

矢量空间数据融合 方法研究

赵东保 著
付宇



中国出版集团



世界图书出版公司

矢量空间数据融合 方法研究

赵东保
付宇 著

中国出版集团
世界图书出版公司
广州·上海·西安·北京

图书在版编目(CIP)数据

矢量空间数据融合方法研究 / 赵东保, 付宇著. —

广州: 世界图书出版广东有限公司, 2015.5

ISBN 978-7-5100-9657-0

I. ①矢… II. ①赵… ②付… III. ①地理信息系统
— 数据融合 — 研究 IV. ①P208

中国版本图书馆 CIP 数据核字(2015)第 117984 号

矢量空间数据融合方法研究

责任编辑 黄利军

封面设计 高 燕

出版发行 世界图书出版广东有限公司

地 址 广州市新港西路大江冲 25 号

印 刷 虎彩印艺股份有限公司

规 格 787mm × 1092mm 1/16

印 张 10.375

字 数 182 千字

版 次 2015 年 6 月第 1 版 2016 年 1 月第 2 次印刷

ISBN 978-7-5100-9657-0/P · 0061

定 价 37.00 元

版权所有, 翻印必究

目 录

1 绪 论	1
1.1 研究背景	1
1.2 研究意义	2
1.3 国内外研究现状分析	5
2 数据融合匹配问题剖析	14
2.1 遥感影像的数据融合	14
2.2 GIS 矢量数据与遥感影像的数据融合	19
2.3 矢量要素自动匹配的类型划分	21
2.4 矢量要素自动匹配的数据预处理	24
2.5 矢量要素自动匹配的常用评判指标	25
2.6 矢量要素自动匹配的匹配策略	28
2.7 本章小结	32
3 多源兴趣点数据的自动匹配	33
3.1 同名兴趣点要素自动匹配的一般方法	33
3.2 基于优化函数的匹配	37
3.3 基于语义相似的匹配	39
3.4 本章小结	43
4 同名线要素的自动匹配——以道路网为例	44
4.1 道路网自动匹配的一般方法	44
4.2 全局寻优的矢量道路网自动匹配	52

4.3	实例验证与分析	60
4.4	矢量道路网匹配结果质量检核	65
4.5	本章小结	72
5	同名面要素的自动匹配	74
5.1	面要素自动匹配的一般方法	74
5.2	面要素的一对多自动匹配方法	83
5.3	实例验证与分析	90
5.4	本章小结	94
6	同名要素的位置纠正	95
6.1	同名要素上同名特征点自动匹配的一般方法	95
6.2	基于改进串匹配法的同名要素同名特征点的自动匹配	101
6.3	实例验证与分析	106
6.4	基于同名要素同名特征点自动匹配的位置纠正	114
6.5	本章小结	118
7	矢量要素的变化检测和更新——以道路网为例	119
7.1	现有道路网更新的一般性方法	120
7.2	基于浮动车轨迹点数据的道路网自动变化发现与更新	121
7.3	算法步骤	127
7.4	实例验证	129
7.5	本章小结	130
8	空间信息内容集成系统的设计与实现——以道路网融合更新为例	131
8.1	系统的总体设计	131
8.2	系统的运行实例——以道路网融合更新为例	134
8.3	本章小结	149

参考文献	150
------------	-----

1 绪 论

1.1 研究背景

随着空间数据采集手段的日新月异,人们已经积累了各种类型的空间数据,如何对这些数据进行有效融合与集成已逐渐成为地理信息科学需要迫切解决的问题之一。针对空间数据的集成,近些年来,人们在文件格式转换、空间数据模型的统一和空间数据互操作等方面已经开展了不少卓有成效的研究工作,然而对空间数据的内容本身,诸如位置、属性以及空间关系等方面的集成则较少涉及。在实践中,不同组织和部门出于不同需要常常会反复采集覆盖相同地区的空间数据,尽管理论上讲,这些空间数据是对同一个客观实体的描述,应当完全相同,然而由于所使用的仪器设备、测量手段不同以及采集过程中侧重点的不同,这些数据总是会存在着各种各样的差异。比较常见的差异,诸如同名地要素位置上的差异、附着在同名要素上的属性信息的差异、由于制图综合原因所造成的各种类型要素详细程度上的差异以及在采集数据时所处时段上的差异等。尽管存在着不少差异,但这些不同版本的空间数据又存在着各自的优势与劣势。因此,如果能够对这些不同版本空间数据进行有效融合与集成,就能够博采各个版本空间数据之所长,以获得一个集各种优势于一体的新的空间数据集,从而节省大量不必要的重复劳动和开支。出于这种对空间数据的集成、复用和共享的强烈需求,20世纪80年代就诞生了地图合并技术(Saalfeld, 1988)。

