

文章编号: 1007-4619 (2004)02-0107-06

城市地图数据库面实体匹配技术

张桥平^{1,2}, 李德仁^{1,2}, 龚健雅²

(¹ 武汉大学遥感信息工程学院, 湖北 武汉 430079;

² 武汉大学测绘遥感信息工程国家重点实验室, 湖北 武汉 430079)

摘要: 同一地区不同来源地图数据库同名面实体的识别或匹配对空间数据库的集成与信息共享非常有意义。该文研究了城市地图数据库同名面实体匹配的有关问题, 提出了基于模糊拓扑关系分类的面实体匹配方法, 该方法充分考虑了源地图数据库的不确定性, 能处理非一对一的匹配情况。匹配结果不仅可以作为城市地图数据库集成和信息共享的依据, 还可以用来分析两个地图数据库之间的差异, 进行变化检测与自动更新。

关键词: 地图数据库合并; 实体匹配; 形态距离; 模糊拓扑关系

中图分类号: P208 **文献标识码:** A

1 引 言

一个地区或城市通常由不同的部门负责采集地图数据并各自建立相应的地图数据库, 如何将这此不同部门的地图数据库集成或合并, 从而达到信息共享、优势互补, 是今后地理信息系统(GIS)工程建设的一个难点^[1,2]。实践表明, 地图数据库的集成和信息共享并不能简单地通过数据格式转换或坐标全局变换等方式来实现, 而需要通过更进一步的地数据库的实体匹配与合并等技术来解决^[3]。简单地说, 地图数据库合并(Conflation of Geographic Databases)是指在同名实体匹配的基础上, 将同一地区两个或两个以上不同的地图数据库合并成一个数据库, 新生成的地图数据库在某些方面如点位精度、详细程度或现势性等比源数据库好^[4]。

同名实体(Identical Entity/Counterpart Element)是指两幅图或两个数据库中反映现实世界同一地物或地物集的空间实体, 同名实体在不同来源的地图通常都存在着差异, 这种差异是由于制图误差, 不同应用目的或不同人的解释差异以及制图综合等因素的影响而产生的^[3-5]。同名实体的识别或匹配(Identical Entity Identification/Feature Matching)就是

通过分析空间实体的差异和相似性识别出不同来源图中表达现实世界同一地物或地物集(即同名实体)的过程。它是地图数据库合并的关键技术之一^[4]。

在大比例尺城市地图中, 面状地物要素往往占有比较大的比例, 如房产图、地籍图、地形图中的建筑物或房屋实体, 它们不仅是地图表达的主要内容, 也是地图使用者关心的主要内容。但是, 已有文献对面实体的匹配和合并的研究还比较少。虽然从几何的角度来说, 点连成线, 线围成面, 处理面实体的匹配可以基于点或基于线实体的匹配来实现, 但是基于点和线匹配的面实体匹配方法并不能完全解决面实体的匹配问题, 比如, 单纯借助点实体或线实体的匹配算法, 就无法处理面实体之间非一对一的匹配情况。另一方面, 直观上同一地区不同来源地图数据库中同一面状地物的图形实体(即同名面实体), 可以根据实体间的位置(距离)、形状、大小、拓扑关系, 属性数据等多方面来识别, 但是总的来说, 面实体之间的相似度或者只适合于一对一的情况(如距离相似度)或者还不成熟(如形状相似度、语义相似度)^[6,7]。因此, 目前还没有能比较理想地解决同名面实体的匹配问题。同一地区不同来源地图数据库中同名面实体可能的匹配关系可以概括为表1的四种情况。图1示意了其中的三种复杂的情况。

收稿日期: 2002-08-13; 修订日期: 2003-09-17

作者简介: 张桥平(1971-), 男, 博士, 副教授, 毕业于武汉大学摄影测量与遥感专业, 主要从事空间数据库集成理论和方法研究, 已发表论文10余篇。

表 1 不同来源地图中同名面实体之间可能存在的匹配关系

Table 1 The matching relationships between identical areal entities from different sources

