

基于 C-Tree 的无级比例尺 GIS 多边形综合技术

田 鹏 郑扣根 张 引 潘云鹤

(浙江大学 CAD&CG 国家重点实验室, 杭州 310027)

(浙江大学计算机系人工智能研究所, 杭州 310027)

摘 要 无级比例尺 GIS 是数字地球和 Web GIS 的核心技术之一,但随着 GIS 的广泛应用和深入发展,现有的 GIS 技术已经不能满足信息社会的需要,其中一个需要解决的重要问题就是 GIS 的空间数据量如何随着比例尺的变化自动增减.针对无级比例尺 GIS 多边形综合中的选取与合并技术,在对选取的数量规律和质量原则以及合并的原则进行充分论述的基础上,提出了一种多边形图层数据组织策略 C-Tree,并给出了基于 C-Tree 的多边形综合算法.对于给定的大比例尺地图多边形图层数据,该算法可以高效率地完成多边形选取与合并的综合操作,输出小比例尺地图图层数据.该算法现已成功应用于时空一体化智能城建信息系统,并获得了满意的结果.

关键词 GIS 无级比例尺 GIS 制图综合 线综合 多边形化简 多边形综合

中图法分类号: P208 TP392 **文献标识码:** A **文章编号:** 1006-8961(2001)08-0765-06

Polygon Generalization Based on C-Tree in Scaleless GIS

TIAN Peng, ZHENG Kou-gen, ZHANG Yin, PAN Yun-he

(CAD&CG National key laboratory, Zhejiang University, Hangzhou 310027)

(Institute of Artificial Intelligence, Department of Computer Science & Engineering, Zhejiang University, Hangzhou 310027)

Abstract Scaleless GIS is one of the key technologies of digital-earth and web GIS. With the development of GIS, the current GIS techniques couldn't meet the requirements of the information society very well. The ability of deriving small scale map from large scale map automatically is very important in modern GIS. To derive small scale map from large scale map in scaleless GIS is related to cartographic map generalization technique. Based on the description of the amount rule of selection, quality principle of selection, combination principle and attribute data generalization. A hierarchical organization strategy of polygon map layer, C-Tree, and the corresponding algorithm of generalization in scaleless GIS is presented in this paper. The algorithm based on C-Tree can complete the polygon selection and combination operation in high efficiency for given large scale polygon map layer and small output scale. This strategy has been implemented into a 863 project of China and the experimental results show that it can complete polygon layer generalization with high efficiency and can maintain the appearance and the topological structure of map well.

Keywords GIS, Scaleless GIS, Cartographic generalization, Line generalization, Polygon simplification, Polygon generalization

0 引言

地理信息系统 GIS 的主要功能是采集、管理、分析和表达与地球相关的信息,随着 GIS 在各领域

应用的不断拓展和深入,特别是随着数字地球和 Internet 技术的发展,现有的 GIS 数据处理技术已经不能满足信息社会的需要,其中一个重要的原因就是 GIS 无法解决矢量空间数据随比例尺变化而产生的信息量增减的问题,即无级比例尺 GIS 空间

基金项目:国家 863 计划资助项目(306-04-03-3)

收稿日期:2000-08-15;改回日期:2000-12-29

信息压缩与复现问题。

所谓无级比例尺 GIS (Scaleless GIS) 就是以一个大比例尺单精度空间数据库为基础数据源,在一定空间区域内,通过空间对象的信息量随比例尺的变化自动增减的方法,来使 GIS 空间信息的压缩和复现与比例尺实现自适应的一种信息处理技术,如今这种无级比例尺 GIS 已是数字地球和 Web GIS 的核心技术之一^[1]。

作为无级比例尺 GIS 的基础数据源,大比例尺单精度空间数据库是由某区域范围(如中国全境)的大比例尺(如 1:1 万)地理基础数据和专题信息所构成的大型空间数据库,该数据库通常由数千或数万幅同比例尺的系列图幅(如地形图、专题图)在统一的地理坐标参照下,无缝拼接而成,而应用所需要的任意小比例尺地图数据是从大比例尺地图数据自动派生生成^[2]。