地图合并(Map Conflation)技术一般是针对矢量地图而言,它是地图编绘过程中的一个常见操作。在现阶段,关于地图合并,其较完备定义为:“在同名要素匹配的基础上,对相关要素的几何位置、空间关系和属性等进行调整,以

实现同一地区不同来源地图数据库的集成和信息融合。”由以上定义可以看出,能否快速和准确地对同名要素进行匹配直接关系到地图合并的质量高低,是其首要和关键所在。

矢量要素的自动匹配技术尽管起源于地图合并领域,但它在其他领域也同样具有重要基础性作用。例如,在地图变化检测领域,只有首先对同名要素进行准确匹配,才可以对地图的变化部分做出判断。又如,在最近蓬勃发展的“人人参与空间数据采集计划”(Goodchild, 2007)中,只有在识别出同名要素后,才可以对同名要素进行一致化处理并实现信息的共享,鉴于要素匹配技术的重要性,故而有必要针对矢量地图中同名要素的自动识别与匹配所涉及的理论与技术开展专门研究。

1.2 研究意义

当前许多 GIS 应用项目所面临的一个主要问题就是如何将不同来源的数据集成在一起 (Devoegele, 1998)。集成不同来源的信息是 GIS 的一个基本特征,但这并不意味着将不同来源的数据简单堆砌在一起就可以解决问题。要素匹配技术可以大大提高 GIS 系统集成不同来源空间数据的能力,可以说是空间数据集成领域的关键技术之一,其理论意义及实践意义归纳起来主要为如下三点。

(1) 基于要素匹配实现空间数据的有效融合

数据的采集和更新是 GIS 工程中耗时耗力的一项工作,尽管人们在这个方面已经花费了大量的时间和金钱。当今的许多 GIS 应用系统一方面仍然存在数据不足和更新周期过长的问题,另一方面现有数据又得不到有效再利用,许多存储在各个数据库的现有数据在语义、存储格式、数据模型、编码方案等方面均可能不同,再加上这些数据有关文档也可能不全,使得这些数据的使用只能局限于原来的应用部门或原来的应用项目,形成一个个封闭的“数据孤岛”。如今,许多 GIS 工程项目所需要的数据仅来自于一个部门是难以满足所有应用要求的,例如同样是道路数据,测绘部门采集的数据位置精度很高,而交通部门采集的数据则会更具有丰富的道路属性。显然,这两者各有所长,又各有所短,舍弃任何一个都非常可惜,而如果一项应用需要结合两方面的数据,就必须克

服上述“数据孤岛”所带来的问题,找到一种合适的空间数据融合与集成的手段。然而,目前大部分的商业 GIS 软件在这个方面的功能还都较弱,只提供了数据格式转换、坐标系统变换等基本的数据集成功能,即便是以专注于空间数据集成而著称的 FME 软件也同样缺乏基于内容集成的能力。当人们需要对空间数据的内容进行融合和集成时,主流的 GIS 平台软件,例如 ArcGIS,还仍然主要依赖于作业员的目视和手工编辑来消除不同来源空间数据的逻辑上的不一致性,因而效率低下。Goodchild (1996) 指出,随着空间数据的日益丰富,用户将经常会为了特定目的而需要使用多种来源的空间数据,在这种情况下,用户会希望将这些数据进行某种形式的平均或匹配。然而,想要实现空间数据的融合,首先要解决的问题就是要确定对哪些数据进行融合,换言之,就是确定不同来源地图同名要素之间的匹配关系,只有确定了同名要素,才可以相互交换各种信息,实现所谓的融合。由此可见,要素匹配问题是空间数据融合问题的关键所在。

