主任：刘 杲 卢鸣谷

总干事：罗见龙 梁祥丰

委 员：（以姓氏笔划为序）

王大中	王为珍	牛田佳	王守武	刘 仁	刘 杲
卢鸣谷	叶培大	朱丽兰	孙宝寅	师昌绪	任新民
杨牧之	杨嘉墀	陈芳允	陈能宽	罗见龙	周炳琨
欧阳莲	张兆祺	张钰珍	张效祥	赵忠贤	顾孝诚
徐修存	谈德颜	龚 刚	梁祥丰		

《全国高技术重点图书·信息获取与处理领域》 编审委员会

主任委员：陈芳允

委 员：汪成为 杨震明 袁宝宗 邓文强

前 言

近些年来,地理信息系统作为信息科学和信息产业的重要组成部分受到世界各国的普遍重视,并得到了快速地发展。数字地理信息是地理信息系统建立和应用的基础。国内外的实践表明,在建立地理信息系统的过程中,地理信息获取在工作量和资金投入两方面均占80%左右。现有各种比例尺的地形图是对地理信息较完整的表示,过去、现在和将来相当长的时期内,地图数字化都是获取数字地理信息的主要途径之一。传统的地图手扶跟踪数字化方法有速度慢、精度低、可靠性差、自动化程度低和劳动强度大等缺点,已远远不能满足地理信息系统发展的需要。从地图扫描图像上识别和提取数字地理信息为地图数字化开辟了新的途径,对促进地理信息系统的建立和应用具有重要意义。

地图信息的识别与提取技术涉及到计算机、模式识别、数字图像处理、人工智能、计算机视觉、地理学、地图学等多个学科。正如视觉计算理论的创始人美国麻省理工学院的 Marr 教授所指出的那样,所谓“看见”,就是指“发现有什么东西,在什么地方”。地图信息识别与提取在技术上主要解决地图上各种地理要素属性信息和空间信息的获取问题。近些年来,国内外的有关学者在该领域内进行了广泛而深入的研究,推动了地图信息识别和提取技术的发展。

本书是在作者博士论文的基础上完成的,是作者近年来在地图信息识别与提取领域所获研究成果的总结,同时还包括了国内外同行的部分有关研究成果。

本书共分七章,第一、二章主要介绍了地图信息识别与提取技术

的发展与现状,第三章阐述了基于结构特征的地图信息识别与提取的基本原理,第四章讨论了地图扫描图像的预处理方法,第五、六、七章分别讨论了地图上主要地物符号、等高线和注记等信息的识别与提取方法。

由于地图信息的识别与提取是一门涉及到多门学科的综合性的技术,并且正处于快速发展之中,加之作者水平有限,书中一定存在不少缺点和错误,敬请广大读者批评指正。

目 录

第一章 绪论	(1)
1.1 地理信息获取的方法与途径	(3)
1.2 地图信息识别与提取技术的发展与现状	(5)
1.3 目前地图信息获取技术存在的主要问题	(11)
第二章 地图信息识别与提取方法综述	(12)
2.1 线划跟踪方法	(12)
2.1.1 细线划的跟踪	(12)
2.1.2 粗线划的跟踪	(14)
2.1.3 区域边线的跟踪	(15)
2.1.4 预测跟踪算法	(15)
2.1.5 交叉点的处理	(16)
2.1.6 评述	(18)
2.2 数学形态学方法	(18)
2.2.1 二值影像的形态变换	(19)
2.2.2 二值影像形态变换的结构元	(21)
2.2.3 地图信息识别与提取的数学形态学典型算法 简介	(23)
2.2.4 评述	(27)
2.3 统计—结构方法	(27)
2.3.1 彩色地图的统计聚类分色处理方法	(28)
2.3.2 点状符号的统计—结构法识别	(30)
2.3.3 评述	(32)
2.4 神经网络方法	(33)