从大比例尺地图数据自动派生生成小比例尺地图数据的本质是自动地图制图综合。地图制图综合即选取主要的,实质性的信息并有目的地加以概括,其目的在于根据地图的用途、主题和比例尺,在地图上表示出客观实际某一部分的典型特征和基本特点^[3]。由于影响制图综合的因素很多,且这些因素很难用简单的数学和逻辑语言来描述,因此要在有限的地图幅面内和满足最佳目视效果的条件下,使空间信息随比例尺的变化而自动增减是非常困难的,目前尚无理想的解决方案^[1]。

1 多边形综合

定义 1 多边形综合 (Polygon Generalization) 是指由大比例尺地图数据自动派生生成小比例尺地图数据时,对多边形数据进行综合的处理技术。它要求在保持多边形的拓扑结构正确和基本几何形态的前提下,结合与多边形对应的地理对象的语义信息,对其进行精简和抽象。这种多边形综合包括化简、选取、合并、象征、夸张和移位等处理技术。

采用拓扑结构的 GIS (如: ESRI ARC/INFO) 的基本几何元素包括点、线和多边形,其中线和多边形由若干弧段组成。一般点的处理较简单,线综合 (Line Generalization) 技术也已经比较成熟,无级比例尺 GIS 数据综合的重点和难点是多边形综合。这种综合的实质是地理信息的精简,而多边形化简 (Polygon Simplification) 基于多边形弧段的线综

合。由于化简处理只是几何对象的数字化节点数量从大到小的过程,因此它不会改变图幅内几何对象(点、线和多边形)的数量,从大比例尺地图数据综合生成小比例尺地图数据时,随着输出图幅面积的减小,输出图幅内的几何对象将变得拥挤,并出现团块,使读图者无法辨识,同时,由于很多对象的输出尺寸也已经小于制图精度所要求的最小尺寸,因此就几何尺寸而言,对这样的对象进行化简已经失去意义。按照制图综合的原则,此时虽可根据地图的用途、主题、制图区域的特点和比例尺来对地物进行取舍,但化简不能解决地图综合的全部问题。

选取 (Selection) 处理,即删除次要的地理特征而保留主要的和实质性的地理特征,但是,由于直接删除多边形对象,可能使相邻多边形之间出现空洞,从而造成多边形图层的拓扑错误,因此此时要进行合并 (Combination) 处理,即将要删除的多边形合并到与之相邻的属性相同或相近的多边形中。

1.1 选取的数量规律与质量原则

定义 2 多边形图层的图幅数据量是图幅范围内多边形的个数。

定义 3 设大比例尺地图数据源多边形图层的图幅数据量为 m , 小比例尺输出图幅数据量为 m' , 则图幅化简尺度 $S = m'/m$, 其中, $m' \leq m$ 。

研究表明,与地图制图综合程度有关的制图方法是符合开方根规律的^[4],即开方根规律适用于与制图综合程度有关的制图数量标准 $N = K \times \sqrt{M}$, 其中, N 是实地尺度下与该要素、方法、地物等有关的数量标准; K 为尺度常数; M 为地图的比例尺分母。

根据开方根规律,即可得到由大比例尺地图来综合生成小比例尺地图的过程中地物的选取规律,其简单的选取规律可采用如下公式:

$$N_2 = N_1 \times \sqrt{M_2/M_1} \quad (1)$$

其中, N_1 为大比例尺图幅数据量; N_2 为小比例尺输出图幅数据量; M_1 为大比例尺的分母值; M_2 为小比例尺的分母值, 其中, $M_1 \geq M_2$ 。由定义 3 可知,