序号	匹配关系	说明
1	1:1	两幅图中都有,并且是一一对应的关系
2	1:0 或 0:1	一幅图中有,另一幅图中没有
3	1:N 或 M:1	一幅图中的一个实体与另一幅图中的多个实体对应
4	M:N	一幅图中的 M 个实体与另一幅图中的 N 个实体对应

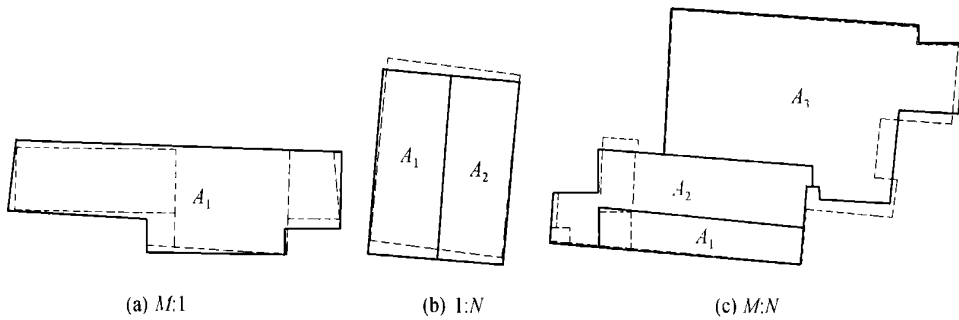


图 1 面实体匹配可能的复杂情况

Fig. 1 The matching relationships between identical areal entities

2 基于模糊拓扑关系分类的面实体匹配方法

2.1 基于重叠面积的候选匹配实体集的确定

候选匹配实体集的确定是地图数据库同名实体中的一个重要过程,它不仅将决定实体匹配的效率,还将决定匹配的正确性和完整性。同一地区不同来源的地图(即使是不同比例尺)在消除整体和局部坐标偏差后,同名面实体总是会有较大的重叠面积;面实体之间的重叠面积大小反映了两个面之间的距离差异以及形状差异等,而且它反映了两个面实体之间的整体相似性,符合人眼的视觉效果。因此,本文提出根据面实体之间重叠面积的大小来确定待匹配面实体的候选匹配实体集。

假设, A_1A_2 为来自不同来源地图数据库的两个面实体,我们用式(1)计算它们之间的相似度:

$$Sim(A_1, A_2) = \frac{Area(A_1 \cap A_2)}{Area(A_1)} \quad (1a)$$

$$Sim(A_2, A_1) = \frac{Area(A_1 \cap A_2)}{Area(A_2)} \quad (1b)$$

当两个相似度中有一个大于一定阈值(比如 0.5)时,意味着它们存在匹配的可能;将 A_2 的多边形标识号连同根据重叠面积计算的相似度值加入到 A_1

本文提出如下匹配策略:先由面实体之间的重叠面积确定两个面实体(集)之间可能的对应关系,然后计算两个面实体(集)之间的形态距离(morphological distance),并由形态距离的模糊分类确定两个面实体(集)之间的模糊拓扑关系,最后根据模糊拓扑关系确定同名面实体匹配结果。该方法的优点是不仅能处理非一对一的复杂情况,而且还充分考虑了面实体本身的不确定性,匹配结果与人眼的视觉判断非常相近。

的候选匹配集中,同样也将 A_1 的多边形标识号以及相应的相似度值加入到 A_2 的候选匹配集中。

由于实体的匹配是一个非常耗时的过程,因此因尽量避免大范围的搜索,本文提出搜索时先判断两个面实体的最小外接矩形框(MBR)是否相交,然后再来求它们的相交多边形,若相交多边形为非空,才用式(1)来计算它们的重叠面积以及相似度。

2.2 面实体(集)之间匹配关系的确定

根据重叠面积确定了待匹配面实体的候选匹配实体集,为了进行后续处理,还需要进一步确认:(1)面实体(集)之间的匹配类型。从直观上来看,当 $Sim(A_1, A_2)$ 接近 1 时,表明 A_1 是整体与 A_2 匹配的;反过来,当 $Sim(A_2, A_1)$ 接近 1 时,表明 A_2 是整体与 A_1 匹配的;当两个相似度都接近于 1 时,表明这两个面实体之间的匹配关系是 1:1 的。其它情况则稍微复杂一些,面实体之间的匹配关系可能是多对多的;(2)根据匹配类型对实体(集)进行重组,即将若干个面实体组成一个整体(采用 GIS 中的多面结构(Multi Polygon)),也就是将所有 1:N、M:1 以及 M:M 的情况处理成 1:1 的对应关系。