2.4.1	神经网络的基本概念	(34)
2.4.2	Kohonen 神经网络与地图分色处理	(36)
2.4.3	BP 神经网络与点状符号识别	(37)
2.4.4	评述	(38)
2.5	人工智能方法	(39)
2.5.1	地图信息识别所需知识及其分类	(40)
2.5.2	地图识别所需知识的表示方法	(41)
2.5.3	地图信息识别中知识的利用	(44)
2.5.4	评述	(45)
2.6	本章小结	(45)

第三章 基于结构特征的地图信息识别方法与策略

		(48)
3.1	地图信息结构特征的种类	(48)
3.2	基于结构特征的地图信息识别与提取的基本原理	(50)
3.2.1	模式基元的选择与提取	(50)
3.2.2	模式基元之间关系的分析与描述	(56)
3.2.3	地图符号的识别与提取	(61)
3.3	特征提取的有关算法	(62)
3.3.1	像素点特征	(62)
3.3.2	收缩变换	(63)
3.3.3	扩张变换	(64)
3.3.4	细化	(65)
3.3.5	直线段的检测与提取	(69)
3.3.6	曲线特征点的提取	(70)
3.3.7	Hough 变换	(71)
3.3.8	RLS 变换	(71)
3.4	地图信息识别与提取的分层处理策略	(73)

3.4.1	地图分色或分版	(73)
3.4.2	地图要素的分类	(74)
3.4.3	已识别地图要素的去除	(74)
3.5	地图信息的属性编码与数据结构	(75)
3.5.1	属性编码	(75)
3.5.2	数据结构	(75)
3.6	本章小结	(77)

第四章 地图扫描图像的预处理 (78)

4.1	地图扫描图像的二值化	(78)
4.1.1	定阈值二值化方法	(79)
4.1.2	可变阈值二值化方法	(79)
4.1.3	自适应二值化方法	(79)
4.1.4	最优阈值二值化方法	(81)
4.1.5	地图扫描图像二值化实例	(82)
4.2	地图图像的自动定向	(84)
4.2.1	十字交点的识别与提取	(84)
4.2.2	图廓点的识别和提取	(87)
4.2.3	定向参数的求取	(88)
4.2.4	实例与结论	(89)
4.3	地图图像上方里网线的自动消除	(95)
4.3.1	方里网线端点高斯坐标的求取	(96)
4.3.2	方里网线端点图像坐标的求取	(99)
4.3.3	方里网线的跟踪与检测	(100)
4.3.4	方里网线的消除	(100)
4.3.5	方里网线消除实例	(101)
4.4	图外信息的消除	(103)
4.5	本章小结	(105)

第五章 地物符号的识别与提取 (107)

5.1 黑块类符号的识别与提取	(107)
5.1.1 收缩变换	(108)
5.1.2 消除细线和孤立点	(108)
5.1.3 扩张变换	(108)
5.1.4 边线的自动跟踪	(109)
5.1.5 黑块类符号属性判别	(110)
5.1.6 黑块类符号定位信息的提取	(110)
5.1.7 已识别黑块类符号的消除	(112)
5.2 带晕线类符号的识别与提取	(115)
5.2.1 图像变换	(115)
5.2.2 开线划的消除	(116)
5.2.3 边界点的自动跟踪	(117)
5.2.4 带晕线类符号属性判别	(117)
5.2.5 已识别带晕线类符号的消除	(119)
5.3 虚线类符号的识别与提取	(126)
5.3.1 两端均为自由端的线分割与提取	(126)
5.3.2 交叉短线的分割与提取	(127)
5.3.3 短线的跟踪与合成	(128)
5.3.4 虚线类符号定位信息的提取	(129)
5.3.5 虚线类符号属性判别	(129)
5.3.6 已识别虚线类符号的消除	(129)
5.4 闭合曲线类符号的识别与提取	(132)
5.4.1 闭合曲线的跟踪	(132)
5.4.2 闭合曲线类符号属性判别	(133)
5.4.3 闭合曲线类符号定位信息的提取	(135)
5.4.4 已识别闭合曲线类符号的消除	(135)
5.5 组合线划类符号的识别与提取	(137)

