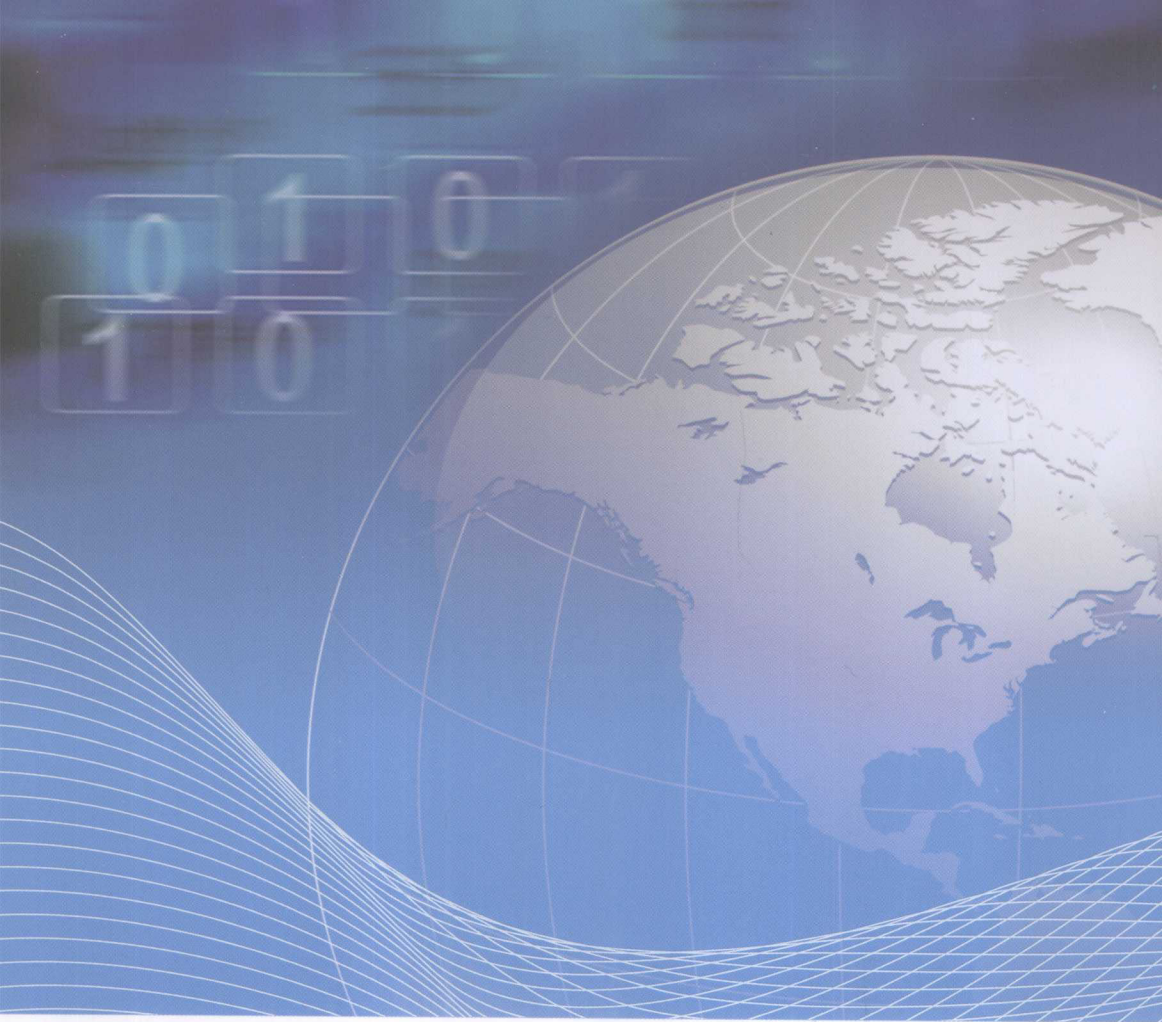




空间数据检索及 快速处理技术研究

杜根远 著



WUHAN UNIVERSITY PRESS

武汉大学出版社



空间数据检索及 快速处理技术研究

杜根远 著



WUHAN UNIVERSITY PRESS
武汉大学出版社

图书在版编目(CIP)数据

空间数据检索及快速处理技术研究/杜根远著. —武汉: 武汉大学出版社, 2015. 10

ISBN 978-7-307-16908-1

I. 空… II. 杜… III. 空间信息系统—数据检索—研究 IV. P208

中国版本图书馆 CIP 数据核字(2015)第 227523 号

封面图片为上海富昱特授权使用(© IMAGEMORE Co., Ltd.)

责任编辑: 鲍 玲 责任校对: 汪欣怡 整体设计: 马 佳

出版发行: 武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件: cbs22@whu.edu.cn 网址: www.wdp.com.cn)

印刷: 湖北恒泰印务有限公司

开本: 720 × 1000 1/16 印张: 14.75 字数: 219 千字 插页: 1

版次: 2015 年 10 月第 1 版 2015 年 10 月第 1 次印刷

ISBN 978-7-307-16908-1 定价: 38.00 元

版权所有, 不得翻印; 凡购我社的图书, 如有质量问题, 请与当地图书销售部门联系调换。

序

空间数据是一种具有空间位置属性的基础数据和专题数据，是国家基础设施建设和地球科学研究的重要资源。随着对地观测技术的发展，人类获取的空间数据越来越多。遥感数据是非常重要的空间数据。目前，遥感数据呈现出高空间分辨率、高光谱分辨率、高时间分辨率和多平台、多传感器、多角度的发展趋势。