(2) 基于要素匹配实现地图的变化检测和自动更新

地图更新问题是地图生产的核心与关键所在,由于以往很长一段时期内空间数据的采集本身就需要耗费大量时间,因此实现地图的自动与快速更新的需求并不十分强烈。近年来,随着空间数据采集技术的迅猛发展,各类空间数据源的获取周期愈来愈短,并正在逐步地达到实时化的目标,此时如何利用新的数据源更新原有地图就日益引起人们的关注。今天, TM 遥感影像数据空间分辨率已经可达 0.6 米的精度,而与此同时,各种多波段,多时相的遥感影像卫星也层出不穷,蓬勃发展的数字摄影测量技术更是大大提高了传统测绘的工作效率,地图的生产周期日益缩短。在 GPS 领域, RTK 测量技术可以快速地获取点位坐标,其点位精度也早已可达 2 个厘米之内,而精密单点定位技术也在蓬勃发展。如此众多的高精度和高现势性的数据正在呈几何级增长,如何将这此新数据融入到已有的数据库中,为用户提供更好的数据服务是数据管理和维护者面临的一个重要课题。正如 Saalfeld (1996) 指出的那样,所有这些新的数据,无论它们精度有多高,只有将它们融入到已有的数据库中,才能显示出它们的价值。

在现实实践中,地图更新更多的是针对 GIS 中的矢量地图数据库而言的,对矢量地图的更新可以分为利用遥感影像更新矢量地图和利用矢量地图更新矢量地图两种,而其中的后者又可以细分为以下四种情况:第一种情况是某部门对某区域进行了空间数据的最新采集,此时需要更新该区域的原有数据。第二种情况是某部门所使用的地图(从数据库)一直是购买自另一个部门(主数据

库),假定从数据库已经在该地图上进行了各种编辑操作,现主数据库发生了更新,那么从数据库又该做何种更新。第三种情况是某部门已拥有某地区不同时段、不同尺度的地图,但在过去并未建立这些不同地图之间的联系,而现在需要对地图的同名要素进行匹配,从而形成地图的版本化,这种情况也属于地图更新的范畴。第四种情况是由大比例尺地图的更新级联推理中小比例尺地图的更新,此时需要获知大比例尺地图出现了哪些更新,而大比例尺地图又和中小比例尺地图之间存在着怎样的要素匹配关系,无论是以上哪种情况,欲实现地图更新,就必须要对地图变化的部分进行自动检测,可以说地图更新和变化检测是一个问题的两个方面,二者相辅相成,而变化检测是建立在要素匹配基础之上的。容易理解,当旧地图中的某个要素可以在新地图中找到与其匹配的要素时,则可认为该要素没有变化,反之就是变化了的要素,这时就可用新的数据将其替换,从而也就实现了地图的自动更新。

(3) 基于要素匹配实现空间数据的质量评估和位置校正

空间数据质量是 GIS 应用系统的生命线,当数据质量无法得到保证时,所做出的各种分析推理都将缺乏可靠的依据。在实践中,可能会遇到这种情况,即虽然手头并不缺乏空间数据,但却不知道所拥有空间数据的质量情况,或者仅仅只是了解一个大致信息。对于这种情况,如果同时还拥有该地区其他版本的空间数据,就可在对同名要素进行匹配后,通过对同名要素进行比对,可利用已知数据质量的空间数据对未知数据质量的空间数据进行质量评估,显然,不同版本的同名要素可以视为不等精度下或者同等精度下的多次独立观测,对于多次独立观测,经由测量误差理论可进行精度评定、消除残差、粗差定位等,从而对不同版本的空间数据做出相对质量评估。

除了可对空间数据质量进行相对评估外,更重要的是还可以利用高精度的空间数据对低精度的空间数据进行位置校正。在识别出同名要素后,考虑各种约束条件,诸如要素的形状、长度、角度等等限制条件,可采用一定坐标转换模型对相关要素进行带限制条件的最小二乘平差,从而实现对矢量要素几何位置的精确纠正。另外,空间数据的质量不仅仅表现在空间位置准确性上,还表现在其他方面,如逻辑一致性上。例如,根据道路网网络分析的要求,加油站应该座落在道路上,但是实测数据可能并不能满足这个要求,此时可按照一定准则,将加油站匹配到道路上的一个最合适的位置上。显然,所有这一切都必须以要素匹配技术作为基础。

