

空间数据计算模式分析与应用

明冬萍¹, 骆剑承¹, 周成虎¹, 沈占锋¹, 李均力², 盛昊¹

(1 中国科学院地理科学与资源研究所资源与环境信息系统国家重点实验室, 北京 100101;

2 武汉大学遥感信息工程学院, 武汉 430079)

摘要:随着传感器和其他数据采集技术的不断进步以及对地观测网络的建设和启动,空间数据的高性能处理和分析成为摆在地理工作者面前的瓶颈。本文以此为出发点,按照不同地理领域(陆地、大气、海洋)的空间数据载体的形态的不同,将空间数据划分为反映固态基质信息的陆地空间数据,反映液态基质信息的陆地水文空间数据,反映液态基质信息的海洋流体空间数据和反映气态基质信息的大气流体空间数据四类,并对每类数据的最小单元问题进行了初步的分析。本文详细阐述了地理空间计算的涵义,并根据计算行为模式及计算的侧重点的不同,将地理计算过程分为深度计算过程与主动计算过程(即“数据→特征→知识”的一般计算过程),并就此进行了阐释。以基于特征的遥感信息提取和目标识别工作为例,对上述理论进行了说明和验证。最后对空间数据计算模式相关问题进行了总结,并对以后的研究做了展望。

关键词:空间数据;计算模式;基元;深度计算;主动计算;遥感

1 引言

目前,利用对地观测高新技术已经建成或正在建设一系列的对地观测和探测地球的各种过程的四维观测网络,如地球观测系统(EOS)、全球陆地观测系统(GTOS)、全球气候观测系统(ECOS)、全球海洋观测系统(GOOS)、全球环境监测系统(GEMS)和平流层变化探测网络(NDSC)等^[1]。全球和国家尺度有关地球系统变化的长期观测、监测与信息网络正在形成,给数字地球的产生提供海量信息的基础。然而,随着空间数据获取技术的不断发展,人们能获得越来越多的数据,例如 NASA 的 EOSDIS 项目在 10 年中以每秒 3~5M、每天数百 G 速度获取各种遥感信息,而现在则达到了每天就要接收 1000GB 以上^[2],而如此大量的数据的处理与分析就成为摆在人们面前一个严峻的问题。面对如此海量的数据,如何从中快速、有效地提取与应用目标直接相关的特征信息来服务于决策分析已经成为目前空间信息处理的瓶颈,本文从这个角度出发,提出了空间数据的深度计算与主动计算模式,并以基于特征的遥感信息提取和目标识别工作为例,对其理论和过

程进行了分析和验证。

2 空间数据模式与计算

2.1 空间数据基本单元与地理计算

传统的 GIS 与空间信息领域应用认为:空间数据(Spatial Data)是用以表示空间实体的位置、形状、大小及其分布特征诸多方面信息的数据^[2],一般是通过遥感、测量、GPS 定位及统计调查等手段获得以反映其空间现象和过程。对于这些以不同手段获得的数据,按照不同地理领域(陆地、大气、海洋)的空间数据载体的形态的不同,可基本上将其分为四大类:反映固态基质信息的陆地空间数据、反映液态基质信息的陆地水文空间数据与海洋流体空间数据和反映气态基质信息的大气流体空间数据。

通过研究表明,无论对于每种形态的空间数据,实际上都存在着基本单元,现代地理学辞典将之定义为地理基质(地理因子)在一定层次上的组合^[3],也就是说在单元内部有某种或某几种属性是相似或均质的,本文将之简称为基元。比如土地利用中的土地单元、景观学中的景观单元、陆地水文