随着 Google Earth、World Wind 等地球空间信息平台的日渐普及，其海量、多尺度、多分辨率地球空间数据的组织管理、实时调度、快捷处理等，均不可避免地需要构造其全球空间数据模型，以便更加高效、快速、无缝地存取地球空间数据。空间数据的爆炸式增长，为全球、海量、多源的空间影像数据的处理和应用提出了新的考验，其中最主要的问题就是如何建立高效的空间数据组织方式，实现海量数据的快速检索和便捷化应用。

地球剖分理论和高性能计算为上述问题的解决提供了全新的思路。剖分是对地理空间进行位置划分的方法，剖分理论是按照一定规则将地球表面逐级划分为大小相似、没有相互依赖及顺序关系的数据集合，数据之间通过空间框架定义和唯一编码进行表达，存储节点和地理空间之间是一一映射关系。高性能计算利用集群的并行计算能力，通过高速网络实现任务分割和作业调度，从而实现海量空间影像数据的高性能并行处理。

《空间数据检索及快速处理技术研究》一书作者所在的研究团队在国家自然科学基金、“地球探测与信息技术”教育部重点实验室开放基金、河南省相关研究课题的资助下，围绕空间数据的快速处理及检索相关的理论与技术问题进行了较为深入细致的研究和探索，取得了一些有益的结论，该书即是他们团队研究成果的总结。书中所介绍

的空间剖分数据并行化处理方案把地球剖分理论与并行计算进行了有机结合，是一种新的尝试和创新，相信会对本领域研究者有所启迪。

笔者深信，本著作的出版，对于推动和发展空间数据检索及快速处理技术的综合交叉应用具有一定的学术和应用价值。祝愿作者及其所在的团队在科学研究的征程上不断取得新的成果并及时与同仁们分享。

A handwritten signature in black ink, consisting of stylized, fluid characters that appear to be '敬' (Jing) followed by a long horizontal stroke.