1.3 国内外研究现状分析

矢量要素的自动匹配,就其匹配方法而言,无外乎是从几何特征或者语义特征两个方面入手。几何特征主要包括位置、形状、拓扑结构等,语义特征主要包括属性信息等。下面将分别从国内外两个角度就基于几何特征的点要素匹配、线要素匹配、面要素匹配以及基于语义的要素匹配等四个方面对矢量要素自动匹配技术的研究现状进行归纳和总结。

1.3.1 国外研究现状

(1) 点要素的自动匹配

这里所指的点要素一方面是指一些独立的点状地物,另一方面也是指具有重要标志意义的特征点,例如道路的交叉点、面状地物的质心等。点要素匹配的一种简单思路是直接根据先验知识进行判断,假定已知两幅地图各自的点位精度以及比例尺等信息,则由测量误差理论可推导出同名点的最大距离误差,将此距离误差作为阈值即可对点要素是否匹配做出判断。不过在实际情况中,由于点要素可能过于密集,再加上同名点的距离误差又较大,导致某个点要素存在多个潜在的候选匹配要素,这个时候就必须寻找一种可靠的匹配方法使得匹配结果尽可能正确。Beerli 等人(2004)针对独立点状地物,依据模糊集理论,提出了一种通过计算可信度值来对点要素进行匹配的方法。随后,Beerli 等人(2005)又对当点要素并非来自于两个图层,而是多个图层时的情况进行了讨论,进一步扩充了基于模糊集理论的点要素自动匹配方法。在地图合并中,点要素更多的是指那些具有重要标志意义的特征点诸如道路交叉点,当对这些点要素进行匹配时,除了可以考虑距离因素外,还可以考虑拓扑结构和方位关系等特征信息。例如 Saalfeld(1988)在对道路结点进行匹配时,就设计了一种可以称之为“蜘蛛编码”的八字节结点拓扑结构二进制编码方案,蜘蛛编码将道路结点所关联弧段的延伸方向分为八个连续不相交的角度区域,并据此对道路结点进行匹配。Samal(2004)给出了一种基于上下文的要素匹配方法,该方法

根据已知点要素的匹配结果,依据其他点要素与它们之间的相对方位关系,推导其他点要素可能的匹配关系。除此之外,在计算机视觉和模式识别领域,点的匹配问题被专门称作“点模式匹配”问题(Point Pattern Matching),其中不少经典算法均可以借鉴,诸如 Ballard (1981), Fischler (1981), Olson (1997), Chui (2000) 等。

(2) 线要素的自动匹配

线要素的自动匹配是地图合并中最为活跃的一个研究方向,甚至可以说是地图合并技术产生的源动力。早在 1980 年,美国人口调查局为了完成十年一次的人口普查,数字化了全国 278 个大中城市的纸质地图,共有 5500 多幅图,每幅的面积为 25 平方英里,这些地图主要是一些线要素,含有丰富的属性信息,如名称,街区地址范围,街区类型等。1983 年,美国人口调查局又获得了美国地质测量局的高精度电子地图数据,其中仍然主要是线要素,但没有属性信息。人口调查局想将已有的数字化地图与新的美国地质测量局的地图进行地图合并,以便能检测并消除两幅地图中存在的不一致现象,并将属性信息传递给美国地质测量局的地图数据库。由于要匹配和合并的数据量很大,又必须保证匹配和合并的正确和快捷,使得有必要建立一个自动的地图合并系统。当时,美国地质测量局的地图数据并没有包含属性信息,因而要素匹配只能依靠一些几何测度来进行,为此 Saalfeld 等学者(1988)开发了世界上第一个数字地图合并系统。他们提出了著名的“橡皮筋技术(Rubber Sheeting)”,并利用该技术对地图中结点的位置不断进行调整和迭代,从而获取同名结点之间的匹配关系,并进而再获取线要素之间的匹配关系。橡皮筋技术后来还被引入到主流商业 GIS 软件中,如 ArcGIS 等。不过当时的数字地图合并系统还只能对线要素 1:1 的匹配关系进行处理。

由于道路网是一类非常重要的线状要素,它和人们的生活息息相关,进入到 20 世纪 90 年代以后,对于线要素的自动匹配的研究大都集中到道路网的自动匹配上。这方面的代表性文献如:Walter 等学者(1999)在对道路网的匹配类型进行全面总结的基础上,直接依据道路网的弧段而非结点进行道路网匹配,其主要思想是:首先利用所谓的缓冲区增长算法确定候选匹配弧段集合,然后给定一个互信息函数,互信息函数的取值大小不但与匹配弧段之间的形状、位置和拓扑等特征的差异有关外,还与它们邻域范围内的其它弧段的匹配情况有关,然后采用 A*算法对互信息函数进行求解,并将最优解所对应的道路匹配关

