

基于地球剖分的空间信息区位标识^①

吕雪锋^{②*} 程承旗^{③***}

(* 民政部国家减灾中心 北京 100124)

(** 民政部灾害评估与风险防范重点实验室 北京 100124)

(*** 北京大学工学院航空航天信息工程研究所 北京 100871)

摘 要 针对空间信息的区域范围和位置标识的统一性和多尺度性问题,基于 2^n 一维整型数组的全球经纬度剖分格网(GeoSOT)框架提出了一种具有全球统一性、多尺度性和结构性的空间信息区位标识编码模型。该区位标识编码模型通过将球面固定格网剖分的空间定位、空间尺度、空间范围和空间结构集于一体的三元组 (C_0, M, N) 来对同一地理对象的区域范围与位置、内部空间区域格网、空间点线面实体、空间属性信息、空间信息网络域五种区位信息进行统一描述与一致性表达。通过对区位标识编码定义、生成方法及应用方法的综合设计,可实现同一地理区域的多元空间信息之间形成一致性区位关联能力,为全球多元空间信息的统一管理和应用提供统一地理编码参考。

关键词 空间区位标识,地球剖分,离散格网,空间信息标识,空间信息组织

0 引 言

随着对地观测技术的发展,当前空间信息综合应用体系下的空间信息获取能力得到了大幅提高,空间数据已逐步呈现多源、多尺度、多时相、全球覆盖和高分辨率等特征,数据量快速增长,形成了 GB 级、TB 级、PB 级的发展趋势,但各类空间信息的组织方式或标准在尺度、空间、编码等方面存在不一致,如基础地理信息数据按照地形图的图幅进行组织,自然灾害的灾情信息按照行政区划进行组织,卫星遥感数据按照轨道条带或分景进行组织,且不同卫星轨道条带或分景数据的划分标准及其标识存在不一致,因而如何更加有序、高效地组织与管理全球海量性和多源异构性的空间数据并形成统一的标准(基准、尺度、时态、语义),以实现跨部门、跨行业多元空间信息的快速共享、整合与分发,已经成为空间信息科学领域和行业应用部门广泛重视的研究课题^[1-4]。空间信息区位标识作为空间信息标识的重要组成部分之一,是统一有序地组织与管理空间信息的一个重要环节,是提高空间信息的获取、利

用、共享、整合效率与水平,消除语义冗余与异构,提高空间信息组织的标准化程度的重要地理编码基础。本文研究了基于地球剖分的空间信息区位标识,采用基于 2^n 一维整型数组的全球经纬度剖分格网(geographic coordinate subdivision grid with one-dimension-integral coding on 2^n -tree, GeoSOT),建立了一种将地理对象的空间尺度、空间位置、空间范围和空间几何形态集于一体的空间信息统一区位标识编码模型,以实现在不同空间尺度下对同一空间信息区位的一致性地理编码和语义描述,将同属一个地理区位的空间数据、地理对象与空间信息存储资源形成一致性区位关联,为多元空间信息的统一区位组织提供地理编码参考。

1 相关知识

在本文中,空间信息区位标识的基本概念是指用地理编码来表达空间信息的区域范围和位置,即对空间信息的区域范围和位置标识的统称。通过对空间信息区域范围和位置的统一地理编码标识,可将局部的与整体的乃至全球的空间信息形成一致性

① 973 计划(61399)和海洋公益性行业科研专项(201305020)资助项目。

② 男,1979 年生,博士,助理研究员;研究方向:空间信息组织与管理研究;E-mail: jackson. Lu@126. com

③ 通讯作者,E-mail: ccq@pku. edu. cn

(收稿日期:2013-07-23)

区位关联。

目前空间信息区位标识的方法大致可以分为基于坐标(经纬度)的区位标识、基于图幅图廓的区位标识和基于地理空间格网的区位标识^[5-7]。其中,基于坐标(经纬度)的区位标识是后两者的地理坐标基础。

基于坐标(经纬度)的区位标识方法通常采用坐标串的方式来表示某一地理区域的位置及范围,即用地理对象的中心点 (λ_0, φ_0) 来描述地理对象的空间位置信息,用地理对象的外围轮廓点 $((\lambda_1, \varphi_1), (\lambda_2, \varphi_2), \dots, (\lambda_n, \varphi_n))$ 来描述其区域范围。但这种描述方式存在一些问题:一是不同的人选择的中心点及轮廓点往往不同,造成在实际应用中同一地理对象的区位标识不统一;二是同一地理对象在不同空间尺度上的坐标描述方式也存在不一致,即同一个地理对象在小比例尺地图上用一坐标点来标识,在大比例尺地图上通常用一串坐标点来标识,两者标识编码不统一。基于图幅图廓的区位标识方法的地形图分幅体系最具有代表性,其通常适用于测绘图幅的区位标识,但空间多尺度区位标识能力有限,目前难以满足地理对象(空间实体)及遥感影像数据块的区位标识编码要求。基于地理空间格网的区位标识方法因受目前一些格网的划分与特殊应用背景相关(行政区划、人口、环境等)的影响,难以满足全局或全球统一规格区位标识的要求,另外在实际操作中不能直接利用格网单元的地址编码对“数据块”进行层次关联与邻近索引。

而在对全球地理空间的统一多尺度固定划分方面,地球剖分提供了一种最具有潜力的解决方案。其基本思想是将地球表面剖分为面积或形状近似、空间无缝无叠、尺度连续的多层次面片(或格网单元),通过对每个剖分面片进行有序的地理时空递归编码,使得大到地球、小到厘米的面片都有唯一的地址编码,从而用每个剖分面片对应的地址编码代替地理坐标在球面上进行各种操作。由于不同层次的每个剖分面片在记录空间位置信息的同时也记录了比例尺和精度信息,这为全球多尺度空间信息或数据的统一表达与管理提供了格网基础^[8-20]。

基于地球剖分的空间信息区位标识研究就是在地球剖分框架下,利用剖分面片及其地址编码来构建空间信息区位标识的统一编码模型,以解决同一地理区域与所对应的空间信息在其组织应用中的尺度、空间、编码等不一致问题,为多元空间信息的统一区位组织提供重要参考依据。本文采用基于 2^n

一维整型数组的全球经纬度剖分格网(GeoSOT)来建立具有全球统一性、多尺度性和结构性的空间信息区位标识编码模型^[22]。通过这套空间信息区位标识编码模型能够对地理对象的区位信息及与其相关联的多元空间信息进行统一描述和一致性区位关联,从而对同一地理区域的各种空间信息进行一致性区位检索与关联调度,以更有利于跨部门、跨行业多源空间数据的区域性整合与分布式系统的共享应用。

2 空间信息区位标识的格网基础

本文从空间信息的组织角度出发,结合我国测绘体系的大量研究成果与全球遥感数据的组织形式,以及现有测绘、遥感数据的集成和主要空间信息组织格网或分幅标准,选择剖分格网 GeoSOT 作为空间信息区位标识的基础格网。

GeoSOT 通过三次地球经纬度空间扩展(将经纬度坐标空间扩展为 512° ,将 1° 扩展为 $64'$,将 $1'$ 扩展为 $64''$),得到一组间隔为 $\{2^9, 2^8, 2^7, 2^6, 2^5, 2^4, 2^3, 2^2, 2^1, 2^0, 2^{-1}, 2^{-2}, 2^{-3}, 2^{-4}, 2^{-5}, 2^{-6}, 2^{-7}, 2^{-8}, 2^{-9}, 2^{-10}, 2^{-11}\}$ 的经纬网剖分体系,实现了整度、整分、整秒的整型二叉树格网剖分,如图 1 所示。每个剖分面片地址编码的四进制 1 维编码形式为 Gdddddmmmmmmssssssuuuuuuuuuuuu,其中 d、m、s、u 分别表示度级、分级、秒级及秒以下编码位,分别取值为 0、1、2、3 的四进制数^[21]。

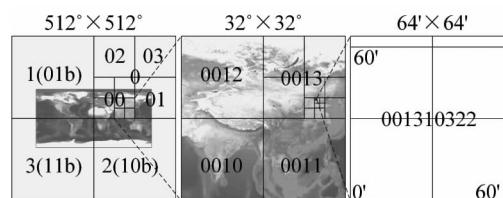


图 1 GeoSOT 剖分格网示意图

基于 GeoSOT 框架建立空间信息区位标识的适应性主要体现在以下几个方面:

(1) GeoSOT 的第 n 级与第 $n+1$ 级之间的父子面片的面积之比大致都为 4:1,满足均匀变化,形成了上至全球、下至厘米级面片的全球二叉树格网结构。

