

线状空间数据传输的几何相似性度量 算法与实验分析

孙金礼^{1,2}, 陈杰², 邓敏³

(1. 山西煤炭运销集团大同有限公司, 大同 037004; 2. 山西大同大学矿业工程系, 大同 037003;
3. 中南大学测绘与国土信息工程系, 长沙 410083)

摘要: 同一地物在不同比例尺或者不同来源的地图上通常存在着相似性, 对于图形几何相似性度量方法的研究有利于地图编制、查询、匹配、更新。线状地物要素在地图中占有很大的比例, 因此, 本文以线状空间目标为例, 在前人的基础上给出了线状空间数据的几何图形相似性度量模型: (1) 以差异距离作为相似性特征的位置相似度; (2) 以分形维数作为相似性特征的形状相似度; (3) 以长度或者周长作为相似性特征的大小相似度。相对于(3)大小相似度而言, (1)位置相似度、(2)形状相似度综合考虑了几何图形整体统计的方法和局部几何特征结构。完成多尺度传输的线状空间数据几何相似性度量实验, 并对数据传输量与几何相似性度量方法进行比较, 实验结果表明: 基于广义 Hausdorff 距离模型的中位数 Hausdorff 距离的位置相似性对于空间数据渐进性传输具有稳定性和可行性。最后, 总结了本文的研究成果, 并展望了该方向进一步研究的若干问题。

关键词: 几何相似性; 分形维数; 广义 Hausdorff 距离

DOI: 10.3724/SP.J.1047.2011.00701

1 引言

如何对多尺度传输与表达的空间数据进行几何图形相似性度量, 对于空间信息渐进传输质量的评价具有非常重要意义^[1]。几何图形相似性是图像查询、更新和匹配的主要方面, 其可概括为几何图形整体统计的方法和局部结构的方法。其中, 前者是从几何图形中取出一组图形的全局特征, 用统计方法描述这些特征的度量模型^[2-7], 如面积、周长、Hausdorff 距离、外接矩形等, 图形整体特征的选取比较容易, 但不能精确描述图形的局部几何结构。基于局部几何结构的方法主要是抽取图形的局部结构^[8-13], 如关键点、曲率、凸凹结构、Hough 变化等, 局部结构的方法有利于描述图形的局部特征, 但不利于抽取图形的整体特征。

本文在前人的基础上, 针对线状空间数据, 给出了线状空间数据的几何图形相似性度量模型: (1) 以差异距离作为相似性特征的位置相似度, 即基于广义 Hausdorff 距离模型位置相似性; (2) 以分

形维数作为相似性特征的形状相似度; (3) 以长度或者周长作为相似性特征的大小相似度; 相对于(3)大小相似度而言, (1)位置相似度和(2)形状相似度综合考虑了几何图形整体统计的方法和局部几何特征结构。最后, 实验完成了线状多尺度传输的空间数据几何图形相似性度量, 并对数据传输量与图形相似度量方法进行比较分析。

2 线状空间数据的相似度与度量算法

2.1 位置相似度及度量算法

空间距离是一种非常重要的空间概念, 可以用来描述空间目标之间的相对位置分布情况, 反映空间邻近目标之间的邻近程度和相似性程度^[14]。比较目标时, 考虑目标的位置很重要; 位置相似度意味着每个要素相对于其他要素而言, 它们之间要用一个距离差异值来表示。对于线状目标来说, 该差异值能准确的表示空间目标间相对位置的差异。很多文献对距离度量方法(如最近距离、最远距离、

收稿日期: 2011-05-03; 修回日期: 2011-08-31.

作者简介: 孙金礼(1964-), 男, 汉族, 山西原平人, 硕士, 副教授, 现主要从事空间地物相似性度量和地理空间数据传输理论与方法。E-mail: chenjie_301@126.com

质心距离)进行了介绍,但是,大部分表达距离度量时并没有顾忌空间目标的形状差异,通过质心的距离定义也难以表达空间目标间相对位置的差异。随后,邓敏等人在 Chen^[15] 空间目标的 Hausdorff 距离度量方法基础上,提出的广义 Hausdorff 距离度量模型和中位数 Hausdorff 距离度量^[16]。本文将该方法中位数 Hausdorff 距离用于度量线目标几何位置相似性度量中。

设 A、B 为两个空间目标集合,并且令

$$\begin{aligned} \tau_A &= \frac{\theta\{(A \oplus S(\epsilon)) \cap B\}}{\theta(B)} \\ \tau_B &= \frac{\theta\{(B \oplus S(\epsilon)) \cap A\}}{\theta(A)} \end{aligned} \quad (1)$$