收稿日期:2005-06-06;修回日期:2006-01-14。

资助项目:中国科学院地理科学与资源研究所知识创新工程领域前沿项目(CXI0G-D02-01);国家自然科学基金项目(40401039)。

作者简介:明冬萍,(1976-),女,博士研究生,研究方向为遥感图像智能处理、遥感信息提取技术等。E-mail:mingdp@reis.ac.cn

学中的水文响应单元,以及海洋中的水团、大气中的气团等。尽管人类的肉眼看不见,但是这些基元是客观存在的。它们在陆地、大气、海洋各种空间数

据的表现形式和其计算分析所涉及的学科领域如表 1 所示。

地学计算是应用计算机技术处理地学相关问

表 1 空间数据地学表现形式

Tab.1 Geo-representation form of spatial data

空间数据	载体表现形态	基元	主要获取手段	数据模型类型	涉及学科领域	典型实例
陆地空间数据	固态	土地单元、景观单元、遥感特征基元等	遥感、测绘	连续数据(多采用对象模型,即矢量、栅格)	地理学、地质学等	土地覆盖、地质与矿藏分布等
	液态	水文响应单元、流域单元等	水文站、遥感、测绘	离散数据与连续数据(多采用对象模型,栅格)	地理学、水文学、流体力学等	河流、水库等水文现象
海洋流体空间数据	液态	水团	遥感、海洋水文台站、船测、浮标、海流计	离散数据与连续数据(多采用场模型)	海洋学、流体力学等	海洋环流、海洋矿产及渔业分布等
大气流体空间数据	气态	气团	遥感、气象台站	多为离散点数据(多采用场模型)	大气学、物理、化学等	大气成分监测及风雨云等气象

题^[6]。一直以来, GIS 中的地学计算主要侧重于图形数据的存贮、检索和分析^[7]。Openshaw 和 Abrahart 在 1996 年将“地学计算”描述为利用计算机技术,尤其是利用那些非传统的聚类分析等方法对自然和人文相关地学问题进行相应的数据处理、分析与解释过程,并通过相应的数据处理分析等挖掘手段来达到地学知识发现,进而为人们更好地认识和解决地学问题提供依据^[8];随着空间数据处理与应用技术的不断发展,“地学计算”也有了更广泛的涵义,Longley(1998)认为“地学计算”还包括了空间数据分析、空间动态建模、可视化以及时空动力学的相关问题^[9]。

本文将地学计算外延为空间计算,即通过对空间数据的处理,提取与应用目标相关的特征信息,并融合形式化的领域知识,建立决策模型,如神经网络、元胞自动机、决策树和遗传算法等,为地理学模拟和分析,以解决复杂的和不确定性问题^[10]。

2.2 空间数据的计算模式

近年来,随着对地学方法论研究的不断重视,关于空间格局与空间过程以及其间的关系的研究也不断地被强调^[12-14],因此,将格局与过程关系的研究作为重要的观点和技术路线^[13]。开展空间数据计算模式的研究旨在为空间计算问题提供一般理论方法上的指导,为大型地学问题的研究和解决提供强有力的支持。

在目前空间计算研究所采用的模型大多都是

对数据本身的计算,而空间对象之间的语义关系、内部关系以及各种基于环境背景的知识往往被忽视,这极大地影响了空间计算在较高层次上的地学分析和辅助决策能力^[15]。实际上,对于相互关系(包括语义关系)的忽视自然是对地理本质规律的忽视,同时,忽视语义关系会使我们在已有的认知水平上对原本为有机整体的地理世界进行僵硬的分割,从而导致基于这种认识空间计算与分析显得被动^[17]。

基于这种知识在计算过程中的参与与否或者参与形式与程度上的差别,并根据空间数据计算行为及计算的侧重点的不同,本文将空间数据计算过程分为深度计算过程与主动计算过程。图 1 为空间深度计算与空间主动计算的一般过程示意图。

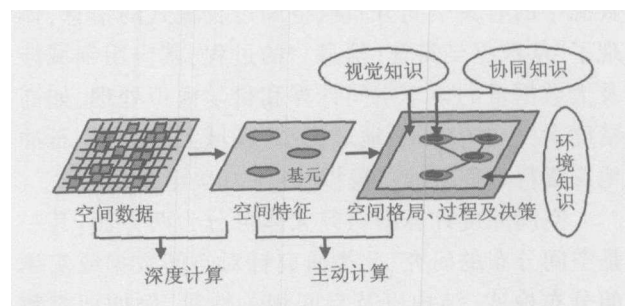


图 1 空间数据计算模式示意图

Fig.1 Computing pattern of spatial data

从图 1 可以看出,空间深度计算主要是通过大量的计算来提取数据特征,从而从海量数据中提取出基元,体现了“从数据到特征(信息)”的过程。后

面的处理与分析也是基于这些基元来揭示隐藏在空间数据资源中的有关空间分布或空间过程模式的信息。这是由于一般的数据处理与计算过程中的原始数据量都较大,直接基于这些原始数据或采用知识对这些原始数据进行决策过程均不可行,所以必须要先对原始数据进行处理,提取出地理基质一致的基本单元,其计算量相对原始数据的直接计算来说要小得多。因此空间深度计算的任务主要是提取特性均质的基元,为空间主动计算的决策分析做准备。