2.3 面实体(集)之间模糊拓扑关系的确定

前面过程中确定了匹配面实体(集)之间的匹配关系,还需要进一步确定这两个面实体(集)是否为同名实体?这牵涉到如何衡量两个实体(集)之间的差异,Winter 在研究不确定性面之间的拓扑关系时引用了 Molenaar 提出的模糊拓扑关系的概念^[8],并提出可以由两个面实体之间的形态距离来计算两个面的拓扑关系的隶属度。本文在此基础上从地图数据库同名面实体匹配的应用角度出发,对 Winter 的方法进行改进,并成功应用于城市大比例尺地图数据库建筑物面实体的匹配。

2.3.1 两个面实体之间形态距离的计算

为了定义形态距离,Winter 先将一个平面根据两个面的关系分成如下三个区域^[8](如图 2):

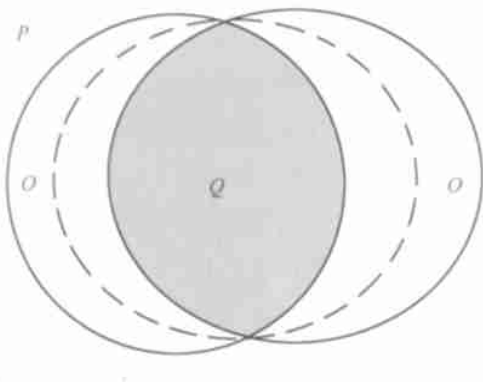


图 2 形态距离计算

Fig.2 The definition of morphological distance between two zones

$$\begin{cases} O = \mathfrak{R}^2 \setminus P, Q \\ P = A^c \cap B^c \\ Q = A^o \cap B^o \end{cases}$$

A^o, B^o 为两个面 A, B 的内部区域, A^c, B^c 为两个面 A, B 的外部区域,即补集。其中的区域 O 反映了两个面之间的不确定带。形态距离就是根据这个区域来定义的。设 s 为 O 区域骨架线(图 2 中的虚线)上的一点, d 为点 s 在 O 区域内的最大圆的直径,则点 s 的形态距离 $\vartheta_{AB}(s)$ 按下式定义:

$$\vartheta_{AB}(s) = \begin{cases} d & \text{if } s \in A^c \\ -d & \text{if } s \in A^o \\ 0 & \text{if } s \in \partial A \end{cases} \quad (2)$$

矢量数据结构下,形态距离的计算比较困难,而在栅格数据结构下,则可以通过形态数学中的“侵蚀”(Erosion)和“扩张”(Dilation)操作来实现,即区域

A “侵蚀”一直到达点 s (当 s 位于区域内部时),或“扩张”到点 s (当 s 位于区域外部时)所用的距离。

2.3.2 两个面实体之间形态距离的分类

为了由两个面之间的形态距离确定它们的拓扑关系,先根据它们的符号分成三类 $\Omega_0 = \{ \omega_-, \omega_0, \omega_+ \}$ (为简便,将 s 点的形态距离 $\vartheta_{AB}(s)$ 记为 ϑ_{AB}):

$$\Omega_0 = \{ \omega_-, \omega_0, \omega_+ \}, \text{ with } \begin{cases} \vartheta_{AB} \rightarrow \omega_-, & \text{if } \vartheta_{AB} < 0 \\ \vartheta_{AB} \rightarrow \omega_0, & \text{if } \vartheta_{AB} = 0 \\ \vartheta_{AB} \rightarrow \omega_+, & \text{if } \vartheta_{AB} > 0 \end{cases} \quad (3)$$