前 言

随着全球立体对地观测系统的逐步形成和完善,空间数据的数量、大小、复杂性和传输速度都在飞速增长,全球、海量、多源是其显著特征,其中,遥感图像数据是应用最为广泛的一种空间数据。目前,遥感应用的水平滞后于空间遥感技术的发展,从而造成空间数据资源的巨大浪费,其应用价值得不到充分利用,形成了空间数据的生产和传输能力远远大于空间数据解析能力的局面。研究海量遥感图像数据的高效组织与快捷应用、快速检索有效空间信息、提高遥感图像分析识别的精度,是目前遥感应用中亟待解决的问题,具有十分重要的科学意义和应用价值。

解决这一问题的关键是发展有效的空间数据组织管理和内容检索方法,空间数据组织效率和处理速度已经成为制约技术应用的瓶颈,目前,地球剖分理论和高性能计算为上述问题解决提供了一种可能途径。地球剖分理论是按照一定的规则将地球表面逐级划分为球面网格单元的方法,这些区域单元无缝、多层次且拟合地球表面,在空间信息表达与管理上具备独特的优势。该理论将空间数据逻辑分割为大小相似、没有相互依赖及顺序关系的数据集合,数据之间通过空间框架定义和唯一编码得以表达,存储节点和地理空间之间是一一映射关系,空间数据按照其所在的空间位置进行存储管理,在地理计算、空间分析及表达过程中,可利用集群的并行计算能力对其进行加速处理,从而为海量遥感影像数据高性能处理提供新的解决思路。

目前,图像内容检索技术取得了一些研究进展,但是针对遥感图像内容检索的研究却相对缓慢,无论是理论体系还是应用系统,都还不成熟,遥感图像具有尺度大、主题不明确、多时相、语义丰富等特点,普通图像中的研究成果不能直接应用于遥感图像内容检索中去。

对于一个完整的遥感图像内容检索系统,其数据组织、存储与管理、特征描述及提取、相似性度量、网络服务模式、系统架构设计及实现等研究工作面临着许多困难与不足,研究所涉及的各项关键技术势在必行。

本书针对空间数据检索及快速处理技术所涉及的关键技术,提出了一些创新性思路和方法,并分别从理论和技术的角度对其价值和实用性予以分析和验证。

本书由许昌学院杜根远博士拟定全书的撰写纲要,并负责各章节的内容安排、终稿审定和全文审校。其中,张火林讲师参与第6章的撰写,熊德兰副教授参与第2章的撰写;史进玲讲师参与第4章的撰写,邱颖豫副教授参与第5章的撰写。

本书研究基础来源于国家自然科学基金委河南人才培养联合基金——基于剖分模型的遥感影像模板并行处理方法研究(U1304403);“地球探测与信息技术”教育部重点实验室开放基金课题——基于数字地球平台的遥感图像内容检索关键技术研究(2008DTKF012);河南省科技攻关(重点项目)计划——Hadoop下剖分遥感影像并行处理平台设计与开发(132102210398),剖分遥感影像模板化并行处理技术研究与应用(112102210079);河南省基础与前沿技术研究计划——云环境下海量遥感影像数据存储机理研究(132300410349),粒计算及其在遥感图像检索中的应用研究(102300410060);河南省教育厅自然科学研究计划——遥感图像内容检索关键技术及数据库存储机理研究(2010A520035)。

感谢成都理工大学苗放教授在百忙之中为本书写序,在此深表感谢!在本书写作过程中,借鉴和参考了国内外同行的研究成果和有益经验,同时也引用了大量的参考文献,谨在此表示深深的敬意和感谢!

由于作者学术视野、专业水平及研究深度有限,书中难免有遗漏和错误之处。对于书中的错漏及不当之处,敬请广大读者批评、指正!

杜根远

2015年5月

目 录

第 1 章 遥感数据检索及处理技术发展	1
1.1 研究背景及意义	1
1.1.1 问题提出	1
1.1.2 研究意义	4
1.2 相关技术研究现状	5
1.2.1 图像内容检索技术	5
1.2.2 遥感图像内容检索技术	9
1.2.3 空间数据组织理论及快速处理技术	13
1.3 主要研究内容	16
第 2 章 空间数据组织及检索技术概述	17
2.1 空间数据组织与管理概述	17
2.1.1 概述	17
2.1.2 地球空间数据组织与管理	19
2.2 遥感图像数据的网络服务模式	27
2.2.1 WebGIS	27
2.2.2 C/S 和 B/S 模式	28
2.2.3 空间信息网络服务 G/S 模式	29
2.3 图像特征描述与提取	31
2.3.1 颜色特征	31
2.3.2 纹理特征	38
2.3.3 形状特征	39
2.3.4 图像高层特征提取	40
2.3.5 综合多特征提取	41