而空间主动计算主要强调计算过程的主动性与智能性,即将大量的地学知识和专业模型融入基于地理空间单元的计算分析过程,体现了“从特征(信息)到知识(决策)”的过程,即由计算到决策的过程,以提高系统在知识层次上的综合与分析能力。在通过深度计算取得数据所对应的基元的基础上,我们需要对这些特征进行进一步的处理与转化,完成最终的决策过程,这一步骤是通过知识辅助来完成,进而指导相应的决策过程。

(1)空间深度计算

深度计算(Deep Computing)的概念最早源于IBM公司的“深蓝”超级计算机^[18],Henry等人将深度计算定义为采用超级处理器或先进分析软件计算系统解决复杂的、大规模模型与分析问题的计算过程^[19]。在各种领域的空间数据的计算与分析中,基本都存在最小的计算分析单元,如景观学中的景观单元、土地学中的土地单元、大气中的气团和海洋中的水团^[17]等。空间深度计算主要是通过对这些基本地理单元的大量的计算来揭示隐藏在空间数据资源中的有关空间分布或空间过程模式的信息,体现了“从数据到特征(信息)”的过程,其应用领域涉及大数据量的地学空间计算和科学模拟处理,如高精度大气环流模拟、地球模拟、区域天气预报、石油地震资料处理与油藏模拟、流体力学计算等^[20-22]。

空间深度计算的研究大体可分为两方面:其一是空间分布的研究,此类研究针对空间数据或实体的分布模式、结构以及空间相关性等,例如地震数据中对地震条带的提取^[23,24];其二是空间过程方面的研究,空间过程的研究则侧重考察时间属性下空间分布的变化过程以及相应的规律,常常涉及到复杂系统与空间动力学等相关理论与方法的介入,如对土地覆被变化(LUCC)的历史过程与未来趋势的分析模拟^[25,26]等。

(2)空间主动计算

主动计算(Active Computing),目的在于能够有效自主地监督业务流程^[27]。这种计算的特点是能够在一定知识的辅助下,主动地去完成人们所需要完成的事情,具有一定的智能性。

由主动计算的定义可以看出,相对于深度计算来说,主动计算并不侧重在大数据量、大计算量的计算本身,而更强调计算过程的主动性与智能性。这一计算过程并不是我们所理解的数据处理、分析过程,而是对网络上的人们的行为的总结与消化的过程,并根据大量的网络操作对其进行知识的总结与归纳,用以对下次相同或相近的任务处理时,进行辅助决策功能,具有智能性、移动性与分散性,体现了协同决策功能,单个网络结点上的空间数据计算过程可以简单用图2表示^[28-30]。如图2所示,主动计算实现方式同样涉及大量的算法过程,可以利用人工智能、智能体和知识表达和网络服务等方法技术通过训练学习和推理的方式进行实现^[31,32]。

一般来说,空间对象是人类知识结构与表达的基本单元,也构成了空间主动计算的基本特征单元,这种基本特征单元往往是通过空间深度计算而得到的(如图1所示)。通过对特征单元或单元组合(单元群)及其相互关系的分析,以及其与环境背景关系的研究,结合具体领域知识,我们能够实现对地理现象的空间分布格局与空间行为过程的描述,因而空间主动计算过程体现了“从特征(信息)到知识”的过程。在实际网络操作中,这种知识参与后的空间数据计算模式不断被总结与归纳,并在下次进行相同或相近的任务处理时提供辅助决策功能。

