

# 对空间数据多尺度表达有关问题的思考

艾廷华<sup>1,2</sup> 成建国<sup>3</sup>

(1 武汉大学资源与环境科学学院,武汉市珞喻路 129 号,430079)

(2 武汉大学地理信息系统教育部重点实验室,武汉市珞喻路 129 号,430079)

(3 深圳市规划国土房产信息中心,深圳市红荔路 8009 号,518034)

**摘要:**多尺度空间数据表达及数据库建立在流媒体网络数据传输、自适应动态可视化、空间认知导航、互操作尺度匹配多个领域都有贡献,但实现该技术面临着数据存储量大、操作响应慢、横向空间一致关系难以维护、尺度变化难以达到真正的连续性等诸多问题。针对这些问题,提出了多级尺度显式存储、初级尺度变化累积、关键尺度函数演变和初级尺度自动综合四种技术策略。

**关键词:**空间数据库;空间数据综合;网络 GIS;流媒体传输

**中图法分类号:**P283.7;P208

对同一区域不同尺度、不同分辨率表达的空间数据进行匹配与集成是空间数据库融合技术的一个重要内容<sup>[1,2]</sup>。“尺度”概念被认为是空间数据表达的一个重要特征,从认知科学的观点出发,它体现了人们对空间事物、空间现象认知的深度与广度。一般意义上,地学领域的“尺度”是指研究对象在空间域上的延展范围或时间域上的覆盖区间,而地图学与地理信息系统中的“尺度”为“比例尺”所取代,定义为表达空间(地图图面)中的距离与实际地理空间距离的比率。在“表达距离”、“实际距离”、“比例尺”三个基本量中,认知表达空间的“表达距离”是固定的(不同比例尺下的序列地形图图幅范围均为 50 cm<sup>2</sup> 左右,计算机显示屏是固定的),这样“空间尺度”与“比例尺”的关系变为分母与商的关系,即大尺度对应小比例尺,小尺度对应大比例尺。

在数字技术、网络传输、多媒体可视化等技术条件下,人们不再满足于静态、单一分辨率的空间表达,提出了从多角度、多视点、多层次对空间认知表达的要求,这是 IT 业日益兴起的“以人为本”、“个性化”、“自适应”需求的表现。作为一种视觉化 IT 产品,GIS 应当为用户提供“连续”“无级变焦”式可视化功能,即具有多尺度、多分辨率的空间表达与应用。观察的视点越深,获得的信

息越多,越详细;观察的视点越远,获得的范围越广,获得的信息越概略<sup>[2,3]</sup>。选择怎样的空间表达尺度不是预先在服务器固定好的,而是不同终端用户根据其个性化的要求,由服务器/终端机制提供自适应的表达尺度。为满足这一需求,建立多尺度空间数据库,提供多尺度空间表达机制成为当前 GIS 理论研究和工程建设中的一个热点问题。

## 1 多尺度空间数据库的作用

空间数据库的服务功能呈多元化趋势,从而使得“多尺度”机制相应地产生多种功能作用。对其区别分析,可以有针对性地建立多尺度技术策略。

### 1.1 在空间认知中辅助从粗到细的信息导航

人们对空间现象的认知表现为从总体到局部、从概略到细微、从重要到次要的层次顺序。在传统地图表达中,通常通过概略图、区位图、索引图等方式,配以主地图内容,实现地物目标的搜索和空间信息的查询。多尺度空间数据库表达了同一区域多分辨率下的空间信息内容,通过比例尺的类似于光学系统的“变焦”调节,可以展示从大范围主体信息内容到局部区域细微信息内容的动

态表达,从而引导用户对该区域的认识,进行信息导航,辅助用户截取其感兴趣的局部区域,并沿着该路径深入到细节内容。对于空间信息导航,传统地图技术中由概略图到主图有大幅度的比例尺跨越,会使两者难以对应,而多尺度空间数据库的比例尺调节接近于连续式变化,没有大的跳跃,较好地满足了思维连续性的要求。