(2) 与美国国家航空航天局(National Aeronau-

tics and Space Administration, NASA) 的科学数据浏览平台 WorldWind、Google 公司的数字地球 Google Earth 的经纬度格网体系相比^[22-25], GeoSOT 具有经纬度整度、整分、整秒的整型特性,能够一致性聚合测绘、气象、海洋及世界各国主要的规格地理格网产品,而 WorldWind 与 Google Earth 的格网不具备这种能力。

(3) GeoSOT 编码将度、分、秒及秒以下格网进行分段编码,使得整度、整分、整秒的规格数据在区位标识过程中,不存在舍入误差,数据可以精确表示。例如经纬度坐标 $20^{\circ}13'35''$, 采用 GeoSOT 二进制 1 维编码时可以精确编码,即 $20^{\circ}13'35'' = 000010100 - 001101 - 100011$; 而 WorldWind、Google Earth 的格网编码由于采用直接的二进制剖分编码,在表达整度、整分、整秒的规格数据时,存在舍入误差。

(4) GeoSOT 编码具有四进制 1 维、二进制 1 维、二进制 2 维与十进制 2 维四种编码形式,能与计算机处理模式及经纬度语义形成一致,各进制编码之间能够进行快速转换,转换关系见图 2。其中,四

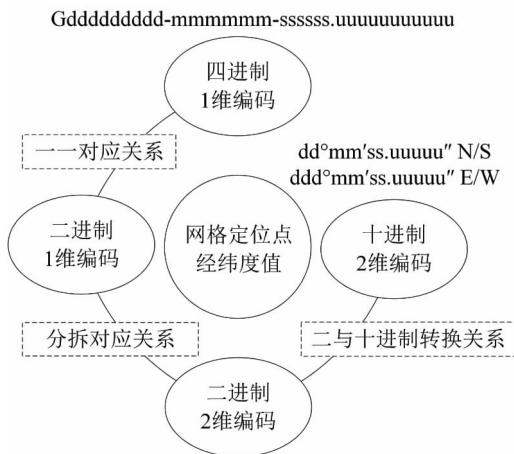


图2 GeoSOT 剖分格网编码的各进制转换关系

进制 1 维编码可用于建立空间索引,其首位数字编码区分四个半球;二进制 1 维和 2 维编码可认为是四进制 1 维编码在计算机内的存储形式,可用于计算机内部存储与空间运算;十进制 2 维编码与经纬度坐标形式一致,用于人机交互,将十进制 2 维编码直接转换为二进制时,便是 GeoSOT 的二进制 2 维编码。

3 GeoSOT 区位标识的编码模型设计

3.1 GeoSOT 区位标识的编码设计思路

基于地球剖分的空间信息区位标识主要涉及两个层面的问题:一是对于地球上任一空间信息,如何利用剖分面片来表示空间信息的区域范围和位置;二是在此基础上如何利用剖分面片的地址编码来记录和标记空间信息的结构信息。那么,在设计 GeoSOT 区位标识时,可以将空间数据的区位、地理对象的区位、空间信息存储资源所管理的业务区划范围都看作地理对象的多边形区位来统一标识,建立一个将地理对象的空间尺度、空间位置、空间范围和空间几何形态集于一体的统一区位标识编码。

但 GeoSOT 是作为一种对全球地理空间的固定划分,在对地理对象进行区位标识时,由于其固定不变性,若直接使用单个剖分面片编码进行区位标识可能会存在以下问题:(1)有些地理对象的区位就不会恰好落在单个面片内;(2)即使落在单个面片内,也可能存在冗余过大的问题;(3)单个面片不足以表达地理对象的空间几何形态。因此,在对地理对象进行区位标识时,仅采用单个面片是不够的,还需通过多个面片的聚合表达来适应复杂地理对象的区位标识。

如图 3 所示,在 GeoSOT 区位标识的编码模型中,可以用 GeoSOT 剖分层级来表示地理对象的空间尺度特征,用剖分面片的位置和大小来表示地理对象的空间位置特征和空间范围特征,用剖分面片的聚合形态来表示地理对象的空间几何形态特征。

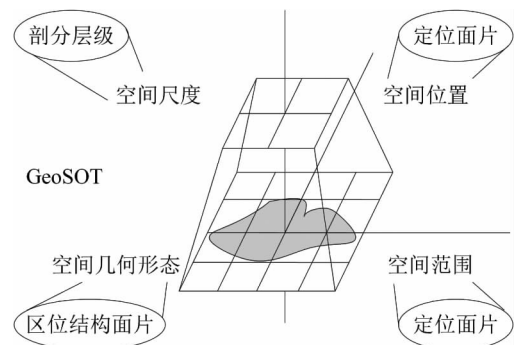


图3 GeoSOT 区位标识编码模型的设计思路

3.2 GeoSOT 区位标识的编码定义

根据 GeoSOT 区位标识编码模型的设计思路, GeoSOT 区位标识编码模型可以采用以地理对象区

位结构面片中的某个角点面片为位置向量,以沿纬向的面片跨度和沿经向的面片跨度为尺度向量的聚合编码,将空间定位、尺度、范围和结构集于一体,即用三元组 (C_0, M, N) 来表示,如图 4 所示。图中 C_0 为角点定位面片编码,用来标识地理对象的基准面片,相当于坐标原点且坐标的方向与角点定位面片所在的位置保持一致; M 为沿纬向的面片跨度, N 为沿经向的面片跨度, M 和 N 取值范围为 $\{1, 2, 3, \dots, n\}$,当 $M=1$ 和 $N=1$ 时,即纬向和经向面片跨度均为 1 时, $(C_0, 1, 1) = C_0$ 。

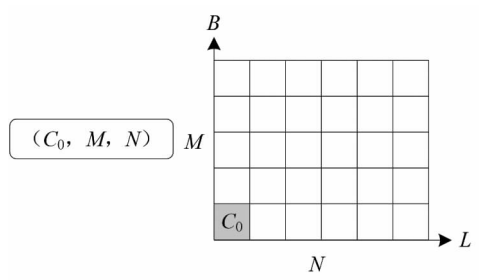


图 4 GeoSOT 区位标识编码模型的定义

对于角点定位面片 C_0 ,各半球区域数据的角点定位面片与 GeoSOT 的定位角点方位保持一致,如图 5 所示。

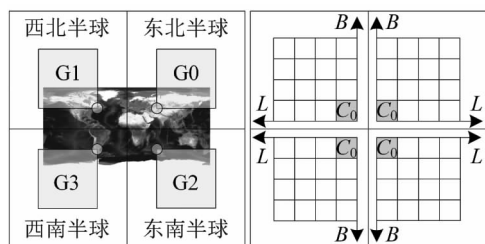


图 5 角点定位面片的确定方法

在 GeoSOT 区位标识编码模型 (C_0, M, N) 下,对于栅格和矢量地图数据, C_0 为左下角点定位面片编码, M 和 N 为单一元素,即 M 表示沿纬向的面片个数, N 表示沿经向的面片个数,如图 6(a) 所示。对于卫星分景数据,其四角点的经纬度坐标连线与经纬度网线虽有一定夹角,但其夹角是可计算的,所以仍可以卫星分景数据的最小外包矩形左下角点所在的面片为角点定位面片,再分别记录纬向面片个数 M 和经向面片个数 N 。对于不规则多边形区域,可确定其最小外包矩形左下角点所在的面片为角点定位面片 C_0 ,再沿纬向方向 M 记录每行面片的起止位置 (j, k) ,且起止位置之差即为该行沿经线方向上的面片跨度 N ,如图 6(b) 所示。

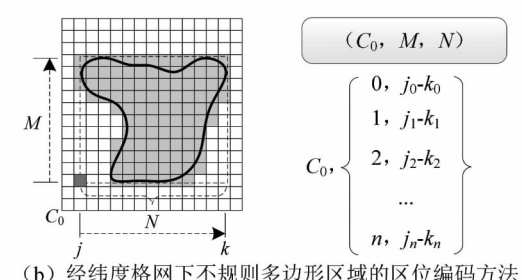
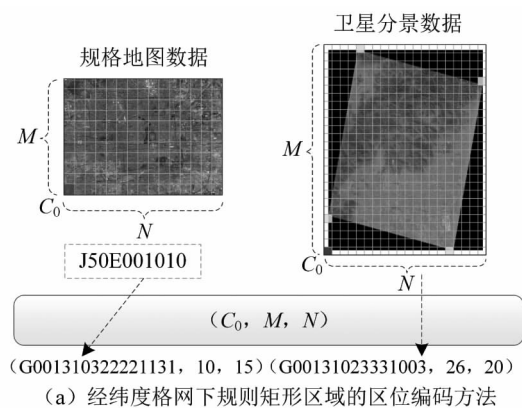