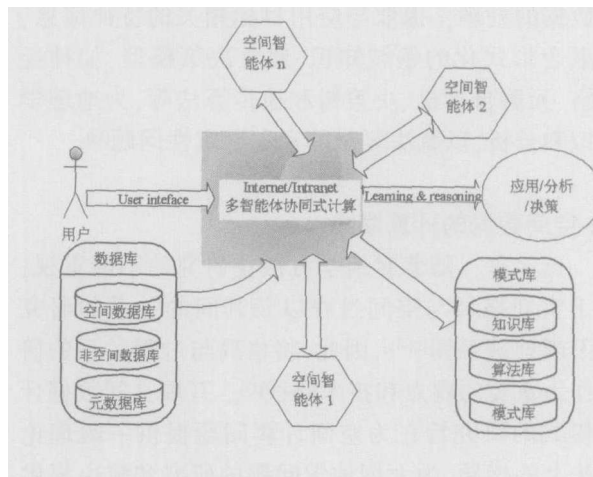


图2 单个网络结点内部空间数据主动计算模式示意图
Fig.2 Computing pattern of spatial data in a web node

3 实例结果与分析

针对高分辨率遥感信息提取与目标识别问题,提出了基于特征基元的影像计算方法体系(图3),来印证前述关于空间数据深度计算与主动计算模式的理论。

面向特征单元的影像信息提取和目标识别的基本流程是从影像-大尺度单元-基本单元-特征单元-复杂目标的识别过程。其中从影像到特征基元是特征挖掘的深度计算过程;主要利用计算机的高性能计算能力自下而上地提取一定尺度下,具有一定空间聚合度的单元,一般来说是通过影像分割或边缘检测来实现;从单元到目标则主要依赖于形式化知识的引导,通过主动计算行为,自上而下地将单元组合为空间的排列,从而形成复杂目标;主要是通过基元的光谱特征、形状、大小、纹理等特征计算,在领域知识的支持下,构建相应的目标库和知识库,计算基元与邻接对象之间的空间关系,通过基于协同语义知识等各种推理机,进行相关的推理活动,实现对目标甚至复杂目标对象的识别,并揭示各类目标对象的空间格局及空间关联等知识,完成主动计算的过程。详细过程可参见文献^[34-36]。

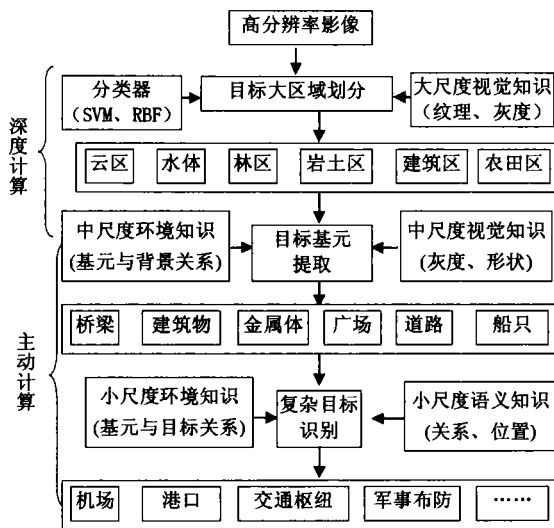


图3 基于特征基元的高分辨率遥感影像多尺度信息提取技术路线

Fig.3 Work flow of parcel based multi-scale information extraction from high resolution RS images

实验所采用的数据是某机场的 0.7m 分辨率的 QUICKBIRD 全色影像,大小为 3545×3265,见图 4

(a)。按照上面技术路线在分布式环境下^[6]进行块状基元和线状基元提取,结果分别如图 4(b)和图 4(c)所示;再进一步利用基于知识的特征进行约束,提取出建筑物(主要为机库)、金属体(飞机)和跑道(为矢量图层,详见参考文献[34]);基于地理本底、空间距离等知识协同判断和综合上述步骤提取到的结果,得到最终的机场矢量图,完成机场提取如图 4(f)所示。

4 结语

(1) 根据计算行为模式及计算的侧重点的不同,将地学计算过程分为深度计算(海量空间数据的地理单元提取等)与主动计算(基于地理单元的计算分析等),体现了“数据→特征→知识”的一般计算过程。进而将为大型地学问题的研究和解决提供强有力的支持。