事实上,我们只需要关心形态距离的最小值和最大值,即 $\min(\vartheta_{AB}(s)), \max(\vartheta_{AB}(s))$, (简记为 $\vartheta_{\min}, \vartheta_{\max}$),因为我们已经可以根据这两个值的分类来确定两个面之间的拓扑关系。Winter 指出若将 Egenhofer 的 9 交拓扑关系模型中的拓扑关系“重叠”区分为“弱重叠”和“强重叠”,则 9 种拓扑关系与 C^i, ω_{\min} 和 ω_{\max} 的不同取值相对应^[8],如下表。

表 2 三元组与 9 交模型拓扑关系的对照

Table 2 The triple $\{C^i, \omega_{\min}, \omega_{\max}\}$ distinguishes the same relations as the intersection model

拓扑关系	大类(C)	ω_{\min}	ω_{\max}
相离(DisJoint)	C^1	ω_+	ω_+
相接(Touch)	C^1	ω_0	ω_+
弱重叠(WeakOverlap)	C^1	ω_-	ω_+
强重叠(StrongOverlap)	C^2	ω_-	ω_+
覆盖(Covers)	C^2	ω_-	ω_0
被覆盖(CoveredBy)	C^2	ω_0	ω_+
包含(Contains)	C^2	ω_-	ω_-
被包含(ContainedBy)	C^2	ω_+	ω_+
相等(Equal)	C^2	ω_0	ω_0

2.3.3 两个面实体之间模糊拓扑关系的确定

两个不同来源的地图中的面实体由于源图中的各种误差,总是存在着不确定性,也就是形态距离的计算结果是受不确定性的影响的,引起不确定性的原因有很多,不同来源的地图存在的差异,可能的原因有:(1)不同精度的采集方法;(2)概念/语义模型不同;(3)现实世界本身发生了变化。因此问题的关键在于如何根据有不确定性的形态距离值来确定它们的分类,并最终得到两个不确定面之间的模糊拓扑关系。

Winter 在假设两幅图的点位中误差已知的前提下,讨论了不确定性面之间形态距离值的分类以及模糊拓扑关系的确定问题。本文认为 Winter 讨论的方法中有两点不足:一是在很多地图数据库实体匹配与合并过程中,我们并不能确切知道两个源图的

点位精度,因此根据源图的点位精度来确定两个面的模糊拓扑关系并不实用;二是根据点位误差的概率密度函数来计算两个面之间属于某种拓扑关系的概率,太过复杂。本文提出在地图数据库实体匹配与合并过程中,可以根据同名点之间的距离值来确

定两幅图的点位相对中误差 σ , 然后根据 σ 来确定形态距离的隶属度函数, 并根据隶属度的大小来确定最小和最大形态距离值的分类。一个形态距离值 ϑ 属于 $\Omega_0 = \{\omega_-, \omega_0, \omega_+\}$ 的隶属度由以下公式计算(如图 3):

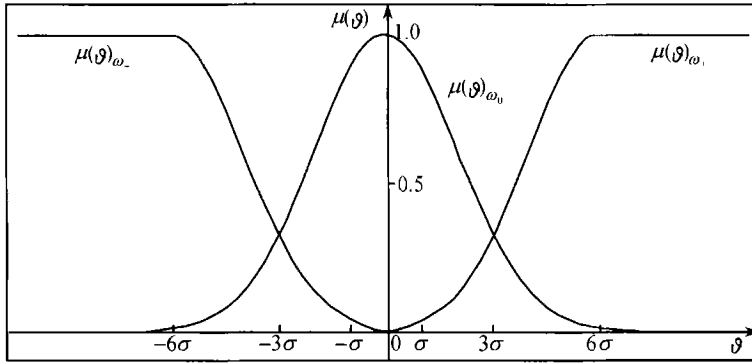


图 3 形态距离隶属度函数曲线

Fig. 3 The membership function of morphological distance

ϑ 属于 ω_- 的隶属度函数:

$$\mu(\vartheta)_{\omega_-} = \begin{cases} 1 & \vartheta \leq -6\sigma \\ e^{-\frac{(\vartheta+6\sigma)^2}{2\sigma^2}} & \vartheta > -6\sigma \end{cases} \quad (4a)$$

$$\vartheta \text{ 属于 } \omega_0 \text{ 的隶属度函数: } \mu(\vartheta)_{\omega_0} = e^{-\frac{\vartheta^2}{2\sigma^2}} \quad (4b)$$