图 6 多边形区域的 GeoSOT 区位标识编码方法

3.3 GeoSOT 区位标识的编码应用方法

在 GeoSOT 区位标识的编码定义下,可以从空间区位索引和区位结构表达两个层面来建立空间信息区位标识的编码应用方法。其中,面向空间区位索引的应用层主要根据地理对象的空间尺度、位置及范围信息确定地理对象是落在哪个或哪几个剖分面片内,以便于对空间信息进行快速的空间定位和区位检索;面向区位结构表达的应用层是利用特定尺度的剖分面片来聚合地理对象的空间几何形态,它是在空间区位索引应用层内对地理对象区位的更进一步精细描述,以满足地理对象的不同应用需求。

以某幅地形图比例尺为 1 : 5 万、图幅编号为 J50E001010 的遥感影像产品或某个规格的地理区域范围为例,它的西南角点的经纬度坐标为 $(39^{\circ}50'0'', 116^{\circ}15'0'')$ 。如图 7 所示,在空间区位索引应用层,它的 GeoSOT 区位标识编码为 $(G00131032222, 1, 2)$;在区位结构表达应用层,它的 GeoSOT 区位标识编码可为 $(G001310322221131, 10, 15)$,即以左下角点面片 G001310322221131,以 1'面片(第 15 级剖分格网)纬向面片跨度 + 10',经向面片跨度 + 15';在用户界面显示时,它的 GeoSOT 区位标识编码可以按照区位结构表达应用层对应的 GeoSOT 剖分格网大小、左下角点面片的十进制 2 维编码、纬向面片个数和经向面片个数格式来表示,即 $(1', 39^{\circ}50' N, 116^{\circ}15' E, 10, 15)$ 。

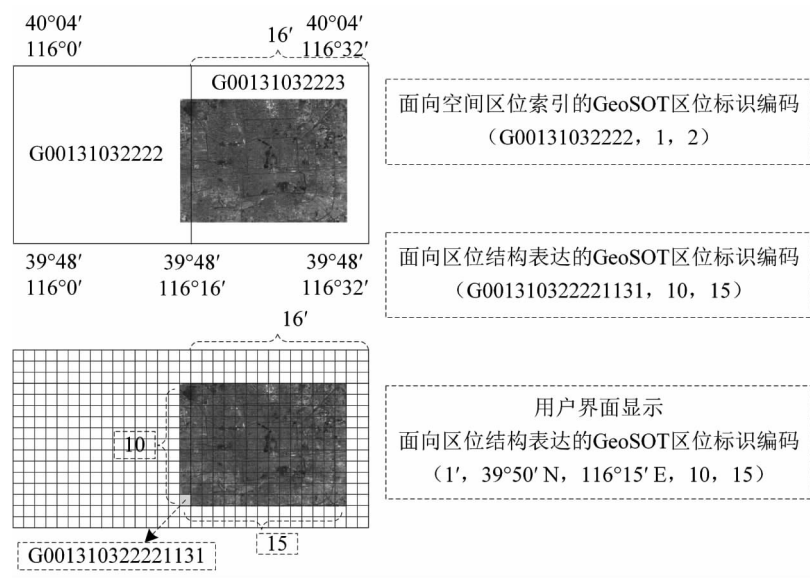


图 7 我国 1 : 5 万地形图幅矩形范围的区位标识

对于不规则的多边形区域,以北京市西郊机场为例,机场的空间区位和描述空间区位的经纬度坐标点如图 8 所示,它的最小外包矩形范围大小为 $50.58'' \times 1'33.77''$ (经差 $\times$ 纬差)。如图 9 所示,在空间区位索引应用层,它的 GeoSOT 区位标识编码为 (G0013103222311,2,1);在区位结构表达应用层,若采用第 19 级 GeoSOT 剖分格网来对机场区域进行区位标识,它的 GeoSOT 区位标识编码为

(G001310322231103122,{(1,9-10),(2,9-10),(3,8-10),...,(24,8-9)}),即以左下角点面片 G001310322231103122,以 4"面片在纬向面片总跨度 + (24 $\times$ 4)",经向面片最大跨度 + (13 $\times$ 4)",用户界面显示的区位标识编码为 (4",39°56'44" N,116°15'48" E,24M,13N)。其中,M 与 N 用于标识不规则多边形区位的纬向面片总跨度和经向面片的最大跨度值。