### 1.2 在网络传输中支持“流媒体”、“渐进式”的新型数据传输

多尺度空间数据库可改变网络环境中的数据传输方式,实现跨尺度变化中的逐步精细化的数据叠加传输。这是一种新型的数据传输方式,不是将研究区域的图形数据从北到南、从西到东完整地由服务器传输到终端,而是经过多次扫描整个区域,按流媒体方式将数据叠加传输<sup>[3-5]</sup>。目前,这种方式在图像数据、栅格数据、TIN 结构数据的网络传输、网上下载中基于 LOD 技术已普遍实现,视觉上表现为图像效果由模糊逐渐变清晰。建立多尺度空间数据库后,可以将这一方式推广到以矢量数据为主的 GIS 数据传输中,按照信息内容的层次性,建立顾及尺度影响的新型空间索引,在多尺度框架下,在服务器端建立层次化数据组织,实现从主体到细节化的渐进式传输。这一渐进式的传输方式可介入用户的互动操作,满足用户的自适应条件,节省传输时间,提高传输效率。有的用户可能只需了解区域的概略信息内容,在获得其符合要求的内容后可随时停止传输,避免全部数据传输而浪费时间。

### 1.3 实现用户自适应的动态可视化

GIS 数据可视化中的放大、缩小是为了从不同层次深度获得空间信息的视觉化,在单一尺度的空间数据库支持下,放大、缩小可得到图形符号化的不同形式,但并没有增加/减少信息内容,比如,到一定比例尺后进一步放大只能得到符号、文字占满整个屏幕的粗大的线划、笔划、马赛克效果等。在新技术条件下,人们不再满足于静态、单一分辨率的空间可视化,提出了从多角度、多视点、多层次对空间可视化表达的要求。在多尺度空间数据库支持下,可根据屏幕当前可视化比例尺,动态地选择对应的尺度内容进行显示,获得适宜的可视化效果<sup>[6,7]</sup>。目前存在两种可视化方式,一是对同一研究区域动态地展示多种尺度下的显示版本,但同一版本不同位置的显示比例尺是相同的;二是同一版本下不同位置显示的比例尺不同。前者类似于用多面具有不同反射能力(可理解为可视化内容的多少)的平面镜来观察区域,后者类

似于用凸凹不平的“哈哈镜”来观察区域,在同一窗口下,有的区域放大,有的区域缩小。

### 1.4 在系统集成中实现横向一致性匹配及便捷的互操作

不同来源、不同时间、不同精度的空间数据集成与融合是 GIS 工程应用中的突出问题,其中语义特征的匹配和尺度特征的匹配是两大关键技术。无疑,多尺度空间数据库的建立增强了不同数据集成匹配的能力,可以实时地将不同尺度的数据调整为一致,或通过临时输出不同版本的数据使得其尺度达到一致或接近,为不同系统间的数据互操作提供条件。

## 2 建立多尺度空间数据库需解决的关键问题

多尺度空间数据库可认为是数据表达在空间域上多层次版本的集成。一方面要获得连续的尺度表达效果,另一方面又不能使数据量无限增大。具体地,多尺度空间数据库的建立面临着以下几个关键技术问题。

### 2.1 比例尺变化粒度适宜

首先是数据库的多尺度表达“多”到什么程度,即在比例尺变化轴上如何划分“刻度”或“粒度”,类似于视频技术中 1 s 存储多少“帧”的确定。比例尺变化粒度划分太大,是离散的跳跃式表达,不能获得连续的效果;比例尺变化粒度划分太小,会大大增加数据量。因此,当比例尺发生微小变化  $d(s)$  时,数据表达在什么层次上发生变化是首先要解决的问题。显然,多尺度空间数据库不可能也没有必要存储无限小粒度的数据表达,比例尺的变化还是离散的、跳跃的,但划分的粒度针对应用需求是合适的。

GIS 数据可看作是在三个水平上的层次结构:要素层、目标层、几何细节层。要素是具有相同语义特征的目标集;目标是具有独立地理意义的表达实体,是构成要素的基本单位;几何细节是几何表达上的划分的结构体,是构成目标的基本单位,如构成河流目标的“弯曲”特征,构成面状目标的三角形剖分单元。当比例尺发生变化时,数据表达发生变化的主体对象可以有三个层次:要素层、目标和几何细节,它们分别对应不同的变化粒度。在要素层上划分粒度过于粗糙,而且不同要素层之间难以根据重要性差距进行排序(除非考虑专业化的特殊要求)。在某一要素层下,目标可根据表达的重要性进行排序,与表达尺度建立