系作为最终解。该方法虽匹配效果较好,且可应对 $M:N$ 的匹配情况,但却非常耗时。Mantel (2004) 又进一步对 Walter 提出的缓冲区增长算法进行了改进,一定程度上加快了搜索候选匹配道路弧段的速度。Walter 的直接利用弧段进行匹配的思想引起了广泛关注,例如, Davis 等人 (2003) 在他们的开源产品 JCS (Java Conflation Suit) 中通过计算弧段之间的一种顶点 Hausdorff 距离来判断道路之间是否匹配,距离越小的则匹配的可能性越大。而 Deng (2007) 更是对传统 Hausdorff 距离进行了扩充,使其具有更强的鲁棒性和抗噪能力,能够容忍目标局部形状的变化,可用来判定非一对一的匹配模式。除了常见的 Hausdorff 距离之外,更为常见的线状要素距离测度还有 Saalfeld (1993) 提出的一种 L2 距离测度, L2 距离实质上是对 Fréchet 距离的一种简化,将两条曲线之间的 Fréchet 距离(也叫皮带距离)定义为所有单调参数化路径中的最小路径长度。这是最自然的距离测度,但由于需要处理所有可能的参数化方法使其计算非常复杂,其时间复杂度高达 $O(pq \log_2(pq))$, 其中 p 和 q 分别为两条曲线的线段数目。而 L2 距离则仅需线性时间,在计算效率上要快捷得多。不过,利用道路弧段之间的距离测度判断道路是否匹配只适用于待匹配的两组道路网在位置上极为接近的情况,当同名道路在位置上存在较大偏差时,很容易导致误匹配。

由于道路网中道路交叉点和道路弧段都蕴含着十分重要的特征信息,事实上人们更多的是将结点和弧段加以综合考虑。例如, Filin 等人 (2000) 所提出的方法,首先根据道路网中结点的位置、拓扑等特征信息对结点进行初匹配,然后再根据结点所关联的弧段采用一种称之为“Round trip walk”的方法对匹配结点做进一步调整。具体做法是从待匹配道路网的一点 A 沿其关联弧段走到其所有邻接点 $B, C \dots$ 等处,找到 B, C 等点在参考道路网中的匹配结点 B', C' 等,如果从 B', C' 等点出发存在关联弧段可返回 A' , 那么 A 与 A' 之间的匹配关系便可得到确认,反之则予以修改。如此这般,反复迭代,从而可以获得更为可靠的结点匹配关系,以此为基础再获得道路的匹配关系。不过该方法严重依赖于结点的初匹配结果,当存在较多结点初匹配不准确时,后续的迭代调整很难进行。Safra (2006) 针对移动环境下对道路网匹配的高效率要求,提出了一种简单、快速的方法,不过其方法显然只能针对待匹配的两组道路网在几何位置上甚为接近的情况。Volz 等人 (2006) 所提出的方法则是首先设置严格的阈值,而后找出满足严格阈值条件下的匹配结点,并将这些匹配结点作为种子点,从这些种子点出发根据弧段之间的形状、距离等特征信息再找出满足严格