2.4 图像检索的相似性度量方法	42
2.4.1 距离度量方法	43
2.4.2 基于粒计算的相似性度量方法	44
2.5 检索方式及相关反馈机制	45
2.6 检索算法的性能评价	47
2.7 本章小结	48
第3章 遥感图像聚类分割方法研究	49
3.1 难点及意义	49
3.2 遥感图像分割方法概述	50
3.2.1 图像分割定义及研究进展	50
3.2.2 遥感图像分割	52
3.2.3 遥感图像聚类分割研究进展	54
3.3 结合 ECM 和 FCM 聚类的遥感图像分割方法	56
3.3.1 模糊 C 均值聚类算法	56
3.3.2 EC-FCM 算法思想	58
3.3.3 实验验证	61
3.3.4 结果分析	65
3.4 基于改进 FCM 的遥感图像序列分割方法	66
3.4.1 颜色空间选择	67
3.4.2 距离测度的选择	68
3.4.3 序列分割策略	70
3.4.4 实验及结果讨论	70
3.5 本章小结	75
第4章 基于粒计算的图像相似性度量研究	76
4.1 相关概念	76
4.2 粒计算理论	77
4.2.1 粒计算基本要素	78
4.2.2 粒计算基本理论	79
4.2.3 粒计算的基本问题	81

4.3 信息系统中的属性约简与多粒度度量	83
4.3.1 基于粗糙熵的信息系统属性约简算法	83
4.3.2 基于知识粗糙熵的序信息系统约简算法	87
4.3.3 信息系统中的多粒度度量	90
4.4 基于粒计算的图像区域相似性度量方法	92
4.4.1 图像特征信息的粒计算表示	92
4.4.2 图像区域相似性度量方法	96
4.4.3 实例分析	97
4.5 本章小结	100
第5章 空间剖分数据存储调度服务模型研究	101
5.1 空间数据组织理论发展	101
5.1.1 地球剖分组织理论	102
5.1.2 EMD 剖分模型	103
5.1.3 剖分面片及其编码	105
5.2 面向客户端聚合服务的 G/S 模式架构	108
5.2.1 基本架构	108
5.2.2 技术理论体系	109
5.2.3 地学信息浏览器	113
5.2.4 分布式空间信息服务器群	114
5.3 剖分面片模板数据模型研究	115
5.3.1 影像数据剖分面片模板	115
5.3.2 基于模板的剖分面片计算模式	118
5.3.3 剖分模板数据库系统构建及应用	120
5.4 空间剖分数据存储调度服务模型构建	128
5.4.1 剖分数据网络服务协议体系架构	129
5.4.2 协议支持下的空间数据访问流程	131
5.4.3 剖分数据存储调度模型总体框架	133
5.4.4 地址编码结构	134
5.4.5 寻址流程	136
5.5 服务应用实例	137

5.6 本章小结	140
第6章 空间剖分数据并行处理方法及平台开发	142
6.1 并行处理技术概述	142
6.1.1 并行处理基本概论	142
6.1.2 并行编程模型	143
6.1.3 并行处理技术在遥感领域的应用	147
6.2 基于剖分面片模板的并行处理技术	147
6.2.1 剖分模板并行处理模式	147
6.2.2 剖分面片基本空间关系	148
6.2.3 剖分模板计算模式	149
6.2.4 基于剖分面片模板的遥感影像并行处理方法	151
6.3 基于 OpenMP 与 MPI 的遥感影像并行分割算法	157
6.3.1 K-Means 算法	157
6.3.2 MPI+OpenMP 混合编程模式	158
6.3.3 基于 OpenMP 与 MPI 的遥感影像并行分割算法	158
6.3.4 具体应用实例	160
6.4 剖分遥感影像并行处理平台	162
6.4.1 开发环境介绍	162
6.4.2 软件开发过程	164
6.5 本章小结	170
第7章 遥感图像内容检索原型系统设计及实现	171
7.1 系统开发背景	171
7.2 分布式 CBRSIR 数据库存储机理	172
7.2.1 遥感图像数据模型分析	172
7.2.2 使用 Oracle 的栅格化空间数据存储	174
7.2.3 构建基于 Oracle 的分布式数据库	179
7.3 原型系统体系结构设计	182
7.3.1 系统设计原则及总体架构	182
7.3.2 基本数据库操作	185