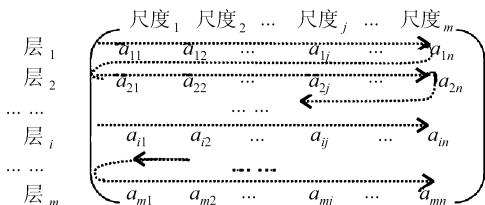
函数关系,在地图综合技术中通过“选取”算子实现。当尺度变化到某一刻时,够条件的目标便显示出来,变化的对象以完整的目标图形出现或消失。更细小的粒度划分则是在“几何细节”层次上,随着比例尺的变化,目标的几何细节有逐步演变的过程,接近于连续式的变化。

在建立多尺度空间数据库时,如何划分合适的比例尺变化粒度取决于应用过程中感兴趣信息内容的层次结构。比如,用户感兴趣的是河流流域的网络结构,划分的粒度达到河流目标层即可,没有必要将河流的弯曲演变出来。如果用户感兴趣的是河流的分布,则要深入到弯曲层次,逐步地将河流的曲线表达由概略到详细的过程演变出来。总体上,对比例尺变化粒度的划分有如下原则:发生变化的对象层次比感兴趣的信息内容的层次低一级。

## 2.2 数据容量最小化

与单一尺度的空间数据库相比,多尺度空间数据库无疑会大大增加数据量,采取必要的技术手段降低数据容量成为需要解决的关键问题之一<sup>[2,3]</sup>。在视频技术中,数据压缩一直是研究的热门课题,受视频压缩技术常用方法的启发,多尺度空间数据库的数据量压缩可以通过以下技术策略:① 只存储变化的数据部分,取代完整的数据存储;② 识别关键变化部分,过滤无关紧要的细小变化;③ 通过几何图形渐变函数导出新尺度下的数据表达。

策略①的实施是顾及到相邻尺度间的数据表达存在大量的相同的事实,提取存储相邻尺度间的变化部分比存储完整数据会大大降低数据量。对国家序列比例尺地形图 1:1 万、1:5 万、1:10 万、1:25 万、1:100 万的同一区域的表达作



多尺度空间数据库建立在数据组织上最终表现为空间尺度上的线性索引结构,通过用户视图截取线性结构上的部分数据集实现某一尺度的表达,这说明多尺度空间数据库中的目标可以用顺序表将其串起来。对矩阵结构下的目标建立线性索引有“按行占优”和“按列占优”两种方式,分别对应“语义占优”和“尺度占优”,如上述二维矩阵。“语义占优”的多尺度表达体现为某语义层目标在

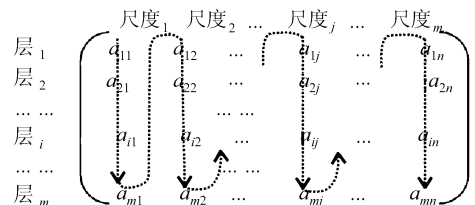
比较会发现,数据表达的变化是缓慢的,“表达生命周期”长的目标在多个比例尺段的图上均出现,而且没有表达形式上的明显差异。

策略②是在策略①的基础上提出的,虽然存在变化,但有性质上的差异,可区别对待变化部分的记录与处理。犹如河流的流动存在阶地瀑布和缓流一样,地物目标在比例尺表达空间上也存在质的突变和量的缓变。如水系的表达,狭长多边形的形状逐步简化是量变,在某一时刻由多边形变成中轴线表达时产生质变,之后曲线形状的化简为量变,在某一时刻河流消失又产生一次质变。数据组织时只存储记录发生质变的数据状态,中间的量变状态通过函数内插导出,从而大大降低数据量。

策略③与策略①和策略②结合应用,由几何图形函数来导出不同尺度的表达犹如对圆的表达通过圆心、半径两个参量实时计算一样,可从实际存储圆的坐标点中大大减少数据量。在计算机图形学动画设计、Flash 设计中,普遍采用了这一思想。空间数据具有不规则性,难以寻找合适的图形函数来拟合,但在 CG 的 Morphing 技术支持下,通过两端图形状态的控制建立合适的图形渐变函数,在较小尺度变化幅度内,得到满意的效果,同时达到减少数据量的目的。

## 2.3 兼顾考虑语义特征

尺度特征与语义特征是空间数据表达的两个主要因素,更一般意义上的多表达空间数据库应同时考虑尺度、语义特征。尺度与语义两者相互独立,某一尺度下包含有各种专题语义的目标,某一语义特征下的目标可以有不同尺度特征的数据表达。数据库表达的目标可以置入“尺度”×“语义”的二维矩阵中:



各种尺度下表达完成后方可跳到另一语义层目标中,类似于专题地图中的背景层与主体层的区分,这种表达顾及了用户的倾向性。“尺度占优”的多尺度表达体现为严格按尺度特征建立目标的顺序表,对语义的考虑是平均化的,不特别倾向于某一专题语义的目标,类似于序列比例尺普通地图的表达,六大要素类型平均考虑,仅由尺度特征决定目标是否表达。

### 2.4 操作响应速度实时

网络传输环境服务器/终端机制下实现多尺度表达的操作有两种：“在线式操作”和“脱线式操作”<sup>[8]</sup>。在服务器端进行预处理(相邻尺度表达间的变化提取、关键变化部分的识别),获得多尺度表达的粗数据,在终端进行实时处理,完成粗数据基础上的精细化多尺度表达(基于关键尺度表达的函数实时内插)。两种操作相互依存,过多的“脱线式操作”对多尺度表达准备得越充分,对显式记录处理好的数据,便越可减轻终端“在线式处理”的负担,但增加了数据的容量。相反,较少的服务器端“脱线式处理”满足了数据量减少的要求,但增加了“在线式”实时处理的负担,从而导致响应速度缓慢。这一矛盾如何协调取决于两种操作的有效性,如采用基于 Morphing 函数内插生成中间尺度的数据表达,如果形变函数复杂,占用时间长,不能达到实时化效果,就要放弃该技术,由预处理显示记录多种尺度版本下的数据表达,终端应用时直接下载处理好的数据,达到快速响应的目的。最终实现数据量小、操作响应实时,则需要寻求高度自动化的快速地图综合技术,在较少已知数据上,导出不同尺度下的新的表达。

### 2.5 横向数据表达一致

多尺度空间数据库通过用户视图输出一定的数据集(或实时生成导出)得到一定尺度下的数据表达,在横向上,该尺度所对应的数据在空间关系特别是拓扑关系上是否一致是需要重点研究的问题,需要采用后处理的方法来调整破坏了的空关系一致性。在多尺度表达中产生不一致的原因在于新尺度下导出的目标表达未能考虑上下文的影响,而且新尺度下的目标组合不是数据预处理可预见到的,由用户实时处理临时组合,不可避免会产生空间冲突<sup>[9,10]</sup>。如图 1,在小比例尺条件下河流要简化表达为曲线,与大比例尺条件下的双线河表达相比,原来的河流多边形与土地利用多边形间的相切吻合关系被破坏,两目标间产生裂隙,这时需通过后处理由土地利用多边形扩展来修正吻合关系。

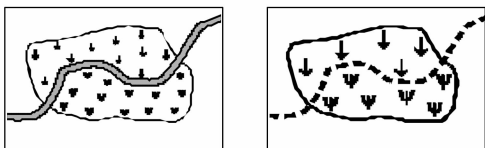


图 1 河流中轴化表达后,土地利用地块扩展来修正其间“吻合”关系

Fig. 1 The Collapse of Road Leads to the Extension of Neighbor Parcels to Fill the Gap Area

## 3 四种技术策略

上述讨论的关于多尺度空间数据库建立的技术需求,很难同时满足,有些本身之间就是矛盾的,比如要达到应用时的实时化,需要预处理充分,将结果以多种尺度版本显式存储,这必然增加数据量。采用何种策略来实现空间数据的多尺度表达及数据库的建立并达到尺度变化的“粒度”精细、数据冗余最小、操作响应速度实时等优化约束条件,还面临着大量理论问题及工程技术问题需要解决。本节从数据组织角度,针对数据量压缩和尺度变化粒度精细(在几何细节层次上变化)的特点,归纳出四种实施多尺度表达的策略<sup>[5]</sup>。

在尺度变化空间,与尺度  $s$  对应的表达为  $f(s)$ ,由尺度  $s_i$  到  $s_{i+1}$  的表达变化为  $\Delta f_i = f(s_{i+1}) - f(s_i)$ ,这里以数据库的存储内容为考察对象,多尺度表达的实现技术有以下四种策略。