其中,⊕为数学形态学中的膨胀算子;0≤τ_A,τ_B≤1;θ(•)为一个度量函数,它的取值取决于 A、B 所表达的空间目标类型(点目标、线目标或面目标),若是点目标取值为 0 或者 1;若是线目标取值为 0 或者交集长度;若是面目标取值为 0 或者交集面积。S(ε)表示半径为 ε 的一个闭球,定义^[15]:

$$\rho(A,B) = \inf\{\epsilon: A \subset B \oplus S(\epsilon), B \subset A \oplus S(\epsilon)\} \quad (2)$$

式中,ρ(A,B)为集合 A、B 的 Hausdorff 度量。分析式(2)可知,ρ 为最小的闭球 S 的半径,以至于集合 A 包含在膨胀集合 B⊕S(ε)中,B 也包含在膨胀集合 A⊕S(ε)中,如图 1。

进而定义

$$\rho_G(A,B) = \inf\{\epsilon: \tau_A \geq \tau_0, \tau_B \geq \tau_0\}, 0 \leq \tau_0 \leq 1 \quad (3)$$

等价表达为:

$$\begin{aligned} \rho_G(A,B) &= \max\{\inf\{\epsilon: \tau_A = \tau_0\}, \\ &\quad \inf\{\epsilon: \tau_B = \tau_0\}\}, 0 \leq \tau_0 \leq 1 \end{aligned} \quad (4)$$

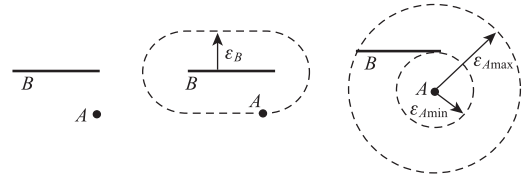


图 1 一个点与一条线的 Hausdorff 距离情况

Fig. 1 Hausdorff distance for a point and a line

称式(3)、(4)为广义 Hausdorff 距离模型。广义 Hausdorff 距离度量模型的实现及具体应用见文献^[16]。显然,当 τ₀=1 时,式(3)、(4)为传统的 Hausdorff 距离^[15];当 τ₀=0.5 时,结合式(3)和式(4)并用内插逼近的方法计算得到的空间目标之间的距离,亦称为中位数 Hausdorff 距离,即为 ρ_M(A,B)。

进而,如图 2,若对 A₁、A₂ 来源不同的同一线状实体进行位置相似性度量:(1)将来源不同的数据进行叠加(不同尺度的同一数据通过尺度变换进行叠加);(2)选择同一参考点状目标 O;(3)计算 O 点到不同来源地图同一地物目标中位数 Hausdorff 距离 ρ_{M1}(O,A₁)、ρ_{M2}(O,A₂)。则位置相似性度量的计算公式可以表达为:

$$WZ(A_1,A_2) = 1 - \left| \frac{\rho_{M1}(O,A_1) - \rho_{M2}(O,A_2)}{\max(\rho_{M1}(O,A_1), \rho_{M2}(O,A_2))} \right| \quad (5)$$

2.2 形状相似度与算法

形状是描述物体的重要特征之一,利用形状特征来区别和检索物体比较直观^[17]。文献^[18]指出,可用面的紧凑度(面积周长比)、边界的描述和面的构成元素来定义面状实体的形状;文献^[19]中的算法使用了类似于线实体形状度量的方法。这些方法虽然简单,但易受各种干扰而影响精度。形状相似度是在一定形状描述方法的基础上计算目标相