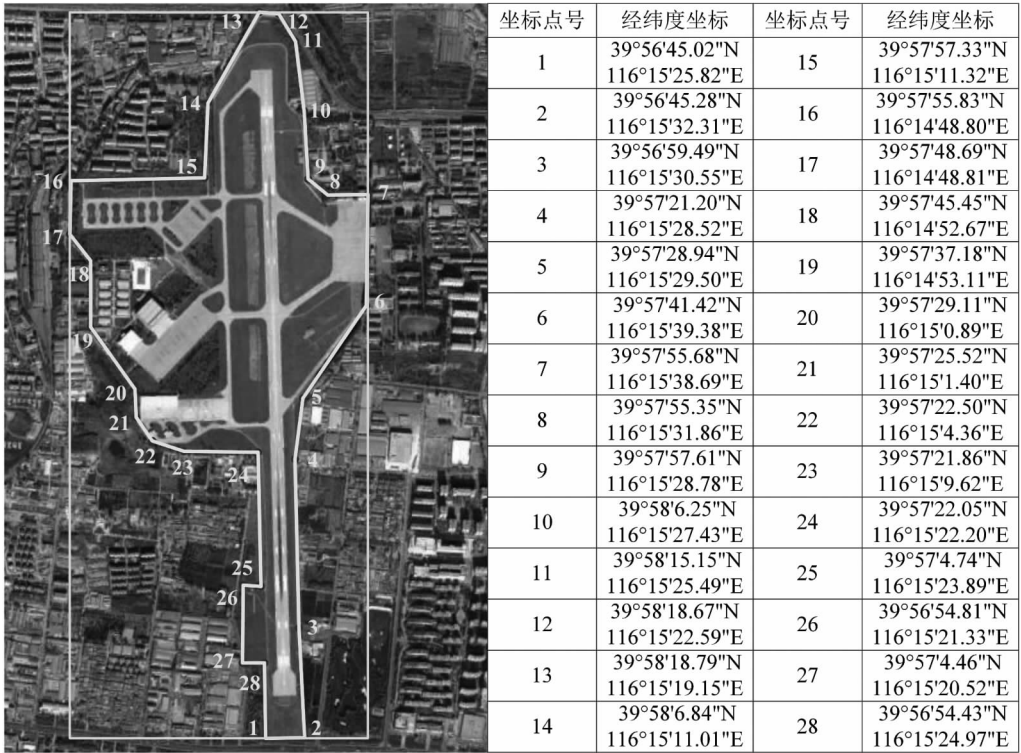


图 8 北京市西郊机场的空间区位与描述空间区位的经纬度坐标点示例

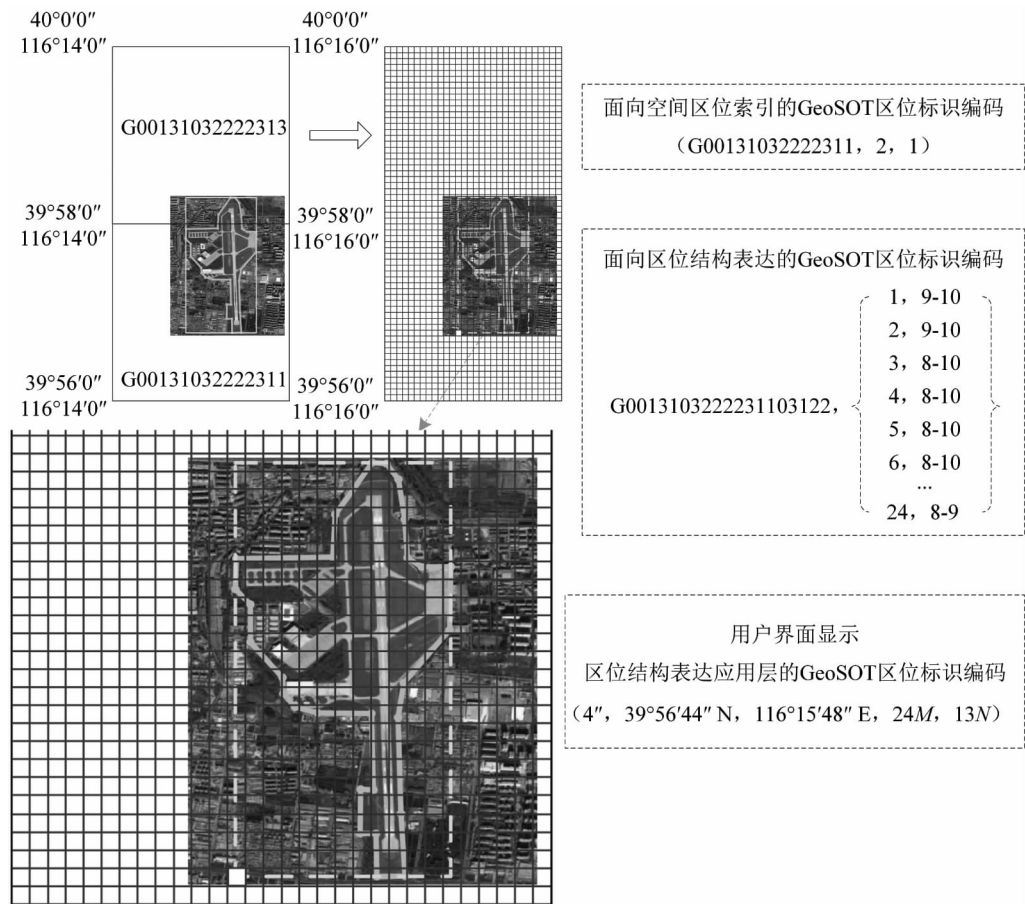


图9 北京市西郊机场的 GeoSOT 区位标识编码示意图

3.4 GeoSOT 区位标识的编码生成方法

(1) 面向空间区位索引的区位面片生成方法

在确定面向空间区位索引的 GeoSOT 区位标识时,选择合适的面片层级是最先需要考虑的问题。为了使得地理对象的空间尺度与 GeoSOT 区位标识的面片尺度相当,我们选择地理对象的最小外包矩形长边作为面片层级的选择依据,即它最接近哪个剖分面片的大小,就将该剖分面片所在的剖分层级作为 GeoSOT 区位标识的面片层级。如图 10 所示,以经纬度弧度为单位,设地理对象最小外包矩形的长度为 D_L , 宽度为 D_W , GeoSOT 的剖分层级为

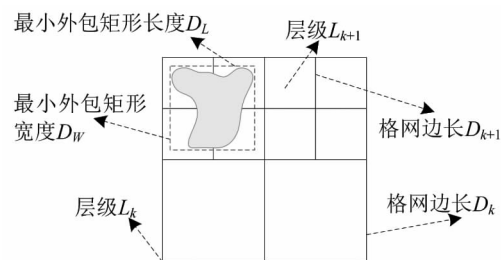


图10 面向空间区位索引的面片层级生成方法

$L_k (k \in [0, 32])$, 其对应的 GeoSOT 格网边长为 D_k ; 若 $D_L \geq D_W$, 当 $D_k < D_L \leq D_{k+1}$ 时, 区位标识的面片层级 $L_A = L_{k+1}$ 。

确定 GeoSOT 区位标识的面片层级之后, 需要判断地理对象是落在哪个或哪几个剖分面片内。在面向空间区位索引的区位面片层级中, 无论地理对象的复杂程度如何, 确定地理对象所对应相当尺度的剖分格网层级后, 它最多会落在不超过 4 个剖分面片内。那么, 在空间区位索引应用层, 确定地理对象的 GeoSOT 区位标识的面片位置和个数的具体方法为: 将地理对象的最小外包矩形的四角点经纬度坐标转换为第 L_A 层级中所对应的剖分面片编码 C_1, C_2, C_3, C_4 (从左下角点开始 Z 序顺序), 判定它们是否同属一个剖分面片。如图 11(a) 所示, 如果 $C_1 = C_2 = C_3 = C_4$ 时, 则地理对象占据 1 个区位面片, 其面片编码 $C_{ref} = C_1 = C_2 = C_3 = C_4$; 如果 $C_1 = C_3$ 且 $C_2 = C_4$ 或 $C_1 = C_2$ 且 $C_3 = C_4$, 则地理对象占据 2 个区位面片, 如图 11(b) 所示; 如果 $C_1 \neq C_2 \neq C_3 \neq C_4$, 则地理对象占据的区位面片个数为 3 或 4, 此时需要根据地理对象区域的经纬度坐标点以剖分边界

线 A 和 B 来判断区位面片的个数,如图 11(c) 和图 11(d) 所示。

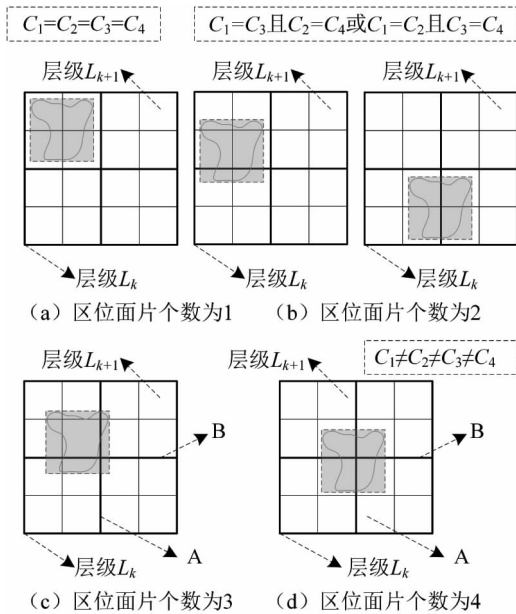


图 11 面向空间区位索引的区位面片个数确定方法

在上述面向空间区位索引的区位面片生成中,地理对象与区位面片关系存在一对多的映射,但某些应用需要地理对象与面片之间满足一一映射关系时,也可以选择地理对象中心点所在的剖片面片或者地理对象主体所在的剖片面片作为面向空间区位索引的区位面片。如图 12 所示,在确定面向空间区位索引的区位面片层级 L_A 后,一是可以根据地理对象的经纬度坐标计算其区域范围的中心点坐标 (λ_0, φ_0) ,将该中心点坐标转换为第 L_A 层级中所对应的剖片面片编码 C_0 ,即面片 C_0 为地理对象中心点所在的区位面片;二是以地理对象最小外包矩形在哪个区位面片中面积占优,则选择哪个区位面片。

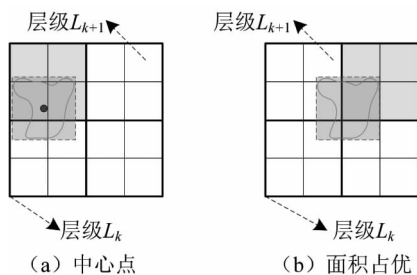
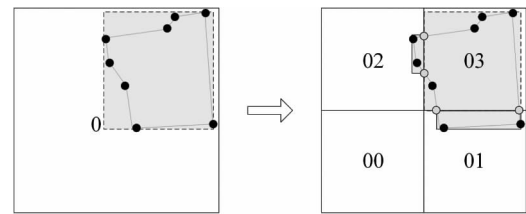


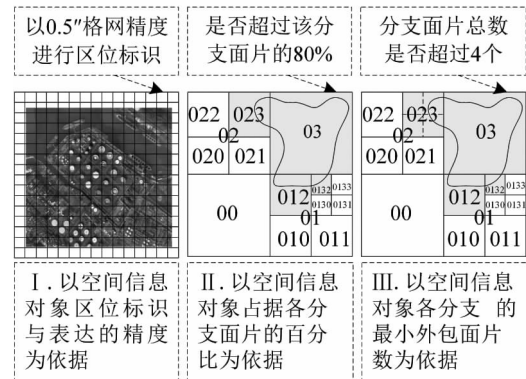
图 12 以中心点或面积占优法确定区位面片

(2) 面向区位结构表达的区位面片生成方法
在确定面向区位结构表达应用层的 GeoSOT 区

位标识时,我们采用逐层分支定界剖分的方式来确定地理对象的区位面片。如图 13(a) 所示,当确定面向空间区位索引的区位面片层级 L_A 后,可逐层向下四叉分支,地理对象都被分割为多个多边形;然后再在各个分支面片内,结合新插入点的经纬度坐标来计算各分支内多边形的最小外包面片;当满足某个条件后便停止向下继续剖分,在各个分支中其剖分层级位于最底层的最小外包面片为地理对象区位标识的最小尺度面片。而在分支定界剖分中,可以采用三种不同的判定截止准则,如图 13(b) 所示。