- 1) 多级尺度显式存储型,数据库存储的数据集为  $\{ f(s_0), f(s_1), f(s_2), \dots, f(s_i), \dots \}$ ;
- 2) 初级尺度变化累积型,数据库存储的数据集为  $\{ f(s_0), \Delta f_1, \Delta f_2, \Delta f_3, \Delta f_4, \dots \}$ ;
- 3) 关键尺度函数演变型,数据库存储的数据集为  $\{ f(s_{k0}), f(s_{k1}), f(s_{k3}), \dots \}$ ,其中  $s_{ki}$  为关键尺度;
- 4) 初级尺度自动综合型,数据库存储的数据集为  $\{ f(s_0) \}$ 。

以上四种策略,从 1)到 4)自动化程度逐步提高,其中 1)和 4)是两种极端状态。策略 1)对各个尺度建立对应的数据表达并显式存储,存储的内容为脱线式操作预处理好的结果,应用时提取某尺度对应的数据集即可,只是一个数据读取调用的操作,但数据冗余最大,尺度变化离散(不是真正的连续性尺度变化)。这是目前在实际应用中普遍采用的方法,将同一区域的数据独立地建立多套比例尺数据库,技术完备一点还在不同比例尺下目标表达间通过目标匹配建立对应关系。策略 4)只存储初级尺度下的数据,由全自动化综合技术在初级尺度数据基础上,实时综合输出任意尺度的数据表达,数据冗余最小,且全部由在线式操作实时产生新尺度数据,是一种理想状态。众所周知,全自动化综合是地图学与 GIS 领域瓶颈问题,目前地图综合研究所取得的成果距全自动化还相距甚远,因此这种策略在目前技术水平下没有实际意义。

1)和 4)这两种策略在空间数据库框架下没

有重要的研究意义。在目前完全自动化技术还没实现,实际应用又很迫切,而且要满足多尺度表达的几个约束条件(数据冗余小、尺度变化精细、操作响应快等)的情况下,策略2)和3)是比较实际的解决方案。

2)和3)两种策略的思想来源于视频数据的压缩,视频数据以“帧”为基本存储单位,其序列构成了时间尺度上的变化过程,而这里空间表达 $f(s)$ 可看作是空间尺度上的变化过程,两者在很大程度上是相似的。为减少数据冗余,视频数据的压缩采用了一些方法,如以变化数据存储取代完全版本的每帧影像存储和基于关键状态由图形变换 Morphing 内插中间连续状态。

对于初级尺度变化累积模型,在尺度空间 $\{s_0, s_1, s_2, \dots, s_i, s_{i+1}, \dots\}$ ,相邻尺度间的表达变化为 $\Delta f_i = f(s_{i+1}) - f(s_i)$ ,则通过递推关系,可计算出任意尺度 $s_i$ 下的表达式:

$$f(s_i) = f(s_0) + \Delta f_1 + \Delta f_2 + \dots + \Delta f_i$$

该公式表明,任意尺度下的表达可通过变化的累积叠加来实现,因此把它称作“初级尺度变化累积模型”,这是顾及尺度特征的空间表达的新型数据模型。这里的初级尺度对应小比例尺低分辨率下的粗糙表达状态。比例尺越大,累积的变化部分越多(上述表达式包含的项数越多),对目标的表达越完备、真实。

该策略实现的关键技术包括:① 相邻尺度表达间“变化”的提取;② “变化”组合的线性结构的建立;③ “变化”绑定实现一定“粒度”的划分(该绑定过程相当于在上述表达式中合适位置加括号);④ “变化”累积后真实表达的恢复<sup>[5]</sup>。

根据尺度变化粒度划分的层次,变化的内容可以是目标实体,也可以是几何细节单元。基于目标实体的变化累积,实际例子中有基于 Horton 编码对水系流域的各河流目标分级,按比例尺建立线性索引,实现流域网络结构的多尺度表达。在地图综合算法现有成果中,基于层次结构的目标表达如 GAP-tree<sup>[6,7]</sup>,可以直接支持变化累积模型的建立。

实现基于目标实体的变化累积模型相对要简单一些,但变化的粒度太粗糙。基于几何细节的变化累积要“细腻”,关键是寻找有效的在尺度特征影响下目标与几何细节间的层次分解与组合,分解的基本单元可看作是前一个尺度与后一个尺度间的变化部分。