$\sqrt{M_2/M_1}$ 就是图幅化简尺度, 并有 $\sqrt{M_2/M_1} \leq 1$ 。

例如,在 1:1000 比例尺的图幅中,小村庄由 7 个建筑物(多边形)构成,如图 1(a)所示。根据式(1),在 1:10000 比例尺地图上保留 2 个建筑物(多边形),如图 1(b)、(c)所示,其选取数据量计算如下: $N_2 = 7 \times \sqrt{1\ 000/10\ 000} = 2.2$ 。

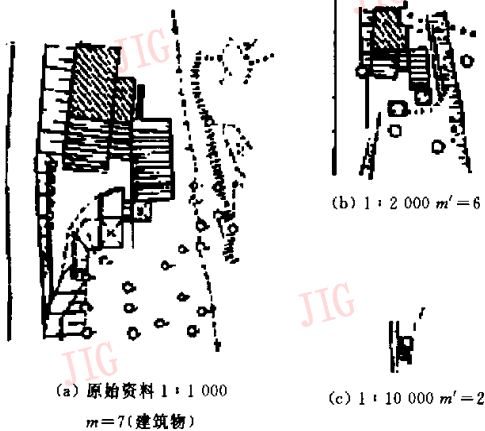


图 1 小村庄建筑物的选取

选取规律是实际进行制图综合的最重要公式,因为它可以有规律地反映地理环境,并且可以在不同的图幅和比例尺系列范围内可靠地保持景观特征,同时,它也为检验和评价制图综合提供了依据^[4]。选取规律是通过对手工制图综合的研究和统计而得出的,它反映的是地理景观特征随比例尺变化的规律,而不是绝对精确的量化标准。如图 1(b)所示,手工综合的结果是在 1:2 000 比例尺地图上保留 6 个建筑物,而不是根据式(1)计算得到的 5 个。此外,定义 2 中,以图幅范围内多边形的个数作为多边形图层的图幅数据量只是一种简单的处理方法,尚有其他因素需要考虑,如多边形边界的弯曲个数、点密度等。毕竟,地图制图技术经过长期的发展,其内涵是十分丰富的,很难用简单的数学公式精确地量化和表达。

式(1)称为无级比例尺 GIS 由大比例尺地图综合生成小比例尺地图的过程中地物选取的数量规律。

定义 4 地物的质量是一个反映地物重要程度的量,它与地图的用途、主题、制图区域的特点以及地物的类型、尺寸等因素相关。

地物 P 的质量描述为函数 $I(P) = f(m_{\text{theme}}, P_{\text{type}}, P_{\text{size}})$ ^[5],其中 m_{theme} 为地图主题; P_{type} 为 P 的类型, P_{size} 为 P 的尺寸。若 P 是多边形地物,则 P_{size} 可以取 P 的面积。根据制图综合的原则,其保留的地物应当是图幅内最主要的和实质性的地物,即保留质量大的地物,而抛弃质量小的地物,且质量的大小是根据地物的 $I(P)$ 函数值来确定。本文将之称为无级比例尺 GIS 由大比例尺地图综合生成小比例尺地图的过程中地物选取的质量原则。一般而言,类型

(P_{type})与地图主题(m_{theme})较密切的地物以及尺寸(P_{size})较大的地物,其质量($I(P)$)也较大。同时,同一类型地物的各个子类本身也具有质量差别。

函数 $I(P)$ 的具体表达式可根据不同的应用目的来制定,但其基础均是面向空间对象的地理实体分类和无级比例尺 GIS 地理信息编码标准。自然界中的任何事物都具有整体上的统一性和内部结构上的差异性,具体表现为时间上的次序性和空间上的层次性,因此构成了完整、严密的地物分类分级系统。通过面向空间对象的地理实体分类和无级比例尺 GIS 地理信息编码,可对空间实体在质量和数量等属性方面进行等级化标定,以便使计算机能够判断出不同空间对象的主次、轻重地位,从而为地理要素的取舍提供客观、准确、科学的依据^[6]。