7.3.3 对标准图形文件的支持	186
7.4 原型系统实现	187
7.4.1 检索接口	187
7.4.2 检索处理	188
7.4.3 检索算法	188
7.4.4 系统实现	191
7.5 本章小结	196
 第8章 结语与展望	 197
8.1 全书总结	197
8.2 研究展望	199
 参考文献	 201

第 1 章 遥感数据检索及处理技术发展

1.1 研究背景及意义

1.1.1 问题提出

空间数据(geospatial data)是指用来表示空间实体的位置、形状、大小及其分布特征等诸多方面信息的数据,可以用来描述来自现实世界的目标,具有定位、定性、时间和空间关系等特性。全球空间数据是指覆盖全球的航天侦察数据、卫星测绘数据、气象海洋环境遥感数据、导航定位数据、地球重力场数据、临近空间和中远距离航空侦察数据等航天航空对地观测数据。其中,遥感图像数据是应用最为广泛的一种空间数据。

由于雷达、红外和多光谱扫描仪、数码相机、高光谱成像仪、全站仪、天文望远镜、电视摄像、电子显微成像、CT 成像等各种宏观与微观传感器或设备的使用,以及常规的野外测量、人口普查、土地资源调查、地图扫描、统计图表等空间数据获取手段的更新和提高,计算机、网络、全球卫星导航系统(Global Navigation Satellite System, GNSS)、遥感(Remote Sensing, RS)和地理信息系统(Geographic Information System, GIS)等技术应用于空间数据,空间数据的数量、大小、复杂性和传输速度都在飞快地增长,全球、海量、多源是其显著特征。其中,遥感技术是人类获取资源环境动态信息的重要手段,随着人类对自身生存环境的日益关注,已经成为社会、政治和经济生活中不可或缺的组成部分。

随着现代科学技术的高速发展,空间信息资源的获取和应用日益

受到广泛重视,我们正从“数字地球”时代向“智慧地球”时代迈进(李德仁,等,2010)。目前,遥感对地观测技术正在形成一个多层次、多角度、全方位、全天候、高中低轨道结合、大中小卫星协同、粗细分辨率互补的全球立体对地观测系统(Earth Observation System, EOS)(李德仁,等,2006; Li, 2009)。QuickBird、SPOT、IRS、IKONOS、EOS、ASTER、WorldView、OrbView-3、CBERS、尖兵系列等产生的新型遥感数据不仅波段数量多、光谱分辨率高、数据速率高、周期短,而且数据量特别大,仅EOS-AM1和PM1每日获取的遥感空间数据就以TB级计算(李德仁,等,2000),Landsat每两周就可获取一套全球卫星遥感图像数据,美国航空航天局(NASA)的数字行星项目每天要产生1000GB新数据,从而造成了“空间数据的生产和传输能力远远大于空间数据解析能力”(李德仁)的局面。