阈值条件的匹配弧段, 然后根据已经获得匹配结果放宽阈值的限制范围, 再次寻找匹配结点和匹配弧段, 如此这般, 反复迭代直至没有新的匹配结点和匹配弧段出现为止。该方法的问题是, 它在各个迭代阶段所设置的阈值完全是针对特定数据进行统计调查后依据经验而得, 并不具有通用性。相比较而言, Zhang (2007) 所提出的方法适用范围更广一些, 其方法在初始阈值确定上仍然是根据先验知识获得, 但随后在匹配过程中利用数理统计分析方法对初始阈值进行优化。但是, 该方法只能适用于待匹配的两组道路网其同名道路在位置上存在大体一致的偏差情况下, 当同名道路位置上的偏差在某些地方大, 而在另一些地方小, 即存在非一致偏差的情况时, 该方法将很难奏效。Zhang (2006) 还专门针对大比例尺地图中的道路匹配问题进行了研究, 在大比例尺道路网中, 道路并非是道路中心线, 而是双线, 且在交叉口的地方道路非常复杂, 再加上还存在立交桥的现象, 在这些情况下道路的匹配要更为困难一些。最近, Mustière (2008) 针对多尺度下道路网匹配问题进行了研究, 他们的方法也是首先对道路结点进行匹配, 在获得了道路结点的匹配关系后, 此时, 道路匹配问题就转化为从多个候选匹配道路弧段中寻找一个路径, 使得这个路径所对应的道路与待匹配道路在某个指标上 (如形状、距离等) 最为相似, 而这个问题实质上对应于图论中求两点之间最短路径的问题。他们所提出的方法不但可以处理道路弧段之间 $1:N$ 的匹配关系, 还可以处理道路结点之间存在的 $1:N$ 匹配关系, 不过在该方法中, 道路结点的候选匹配结点通常数目很少, 即待匹配的两组道路网数据在位置上也同样是甚为接近的, 当道路结点的候选匹配结点过多, 他们的算法也将难以应对。此外, Kang (2004) 对多尺度道路网匹配问题也进行了研究, 他们所提的方法是首先根据大比例尺地图和中小比例尺的最小显示范围, 对大比例地图进行制图综合, 合并或者简化一些道路结点与弧段, 综合后再与中小比例尺的道路网进行匹配, 但是其中又将面临制图综合这个国际难题。

除了道路网的自动匹配外, Blasby 等人 (2003) 在他们的开源产品 JCS 中也对河流网的匹配问题进行了研究, 他们的方法充分利用了河流网在拓扑结构上非常类似于一个树结构这一重要特征, 其匹配首先从河流网的主干河流开始, 然后逐步扩散到分支河流。不过, 河流网并不一直都是树结构, 相对而言, Kieler (2009) 所提出的方法更具通用性, 该方法首先对呈面状的河流构建 Delaunay 三角网, 然后据此提取河流中心线, 将河流组成一个网络, 再采用类似 Mustière (2008) 所述方法进行多尺度下河流网的自动匹配。

(3) 面要素的自动匹配

面要素的自动匹配方法可归纳为以下三类。

第一类方法是将面要素的匹配转换为点的匹配,通过对比面要素对应顶点或者质心之间的距离、点与点以及点与面的拓扑结构来确定面要素的匹配关系,例如,Yuan (1999)利用数学形态学方法提取面要素的质心,当两个待匹配的面要素的质心分别可以落在对应的面要素区域之内,且两个面要素在一些形状特征指标上,例如致密度等又较为接近时,即可认为二者为匹配面要素,Masuyama (2006)所提出的方法则将面要素的最大内切圆圆心作为面要素的代表点,通过判断待匹配面要素与匹配面要素的代表点是否分别包含在对方内部来进行匹配。

第二类方法是考察面要素相互之间的重叠程度,或者是对面要素进行缓冲区提取后,它们缓冲区的重叠程度。显然两个面要素重叠面积越大,它们是同名地物的可能性就越大,采用这种方法不但能实现面要素的一对一匹配还能实现更为复杂的一对多和多对多匹配,多数文献都是采用这种方法进行面要素的自动匹配,例如, Van Wijngaarden (1997)在对多种尺度的地图进行级联更新时, Von Goesseln 和 Sester (2003)在对 ATKIS 数据源中的空间数据进行变化检测、集成和自动更新时都采用了此类方法,第二类方法的问题在于:它需要对矢量多边形进行求交运算,算法不但复杂,而且也较为耗时,且该类方法对面要素的位置准确性依赖过大,只适用于两幅地图在位置上极为接近的情况。

第三类方法则是计算面要素在形状上的相似程度,其相似性测度主要是一些简单的形调函数,形状上愈相似的,则愈可能匹配,VIVID 公司开发的 JCS 软件就提供了多种面要素的形状相似度量方法,可应用于面要素的自动匹配中。计算多边形之间的形状相似度的方法在计算机视觉中更是得到了广泛研究,代表性方法如 Zhang (2003)所介绍的各种傅里叶描述子算法、Arkin (1991)所提出的正切形调函数算法以及 Liao (1996)所介绍的各种矩描述子算法。第三类方法虽然对面要素位置准确性的依赖程度有所减低,但现有算法大多是通过面要素的轮廓信息计算面要素的形状相似度,这就使得这些算法难以胜任面要素之间的非一对一匹配。另外,现有的各种形状匹配算法还不够成熟,与人眼的识别能力相比尚有较大差距,当形状较为复杂、存在较多噪音时,形状相似度的计算还不够可靠。