(2)海量数据的处理与分析所需要的高性能计算技术(分布式和并行计算技术);此外,知识的形式化表述和协同知识推理技术作为主动计算的核心问题,其实现的技术手段在目前研究中还没能很好地设计和实现,有待进一步研究。

(3)在高分辨率遥感信息提取和目标识别具体研究中,深度计算过程中特征基元提取的精度和效率是决定目标识别效率和精度的基本前提,更是主动计算能否高效顺利进行的保障,因此设计新型的以及改进现有的特征基元提取算法是今后研究的另一重点。

参考文献

- [1] Shashi Shekhar, Sanjay Chawla 著, 谢昆青, 马修军, 杨冬青等译. 空间数据库. 北京: 机械工业出版社, 2004.
- [2] Spatial Data Concepts and Issues. http://www soi.city.ac.uk/~dk708/pg2_2.htm (Access: May 2005).
- [3] 肖庆辉. 二十一世纪地球科学展望. <http://www.cgs.gov.cn/NEWS/Geology%20News/20020806/20020806a.htm>
- [4] Ning An, Liujian Qian, Anad Sivasubramaniam et al. Evaluating Parallel R-Tree Implementations on Network Workstations. Technical Report CSE-98-006, Pennsylvania State University, 1998.
- [5] 左大康主编. 现代地理学辞典. 北京: 商务印书馆, 1990.
- [6] Camara G, Monteiro A M. Geocomputation techniques for spatial analysis: are they relevant to health data? Cad Saude Publica. 2001 Sep-Oct;17(5):1059-1071 (<http://epi>).

