

复杂面状矢量要素快速形状匹配方法

付仲良^{1,2}, 邵世维¹

(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)

Methods of Complex Polygon Element Fast Shape Matching

FU Zhongliang, SHAO Shiwei

摘要: 矢量要素匹配是数据库合并和数据更新的核心问题。在分析现有匹配方法不足的基础上, 针对复杂面状要素匹配问题, 提出先对复杂面要素进行基于 Douglas-Peucker 方法的形状简化, 然后对简化后的形状再进行形状匹配。其中, 形状匹配通过正切空间的方法对要素进行描述, 然后利用形状匹配距离计算出形状差异。通过试验表明该方法能够有效提高矢量形状匹配的速度以及正确率, 较好地解决复杂情况面要素匹配的问题。

关键词: 矢量匹配; 形状特征; 空间相似性; Douglas-Peucker; 正切空间

一、引言

矢量要素的匹配是通过目标实体的几何、拓扑和语义的相似性度量, 识别出同一地区不同来源的空间数据集中的同一地物, 从而建立两个空间数据集之间同名目标的联系, 并探测出不同空间数据集之间的差异或变化^[1]。

矢量要素的匹配方法按照判别依据一般分为几何匹配、拓扑匹配和语义匹配。拓扑匹配属于弱条件匹配, 微小的差异都将导致匹配失败; 语义匹配常常依赖于数据模型、属性数据类型及数据完整性, 它们都不足以确定两个面实体为同名实体, 所以实际应用中通常使用几何匹配进行目标之间的相似识别。矢量要素的几何匹配是通过计算参照目标与源目标之间的几何相似度进行的一种匹配方法。现阶段针对面实体的几何相似度提出一些解决方法, 如文献 [2] 根据两个面目标的重叠面积比值来计算其匹配的可能性, 但会出现误匹配的情况; 文献 [3] 通过傅里叶形状描述子, 来进行多边形形状比较以及形状查询, 但涉及大量运算并且匹配效率不高; 文献 [4] 提出基于空间实体特征(位置、形状及大小)的相似性确定同名面实体匹配总相似度的方法, 这种方法利用计算向量间绝对距离的方式来计算形状相似度, 但未考虑向量、数量不一致等情况。

为了提高面要素匹配的效率和精度, 同时又兼顾匹配的准确度, 首先将复杂面要素进行形状简化, 这样既能降低噪声的影响, 又能排除不重要的形状特征, 保留

其重要特征, 从而提高匹配的速度; 在形状匹配时, 利用正切空间的形状描述方法对简化后的要素进行描述, 然后利用形状匹配距离再计算出形状差异值。

二、复杂面状矢量要素快速匹配方法

1. 基于 Douglas-Peucker 的复杂面要素形状简化

定义: 设简化距离阈值为 T ; C 为实平面上的闭合多边形, P_0, P_1, \dots, P_n 为该闭合多边形上顶点, 并沿顺时针方向分布。计算出与 P_0 距离最长的顶点 P_k (如有几个最长值取 k 值最小的 P_k), 连接 P_0 与 P_k 。利用直线 P_0P_k 分别对两段复合线 $P_0 - P_k$ 和 $P_k - P_n$ 上的节点计算到直线 P_0P_k 的距离 D_i , 选取其中距离最大的点 P_j , 如果 D_i 大于限差阈值, 则保留点, 反之剔除该点。利用保留的最大距离点 P_j 将原复合线分为两段, 并用同样的方法对位于它们之间的节点进行检测, 重复此操作, 直至节点到两端点连线的距离最大值小于限差阈值为止, 如图 1 所示。