5.5.1	短线划的跟踪与消除	(137)
5.5.2	长线划的跟踪提取	(138)
5.5.3	组合线划类符号属性判别	(138)
5.5.4	组合线划类符号定位信息的提取	(140)
5.5.5	已识别组合线划类符号的消除	(140)
5.6	长实线类符号的识别与提取	(144)
5.6.1	长线划的跟踪提取	(144)
5.6.2	长实线类符号属性判别	(144)
5.6.3	长实线类符号定位信息的提取	(145)
5.6.4	已识别长实线类符号的消除	(145)
5.7	本章小结	(147)

第六章 等高线的识别与提取

6.1	计曲线的分割与提取	(150)
6.1.1	等高线图像的收缩变换	(150)
6.1.2	细线的消除	(150)
6.1.3	等高线图像的细化	(151)
6.1.4	短线和黑色区域的消除	(151)
6.1.5	计曲线断点的拼接	(151)
6.1.6	计曲线矢量信息的提取	(153)
6.1.7	计曲线的消除	(153)
6.2	首曲线的分割与提取	(155)
6.2.1	首曲线图像的细化	(155)
6.2.2	细线与粘连区域的分离	(156)
6.2.3	细线断点连接	(156)
6.2.4	首曲线矢量信息的提取	(160)
6.3	等高线高程值的获取	(162)
6.3.1	闭合等高线间的拓扑关系	(162)
6.3.2	等高线拓扑关系的树结构表示	(165)

6.3.3	闭合等高线高程之间的关系	(166)
6.3.4	等高线间包含关系的判别	(168)
6.3.5	不闭合等高线的闭合化处理	(169)
6.3.6	等高线高程值的求取	(173)
6.4	本章小结	(175)

第七章 注记信息的识别与提取 (176)

7.1	文字识别的一般方法	(176)
7.1.1	文字识别的统计决策方法	(177)
7.1.2	距离	(178)
7.1.3	类似度	(179)
7.1.4	复合类似度和发胖距离变换	(180)
7.1.5	文字识别的句法结构方法	(183)
7.1.6	两种文字识别方法的比较	(184)
7.2	注记信息识别区域的确定与字切分	(186)
7.2.1	注记信息识别区域的确定	(186)
7.2.2	字切分	(186)
7.2.3	最大宽度回溯字切分算法	(188)
7.3	字的规范化处理	(188)
7.3.1	位置规范化	(188)
7.3.2	尺寸规范化	(189)
7.3.3	笔划粗细规范化	(190)
7.4	印刷体文字识别的粗分类	(190)
7.4.1	粗分类的目的和要求	(190)
7.4.2	重叠区分类	(192)
7.4.3	中心提取分类	(192)
7.4.4	判定树分类	(193)
7.4.5	引导树分类	(194)
7.5	印刷体文字分类的主要特征和方法	(195)

7.5.1	复杂指数和四边码	(195)
7.5.2	粗外围特征	(196)
7.5.3	粗网格特征	(197)
7.5.4	笔划密度特征	(198)
7.5.5	R 变换	(198)
7.5.6	文字特征点	(199)
7.5.7	平均线密度	(201)
7.6	注记信息识别结果的后处理	(201)
7.6.1	单字向字串的转换	(201)
7.6.2	数字字串向数的转换	(202)
7.6.3	注记信息位置的确定	(203)
7.6.4	注记信息的数据结构	(203)
7.7	本章小结	(204)
参考文献		(205)
后记		(210)