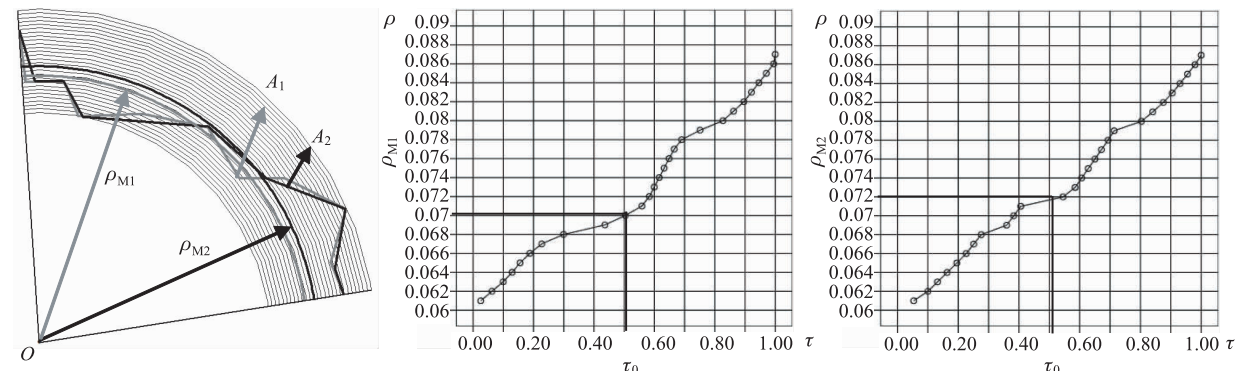


图 2 A₁ 和 A₂ 空间目标与参考目标 O 之间中位数广义 Hausdorff 距离模型的实现

Fig. 2 Computation of generalized Hausdorff distance between spatial objects A₁, A₂ and reference object O

似性或相异性。

分形理论中的分形维数,则能够比较好地表达这种直观上的形态特征。分形维数作为图像表面不规则度的一个量度,与人类视觉对图像表面纹理粗糙度的感知是一致的,即分形维数值越大,对应的图像表面越粗糙;反之,分形维数值越小,对应的图像表面越平滑,因此,分形理论广泛应用于纹理分析、图像分割等领域中。分形的计算方法比较多,本文借助地毯覆盖法来度量灰度图像分形维数,进而,通过分形维数差异值来度量目标间形状相似度。



图 3 一条曲线及其对应的分形维数计算图表

Fig. 3 Computation graph of the fractal dimension for a curve

分形维数计算模型具有旋转、平移、比例不变性,与人的视觉判断较为一致,而且根据分形维数特征,它本身顾及目标局部和整体自相似性,是一种度量形状相似度的很好方法。鉴此,本文以分形维数差异来表达线目标间的形状相似度。进而,对 A_1, A_2 来源不同的同一线状实体进行了形状相似性度量:(1)将来源不同的数据转换成栅格数据;(2)根据上式(7)并结合分形软件 FractalFox 分别计算各自的分形维数 $FD(A_1), FD(A_2)$,则形状相似性度量的计算公式可以表达为:

$$XZ(A_1, A_2) = 1 - \left| \frac{FD(A_1) - FD(A_2)}{\max(FD(A_1), FD(A_2))} \right| \quad (8)$$

2.3 大小相似度与算法

在进行几何相似性比较中,除了考虑线状实体的位置和形状,其大小也是一个要考虑的因素,这里主要是通过线实体的长度或者周长来体现。

因此,如对 A_1, A_2 来源不同的同一线状实体进行大小相似性度量,其大小相似度计算公式可表达为:

$$DX(A_1, A_2) = 1 - \left| \frac{L(A_1) - L(A_2)}{\max(L(A_1), L(A_2))} \right| \quad (9)$$

由 Minkowski-Bouligand 维数定义,对于图像灰度表面,图像的表面积测度,可以表示为

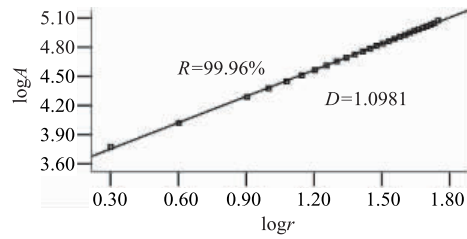
$$A(r) = kr^{2-D} \quad (6)$$

其中, k 为一常量, D 表示图像表面的分形维数 Fractal Dimension, 简称为 FD , r 为测量尺度, $A(r)$ 为图像的表面积。

则式(6)可以表达对数形式为:

$$\log(A(r)) = \log C + (2-D)\log r \quad (7)$$

式(7)可视为以 $\log r_i, \log A(r_i)$ 为变量,以 $2-D$ 为斜率的直线方程,见图 3。



3 线状空间数据相似性度量的实验与结果分析

3.1 相似性度量算法实验

本文相似性度量实验是在全国不同尺度道路网上选取不同表达同一线目标来进行的。度量时,首先,选取 1:500 万比例尺图上一线目标作为原始曲线,然后,在其他各种比例尺图上选取同一线目标与原始曲线进行相似性比较,如图 4。进而,根据本文三种相似性度量计算公式,分别选择实体间的位置(距离)、形状、长度等指标来计算相似度。