ϑ 属于 ω_+ 的隶属度函数:

$$\mu(\vartheta)_{\omega_+} = \begin{cases} 1 & \vartheta > 6\sigma \\ e^{-\frac{(\vartheta-6\sigma)^2}{2\sigma^2}} & \vartheta \leq 6\sigma \end{cases} \quad (4c)$$

其中, σ 为两幅图的相对点位中误差, 该值由同名点匹配过程中根据同名点的位置偏差计算。

根据求得的 ϑ_{\min} , ϑ_{\max} 值, 分别计算属于 $\Omega_0 = \{\omega_-, \omega_0, \omega_+\}$ 的隶属度值 $\mu(\vartheta_{\min})_{\omega_-}$, $\mu(\vartheta_{\min})_{\omega_0}$, $\mu(\vartheta_{\min})_{\omega_+}$, $\mu(\vartheta_{\max})_{\omega_-}$, $\mu(\vartheta_{\max})_{\omega_0}$, $\mu(\vartheta_{\max})_{\omega_+}$, 然后根据表 2 确定两个不确定面拓扑关系的隶属度(式(5)):

$$\mu \begin{cases} \text{StrongOverlap} \\ \text{Covers} \\ \text{CoveredBy} \\ \text{Contains} \\ \text{ContainedBy} \\ \text{Equal} \end{cases} = \begin{cases} \min(\mu(\vartheta_{\min})_{\omega_-}, \mu(\vartheta_{\max})_{\omega_+}) \\ \min(\mu(\vartheta_{\min})_{\omega_-}, \mu(\vartheta_{\max})_{\omega_0}) \\ \min(\mu(\vartheta_{\min})_{\omega_0}, \mu(\vartheta_{\max})_{\omega_+}) \\ \min(\mu(\vartheta_{\min})_{\omega_+}, \mu(\vartheta_{\max})_{\omega_-}) \\ \min(\mu(\vartheta_{\min})_{\omega_-}, \mu(\vartheta_{\max})_{\omega_-}) \\ \min(\mu(\vartheta_{\min})_{\omega_0}, \mu(\vartheta_{\max})_{\omega_0}) \end{cases} \quad (5)$$

式(5)中, $\min(\mu_1, \mu_2)$ 表示 (μ_1, μ_2) 中的最小值, 实际上是两个模糊度的交运算。最后我们可以根据隶属度最大原则确定两个不确定面之间最有可能的拓扑关系。

3 实验结果

为了验证上述面实体匹配方案的有效性和正确性, 本文试验了城市大比例尺地图数据库之间基于模糊拓扑关系分类的面实体匹配方法。试验图分别为房产图、地籍图、地形图, 比例尺都是 1:500, 且三种图具有相同(或相近)的时相。匹配的面实体为建筑物/房屋实体。三种图之间同名面实体的差异主要包括两方面: 一是对同一房屋要素(集)的解释不同, 导致对房屋边界线的合并和分割的不同, 在房屋群的内部这种差异尤其明显; 解释的差异还表现在同一个房屋在一个图中包含了小的附属结构, 而另一幅图中则没有; 二是同名面实体的位置差异。从面实体匹配的角度而言, 第一种差异更难确定, 因为有时它与第二种差异往往混肴在一起。表 3 列出了部分试验结果的统计情况。图 4 为相应的部分匹配结果的图形叠加显示。

表 3 房产图和地形图数据库建筑物面实体匹配结果统计
Table 3 Result of areal entity matching among topographic map and estate map

匹配类型	覆盖	被覆盖	相等	合计	说明
1:1	1	0	14	15	房产图中建筑物面实体总数为 44, 有 4 个匹配关系为 1:0, 这里没有统计在内
1:N	3	3	4	11	
M:1	0	0	6	6	
M:N	0	8	0	8	
合计	4	11	24	40	