(a) 面向区位结构表达的逐层分支剖分示意图



(b) 区位结构面片标识层级的通用判定依据

图 13 面向区位结构表达的区位面片
逐层分支定界剖分

第一种是利用地理对象的精度、比例尺、分辨率等客观信息来确定剖分的下限,这种方式主要针对影像、地形图图幅等空间数据,例如将遥感数据的分辨率与某个应用标准地图的比例尺最高精度相对应,进而与能够一致性聚合该标准地图范围的 GeoSOT 剖分层级相关联。

第二种是以地理对象分支所占当前剖片面片的比例作为该分支的截止条件,例如是否超过该分支面片的 80%。

第三种是利用地理对象分支后,其剖片面片的数目作为截止条件,例如分支面片总数是否超过 4 个。后两种方式主要是针对各种不规则区位的地理对象。

4 GeoSOT 区位标识对地理对象的区位表达

在 GeoSOT 区位标识编码模型下, GeoSOT 区位标识编码能够对于同一地理对象的区域范围与位置、内部空间区域格网、空间点线面实体、空间属性信息区位及空间信息网络域进行统一描述与一致性表达, 形成一致性区位关联。

4.1 地理对象的区域范围和位置描述

对于地理对象的区域范围大小和在什么位置, 可以利用 GeoSOT 区位标识编码 (C_0, M, N) 包含的剖分面片尺度、角点定位面片的位置及区位面片的跨度实现区域范围的计算、区域角点的定位、不同尺度信息关联以及中心点面片位置的确定。其中, 角点位置的定位可由角点定位面片编码 C_0 来描述, C_0 的面片编码长度对应着 GeoSOT 剖分层级及该层级的格网大小 $Cell$, 其区域范围大小可由 $(M \times Cell, N \times Cell)$ 来描述; 对于地理对象的中心点面片, 可以由 $(\lfloor M/2 \rfloor + 1, \lfloor N/2 \rfloor + 1)$ (规则矩形区位) 或 $(\lfloor M/2 \rfloor + 1, j + \lfloor (k - j)/2 \rfloor)$ (不规则多边形区位) 来确定中心点面片的位置。

例如, 对图 7 中图幅编号为 J50E001010 的 1:5 万地形图幅矩形区域, 其面向区位结构表达应用层的 GeoSOT 区位标识编码为 $(G001310322221131, 10, 15)$, 它的区域范围大小为 $(10 \times 1', 15 \times 1')$, 中

心点面片的位置为 $(6, 8)$, 中心点面片编码为 G001310322230332。

4.2 地理对象的内部空间区域格网描述

对于地理对象的内部空间区域格网描述, 可由 GeoSOT 区位标识编码 (C_0, M, N) 直接来确定在空间域中地理对象的内在格网组成、格网尺度大小和区域范围大小, 并建立剖分格网与地理对象影像数据块之间的面片映射关系。其中, 角点定位面片编码 C_0 的编码长度对应着 GeoSOT 剖分层级及该层级的格网大小, 其可以描述格网尺度大小; 纬向和经向区域格网范围大小可由 (M, N) 来描述。

以图 7 中图幅编号为 J50E001010 的 1:5 万地形图幅矩形区域为例, 在区位结构表达应用层, 它的内部空间区域格网组成可由以 1' 面片 (第 15 级剖分格网)、纬向面片个数为 10、经向面片个数为 15 来描述。并且, 对于每个内在格网单元 $C_i \in (C_0, M, N)$ 都可以建立区位结构面片与实体影像数据块之间的区位关联关系。

4.3 地理对象的空间点线面实体描述

在 GeoSOT 区位标识编码 (C_0, M, N) 下, 可以利用多层次 (不同精度) 的剖分面片和递归性的面片地址编码对不同空间尺度的空间实体区位进行统一、多尺度描述。如图 14 所示, 以北京市西郊机场为例, GeoSOT 区位标识编码可以对空间点状实体、线状实体及面状实体进行剖分面片表达。

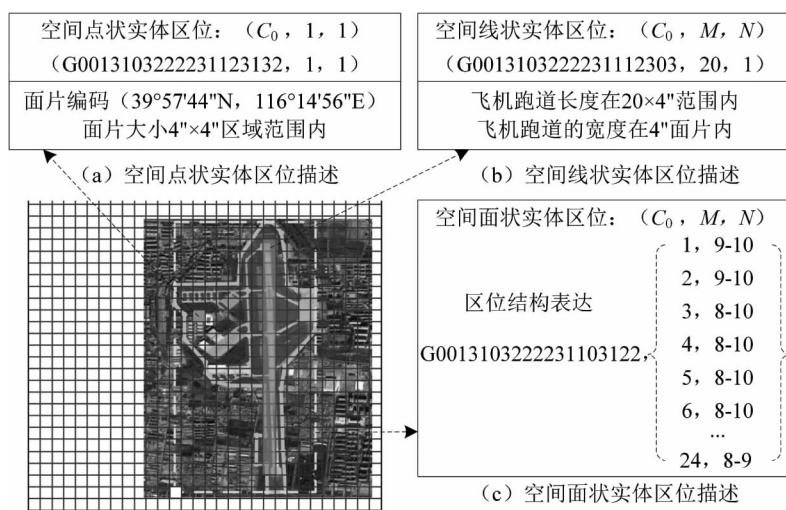


图 14 空间点线面状实体的 GeoSOT 区位标识编码描述

(1) 对于空间点状实体描述, 它的 GeoSOT 区位标识编码为 $(C_0, 1, 1)$, 即用 1 个区位面片编码 C_0

既可描述点状实体的空间位置, 同时也可在一定精度下表达点状实体的区域范围。

以图 14 中西郊机场内某个台站(经纬度坐标为 $39^{\circ}57'45.38''\text{N}$, $116^{\circ}14'58.01''\text{E}$) 为例,就可以用 1 个区位面片编码 G0013103222231123132 (19 位四进制或 38 bit) 来表达,即空间位置为面片编码 ($39^{\circ}57'44''\text{N}$, $116^{\circ}14'56''\text{E}$),面片大小 $4'' \times 4''$ 区域范围内。并且,对于 1 个 GeoSOT 区位标识编码 C_0 ,当选择第 32 级剖分面片(在赤道附近大致尺度为 1.5cm 格网)时,其编码长度为最长,即 64 位二进制 (64bit)。而在经纬度坐标体系下表示该空间点状实体时需一个坐标对 (λ_0, φ_0) ,其占用 2 个 double 类型 (128bit);另外,1 个经纬度坐标对仅代表一个逻辑点,若需要表达该点状实体的区域范围,还需要另加区域范围大小或经纬度坐标对来表达。

(2)对于线状实体,以图 14 中西郊机场内的飞机跑道为例,其面向区位结构表达的 GeoSOT 区位标识编码为 (G0013103222231112303, 20, 1),即以定位面片编码 G0013103222231112303,沿着纬向面片个数为 20,飞机跑道的长度在 $20 \times 4''$ 范围内,飞机跑道的宽度在 $4''$ 面片内。

(3)对于面状实体,例如图 14 中整个西郊机场区域,可以利用面向区位结构表达的 GeoSOT 区位标识编码 (C_0, M, N) 来沿着纬向方向的面片跨度 M ,依次通过位于经向起止位置 j 和 k 的面片聚合表达西郊机场的空间几何形态结构。同时在大比例尺地图尺度,整个西郊机场区域可被抽象为一个点 $(C_0', 1, 1)$,且区位结构面片编码 C_0 是 C_0' 的子集。

4.4 地理对象的空间属性信息区位描述

对于地理对象的空间属性信息的区位描述,如果是针对地理对象的整体,比如整个西郊机场的属性,那么它的区位标识编码采用的就是西郊机场在空间区位索引应用层的区位标识编码;如果空间属性的描述是针对地理对象中的某一处细节,例如西郊机场中的一个油库,那么它的属性数据的区位标识编码采用的就是在西郊机场的区位结构表达应用层中该油库所占据的区位结构面片编码。

例如图 15 中的西郊机场,描述该机场的空间属性信息的 GeoSOT 区位标识编码为 (G00131032222311, 2, 1),若对于西郊机场内某个台站,其在机场的区位结构表达应用层中的区位结构面片编码为 G0013103222231123132,那么描述该台站的空间属性信息的 GeoSOT 区位标识编码为 (G0013103222231123132, 1, 1)。