第一章 绪 论

自公元 105 年纸张在中国问世以来，人们便把它作为最主要的信息载体，记录并传播着在生产活动和社会活动中获取的各种信息，使人类文明得以不断地继承和发扬。人类认识自然、改造自然的历史是首先从认识人类赖以生存和发展的地球开始的，人们把地球表面的自然地理和人文地理信息按一定的数学法则以图形、数字和文字的形式表示在图纸上，这就形成了地图。千百年来，地图作为地理信息主要的表示形式，在人类社会经济活动的许多方面发挥着不可替代的作用。我国是世界上最早绘制和使用地图的国家之一，据史书记载，早在公元前 1100 年的西周时期，周、召二公在修建洛邑时就绘制和利用了洛邑城址附近的地形图。在现代社会中，地图更加成为人类进行社会经济活动不可缺少的重要工具。许多学科如地球科学、环境科学、地理学、土壤学、地质学甚至管理科学等都离不开地图，许多行业如城市规划、工程设计、资源勘察、土地管理、抢险救灾、军事指挥和观光旅游等也都离不开地图。正是由于地图的广泛用途，经典的测量学、摄影测量学和地图制图学等学科把测绘地图的理论、方法和技术作为最主要的研究内容；传统的测绘业也把测制各种地图作为主要任务；同时，地图也一直是广大测绘工作者奉献给社会的主要产品。

然而，本世纪 50 年代电子计算机的出现和其在近 20 年来的飞速发展把当今社会带入了信息化时代，长期以来测绘科学以一纸地图为研究中心的宁静局面也随之被打破。在信息社会里，以纸质地

图来存贮和表示地理信息的方式越来越难以满足人们的需要，其缺点主要表现在以下几个方面：①地球表面包含着几乎是无限的信息，针对不同的用途又需要测制不同品种和不同比例尺的地图，尤其是像我国这样幅员辽阔的国家，地图的数量是十分巨大的，这就给地理信息的保存、管理和使用带来了不小的困难。②随着社会的不断发展，自然地理信息和人文地理信息也在不断地改变和增长，这就需要对地图进行及时的更新，而纸质地图不利于对地理信息进行快速的修改和更新。③地理信息作为宝贵的信息资源，在使用中经常要进行用户和地理信息库以及用户和用户之间的调用，而纸质地图不利于地理信息的快速传输。④随着计算机技术的发展和地理信息应用范围的拓广，人们往往要利用地理信息进行各种空间分析和辅助决策，而纸质地图难以满足对地理信息的这种高层次的需求。目前，存贮容量不限且可重复使用的磁介质（磁盘、磁带）、光盘等技术已日益发展成熟，成为各种信息的理想载体，这就要求用数字的形式来表示地理信息。数字化的地理信息有利于利用计算机技术进行地理信息的存贮、管理、更新、传输和应用。随着计算机技术的发展和数字地理信息需求的增加，数字地图（也称电子地图）应运而生，并逐步成为各种信息系统的重要组成部分。

地理信息系统（GIS）是一种集地理信息获取、存贮、管理、分析和输出为一体的空间信息系统，它由计算机硬件、软件、地理信息和用户四大要素构成。作为信息科学和信息产业的重要组成部分，地理信息系统的研究和应用受到了各行各业的普遍重视。实践表明，在建立 GIS 的过程中，地理信息获取在工作量和资金投入两方面均占 80% 左右，地理信息获取不仅是 GIS 的重要组成部分，而且已成为制约 GIS 发展的一个瓶颈。因此，研究和探索作业效率和自动化程度较高的地理信息获取方法对促进 GIS 的发展和应用具有重要意义。