1.2 合并的原则

定义 5 岛分为孤岛和多岛。孤岛是由一条弧段构成的多边形;多岛是由多个相邻多边形 $\{P_1, P_2, \dots, P_m\}$ 构成,若将它们之间的公共弧段删除,则变成一个孤岛。多岛不包括组成它的多边形中的岛,若这些多边形完全落在另外一个多边形 P 内,则称 P 为多岛 $\{P_1, P_2, \dots, P_m\}$ 的湖。

定义 6 岛多边形是指构成岛的多边形。孤岛的岛多边形是其自身,而多岛则是由多个多边形构成。

由定义 5,多边形图层由若干多岛和孤岛构成,而且由于在选取处理的过程中质量较小的多岛多边形被删除,可能会使相邻多边形之间出现空洞,从而造成多边形图层的拓扑错误,为此,必须要进行合并(Combination)处理,即将要删除的质量较小的多岛多边形并入与其相邻的属性相同或相近的质量较大的多岛多边形。显然,上述删除操作只能施加于孤岛或整个多岛。

定义 7 满足以下条件的多边形图层称为是“镶嵌的”,即如果

- ①任意两个多边形之间不存在相互覆盖;
- ②图层内任意一点或者落在多边形边界上,或者落在唯一多边形内部,即图层内不存在无属性的“空洞”。

GIS 中的多边形图层总是满足条件①的,而且对于不满足条件②的图层,总可以通过添加一个或多个“背景多边形”使之成为“镶嵌的”。例如,全球范围内的陆地多边形图层可以通过添加一个海洋背景多边形使之成为“镶嵌的”。称此处理过程为多边形

图层的“镶嵌化处理”。这样通过多边形图层的“镶嵌化处理”，岛的删除操作可转化为与背景多边形的合并操作。以下认为所讨论的多边形图层总是通过“镶嵌化处理”成为“镶嵌的”，且背景多边形的质量取无穷大。

设在一个多边形图层中，与多边形 P 相邻的多边形为 $\{P_1, P_2, \dots, P_m\}$ ，有 $I(P) < I(P_i)$ ，其中， $P_i \subset \{P_1, P_2, \dots, P_m\}$ ，若要将 P 并入 $\{P_1, P_2, \dots, P_m\}$ 中的某一多边形，可采用以下合并原则，即计算合并函数 $C(P, P_i) = g(L(P, P_i), \text{Compatible}(P, P_i), I(P_i))$ [6]，其中， $L(P, P_i)$ 为 P 与 P_i 的公共边长度， $\text{Compatible}(P, P_i)$ 为 P 与 P_i 的类型相容性， $I(P_i)$ 为 P_i 的质量。将 P 并入函数取最大值的多边形 P_j ，其中， $P_j \subset \{P_1, P_2, \dots, P_m\}$ 。一般而言， P 总是并入与其公共边 ($L(P, P_j)$) 较长的、类型相容性 ($\text{Compatible}(P, P_j)$) 较大的和质量 ($I(P_j)$) 较大的相邻多边形。

函数 $\text{Compatible}(P, P_i)$ 是计算类型相容性的基础，是面向空间对象的地理实体分类和无级比例尺 GIS 地理信息编码标准。如“苔原”与“冻土带”的类型相容性要大于“苔原”与“工厂”的类型相容性。

1.3 属性数据的综合

在对多边形空间数据进行选取、合并等综合处理的同时，其属性数据也要进行相应的综合，即在保证数据完整性的前提下精简属性数据。这里的数据完整性是指地理特征的空间对象与属性数据记录保持一一对应。

当地理特征被删除时，其属性数据记录相应地也要做删除处理，而当地理特征发生合并时，若合并后的地理特征的类型取质量较大者，则可以直接删除质量较小者的属性记录；若合并后的类型本身也进行了综合，如“有林地”与“疏林地”合并后，其类型转化为“林地”，也就是说，若子类地理特征（“有林地”与“疏林地”）合并后，转化为父类地理特征（“林地”），则属性数据也要进行相应的综合处理，即保留父类地理特征的属性字段，而删除只属于子类地理特征的属性字段，但两者的基础都是面向空间对象的地理实体分类。