(4) 基于语义的要素匹配

除了通过几何和拓扑等信息来完成同名要素的自动识别与匹配外,在实践

中,很多时候要素还包含丰富的语义信息(如名称、类别等)可以加以利用,如果再综合考虑语义信息则可进一步提高要素匹配的正确率。为此,Cobb(1998)基于模糊逻辑和证据理论对文本型属性和数值型属性的语义相似度分别给出了具体的度量方法,并将其成功运用到美国标准矢量文件格式 VPF 数据的自动匹配中,不过该方法有一个限制前提,那就是待匹配的两幅地图其要素的语义模式必须是一致的。Uitermark(1999)针对两幅不同来源的地图,根据本体论分别对它们的语义模式进行分析,并推导出语义之间的匹配关系,基于此再进行要素的匹配。Foley(2000)则建立了一个基于知识的模糊推理体系,通过一系列预先设定的知识和规则,对要素的语义是否匹配做出判断。Rodriguez 和 Egenhofer(2003)提出了一个基于本体的语义匹配相似性度量(即语义距离)方法,随后,Samal(2004)等人对其进行了改进,提出了一个改进的语义距离度量方法,并将其应用于空间数据集成领域,改进后的方法不仅能度量语义表达中单个字符串的相似性,而且还能度量多个字符串的相似性。Volz(2005)则首先利用几何信息对要素进行匹配,根据要素匹配的结果再推导要素之间语义的联系,从而可以获得不同来源地图语义模式的关系,实现了不同来源地图语义上的集成。Kieler(2008)也采用了类似的方法,即首先选出一部分样区仅根据其几何信息或者人工辅助来进行要素匹配,然后利用空间挖掘技术推导出要素之间语义上的关联,利用该关联关系可辅助推导其他要素之间的匹配关系。由于不同领域有着各自的术语、命名习惯和分类体系,导致对同一个概念,同一个要素常常存在语义冲突,相对而言,在不少情况下利用语义进行要素匹配就不如利用位置、形状和拓扑结构等几何信息对要素进行匹配更为可靠,因此,在要素匹配中,语义信息通常都是起辅助性作用。对其研究也相对少得多。不过,语义匹配技术在本体论以及自然语言理解等领域得到了较为充分的研究,相关内容可资借鉴。

1.3.2 国内研究现状

与国外相比,国内对于地图合并的研究起步较晚,大概从2000年左右才开始引起人们的注意,总体而言,研究文献还不够丰富。不过最近几年,越来越多的学者开始关注这个问题,文献的数量和研究深度均有明显进步。早在2001年,李德仁和张桥平等就对数字地图合并问题进行了较为全面的论述,从宏观角度指明了地图合并中需要重点解决的问题。而张桥平(2002)还对数字地图