在上述这种空间属性信息描述中,我们通过角点定位面片编码的继承关系,可以看出这个油库是

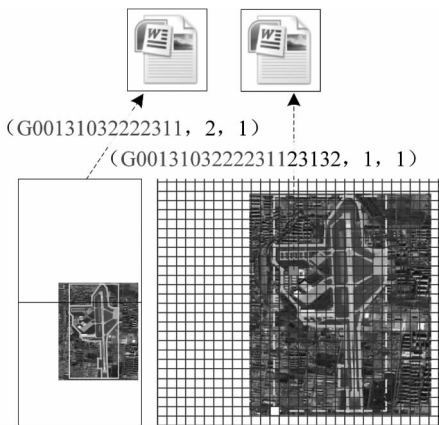


图 15 地理对象的空间属性信息区位描述

包含在机场面内的,是它的一个子集(具体空间区位包含关系的计算方法见 4.7 小节)。而在目前经纬度区位标识体系下,将描述某个地理对象的空间属性信息存入数据库时,则用该地理对象的 ID 来标识其空间属性信息;不同的系统对同一地理对象的 ID 标识可能存在不一致,这就需要多层元数据来解译,而采用 GeoSOT 区位标识编码法可以有效地规避这一问题。

4.5 地理对象的空间信息网络域描述

对于地理对象的空间信息网络域描述,可在现有计算机网络空间地址协议基础上嵌入 GeoSOT 区位标识编码,构建具有地理涵义的空间信息存储资源网络空间地址,即 GeoIP 地址。由此通过对空间信息存储资源所管理的业务区划范围进行 GeoSOT 区位标识,以便于对属于同一地理对象的空间信息存储资源进行动态调度。

(1)GeoIP 地址编码结构。考虑到当前互联网运行环境及发展情况^[26,27],GeoIP 地址的编码结构采用 IPv4 和 IPv6 地址的嵌入方式,如图 16 所示。

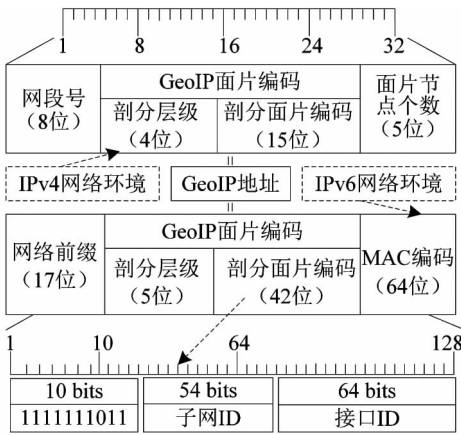


图 16 GeoIP 地址编码结构图

在 IPv6 网络环境下,GeoIP 地址编码主要由网络前缀、GeoIP 面片编码和主机 MAC 地址编码构成。在 IPv4 网络环境下,GeoIP 地址编码主要由网段编号、GeoIP 面片编码和该区位标识所对应的存储单元个数构成。其中,GeoIP 面片编码即为 GeoSOT 区位面片标识编码,由剖分层级编码和对应的剖分面片编码构成。依据 GeoSOT 格网框架,第 21 层级剖分面片的空间范围为 $1'' \times 1''$,其对应赤道附近边长大约为 32m 的区域范围,可以满足实际应用中空间信息存储组织的最小尺度需求,因此,对应 IPv6 网络环境,GeoIP 面片编码的剖分层级可设计到第 21 级,面片编码长度 42 位二进制。而对于 IPv4 网络环境,由于其 IP 地址本身的长度限制,设计到第 15 级(面片编码长度 15 位二进制),即 $1' \times 1'$,对应赤道附近边长约为 2km 的区域范围。

(2)在 GeoIP 地址设计中,每台主机(计算机服务器)的 GeoIP 地址都是唯一的,且 GeoIP 地址的 GeoIP 面片编码与主机都是一对多的映射关系,即一个业务区划可以对应多台计算机服务器。这种一对多的映射关系主要是通过面片节点个数和主机的 MAC 编码(全球唯一)来实现。另外,在进行 GeoIP 面片编码时,均采用 GeoSOT 编码的二进制 1 维编

码。其中,在 IPv4 地址结构中,15 位的剖分面片编码可按照“分-度”的格式(ssssssssssss-dddddddddddddddd)嵌入;在 IPv6 地址结构中,42 位的 GeoIP 面片编码可按照“秒-分-度”格式(uuuuuuuuuuuu-ssssssssssss-dddddddddddddddd)嵌入,其度级编码与 IPv6 地址中第 64 位对齐。

4.6 地理对象的区位标识编码关联结构

根据上述 5 种区位标识信息的描述,可以建立地理对象的五元素关联结构,将地理对象的区域范围检索数据模型、影像数据模型、对象数据模型、网络域调度模型及属性数据模型形成一致性区位标识编码关联。

如图 17 所示,区域范围检索数据模型可建立面向空间区位索引的 GeoSOT 区位标识编码与元数据记录之间的区位关联关系,影像数据模型可建立区位结构面片与影像数据块(行列号)之间的区位关联关系,对象数据模型可建立区位结构面片与对象内部区位结构之间的关联关系,网络域调度模型可建立面向空间区位索引的 GeoSOT 区位标识编码与 IP 地址之间的区位关联关系,属性数据模型可建立空间属性信息的区位面片与属性描述文件之间的区位关联关系。

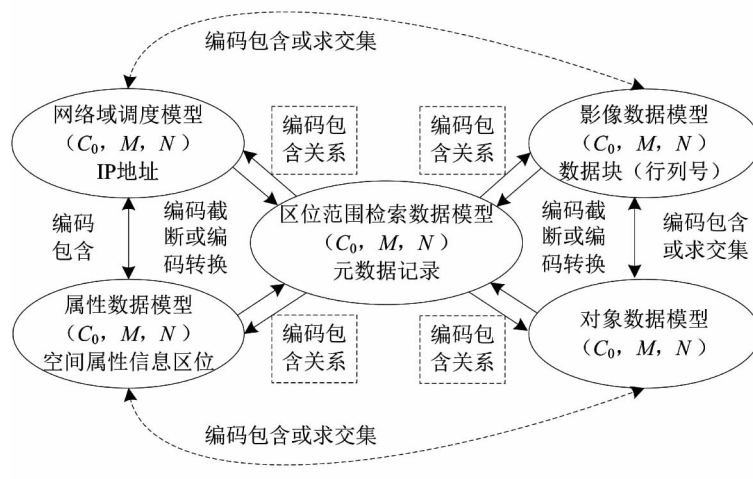


图 17 GeoSOT 区位标识的编码关联结构

4.7 地理对象的空间区位包含关系计算方法

针对上述地理对象的 5 种区位信息之间的区位关联,假设地理对象 A 的 GeoSOT 区位标识编码为 (C_A, M_A, N_A) ,地理对象 B 的 GeoSOT 区位标识编码为 (C_B, M_B, N_B) ,基于 GeoSOT 区位标识编码的空间区位包含关系计算方法可分为以下两种情况。

(1)当 $M = N = 1$ 时,即对于单个 GeoSOT 区位

面片 $(C_0, 1, 1) = C_0$ 时,如图 18 所示,GeoSOT 区位面片 A 包含面片 B 的关系判定方法为:设区位面片 A 的四进制 1 维编码 C_A 的编码长度为 L_A ,B 的四进制 1 维编码 C_B 的编码长度为 L_B ;如果 $L_A < L_B$,且面片编码 C_B 的前 L_A 位编码与面片编码 C_A 的编码相同,则区位面片 A 包含面片 B,即 $A \Leftrightarrow B \Leftrightarrow L_A < L_B$ 且 $C_A \oplus C_B(1, L_A) = 0$ 。其中, $C_B(1, L_A)$ 表示区位

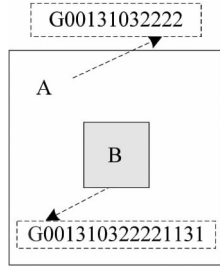
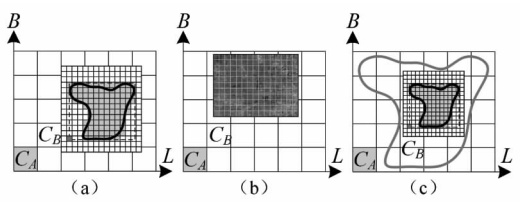


图 18 单个 GeoSOT 区位面片包含关系示意图

面片编码 C_B 中从度级编码位的第 1 位开始到第 L_A 位之间的编码。若 $L_A = L_B$ 且 $C_A \oplus C_B = 0$ 时,则区位面片 A 与区位面片 B 为同一面片。