针对位置相似度量,实验时,需将其他各种比例尺的数据与原始曲线进行叠加(不同尺度的同一数据通过尺度变换进行叠加);然后,选择同一参考点状目标,分别计算不同尺度曲线相对于参考点的中位数 Hausdorff 距离,见表 1。进而,以原始曲线的中位数 Hausdorff 距离为依据,按照式(5)分别进行其他尺度的曲线的中位数 Hausdorff 距离相对于与原始曲线中位数 Hausdorff 距离位置相似度的计算,实验结果见表 2。针对形状的相似性度量算法和大小的相似性度量算法按照式(8)和式(9)进行计算,实验的原始数据及结果,见表 1 和表 2。

为了比较 3 种方法的优越性和可行性,本文还进行了不同尺度间数据的传输量与相似度之间关系的实验,其结果见图 5。

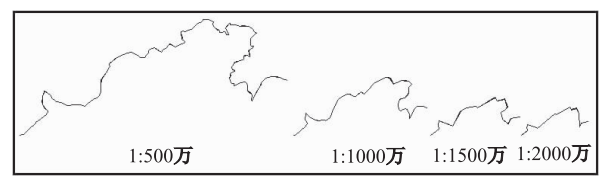


图 4 不同比例尺的同一条曲线
Fig. 4 The same curve on different scales

表 1 实验原始数据
Tab. 1 Row experimental data

实验数据 比例尺	数据量	中位数 Hausdorff 距离 (m)	曲线 长度 (m)	分形 维数 (D)
1 : 500 万	1413	34 920	138 718. 9699	1. 0981
1 : 1 000 万	632	35 185	124 811. 1269	1. 0674
1 : 1 500 万	234	35 135	116 586. 8955	1. 0558
1 : 2 000 万	182	33 860	110 502. 9167	1. 0528

3. 2 结果分析

从表 2 和图 5 可以看出:(1)随着比例尺的逐渐减小,数据传输量逐渐减少;(2)传输同样的数据量,位置相似度变化很小,形状相似度次之,大小相似度变化最大。实验结果表明,位置相似度与人眼判断结果相一致,说明该算法能很好地解决同名实体的相似性问题。

表 2 实验结果数据
Tab. 2 The experimental result data

实验数据 比例尺对照	数据传 输量 (%)	位置相似 度 (WZ) (%)	大小相似 度 (CD) (%)	形状相似 度 (XZ) (%)
1 : 500 万-1 : 500 万	100	1	1	1
1 : 500 万-1 : 1000 万	45	99. 39	89. 97	97. 20
1 : 500 万-1 : 1500 万	17	99. 25	84. 05	96. 14
1 : 500 万-1 : 2000 万	13	96. 96	79. 66	95. 87

4 结语

本文将传输量与相似性度量方法相结合,提出线状空间数据的几何图形相似性度量模型,以多尺度传输的线状地物为实验数据,进行空间数据几何图形相似性度量实验,验证空间数据图形相似性度量模型的正确性。本文最后对不同传输量阶段的渐进传输与图形相似度方法进行比较,其实验结果

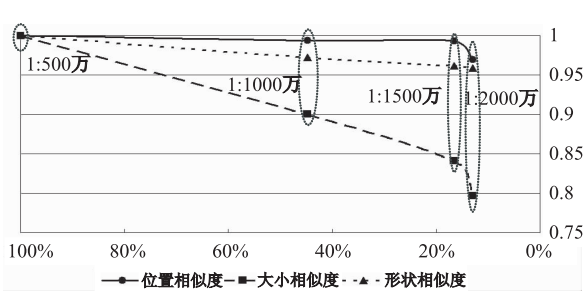


图 5 传输的数据量与相似度关系图
Fig. 5 Relationship between data transmission
volume and similarity

表明:广义 Hausdorff 距离模型位置相似性与人眼判断结果相一致,结果比较稳定,传输同样数据量位置相似度变化相比其他相似性度量方法要小,说明该算法可很好地解决同名实体的相似性问题。后续研究将根据空间数据几何图形相似度方法,在多尺度空间数据匹配、查询、更新等方面进行进一步的研究。

参考文献:

[1] 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.

[2] Pavlidis T. A Review of Algorithms for Shape Analysis [J]. Computer Graphics and Image Processing, 1978, 7:243 - 258.

[3] Yang L, Albrechtsen F. Fast Computation of Invariant Geometric Moments: A New Method Giving Correct Results[C]. Proc 12th IAPR Int. Conf. Pattern Recognition, 1994(1): 201 - 204.

[4] Persoon E, Fuk S. Shape Discrimination Using Fourier Descriptors[J]. IEEE Transaction on System, 1977, 7 (3): 170 - 179.