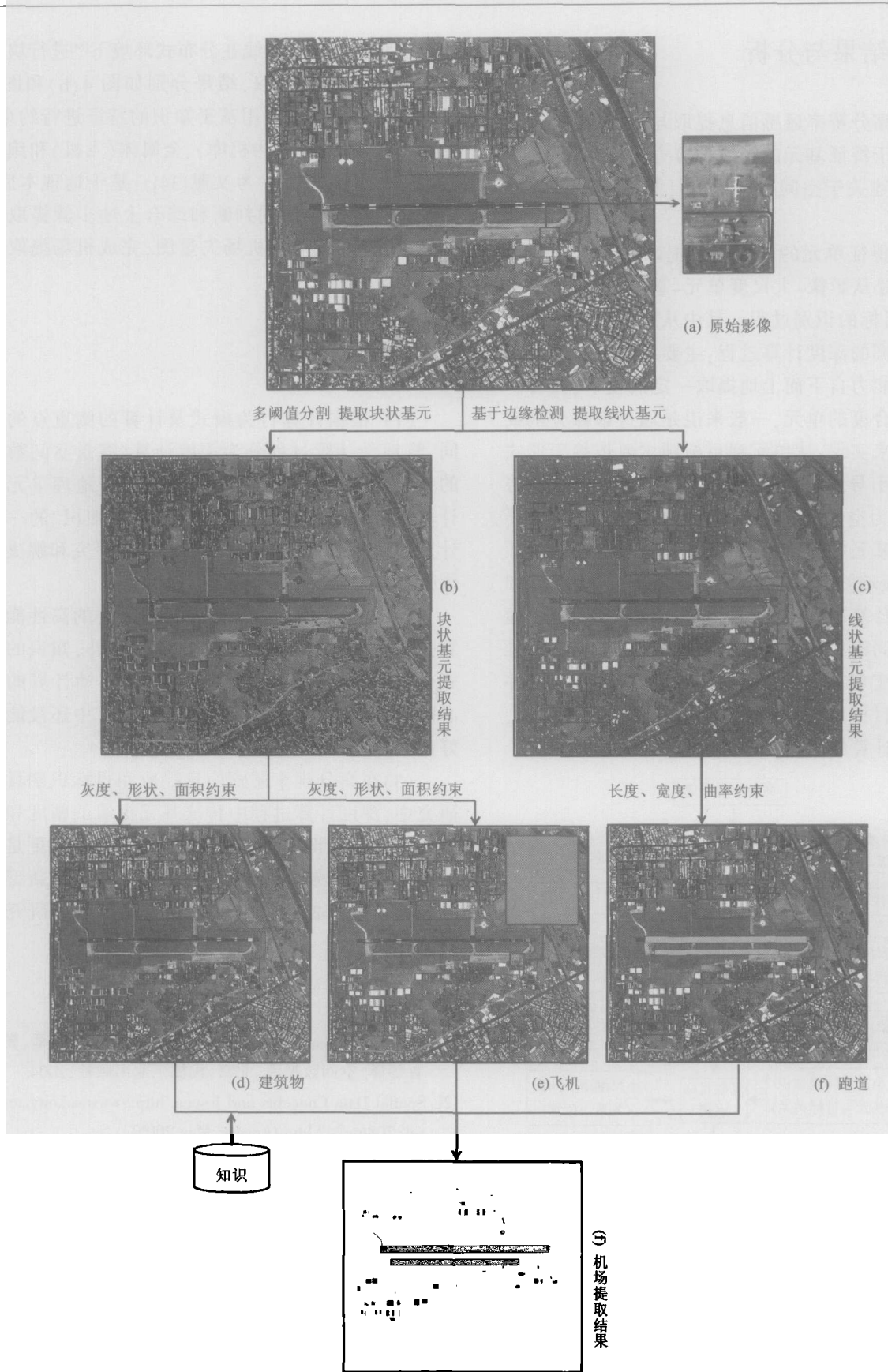


图 4 基于特征的高分辨率遥感影像机场识别

Fig.4 Airport recognition from high resolution remote sensing images based on parcels

- minsai.cl/SigEpi/doc/geo.pdf) (Access: May 2005).
- [7] M Egenhofer, J Glasgow, O Gnther, J Herring, D Peuquet. Progress in computational methods for representing geographical concepts. *International Journal of Geographical Information Science*, 1999, 13, 775~796.
- [8] Openshaw S, Abrahart R J. Geocomputation. In: *Proceedings of the 1st International Conference on GeoComputation* (R. J. Abrahart, ed.). Leeds: University of Leeds, 1996, 665~666.
- [9] Longley P. *Geocomputation: A Primer*. New York: John Wiley and Sons, 1998.
- [10] Gahegan M. On the application of inductive machine learning tools to geographical analysis. *Geographical Analysis*, 2000, 32(2): 113~139.
- [11] 王咏武, 王咏刚. 道法自然——面向对象实践指南. 北京: 电子工业出版社, 2004.
- [12] 郑 度, 陈述彭. 地理学研究进展与前沿领域. *地球科学进展*, 2001, 16(5): 599~606.
- [13] 陆大道, 蔡运龙. 我国地理学发展的回顾与展望——地理学: 方向正在变化的科学. *地球科学进展*, 2001, 16(4): 467~472.
- [14] 葛全胜, 吴绍洪, 朱立平, 张雪芹. 21 世纪中国地理学发展的若干思考. *地理研究*, 2003, 22(4): 406~415.
- [15] Tang A Y, Adams T, Usery E L. A spatial data model design for feature-based geographical information systems. *Int. J. Geographical Information Systems*, 1996, 10(5): 643~659.
- [16] 肖乐斌, 钟耳顺, 刘纪远, 宋关福. GIS 概念数据模型的研究. *武汉大学学报: 信息科学版*, 2001, 26(5): 387~392, 418.
- [17] 李天峻. 分布式地理空间对象模型研究, 北京大学博士学位论文, 1997.
- [18] I Wladawsky-Berger. Turning points in information technology. *IBM Systems Journal*, 1999, 38(2&3): 449~452.
- [19] Henry T, Petri M, Tomi S. *Computationally Efficient Methods for Deep Computing (DeepC)*. 2001.
- [20] IBM Research, Deep Computing Institute. <http://www.research.ibm.com/dci/applications.shtml>. (Access: May 2005).
- [21] Keiko Takahashi et al. Development of nonhydrostatic coupled ocean-atmosphere simulation code on the earth simulator. *Proc. of 7th International Conference on High Performance Computing and Grid in Asia Pacific Region*, July 2004.
- [22] <http://ct.gsfc.nasa.gov/ctreqreport.pdf>. Report from the Earth Science Enterprise Computational Technology Requirements Workshop, 2002. (Access: May 2005).
- [23] 秦承志. 面向点数据的地质数据可视化分析. 中国科学院地理科学与资源研究所.[博士学位论文] 2003.
- [24] 汪 闽. 空间聚类挖掘方法研究. 中国科学院地理科学与资源研究所.[博士学位论文] 2003.
- [25] 刘纪远, 于贵瑞, 王绍强, 岳天祥, 高志强. 陆地生态系统碳循环及其机理研究的地球信息科学方法初探. *地理研究*, 2003, 22(4): 397~405.
- [26] 王绍强, 许君等. 土地覆被变化对陆地碳循环的影响——以黄河三角洲河口地区为例. *遥感学报*, 2001, 5(2): 142~148.
- [27] Want R, Pering T, Tennenhouse D. Comparing autonomic and proactive computing. *IBM Systems Journal*, 2003, 42(1): 129~135.
- [28] 罗英伟. 基于 Agent 的分布式地理信息系统研究. 北京大学博士论文, 1999.
- [29] IEICE Communication Society. *IEICE Communications Society GLOBAL NEWSLETTER*, 2003, 6. <http://www.ieice.org/cs/GNL/GNLvol6.pdf> (Access: May 2005).
- [30] Chao Tang, Li Da-xu, Shan Feng. An agent-based geographical information system. *Knowledge-Based Systems*, Volume 14, Issues 5-6, August 2001, 233~242.
- [31] 毛新军, 王怀民, 陈火旺等. 多 Agent 系统的非终止主动合作计算模型. *计算机研究与发展*, 1999, 7: 769~775.
- [32] 张遂征. 主动计算环境 (ACE) 构造及其关键技术研究. *计算机应用与软件*, 2000, 46~51.
- [33] 明冬萍, 骆剑承, 沈占锋, 汪 闽, 盛 昊. 高分辨率遥感影像信息提取与目标识别技术研究. *测绘科学*, 2005, 30(3): 18~20.
- [34] 明冬萍, 骆剑承, 周成虎等. 高分辨率遥感影像信息提取及块状基元特征提取研究. *数据采集与处理*, 2005, 20(1): 34~39.
- [35] 沈占锋. 分布式影像计算模型及其在高分辨率遥感目标识别中的应用研究. 中国科学院地理科学与资源研究所博士学位论文, 2005.
- [36] 郑 江. 并行遥感图像信息提取与分析方法研究. 中国科学院地理科学与资源研究所 博士学位论文, 2004