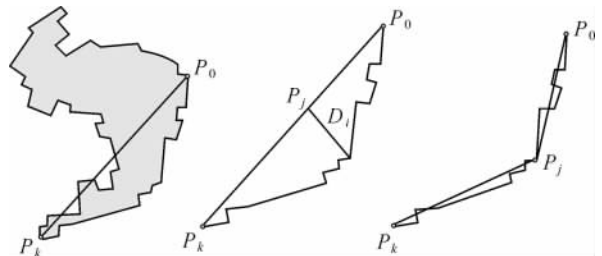


图 1 基于 Douglas-Peucker 的形状简化

收稿日期: 2010-08-24

作者简介: 付仲良(1965—), 男, 湖北麻城人, 教授, 博士生导师, 研究方向为图形图像处理、GIS 等。

2. 面要素形状描述方法

假设多边形的某一顶点作为参考点 P_0 , θ_1 表示起始边 P_0P_1 的方位角, φ_1 表示从起始边 P_0P_1 到 P_1P_2 的转角, φ_k 表示沿着 $P_{k-1}P_k$ 到 P_kP_{k+1} 的转角, 多边形的正切空间形状描述函数为 $\theta(l)$, x 轴代表从起点 P_0 沿着多边形周边到多边形上各点 P_k 的归一化距离 $L_k = \sum_{i=1}^k L_i/L$, y 轴代表各点沿着周边的转角(以顺时针为正方向)的累加 $\theta_k = \theta_{k-1} + \varphi_{k+1}$ ($k = 3, 4, \dots, n$), 如图 2(a) 所示。由于不同起始点、不同的方向所得到的面实体正切空间函数不同, 在源匹配多边形中, 定义 $P_0(x_0, y_0)$ 为起始点, 其中 $x_0 = \max\{x | (x_i, y_i) \in A\}$, $y_0 = \min\{y | x = x_0, (x_i, y_i) \in A\}$ 。从起始点 P_0 沿着多边形顺时针旋转方向为正方向, P_0P_1 为起始方向线。以 P_0 为原点、距离阈值 T 为半径搜索包含在圆内的目标多边形 B 的节点集合 Q , 选取与 P_0P_1 的方位角 θ_1 差异最小的结点为目标匹配多边形的起始点 P'_0 (如图 2(b) 所示)。

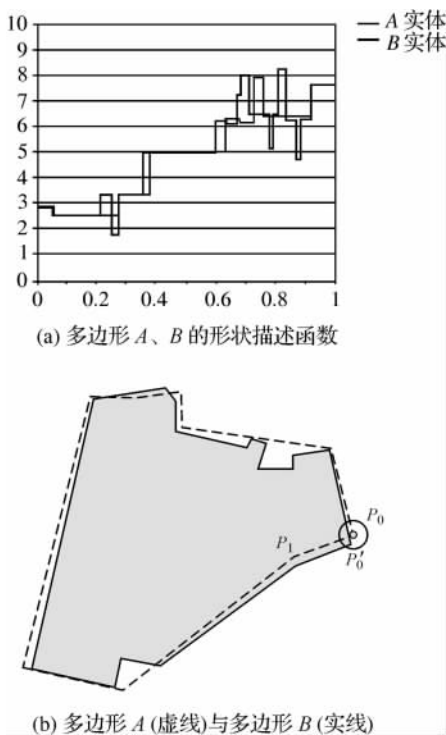


图2 基于正切空间的形状描述函数

3. 面要素形状匹配方法

通过以上的形状描述方法对多边形 A, B 要素进行形状描述, 将其形状化为正切空间表达式, 分别为 $\theta_A(l)$ 和 $\theta_B(l)$, 其中 s 为 x 轴坐标, θ 为 y 轴坐标。通过计算两个矢量要素之间的形状匹配距离来确定它们的相似性, 进而判断两要素是否匹配。定义其匹配距离为

$$D_{AB} = \int_0^1 (1 - |\frac{\theta_A(l) - \theta_B(l)}{\max(\theta_A(l), \theta_B(l))}|) dl \quad (1)$$

D_{AB} 的值越趋近于 1, 表示多边形 A 和 B 的形状越相似, 匹配的程度越好。

三、试验与结论

1. 形状简化试验与分析

采用上述方法对图 3 中数据进行形状简化试验, 给出一个简化效果示意图(简化阈值分别为 8 m、10 m 和 15 m)。

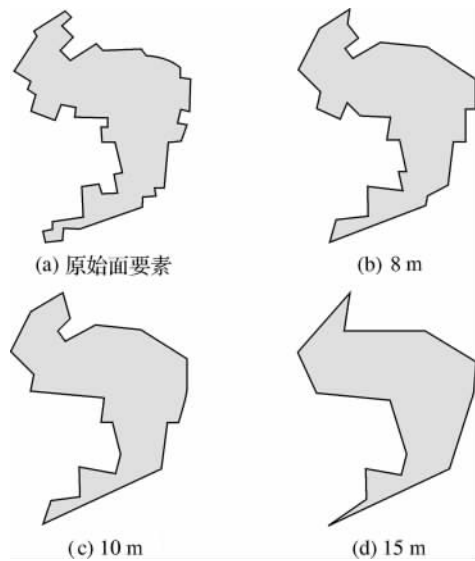


图3 形状简化效果

利用文中基于正切空间的形状描述函数, 对比原要素与三种简化后要素的形状匹配距离、面积匹配距离和周长匹配距离(如表 1 所示), 可以看出, 通过形状简化, 要素节点数大大减少, 要素的形状、面积、周长的变化在 8 m 和 10 m 的简化阈值下可以满足匹配要求。