遥感影像作为一种实时性高、覆盖范围广、信息丰富的空间信息资源,已经成为国家空间数据基础设施建设的重要基础数据,在航空航天、军事侦察、灾害预报、环境监测、土地规划与利用、农作物估产等诸多军事及民用领域发挥了重要作用(Mather, 2004)。随着对地观测技术、遥感技术、计算机及通信技术的迅猛发展,空间信息的数据量急剧膨胀,这为空间信息应用服务的自动化、实时化、智能化创造了有利的前提条件,也给空间信息的组织和管理带来了严峻挑战。同时,在信息全球一体化和大众化背景下,应用领域对遥感影像的兼容性、实时性、精度和可靠性要求也越来越高,处理速度已经成为遥感影像快捷应用的瓶颈(李德仁,等,2010)。因此,需要发展更高效的全球空间信息组织和管理的技术体系,实现海量空间数据的高性能处理和应用服务,以应对蓬勃发展的应用需求。

在过去的几十年里,遥感应用的水平滞后于空间遥感技术的发展,从而造成空间数据资源的巨大浪费,其应用价值得不到充分利用。研究如何从海量的遥感图像数据中获取有效空间信息,提高遥感图像分析识别的精度,是目前遥感应用中亟待解决的问题之一,如何从海量遥感图像数据库中快速准确地检索到所需信息具有十分重要的意义。解决这一问题的关键是发展有效的空间数据管理和检索方法,其核心是图像检索(Image Retrieval, IR)技术,这也是近年来海量遥

感图像检索所面临的瓶颈之一。

自 20 世纪 70 年代以来, 图像检索技术得到了迅速发展。早期被称为基于文本的图像检索(Text-based Image Retrieval, TBIR), 一般采用对图像注释关键词或者文本信息, 通过数据库管理技术实现基于文本的图像检索。传统的 TBIR 已经非常成熟, 如文本搜索引擎 Google、百度等均采用这种方式进行查询。但是随着 Internet 和网络技术的发展, 海量图像数据的出现, 其不足和局限性日益显现, 主要表现如下: 图像信息描述有歧义或无描述而造成检索结果不理想; 丰富的图像可视化信息无法用文字进行完整、全面、准确的描述; 若不具备专业知识无法应用于遥感、医学等特殊领域; 手工注释文本信息的工作量太大以及存在主观性、不确定性; 检索系统无法正确解释用户自然语言等。

20 世纪 90 年代, 为了克服文本检索存在的问题, 人们开始试图从理解图像本身特征的角度实现图像检索, Kato(1992)正式提出基于内容的图像检索(Content-based Image Retrieval, CBIR)概念, 它涉及对图像的查询、索引、浏览、搜索、提取等处理, 利用图像本身所包含的客观视觉特征(颜色、纹理、形状等), 通过这些特征的相似性度量从数据集中检索到目标图像。CBIR 尝试通过对内容的理解来有效利用图像数据库信息, 它是一门包括计算机视觉、图像处理、数据库技术和人工智能等在内的众多学科的新兴交叉学科技术。

近些年来, 中外许多专家学者对 CBIR 进行了大量深入的研究, 也取得了很大的进展, 甚至在某些领域已经商用, 但在图像特征选取、相似性度量、图像数据库组织与存储、用户交互、网络访问模式等方面还存在许多问题, 仍需要进一步研究。同时专门针对遥感图像内容检索的研究进展却相对缓慢, 无论是理论体系还是应用系统, 都还不成熟, 远远无法达到商用的目的。在医学、多媒体(如风景、动植物、商标、邮票等)等普通图像中的研究成果虽然可被遥感图像内容检索所借鉴, 但是, 遥感图像具有尺度大、主题不明确、多时相、语义丰富、海量等特点, 决定了上述研究成果不能直接应用于遥感图像内容检索中去, 对于一个完备的 CBRISIR(Content-based Remote Sensing Image Retrieval)系统, 其系统架构设计、数据组织与管理、

特征描述、网络服务模式、查询机制、相似性度量等诸多方面的准确性和有效性都不同于一般 CBIR 系统,研究工作面临着许多困难与不足,研究遥感图像内容检索所涉及的各项关键技术势在必行。