1.1 地理信息获取的方法与途径

目前,数字地理信息的来源和获取途径主要有以下三种:①全数字野外测量;②利用数字摄影测量和遥感的方法从地面的数字影像上获取;③将现有的各种比例尺的地形图数字化。

自 80 年代初以来,野外直接测量的仪器有了比较快速的发展,全站式电子速测仪逐步替代了传统的光学经纬仪、水准仪和平板仪,使得全数字野外测量成为可能,世界上主要测量仪器生产商都相继推出了各自的全站式经纬仪,如瑞士 Wild 厂的 T2000 和日本 Sokkisa 公司的 SET2 等。近几年来,随着 GPS 技术的发展,轻小型的 GPS 接收机也被成功地用于野外测量中,这方面代表性的产品有瑞士 Leica 公司的 Wild200 等。但是由于全数字野外测量方式本身野外作业和所需人力大等因素的局限,一般适用于小范围、大比例尺和高精度的应用场合。

摄影测量在经历了模拟摄影测量和解析摄影测量的发展阶段之后,现在已进入数字摄影测量时代。数字摄影测量技术的发展主要取决于数字摄影测量工作站和数字影像两个方面。一方面,随着计算机技术和数字图像处理技术的发展,各种型号的数字摄影测量工作站已走向市场,代表性的产品有美国 I²S 公司的 PRI²SM 系列,Intergraph 公司的 ISI 系列、瑞士 Kern 公司的 DSP 系列、Leica 公司的 DSW 和 DPW 系列以及我国武汉测绘科技大学的 Virtuozo 等;另一方面,随着航天技术和传感器技术的发展,仅在 1990~1995 年期间,世界各国就发射了 30 多颗卫星对地进行观测,根据已公布的计划,今后还会发射更多的卫星,这为快速及时地获取地面的数字影像提供了可靠保证,目前在市场上可以并不困难地获得美国 Landsat 和法国 SPOT 卫星提供的数字影像。但是,由于数字摄影测量工作站的价格比较昂贵且作业过程相对复杂等原因,数字摄影测量和遥感获取地理信息的方法更适合于进行数字地理信息的更新。

现有的各种比例尺的地形图是广大测绘工作者长期辛勤工作的结晶,是对地理信息比较完整的表示。尽管多数现有基本比例尺的地形图都不同程度地具有现势性较差的缺点,但不可否认,这些地形图毕竟是一种极其宝贵的地理信息资源。要从现有地形图上获取数字地理信息,必不可少的工作是进行地图数字化。过去人们一直采用在数字化板上进行手扶跟踪的数字化方法。大量的作业实践表明,这种方法具有速度慢、精度低和劳动强度大等明显缺点,同时,由于缺乏科学的检查手段,地图要素漏采和重采的现象在所难免,严重影响了数字化结果的可靠性,已远远不能适应对大量地形图进行数字化任务的要求。近些年来,为了克服手扶跟踪数字化方法的上述缺点,人们开始致力于地图扫描数字化方法的研究和开发。由于地图经扫描仪扫描后,可以获得数字地图图像,这为逐步实现地图数据获取的智能化和自动化创造了有利条件,一些业已成熟的计算机技术、数字图

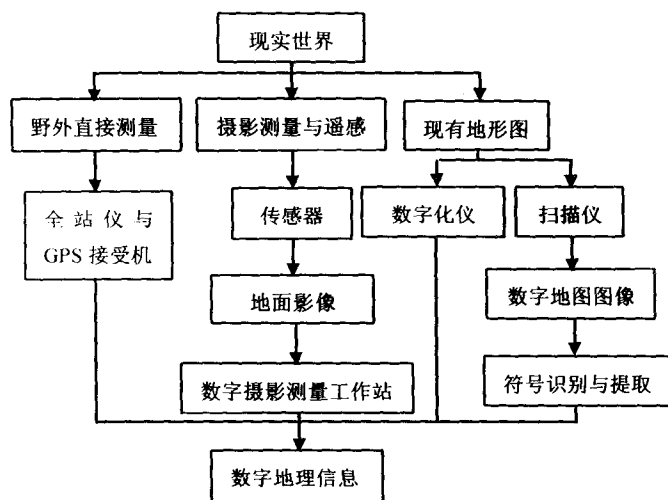


图 1-1 数字地理信息的获取方法与途径