(2) 当 $M > 1$ 或 $N > 1$ 时,若设地理对象 A 和 B 的区位标识编码的剖分层级间隔数 $n = L_A - L_B$, C_A 与 C_B 的四进制 1 维编码长度的差值 $L = \max(L_A, L_B)$, M_{A_max} 、 M_{B_max} 表示纬向面片总跨度值, N_{A_max} 、 N_{B_max} 表示经向面片最大跨度值,那么,若在同一剖分层级(以 C_A 与 C_B 的剖分层级较高者为参考层),地理对象 B 的角点定位面片 C_B 的纬向编码 $C_{B_纬向编码}$ (格网的行号)、经向编码 $C_{B_经向编码}$ (格网的列号)及纬向面片总跨度值与经向面片最大跨度值都在地理对象 A 角点定位面片 C_A 的纬向编码 $C_{A_纬向编码}$ (格网的行号)、经向编码 $C_{A_经向编码}$ (格网的列号)及纬向面片总跨度值与经向面片最大跨度值的范围内时,即满足条件 $0 \leq (C_{B_纬向编码}(1, L) - C_{A_纬向编码}(1, L)) \leq (M_{A_max} \cdot 2^{(1nl-n)} - M_{B_max} \cdot 2^{(1nl+n)})$ 且 $0 \leq (C_{B_经向编码}(1, L) - C_{A_经向编码}(1, L)) \leq (M_{A_max} \cdot 2^{(1nl-n)} - M_{B_max} \cdot 2^{(1nl+n)})$ 。

若地理对象 A 的空间区位为规则矩形区位,地理对象 B 的空间区位为规则矩形区位或不规则多边形区位时,通过该条件即可判定 $A \Leftrightarrow B$,如图 19 (a)和(b)所示;当地理对象 A 的空间区位为不规则多边形区位时,则以区位标识编码层级较高的地理对象为参考层,依次沿着它的纬向和经向面片跨度进行面片位置比较判断包含关系。例如图 19 (c)所

图 19 当 $M > 1$ 或 $N > 1$ 时地理对象的区位包含示意图

示,若 $n < 0$,则以 C_A 为参考点,在 (C_B, M_B, N_B) 中沿着纬向面片跨度 M_B 依次计算位于经向位置 j 和 k 的面片编码 C_{j_B} 、 C_{k_B} 在 (C_A, M_A, N_A) 中 M_{j_B}' 、 N_{j_B}' 、 N_{k_B}' 值,并判断 (N_{j_B}', N_{k_B}') 是否在 (N_{j_A}, N_{k_A}) 范围内;若全部在,则 $A \Leftrightarrow B$ 。

5 GeoSOT 区位标识对空间数据的区位索引

5.1 以 GeoSOT 区位标识进行区位索引的实现

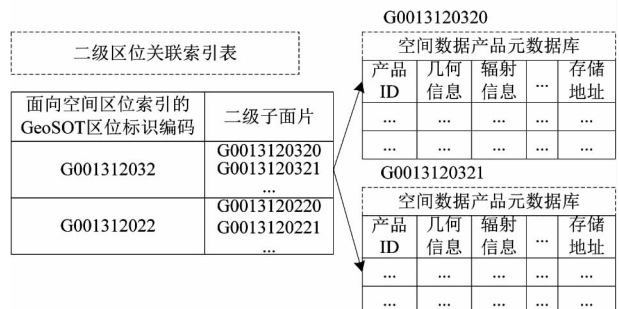
以空间数据的 GeoSOT 区位标识进行区位索引时,可在保持现有空间信息存储管理系统不变的情况下,在每条数据项中添加两个区位标识编码字段(面向空间区位索引的区位标识编码字段和面向区位结构表达的区位标识编码字段)或在其编目数据层之上增加一个面向空间区位索引的区位标识编码关联层,建立以面向空间区位索引的 GeoSOT 区位标识编码为检索字段的区位索引,如图 20 (a)和(b)所示。并且,在空间区位索引应用层,当空间数据的编目数据量不断增长或非常庞大时,可以按照各个 GeoSOT 区位标识的面片聚合和分裂特性形成多级区位面片关联索引,例如图 20 (c)所示。第一级为面向空间区位索引的 GeoSOT 区位标识编码与其子面片编码的关联索引,第二级为各子面片与空

空间数据产品元数据库							
数据 ID	几何信息	辐射信息	时间信息	...	存储地址	面向空间区位索引的区位标识	面向区位结构表达的区位标识
...	...	...	...	...	...	G0013120332	(C_0, M, N)
...	...	...	...	...	...	...	...

(a) 在空间数据产品元数据库的数据项中增加两个的区位标识编码字段

面向空间区位索引的区位标识编码	空间数据产品ID	空间数据产品ID
G0013120332	ID1, ID2, ID3...	ID1, ID2, ID3...
.....		

(b) 在空间数据的编目数据层之上增加一个 GeoSOT 区位标识编码关联层



(c) 基于 GeoSOT 区位标识的两级区位关联索引示意图

图 20 以 GeoSOT 区位标识进行区位索引的实现方式

间数据产品的区位关联索引,那么,在这种区位组织方式下,既可以选择现有关系数据库或空间数据库已有的字符串检索方法对 GeoSOT 区位标识编码进行编码匹配检索,也可以在此基础上建立空间区位索引方法。

5.2 以 GeoSOT 区位标识进行区位索引的构建

根据前文的分析,由于 GeoSOT 区位标识是基于球面空间的固定格网划分,在空间区位索引应用层,空间数据的 GeoSOT 区位标识的面片个数存在 1、2、3 或 4 个,即存在某一空间数据恰好落在一个与其空间尺度相当的剖分面片内或空间数据的区位横跨剖分边界线而占据多个同层级面片的情况。

对于第一种情况,GeoSOT 区位标识的面片个数为 1 时,若以此来建立空间数据的区位索引,在本质上就是建立基于格网的四叉树索引(Quadtree, Q-tree),并且,由于 GeoSOT 区位标识采用固定的、规格的、无缝无叠的空间划分方式,其索引空间的树结构也是固定的,索引结构与对象的插入顺序无关,不会导致索引空间的重叠,可以快速更新局部对象,减少查询过程中的无用路径,有效提升索引效率。

对于第二种情况,GeoSOT 区位标识的面片个数为 2、3 或 4 时,直接采用四叉树索引会带来冗余或索引分布的不平衡。因此,我们将 R-tree 的思想引入,设计了 Q&R-tree 结构。该结构以 Q-tree 索引为主体,将 R-tree 嫁接到 Q-tree 中,形成一棵混合的树型结构,如图 21 所示。其中,用 Q-tree 来管理其

GeoSOT 区位标识的面片个数为 1 的空间数据(后文称为第一类空间范围体)以及若干嫁接到该枝上的小 R-tree;而用 R-tree 来管理其 GeoSOT 区位标识的面片个数为 2、3 或 4 的空间数据(后文称为第二类空间范围体)。

(1)这种区位索引方式的核心是将一颗大的 R-tree 分解成若干小的 R-tree,将每一棵小的 R-tree 嫁接到 Q-tree 的对应枝节点上,这样就合成了 Q&R-tree。一方面,Q&R-tree 既保持了 Q-tree 固定空间索引的优势,又利用 R-tree 格网划分的自适应特性将那些横跨格网的对象进行有效管理;另一方面,Q&R-tree 在每个 Q-tree 节点上生成了一颗小的 R-tree,将整体复杂的 R-tree 分散在小范围内进行管理,降低了每一棵 R-tree 的复杂度,便于管理与更新。

(2)在空间查询的过程中,首先在 Q-tree 中查询第一类空间范围体,然后再查询该节点下的小 R-tree,进行第二类空间范围体的查询,最后将两类空间范围体的集合进行合并,就是最终的空间查询结果。并且,这两种查询方式,在效率上都可以得到保证^[28]。

(3)在更新方式方面,Q-tree 本身采用静态桶排序的方式,它的更新比较快,而 R-tree 族索引由于采用动态生成树的结构,它的更新和维护比较困难。Q&R-tree 采用局部小规模 R-tree 嫁接方式,它的更新也只是在局部范围内的数据更新,不影响其它 R-tree 的结构,由此可以有效地提升更新效率。

(4)在建立区位索引的过程中,对于第一类空间范围体,它的空间索引生成无需聚类,可直接确定它们与 Q-tree 节点格网的归属关系。对于第二类空间范围体,需要进行聚类以确定子空间的分布,但是 Q&R-tree 无需进行全局范围内的聚类,只需要在一个 Q-tree 的小空间内对第二类空间范围体进行局部聚类,这样就可以降低聚类的复杂度,并且不同小空间内的聚类是相互独立的,可以采用并行的方式快速生成,有效地提高索引生成的效率。

6 结 论

本文在 GeoSOT 地球剖分框架下提出了一种将地理对象的空间尺度、空间位置、空间范围和空间几何形态集于一体的空间信息统一区位标识编码模型,即 GeoSOT 区位标识编码。这种区位标识编码建立了固定的区位标识编码技术,同时具有较强对