2 多边形图层数据组织策略 C-Tree

2.1 C-Tree 生成算法

根据选取的数量规律和质量原则以及合并的原则，以下给出针对选取与合并技术而设计的多边形图层的空间数据组织策略“合并树”(Combination-

Tree, 简称 C-Tree)。设多边形图层共包含 n 个多边形 $\{P_1, P_2, \dots, P_n\}$ 和背景多边形，记为 Q 。C-Tree 是一种二叉树结构，每个 C-Tree 节点的结构如下：

```
struct{
    Id/ * 多边形标识码 * /
    I/ * 多边形质量 * /
    TLson, TRson/ * 左右儿子指针 * /
    Flag/ * 标志域, 取值 0 或 1 * /
};
```

算法 1 多边形图层的合并树 C-Tree 生成算法

(1) 若 Q 中只有一个元素，则算法结束。

(2) 取 Q 中质量最小的元素 P_i ，即 $I(P_i) \leq I(P_k)$ ，其中， $i \neq k, 1 \leq k \leq n$ 。

(3) 设 P_i 的相邻多边形为 $\{P_1^i, P_2^i, \dots, P_m^i\}$ ，其中， $\{P_1^i, P_2^i, \dots, P_m^i\} \subset \{P_1, P_2, \dots, P_n\}$ ，首先计算合并函数 $C(P_i^j, P_i)$ ， $1 \leq j \leq m$ 。若 $\{P_1^i, P_2^i, \dots, P_m^i\}$ 中元素个数不唯一，且包含背景多边形，则背景多边形不参与计算，设 $C(P_j, P_i)$ 取得最大值，其中， $P_j \in \{P_1^i, P_2^i, \dots, P_m^i\}$ ；然后生成 C-Tree 节点 P_{ij} ， P_{ij} 的类型取 P_i 和 P_j 中质量较大者的类型，质量值取 $I(P_i) + I(P_j)$ ，最后将 P_{ij} 加入 Q ，并将 P_i 和 P_j 从 Q 中删除。

(4) 转步骤(1)执行。

经算法 1 处理，给定的多边形图层即生成一棵二叉树，称之为多边形图层的合并树 C-Tree。C-Tree 的每个叶节点对应于一个图层多边形，每个非叶节点也对应于一个多边形，后者是由以它为根节点的 C-Tree 子树的叶节点对应的图层多边形合并而成。

由此可见，算法 1 中 C-Tree 的生成过程实际上是多边形合并的过程，即算法经 $n-1$ 步结束，每一步一个多边形被合并，并产生一个 C-Tree 的非叶节点，而且被合并的多边形总是当前图层中质量最小的多边形，顺序记录此过程，即可得到一个包含 $n-1$ 个元素的线性表 C-Index，表中元素用于记录指向新产生的 C-Tree 非叶节点的指针。

由算法 1 易知，C-Tree 具有以下性质：

性质 1 C-Tree 的每个非叶节点一定有两个儿子，若多边形图层由 n 个多边形构成，则其 C-Tree 共包含 $2n-1$ 个树节点。

性质 2 C-Tree 的任一非叶节点的后代非叶节点在 C-Index 线性表中的位置肯定在此节点位置的前面。

在大比例尺单精度空间数据库中，每个多边形

图层空间数据对应于一个存储 C-Tree 的数据库表和一个存储 C-Index 的数据库表,它们可通过一次计算来完成并将结果保存在数据库中,以避免因生成数据结构和临时计算所造成的时间开销,而且每个 C-Tree 树节点在数据库表中仅保存多边形和左右儿子指针.显然,该 C-Tree 和 C-Index 的数据库存储开销和算法 1 的时空开销是非常小的.而多边形数据可仍采取传统的拓扑结构存储,如采用 ESRI ARC/INFO 的数据组织方式.