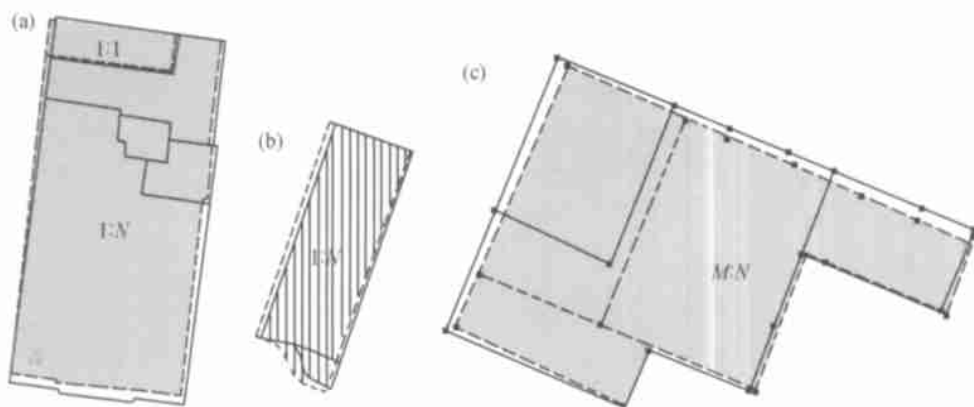


图4 房产图(虚线)与地形图(实线)建筑物面实体部分匹配结果

(a) $1:1$ 、 $1:N$ 相等; (b) $1:N$ 覆盖; (c) $M:N$ 相等

Fig. 4 Result of areal entity matching among topographic map (solid line) and estate map (dashed line)

(a) $1:1$ 、 $1:N$ equal; (b) $1:N$ overlay; (c) $M:N$ equal

从表3中可以看出,具有相同或相近时相的不同来源的城市大比例尺地图数据库,大部分的建筑物面实体是可以匹配到的;而且四种匹配关系($1:1$, $1:N$, $M:1$, $M:N$)都有,说明不同来源的地图对建筑物的合并和分割存在着差异,反映了不同来源的城市大比例尺地图制图的目的不同;匹配建筑物面实体之间的拓扑关系主要集中在“相等”、“被覆盖”和“覆盖”三类中,其中多数是“相等”类型,这一方面说明本文提出的模糊拓扑分类方法可以有效地克服图和图之间的位置差异的影响,另一方面也说明不同来源的城市大比例尺地图数据库之间有很多同名实体,通过地图数据库实体匹配和合并技术,完全有可能实现不同来源地图数据库之间的集成和信息共享。

上述分类结果中,“强重叠”的拓扑关系往往意味着两个面实体之间可能为非同名实体,两个图所描述的地物可能已经发生了变化;“被覆盖”与“覆盖”的情况则说明这两个实体可能是同名实体,但是两个图对同一地物(集)描述的详细程度不同,也就是说其中一幅图中可能包含的细节更多一些。若相应的隶属度比较大,还说明实地地物可能发生了变化,比如在原建筑物边上新盖了房子或拆除了边上的小房子;“被包含”以及“包含”的拓扑关系往往意味着两个面实体之间可能为非同名实体,这可能是由于变更引起的(比如,房屋拆迁后在原来位置重新盖了房子),也有可能是由于其中一幅图中对某些地物的遗漏;而“相等”类型的拓扑关系则表明是两个同名实体,也就是说两个面实体(集)表达的是同一

地物(集),两个实体之间的差异主要是由于制图误差的影响引起的。

本文提出的方法可以用来进行地图数据库比较,地图数据库变化检测与更新以及地图数据库质量评价。比如,来自两个不同地图数据库的面实体被确认为“相等”,则可以据此评估两个数据库的质量,其中一个库可能来自高精度数据源;若两个同名面实体被确认为“强覆盖”则表明其中一个库数据的完整性可能不够。

4 结论

本文提出的不同来源地图数据库之间同名面实体的匹配方案具有如下优点:

(1) 通过形态距离来确定两个不确定面之间的模糊拓扑关系,克服了由于不同来源地图之间可能存在的同名面实体之间的几何位置、形状、大小等方面差异而引起的面实体空间相似度计算的困难。