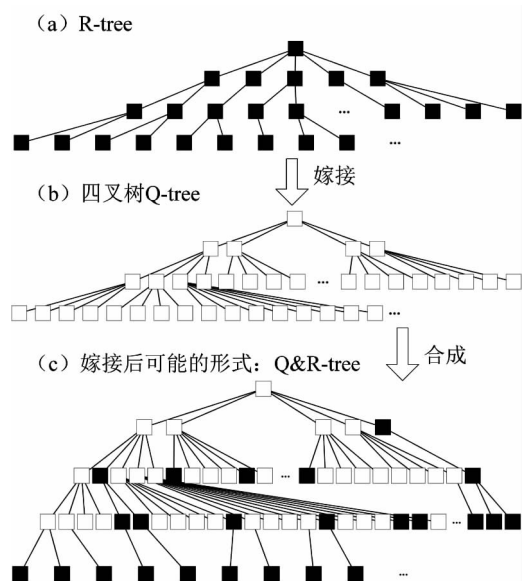


图 21 将 R-tree 嫁接到 Q-tree 中的混合结构示意图

复杂对象的区位标识能力。

GeoSOT 区位标识编码模型继承了经纬度坐标的精确定位能力,发展了将空间定位、尺度、范围和结构集于一体的多尺度剖分面片编码,实现了在不同空间尺度下对同一个空间信息区位的一致性地理编码和语义描述,将同属一个地理区域的空间数据、地理对象与空间信息存储资源形成一致性区位关联,为多元空间信息的统一区位组织提供了地理编码参考。同时,GeoSOT 区位标识编码模型的建立将为深化 GeoSOT 区位标识编码与各类空间区划关联、空间关系计算、遥感数据检索、点线面剖分数据模型、地理域名调度、剖分并行计算等方法研究提供模型基础。并且,对 GeoSOT 区位标识编码模型的研究,将有利于促进基于地球剖分的空间信息区位标识编码技术的发展,为多元空间信息的智能数据管理和服务提供空间信息区位编码理论和方法支持,为实现地理对象的全球唯一区位标识和多元空间信息的统一编码提供一种技术途径,同时也为影像地理信息系统的发展提供一种新型的数据模型。

参考文献

- [1] Li D R, Shao Z F. The new era for geo-information. *Sci China Ser F-Inf Sci*, 2009, 52(7): 1233-1242
- [2] 王家耀. 地图制图学与地理信息工程学科发展趋势. *测绘学报*, 2010, 39(2): 115-118
- [3] Li D R, Tong Q X, Li R X, et al. Current issues in high-resolution earth observation technology. *Sci China Earth Sci*, 2012, 55(7): 1043-1051
- [4] Fan Y D, Wen Q, Chen S R. Engineering survey of the environment and disaster monitoring and forecasting small satellite constellation. *International Journal of Digital Earth*, 2012, 5(3): 217-227
- [5] 胡鹏, 刘沛兰, 胡海等. 地球信息的度量空间和 Global GIS. *武汉大学学报(信息科学版)*, 2005, 30(4): 317-321
- [6] 全国地理信息标准化技术委员会(SAC/TC). GB/T 13989-2012 国家基本比例尺地形图分幅和编号. 北京: 中国标准出版社出版, 2012
- [7] 李德仁, 彭明军, 邵振峰. 基于空间数据库的城市网格化管理与服务系统的设计与实现. *武汉大学学报(信息科学版)*, 2006, 31(6): 471-475
- [8] Fekete G. Sphere quadrees: a new data structure to support the visualization of spherically distributed data. In: *Proceedings of the SPIE 1259, Extracting Meaning from Complex Data: Processing, Display, Interaction*, Santa Clara, 1990. 242-253
- [9] Goodchild M. Discrete global grids for digital earth. In: *Proceedings of International Conference on Discrete Global Grids*. California: Santa Barbara, 2000
- [10] Dutton G. Universal geospatial data exchange via global hierarchical coordinates. In: *Proceedings of International Conference on Discrete Global Grids*, California: Santa Barbara, 2000
- [11] Ottoson P, Hauska H. Ellipsoidal quadrees for indexing of global geographical data. *International Journal of Geographical Information Science*, 2002, 16(3): 213-226
- [12] Sahr K, White D, Kimerling A J. Geodesic discrete global grid systems. *Cartography and Geographic Information Science*, 2003, 30(2): 121-134
- [13] 袁文, 程承旗, 马蔼乃等. 球面三角区域四叉树 L 空间填充曲线. *中国科学: E 辑 工程科学 材料科学*, 2004, 34(5): 584-600
- [14] 李德仁, 肖志峰, 朱欣焰. 论空间信息多级网格的划分方法及编码研究. *测绘学报*, 2006, 35(1): 52-56
- [15] 赵学胜, 崔马军, 李昂等. 球面退化四叉树格网单元的邻近搜索算法. *武汉大学学报(信息科学版)*, 2009, 34(4): 479-482
- [16] 周成虎, 欧阳, 马廷. 地理格网模型研究进展. *地理科学进展*, 2009, 28(5): 657-661
- [17] 程承旗, 关丽. 基于地图分幅拓展的全球剖分模型及其地址编码研究. *测绘学报*, 2010, 39(3): 295-301
- [18] 童晓冲, 贲进, 张永生. 全球多分辨率六边形网格剖分及地址编码规则. *测绘学报*, 2007, 36(4): 428-435
- [19] Ma T, Zhou C H, Xie Y C, et al. A discrete square global grid system based on the parallels plane projection. *International Journal of Geographical Information Science*, 2009, 23(10): 1297-1313
- [20] Tong X C, Ben J, Wang Y, et al. Efficient encoding and spatial operation scheme for aperture 4 hexagonal discrete global grid system. *International Journal of Geographical Information Science*, 2013, 27(5): 898-921
- [21] 程承旗, 任伏虎, 濮国梁, 等. 空间信息剖分组织导论. 北京: 科学出版社, 2012. 51-54
- [22] Boschetti L, Roy D P, Justice C O. Using NASA's World Wind virtual globe for interactive internet visualization of the global MODIS burned area product. *International Journal of Remote Sensing*, 2008, 29(11): 3067-3072
- [23] Martino S D, Bimonte S, Bertolotto M, et al. Spatial online analytical processing of geographic data through the Google Earth interface. *Geocomputation, Sustainability and Environmental Planning*, 2011, 348: 163-182
- [24] Gibin M, Singleton A, Milton R, et al. An exploratory cartographic visualisation of London through the Google Maps API. *Applied Spatial Analysis and Policy*, 2008, 1(2): 85-

97

- [25] Lü X F, Cheng C Q, Gong J Y, et al. Review of data storage and management technologies for massive remote sensing data. *Sci China Tech Sci*, 2011, 54 (12) : 3220-3232
- [26] Li X, Bao C X. Address switching: Reforming the architecture and traffic of Internet. *Sci China Ser F-Inf Sci*, 2009, 52 (7) : 1203-1216
- [27] Hinden R, Deering S. IP Version 6 Addressing Architecture. RFC 4291, 2006
- [28] 郭薇, 郭箐, 胡志勇. 空间数据库索引技术. 上海: 上海交通大学出版社, 2006, 100-124

Spatial information location identification based on global subdivision

Lü Xuefeng^{* **}, Cheng Chengqi^{***}

(^{*} National Disaster Reduction Center of China, Beijing 100124)

(^{**} Key Laboratory of Integrated Disaster Assessment and Risk Government of the Ministry of Civil Affairs, Beijing 100124)

(^{***} Institute for Aeronautics and Astronautics Information Engineering, College of Engineering, Peking University, Beijing 100871)

Abstract

In order to identify spatial information's area coverage and spatial position in a globally unified and multi-scale way, a kind of spatial information location identification coding model based on the geographic coordinate subdivision grid with one-dimension-integral coding on 2^n -tree (GeoSOT) is presented. By means of the globally fixed grid division and structural grid representation, the GeoSOT-based model adopts a triple (C_0, M, N) that integrates the geospatial orientation, scale, range and structure in to one set to describe five kinds of location information belonging to a same geographic object, such as the geographic object's area coverage and spatial position, internal spatial grid, spatial point-line-area entities, spatial attribute information, and spatial information network domain. By the comprehensive design of the methods for this location identification's coding definition, generation and application, the model can make it possible to achieve a consistent location association relation of the multi-type spatial information that belongs to a same geographic area, and provide a unified geocoding reference for the management and application of global multi-type spatial information.

Key words: spatial location identification, global subdivision, discrete grid, spatial information identification, spatial information organization