2.2 基于 C-Tree 的多边形综合算法

对于给定大比例尺多边形图层,设其比例尺为 M ,图幅数据量为 N ;而对于任意输出小比例尺 m ,则根据选取的数量规律,由式(1),即可计算出小比例尺输出图幅的数据量 n ,若其中 $m \leq M$,则 $n \leq N$.算法 2 用于完成 $N-n$ 个多边形的合并.

算法 2 基于 C-Tree 的多边形图层综合算法

(1) 将 C-Tree 全部节点的 *flag* 域标记为 0.

(2) 将 C-Index 的前个元素指向的 C-Tree 节点的域标记为 1.

(3) 算法结束.

算法 2 中的第 2 步标记了 C-Tree 的 $N-n$ 个非叶节点,若每个非叶节点对应了一个多边形的合并操作,则这 $N-n$ 个非叶节点就对应了算法 1 中的前 $N-n$ 个合并操作.另外,根据性质 2,算法 2 结束后,若 C-Tree 中某非叶节点的 *flag* 域标记为 1,则以此节点为根节点的 C-Tree 中的所有非叶节点的 *flag* 域的值都为 1.

算法 3 小比例尺多边形图层输出算法(以 C-Tree 的根节点指针作参数)

```
void Output(*p)
Begin
  If (p=NULL) Then Return;
  If (p->flag=1) Then
```

将与以 p 为根节点的 C-Tree 的全部叶节点对应的多边形合并,并输出.

Return;

EndIf

If ($p \rightarrow LSon = NULL$ AND $p \rightarrow RSon = NULL$)

Then

输出多边形 p ;

Return;

EndIf

Output($p \rightarrow LSon$);

Output($p \rightarrow RSon$);

End

算法 3 采用二叉树的深度搜索策略查询需要合并的多边形集合,其中,与背景多边形的合并实际上是岛的删除.在采用拓扑结构的 GIS 的空间数据组织方式下,这种合并操作很容易完成.对于包含 N 个多边形(不包括背景多边形)的多边形图层,若根据给定的任意小比例尺进行输出,则输出图层可能包含 0 个,1 个, ..., N 个多边形,这样即达到了最精细的输出粒度,也就是真正实现了无级比例尺处理.

C-Tree 是一种面向无级比例尺 GIS 的多边形图层数据的层次组织策略.C-Tree 的上层对应多边形图层数据的粗略表达,随着 C-Tree 的向下扩展,多边形图层的细节将逐步显露出来.C-Tree 的思想与 BLG-Tree^[6]、Strip-Tree^[7]等线综合数据结构的层次组织策略在思想上是相似的.

2.3 实验结果与结论

多边形图层数据的 C-Tree 组织策略与综合算法在时空一体化智能城建信息系统中得到了成功应用.如图 2 所示,该系统对多边形图层进行综合处理.设图 2(a)比例尺为 M ,在输出比例尺为 $0.6M$ 、 $0.3M$ 、 $0.15M$ 条件下,其综合结果分别见图 2(b)、(c)、(d).各输出图幅多边形统计数据见表 1.

表 1 输出图幅数据统计

多边形图层	比例尺	根据开方根规律计算的输出多边形数	图幅实际输出多边形数	被合并的多边形数	被删除的多边形数
原始比例尺地图数据	M	14	14	0	0
输出小比例尺地图数据 1	0.60	10.84	11	1	2
输出小比例尺地图数据 2	0.30	7.67	8	4	2
输出小比例尺地图数据 3	0.15	5.42	5	7	2