The Preliminary Study of Spatial Data Computing Pattern Based on Basic Unit

MING Dongping¹, LUO Jiancheng¹, ZHOU Chenghu¹, SEHN Zhanfeng¹, LI Junli¹, SEHNG Hao¹

(1 The State Key Laboratory of Resources and Environmental Information System, IGSNRR, CAS, Beijing 100101, China;

2 School of Remote Sensing Information Engineering, Wuhan University, Wuhan 430079, China)

Abstract: With the developments in satellite sensor technology, data acquisition technology developed rapidly; and with the start of a series of space-based observation network for Earth science, such as EOS, GTOS, ECOS, GOOS etc., high performance processing and analysis of tremendous data becomes the bottleneck faced by us. According to the shape differences between different data carriers of terrene, ocean and atmosphere, this paper divides spatial data into four classes: terrestrial-solid based spatial data, terrestrial-liquid based spatial data, marine-floating based spatial data and atmospheric-floating based spatial data. Then this paper introduces the concept of the basic unit and proposes their actually existing style in the four types of spatial data mentioned above. Furthermore, this paper simply reviews geocomputation and expands it to geo-spatial computation. Then this paper discusses the connotation and classification of geo-spatial computation and summarizes the general computing procedure: data → features → knowledge. According to the differences of the computational behavior and the computing emphasis, this paper divides geo-spatial computation into two classes: deep-computation and proactive-computation. Consequently, this paper explains the computing pattern of deep-computation and proactive-computation. What's more, a case study of information extraction and target recognition from remote sensing image based features was done to illustrate and testify the ideas mentioned above. Finally, this paper summarizes the relative problems about spatial data computation and expects the direction of future researches.

Key words: spatial data; computing pattern; terrene; hydrology; ocean; atmosphere; remote sensing