目前遥感影像的相关处理技术日趋成熟,已经取得了一大批理论和算法成果。但并行处理技术仍然是提高遥感影像处理速度和效率的最有效方法之一。高性能计算技术、高性能集群、各种专用硬件设备等均在遥感影像高性能处理领域得到了广泛应用。但是,由于缺乏有效的数据组织机制,多源异构的影像数据要实现动态、高效关联比较困难,此外,遥感影像并行处理算法需要采取合理的策略进行粒度分割,以充分发挥集群的计算能力,否则会造成计算集群各节点额外的数据访问开销。而且,遥感影像并行算法的实际运行效率往往达不到理想的加速比,严重依赖于集群的硬件配置情况,不具有通用性。如何科学高效地对空间数据进行组织、合理设计并行算法模型、充分利用集群 CPU(Central Processing Unit)、GPU(Graphic Processing Unit)及 MIC(Many Integrated Core)的计算能力对其进行优化,对编程者而言是一个难点,也给空间数据应用带来了困难。

1.1.2 研究意义

随着各种军用、民用卫星的日益增多,空间数据特别是遥感图像数量急剧增加,对于依赖于遥感图像进行环境监测、资源调查、地面目标监视等任务来说,从大量的空间数据中找出感兴趣图像以及从复杂遥感图像中找出感兴趣目标,日益成为一项繁琐而沉重的工作(李德仁,等,2006)。遥感图像包含两个方面的含义,地物目标的几何空间信息和光谱信息。由于存在比例尺的差异,不同比例尺的图像应用于不同的领域,其中查询使用的重点图像特征也不同。对于遥感图像内容检索,除了像普通图像检索一样进行特征提取之外,还需要综合考虑元数据如地理位置、波段、不同传感器参数、比例尺等因素与图像内容的关系,以及考虑遥感图像幅面大、细节多而引起的存储开销和检索效率等问题。

遥感图像内容检索在特征选取、相似性度量、查询机制和效率、数据快捷处理应用、网络服务模式、数据库机理等多方面都有其自身

特点, 如何对遥感图像进行有效的组织与管理、提高遥感图像快捷处理和应用的效率, 是本书讨论的重点之一。

基于地球剖分模型将空间数据逻辑分割为大小相似、没有相互依赖及顺序关系的数据集合, 数据之间通过空间框架定义和唯一编码得以表达, 由于存储节点与地理空间一一映射, 空间数据按照空间位置进行存储管理, 具有全球多尺度唯一基准, 在空间运算和分析中, 可以利用集群的并行计算能力对其进行加速处理, 从而为海量遥感影像数据高性能处理提供了新的解决思路。

本书作者在相关课题的资助下, 以军民重大需求为牵引, 通过对空间数据组织、管理及快速处理应用, 遥感图像内容检索若干理论和技术问题的研究与探索, 以期在一定程度上推动内容检索技术在空间信息领域的发展和应用, 进而推动空间数据的快捷应用。

1.2 相关技术研究现状

1.2.1 图像内容检索技术

图像内容检索是一门快速有效地从海量图像中检索出相关图像的技术。十余年来, 随着人工智能、图像数据库技术、图像工程、计算机视觉、模式识别、数据挖掘、自然语言理解和 Internet 技术水平的不断发展, 国内外学者对 CBIR 和 CBVR (Content-based Video Retrieval) 的研究也日益深入和广泛, 并已取得了一些重大成果, 一些原型系统也已得到初步应用。如 IBM QBIC (Niblack, et al., 1993)、Columbia 大学的 VisualSeek (Smith, et al., 1996)、Virage 公司的 Virage 系统 (Bach, et al., 1996)、UIUC 开发的 MARS (Mehrotra, et al., 1997)、MIT Photobook (Pentland, et al., 1996)、UC Berkeley 的 Digital Library Project (Ogle, et al., 1995)、CIRES: Content-based Image REtrieval System (Iqbal, et al., 2002)、UCSB Alexandria Digital Library Project 的 Netra (Ma, et al., 1999) 等。国内相应研究包括清华大学计算机系的“Web 上基于内容图像检索”、中国科学院计算技术研究所开发的“基于特征的多媒体信息检索系统” (Multimedia