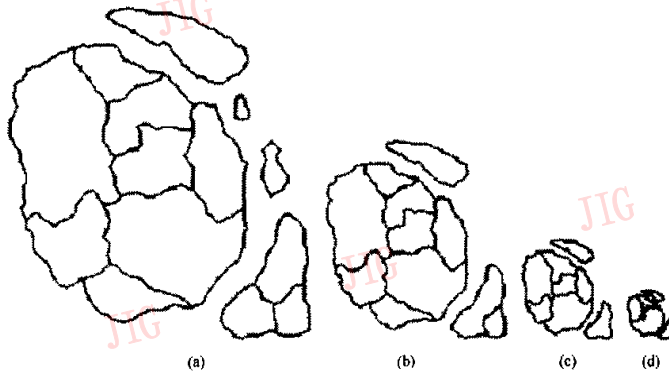


图 2 大比例尺多边形图层经算法 2 和算法 3 综合输出的小比例尺图幅

3 结 语

本文提出的多边形图层数据组织方法 C-Tree 是针对多边形综合中的选取与合并技术设计的,其组织策略是:每个多边形图层在大比例尺单精度空间数据库中对应一个存储 C-Tree 的数据库表和一个存储 C-Index 的数据库表,而组成多边形的弧段数据则根据需要,采用不同的线综合数据组织方法,这种策略在时空一体化智能城建信息系统中得到成功应用.在该系统中,根据给定的输出比例尺,首先按算法 2 和算法 3 对多边形图层进行选取与合并处理,然后对准备输出的多边形弧段进行线综合处理,从多边形对象个数和多边形弧段节点个数两方面精简多边形图层空间数据,其结果不仅十分理想,而且算法的时间和空间开销都非常小.

如前所述,地物质量和类型相容性的计算以及属性数据的综合都是以面向空间对象的地理实体分类和无级比例尺 GIS 地理信息编码标准为基础的,这也是国家空间信息基础建设的重要内容之一.我国目前正在大力推动此项建设的进行.

但由于多边形综合技术是十分复杂的,因此关于多边形综合中的夸张、移位、符号化等技术仍有待进一步研究.

参 考 文 献

- 1 承继成,李琦,易善慎.国家空间信息基础与数字地球[M].北京:清华大学出版社,1999:124~125.
- 2 Beard M K. How to survive on a single detailed database[A]. In: N. R. Chrisman edi, Proceedings of Auto-Carto 8 [C]. Baltimore, MD: ASPRS/ACSM, 1987, 211~220.
- 3 K A 萨里谢夫著.地图制图学概论.李道义,王兆彬译.北京:测

绘出版社,1982:129.

- 4 弗·特普费尔著.制图综合.江安宁译.北京:测绘出版社,1982,5.
- 5 Peter Van Oosterom, Vincent Schenkelaars. The development of an interactive multi-scale GIS[J]. IJGIS, 1995, 9(5):489~507.
- 6 承继成,林晖,周成虎等.数字地球导论[M].北京:科学出版社,2000:162~164.
- 7 Ballard Dana H. Strip trees: A hierarchical representation for curves [J]. Communications of the ACM, 1981, 24 (5): 310~321.



田 鹏 1972 年生,1995 年毕业于山东大学,1997 年获浙江大学计算机系硕士学位.主要研究领域为地理信息系统、空间数据结构和算法、空间数据库、自动地图制图综合等.



郑扣根 1964 年生,副教授,1991 年博士毕业于 University Of Warwick, United Kingdom, 1995 年在浙江大学博士后流动站,出站后为浙江大学计算机系副教授.主要研究领域为地理信息系统、计算机图形学、操作系统等.



张 引 讲师,1970 年生,1992 年、1995 年先后获兰州大学学士学位、硕士学位,1999 年获浙江大学博士学位.主要研究领域为地理信息系统、计算机图象与图形、智能 CAD 等.



潘云鹤 1946 年生,浙江大学计算机系教授,博士生导师,中国工程院院士,浙江大学校长.主要研究领域为人工智能、认知科学、计算机图形学、智能 CAD、地理信息系统等.