[5] Arkin E, Chew P, et al. An Efficiently Computable Metric for Comparing Polygonal Shapes [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991, 13(3): 209 - 215.

[6] Atallah M J. A Linear Time Algorithm for the Hausdorff Distance between Convex Polygons[J]. Information Processing Letters, 1983, 17:207 - 209.

[7] 杨春成,张清浦,等. 顾及几何形状相似性的简单多边形最近距离计算方法[J]. 测绘学报, 2004, 33 (4): 311 - 318.

[8] Mount D M, Netanyahu N S, Moigne J L. Efficient Algorithms for Robust Point Pattern Matching and Appli-

- cations to Image Registration[J]. Pattern Recognition, 1999, 32:17-38.
- [9] Perkins W. A Model-based Vision System for Industrial Parts[J]. IEEE Transaction on Computer, 1978, 27: 126-143.
- [10] Constabile M F, Guerra C, *et al.* Matching Shape: A Case Study in Time-varying Images[J]. Computer Vision Graphics & Image Processing, 1985, 29: 296-310.
- [11] Ballard D H. Generalizing the Hough Transform to Detect Arbitrary Shapes [J]. Pattern Recognition, 1981, 13 (2): 111-122.
- [12] Young S S, Scott P D, *et al.* Object Recognition Using Multilayer Hopfield Neural Network[J]. IEEE Transaction on Image Processing, 1997, 6 (3): 357-372.
- [13] Tan G Z, Gao W, *et al.* Similarity Measures for Polygons Representation[J]. Chin J CAD & CG, 1995, 7 (2): 96-102.
- [14] Frank A U. Qualitative Spatial Reasoning About Distances and Directions in Geographic Space[J]. Journal of Visual Languages and Computing, 1992, 3(4):343-371.
- [15] Chen X Y, Doihara T, Nasu M. Spatial Relations of Distance between Arbitrary Objects in 2D/3D Geographic Spaces Based on the Hausdorff Metric [C]. LIESMARS95, Wuhan, 1995.
- [16] 邓敏, 钮沐联, 李志林. GIS空间目标的广义 Hausdorff 距离模型[J], 武汉大学学报: 信息科学版, 2007, 7 (32): 641-645.
- [17] 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): 501-505.
- [18] Wentze E A. Shape Analysis in GIS[C]. Proc of AC2SM/ASPRS, 1997, 204-213.
- [19] Foley H. A Multiple Criteria Based Approach to Performing Conflation in Geographical Information Systems[D]. New Orleans: Tulane University, 1997.

The Algorithms of Geometry Similarity Measurement and Experimental Analysis for Linear Spatial Data Transmission

SUN Jinli¹, CHEN Jie¹, DENG Min²

(1. Shanxi Coal Transportation and sales Group Datong Co., LTD. Datong 037004, China;

2. Department of Mining Engineering, University of Shanxi Datong, Datong 037003, China;

3. Department of Surveying and Geo-informatics, Central South University, Changsha 410083, China)

Abstract: Transmission and expression of multi-scale spatial data is of greatly significance for the quality of the evaluation of progressive transmission of spatial information by geometric similarity measurement, which has plagued GIS researchers for a long time. The same feature from different sources or different scales is usually similar on the map, so the similarity measurement of geometry shape is conducive to the preparation, query, match and update of the map. The line features on the map are in many high proportions, so this paper proposes the line features geometry similarity measurement models of spatial data on the basis of the predecessors. And then, a reclassification of geometric similarity measurement is made from the consideration of distance, shape, length. The corresponding geometric similarity measurements are further developed, including: (1) differences in distance as the similarity, considering shape differences and relative positions of spatial objects; (2) the fractal dimension as a similarity of shape similarity, to a large extent, being able to express morphological characteristics of spatial objects; (3) the length or girth size as the similarity. Relative to (3) length similarity, (1) position similarity and (2) shape similarity are both considered the overall statistical methods and local geometry structure. To complete the linear multi-scale geometric similarity measurement experiments of spatial data, and compare the relationship of different transmission volume and geometric similarity measurement by multi-scale linear features as the experimental data, the experimental results show that location similarity based on generalized Hausdorff

distance model is feasibility to progressive transmission of spatial data. The change for location similarity measurement is smaller than those with other methods based on the same amounts of data transmission, indicating that the algorithm (generalized Hausdorff distance model) solves the similarity problem of the same feature. Finally, this paper summarizes main findings, and highlights further research directions in the near future, such as match, query, update and other issues for multi-scale spatial data.

Key words: geometric similarity; fractal dimension; generalized Hausdorff distance