表1 形状简化试验

| 匹配要素 | 节点数 | 形状匹配距离 | 面积/m ² | 面积匹配距离 | 周长/m | 周长匹配距离后 |
|---------|-----|-----------|-------------------|-----------|---------------|-----------|
| 原要素 | 126 | 1.000 000 | 22 103.052 709 | 1.000 000 | 1 116.755 958 | 1.000 000 |
| 8 m 简化 | 33 | 0.945 360 | 22 192.898 973 | 0.995 935 | 1 018.678 339 | 0.912 176 |
| 10 m 简化 | 24 | 0.890 359 | 21 785.185 085 | 0.985 619 | 970.717 558 | 0.869 230 |
| 15 m 简化 | 14 | 0.623 541 | 21 060.901 922 | 0.952 850 | 907.635 205 | 0.812 743 |

2. 形状匹配实例与分析

本文以某地区 2003 年与 2008 年的图斑要素 (见图 4) 进行实例匹配,采用 ArcGIS Engine 9.3 与 VS. NET 2008 为开发平台,进行了形状匹配试验 (如表 2、表 3 所示)。其中简化阈值分别为 15 m 与 20 m,匹配阈值选取为 0.85。

通过本文方法与文献 [2] (利用两个面目标的重叠面积比值) 和文献 [3] (利用向量间绝对距离计算的方法) 进行比较 (如表 3 所示)。



图 4 形状简化效果

表 2 多尺度面要素形状匹配

| 算法类型 | 匹配要素 | 简化阈值/m | 节点个数 | 简化后节点个数 | 匹配距离 | 算法时间/s | 是否匹配 |
|------|--------------|--------|------|---------|-----------|--------|------|
| | A1548 与 B114 | 15 | 752 | 58 | 0.973 580 | 1.053 | 匹配 |
| | | 20 | | 48 | 0.968 560 | 0.555 | 匹配 |
| 本文算法 | A1132 与 B12 | 15 | 660 | 49 | 0.763 251 | 0.682 | 不匹配 |
| | | 20 | | 42 | 0.795 233 | 0.414 | 不匹配 |
| | A1003 与 B150 | 15 | 201 | 22 | 0.132 515 | 0.379 | 不匹配 |
| | | 20 | | 17 | 0.209 836 | 0.105 | 不匹配 |

表 3 匹配算法比较

| 算法类型 | 匹配要素 | 匹配值 | 算法时间/s | 是否匹配 |
|---------|--------------|-----------|--------|------|
| 文献 2 算法 | A1548 与 B114 | 0.984 250 | 1.773 | 匹配 |
| | A1132 与 B12 | 0.863 624 | 1.330 | 匹配 |
| | A1003 与 B150 | 0.813 695 | 0.732 | 不匹配 |
| 文献 4 算法 | A1548 与 B114 | 0.861 642 | 0.577 | 匹配 |
| | A1132 与 B12 | 0.463 624 | 0.423 | 不匹配 |
| | A1003 与 B150 | 0.336 248 | 0.091 | 不匹配 |

从表 2 和表 3 可以看出,利用本文的快速匹配方法在速度上明显快于文献 [2] 中的算法。文献 [2] 中计算两面目标的重叠面积,会消耗太多时间,而且通过重叠面积比值的方法匹配准确率不高,会出现误匹配情况。而文献 [3] 中的速度介于简化阈值 15 m 和 20 m 之间。综上所述,选择适当的简化阈值,可以明显提高形状匹配效率,并且准确率高于文献 [2] 和文献 [3] 中的算法。

四、结束语

实体匹配是多数据源多尺度数据集成与更新的关键技术,匹配效果的好坏直接影响到数据集成或更新的效果。本文将形状相似性的距离观与形状特征简化相结合,以形状匹配距离作为相似性特征,通过 Douglas-Peucker 算法对复杂面要素进行简化,大大提高了匹配速度,并提出一种基于正切空间的面状矢量要素形状相似性度量模型,利用形状描述函数较好地解决了匹配的准确率问题。最后通过对相同数据不同匹配算法进行试验比较,在匹

配的速度和准确率上有明显提高,说明本文的方法是有效的。

参考文献:

- [1] 张桥平,李德仁,龚健雅. 城市地图数据库面实体匹配技术 [J]. 遥感学报, 2004, 8(2): 107-112.
- [2] 吴建华,付仲良. 数据更新中要素变化检测与匹配方法 [J]. 计算机应用, 2008, 28(6): 1612-1615.
- [3] 郝燕玲,唐文静,赵玉新,等. 基于空间相似性的面实体匹配算法研究 [J]. 测绘学报, 2008, 37(4): 501-506.
- [4] 唐炉亮,李清泉,杨必胜. 空间数据网络多分辨率传输的几何图形相似性度量 [J]. 测绘学报, 2009, 38(4): 336-340.
- [5] 杨得志,王杰臣,闫国年. 矢量数据压缩的 Douglas-Peucker 算法的实现与改进 [J]. 测绘通报, 2002(7): 18-20.
- [6] 何磊,蒋大为,周敏. 基于简化多边形类正切空间表示的图形渐变算法 [J]. 计算机辅助设计与图形学学报, 2007, 19(3): 304-310.