## 4 结 论

多尺度空间数据库的建立与地图综合技术密切相关,在网络传输、多源数据集成、系统互操作背景下,空间数据多尺度表达向传统的地图综合技术提出了更高的需求挑战。网络环境下完成多尺度表达存在两种综合操作:脱线式综合与在线式综合,过多的脱线式综合预处理可提高多尺度数据操作的响应速度,但减少了数据的用户适应性,增加了数据量,过多的在线式综合实时处理正好是相反的效果。一直以来,“地图综合”被认为是一个国际难题,尽管目前在理论、技术上取得了不少成果,但距全自动化还相距甚远,在这种背景下,将脱机预处理与在线实时处理结合,纳入人机协同作业机制不失为一种现实的解决方案。

受应用需求的驱动,多尺度空间数据库的建立目前在国内多个部门已广泛开展起来,但都是独立地数字化建设各连续比例尺段的数据库,没有顾及其间的联系,建库工作投入了大量人力、物力与资金。如果考虑尺度间的联系,设计跨尺度的建库方案,可在一定程度上节省投入。建库完成后,如何将它们整合到一起,在尺度轴上建立目标间的联系,包括目标映射关系的匹配、渐变函数的置入等都是需要研究的问题。GIS 数据表达在横向上的空间关系一直是研究的热点,当纳入到多尺度表达机制下,在纵向上的尺度关系也应当受到重视。

## 参 考 文 献

- 1 Ruas A. Multiple Paradigms for Automating Map Generalization: Geometry, Topology, Hierarchical Partitioning and Local Triangulation. ACSM/ASPRS Annual Convension and Exposition, 1995(4): 69~78
- 2 Buttenfield B P. Transmitting Vector Geospatial Data Across the Internet. In: Egenhofer M J, Mark D M, eds. Proceedings GIScience 2002. Berlin: Springer-Verlag, 2002
- 3 Bertolotto M, Egenhofer M. Progressive Transmission of Vector Map Data over the World Wide Web. GeoInformatica, 2001, 5 (4): 345~373
- 4 Han H, Tao V, Wu H. Progressive Vector Data Transmission. The 6th AGILE, Lyon, France, 2003
- 5 Ai T, Li Z, Liu Y. Progressive Transmission of Vector Data Based on Changes Accumulation Model. The 11th International Symposium on Spatial Data Handling. Berlin: Springer-Verlag, 2004

- 6 Ai T, Oosterom P van. GAP-tree Extensions Based on Skeletons. In: Richardson D, Oosterom P van eds. *Advances in Spatial Data Handling*. Berlin: Springer-Verlag, 2002
- 7 Oosterom P Van. *Reactive Data Structure for Geographic Information Systems*. Oxford: Oxford University Press, 1994
- 8 Cecconi A. *Integration of Cartographic Generalization and Multi-Scale Databases for Enhanced Web Mapping*. [Ph. D Thesis]. Zurich: University of Zurich, 2003
- 9 Harrie L. The Constraint Method for Solving Spatial Conflicts in Cartographic Generalization. *Cartography and Geographic Information Science*, 1999, 26(1): 55~69

---

**第一作者简介:**艾廷华,教授,博士生导师。研究兴趣为地图综合、空间数据挖掘、移动 GIS。  
E-mail:tinghua-ai@tom.com

## Key Issues of Multi-Scale Representation of Spatial Data

AI Tinghua<sup>1,2</sup> CHENG Jianguo<sup>3</sup>

(1 School of Resource and Environment Science, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

(2 Key Laboratory of Geographic Information System, Ministry of Education, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

(3 Urban Planning and Real Estate, Shenzhen Municipal Center of Land, 8009 Hongli Road, Shenzhen 518034, China)

**Abstract:** Multi-scale representation and multi-scale spatial database plays an important role in such fields as the transmission of streaming media data over web, the self-adaptable visualization of spatial information, the navigation in spatial cognition, the scale match during inter-operation and other applications. Realizing this technology has to resolve the questions including the large data volume, the slow response, the conflicts between data representations and the steep change in scale range. This study aims at the questions above presenting four strategies: storing data of multi-scale versions, accumulating changes based on the initial version, transferring between key state versions and generalizing to derive any version based on the initial state. Among four strategies, the changes accumulation model has been attached importance to be a new method which borrows the ideas from the technology of video data compress.

**Key words:** spatial database; spatial data generalization; Web GIS; streaming media transmission

---

**About the first author:** AI Tinghua, professor, Ph.D supervisor, with the research interests of map generalization, spatial data mining and mobile GIS.

E-mail: tinghua-ai@tom.com