合并技术进行了较系统的研究：针对线要素的匹配，提出了基于中间面积法的线要素距离测度和基于方向变化角的线要素形状相似性测度；对面要素的匹配，通过计算面要素之间的重叠面积来确定面要素之间的对应关系，而后根据形态距离再确定面要素之间的模糊拓扑关系，借此刻画面要素之间互为同名地物的模糊程度。此外，他还提出了一种基于同名点拓扑关联的局部坐标变换方法，较好地避免了矢量要素合并过程中可能带来的变形或者失真等问题。童小华等（2007，2008）也是国内对数字地图合并技术开展较多研究的学者之一，对于要素匹配，他将 Beerli 等人（2004）针对点状要素所提出的基于概率可信度的匹配方法扩展到其他几何类型的矢量要素，通过综合考虑各种几何特征指标，使得匹配结果更为准确。他还将最小二乘平差理论引入到地图合并中，在对实体坐标进行局部坐标变换时，同时顾及实体自身的面积、长度、角度和形状等限制条件，使其在坐标变换后这些特征尽可能保持不变，避免了地图合并后图形过度变形等现象的发生，对于误匹配则采用测量平差理论中的抗差估计对误匹配点进行处理，可实现对粗差也即误匹配的有效定位和消除。除张桥平、童小华等人外，国内的其他学者也对数字地图合并技术的各个侧面开展了研究，例如，在道路网匹配方面，眭海刚（2002）讨论了同名线要素的匹配问题，提出了基于缓冲区的同名线要素匹配技术，借此可实现对道路网变化部分的自动检测，陈军等（2007）专门针对道路数据的缩编更新问题开展了研究，他们提出了基于网眼密度的道路选取方法以及顾及层次分析的 1:10000 和 1:50000 两种比例尺道路网的匹配方法，以此为基础实现了道路网数据缩编更新的自动综合处理。陈玉敏等（2007）针对多尺度道路网的匹配问题，提出了一种结点距离匹配方法，将复杂的折线与折线之间的几何相似度计算转换为求结点到折线的距离，从而较大地降低了计算复杂度。张韵等（2008）针对矢量格式的道路网变化检测问题（实质上这也是一种匹配问题），通过对传统的平面扫描线法进行改进，提出了一种高效的计算道路网几何信息差异的算法。在面要素匹配方面，郝燕玲等（2008）给出了一种通过计算面要素之间的形状相似度来对面要素进行匹配的方法，形状相似度是借助所谓形状描述函数来进行计算，并综合考虑了面要素的位置、尺寸等几何特征，章莉萍等（2008）在研究相邻比例尺的居民地要素变化规律的基础上，提出了一种所谓“增量式凸壳匹配方法”，该方法可以较好地对多尺度的居民地要素进行自动匹配。郭黎等（2008）采用方向关系矩阵模型进行空间方向关系的描述和计算，据此给出了基于空间方向相

似性的面要素匹配算法流程。在矢量要素的边界匹配方面,邓敏(2005, 2007, 2008)基于分解与组合的思想,在描述“线/线要素”“线/面要素”以及“面/面要素”拓扑关系的基础上又纳入了局部方向和局部距离关系信息,实现了对矢量要素空间关系的集成表达,其方法可以更好地探测矢量要素边界信息的不一致现象。艾廷华(2000)针对相邻多边形共享边界的一致化改正问题进行了研究,依据相邻多边形之间的空间关系,将共享边界的不一致现象共区分为六种类型,并通过提取 Delaunay 三角网的骨架线来对共享边界的不一致现象进行改正。在矢量要素的合并变换方面,臧德彦等(2007)针对矢量地图合并时同名点是否准确和可靠的问题进行了研究,给出了同名点可靠性统计检验的方法。唐文静等(2008)基于多评价因素分别给出了点要素、线要素的合并变换方法。除以上外,还有一些学者的研究工作也与要素匹配技术有较密切的联系,例如,丁虹(2004)通过引入本体论和空间认知的思想,对空间相似性的理论进行了研究,分别建立了空间方向相似性、空间拓扑相似性、空间语义相似性和空间场景相似性等计算模型。刘东琴等(2005)以国家基础地理信息数据库中全国 1:250000 和全国 1:1000000 比例尺的公路和河流要素为实例,阐述了利用位置匹配来建立多个数据库相互联系的思路和方法。王育红等(2008)针对 GIS 中生产单位所提供的主数据库与应用单位的从数据库之间的模式匹配问题进行了研究,提出了一种改进的基于实例的层次式模式匹配方法。艾廷华(2009)利用傅里叶描述子提出了面向形状相似性的空间查询算法,增强了 GIS 空间分析的功能。闫浩文(2009)对多尺度地图空间相似关系的一些基本问题进行了探讨。应申(2009)基于要素匹配技术给出了基于版本数据库的变化信息提取和更新发布的技术路线。徐枫(2009)对已有的各种要素匹配技术进行了分析、比较和综述。此外,在语义匹配方面,张雪英(2008)、李红梅(2009)和徐爱萍(2009)等都对如何计算语义相似度给出了一些方法,尽管这些方法并不是特别针对要素匹配领域的,但对于该领域的研究都具有借鉴价值。

1.3.3 当前研究中存在的问题

以上对国内外研究现状进行了较为全面的归纳和总结,所分析的主要是国内外与地图合并技术有直接和密切关联的文献。当然,其他研究领域中的有关文献也值得借鉴,这些领域包括地图更新(蒋捷,2000)、变化检测(吴建华, 此为试读,需要完整PDF请访问: www.ertongbook.com)