(2) 可以处理不同来源地图数据库之间同名面实体非一对一匹配的复杂情况。

(3) 匹配结果中包含了所属关系的隶属度,为用户进一步分析确定面实体之间的实际匹配关系提供了依据。

(4) 由同名点的统计信息得到两幅图之间的相对点位中误差,并据此确定模糊拓扑关系分类准则,这样在不知道源图数据质量的情况下,仍然能得到符合要求的匹配结果。

由前述讨论可知,本方法的关键是两个面实体之间形态距离的确定,因此能否成功将本方法应用

于地图数据库合并的生产实践中还决定于形态距离计算的效率和有效性。另外,本文讨论的算法和思路主要着眼于城市大比例尺地图数据库中的建筑物面实体,要应用于其他类型的面实体(如湖泊)或不同比例尺之间的面实体的匹配还需要进一步地完善。

参 考 文 献 (References)

- [1] GoodChild M. F. Uncertainty: The Achilles Heel of GIS? [E]. [Http: www. findarticles. com/cf_0/m0edm/1998_nov/53366331/print. jhtml](http://www.findarticles.com/cf_0/m0edm/1998_nov/53366331/print.jhtml), 1998.
- [2] Saalfeld A. Spatial Data Acquisition and Integration [E]. [http: www. ncgia. ucsb. edu/other/ucgis/research _ priorities/old3%20%287-07-97%29/acquisition. Html](http://www.ncgia.ucsb.edu/other/ucgis/research_priorities/old3%20%287-07-97%29/acquisition.html). 1996.
- [3] Walter V. . and Fritsh D. Matching Spatial Data Sets: A Statical Approach[J]. *International Journal of Geographical Information Systems*, 1999, **13**(5): 445—473.
- [4] Cobb M. , Chung M. , Foley H. A Rule-based Approach for the Conflation of Attributed Vector Data [J]. *GeoInformatica*, 1998, **2**(1):7—35.
- [5] Saalfeld A. Automated Map Conflation [D]. Dept. Of Computer Sci. , The University of Maryland, Washington D. C. , USA, 1993.
- [6] Lemarie C. , Raynal L. Geographic Data Matching: First Investigations for a Generic Tool[C]. In: Proceedings of the GIS/LIS '96 annual Convention, Denver, Colorado, USA, 1996;333—341.
- [7] Wentz E. A. Shape Analysis in GIS [C]. ACSM/ASPRS, 1997, Seattle, Washington. Vol 5. (Auto-Carto 13);204—213.
- [8] Winter S. Uncertain Topological Relations between Imprecise Regions [J]. *International Journal of Geographical Information Systems*, 2000, **14**(5):411—430.
- [9] Molenaar M. An introduction to the theory of spatial object modeling for GIS[M]. London: Taylor & Francis, 1998.

Areal Feature Matching among Urban Geographic Databases

ZHANG Qiao-ping¹, LI De-ren^{1,2}, GONG Jian-ya²

(1. School of Remote Sensing and Information Engineering, Wuhan University, Hubei, Wuhan 430079, China;

2. Key State Lab of Information Engineering in Surveying Mapping and Remote Sensing, Wuhan University, Hubei, Wuhan 430079, China)

Abstract: This paper presents an idea to determinate the variant matching relationships among areal entities based on the overlay areas. Then a methodology to identify the counterpart elements based on the fuzzy classification of topological relationship is illustrated, which applies the concept of morphologic distance between areas introduced by Winter and the concept of fuzzy topological relationship theory. The resulted topological relationship between areas indicates the discrepancy between the considered areal features and also shows the degree of being or not being belonged to the identical features, which is important for further analysis or decision making. The methodology presented here enhances the technology of feature matching in the area of map conflation. Tests show that the method is well performed and has the ability to identify the non-one-to-one matching relationship and is well suited for areal feature matching among urban geographic databases with the same scales as well as similar scales.

Key Words: conflation of geographic database/map conflation; feature matching, morphologic distance; fuzzy topological relationship