

网络多分辨率传输中曲线集的相似性度量模型研究

操震洲¹

1 南京工业大学测绘学院,江苏 南京,211800

摘要:针对空间数据在网络多分辨率传输中的相似差异,提出了一种适用于曲线集的相似性度量模型。模型根据拓扑关系和几何特征差异计算曲线相似度,然后以曲线最小凸包面积为权,计算曲线集的加权相似度和相似精度,以加权相似度评估曲线集的整体相似程度,以相似精度评估曲线集各曲线间的相似差异。最后通过实验验证了该模型的有效性。

关键词:相似度模型;渐进传输;多分辨率;形状相似度;拓扑相似度

中图法分类号:P208

文献标志码:A

空间相似性是空间认知的基础,空间数据相似性度量模型是一个研究热点,可应用于实体匹配、同一实体跨尺度变换后的相似性度量等领域^[1]。空间实体的相似性体现在几何图形、空间关系、语义等方面,目前的研究侧重于几何图形相似性度量。图形相似度的计算可基于某种具有平移、旋转、缩放不变性的形状描述子,如傅里叶形状描述子^[2-4]、夹角链码法^[5]等,也可基于位置、面积、差异距离等参数,文献^[6-8]据此计算图形相似度以进行空间实体匹配、道路数据变化检测等,对空间关系相似性的研究相对较少,描述空间关系的有 4 交模型^[9]、9 交模型等;文献^[10,11]基于拓扑关系、距离关系、方位关系提出了空间场景相似性概念及其度量模型;文献^[12]在此基础上给出了线群相似度的计算方法,该方法同时顾及了空间关系和几何图形两方面的相似性。

矢量数据网络多分辨率传输最早由 Bertolotto 提出^[13],它是提高网络传输效率的有效途径之一。度量网络传输中的多分辨率矢量数据相对于原始矢量数据的相似性具有重要意义,一方面可用于评估客户端的图形质量,另一方面可作为服务器端数据多分辨率分解的参考指标。文献^[14]提出了一种空间数据网络多分辨率传输的相似性度量方法,该方法适用于单要素的几何图形相似性度量,但未考虑要素的空间关系相似。多分辨率矢量数据的生成通常基于小波变换、傅里叶变换、线化简等方法,这些方法都不可避免地产生拓扑异化,表现为要素间原有相邻、相离、相交、

重合等位置关系的变化。因此,相似性度量模型应顾及空间关系相似。另外,对于一组要素,既要考察其整体相似度,也要考察该组目标个体间的相似差异。本文提出了一种适用于网络多分辨率传输环境下的曲线集相似性度量模型。

1 曲线集相似性度量模型

1.1 单曲线的相似性度量

记原始曲线为 L , L 在多分辨率传输中的图形为 L' ,单曲线的相似性是指 L' 相对于 L 的相似程度。由于 L 与 L' 为同一曲线,两者的相似性主要体现在几何图形和空间关系方面。

在网络多分辨率传输过程中, L' 的图形会越来越逼近 L ,如图 1 所示。图中细虚线为 L ,实线为 L' ,粗虚线为曲线 L 的最小凸包边界。曲线 L' 、 L 总是位于凸包内,且两曲线图形差异可通过两者所围区域的面积大小体现。记曲线 L' 、 L 所围区域的面积为 ΔS ,曲线 L 的最小凸包面积为 S ,曲线 L' 的几何图形相似度 $Sims$ 可表示为:

$$Sims = 1 - (\Delta S / S) \quad (1)$$

空间关系主要有拓扑、方位和距离关系,皆为要素间的空间位置关系。方位、距离关系可视为空间位置关系的定量表达,而拓扑关系可视为空间位置关系的定性表达,本文侧重考虑拓扑关系的相似性。曲线间的拓扑关系可分为三类:相离、相交(含相邻)、重合。拓扑关系的变化可用图 2 所示的拓扑关系邻域图来描述。定义跨类间拓扑

收稿日期:2012-12-31

第一作者:操震洲,博士生,讲师。主要从事空间数据的多尺度组织与渐进传输方面的研究。E-mail: czz@njtech.edu.cn

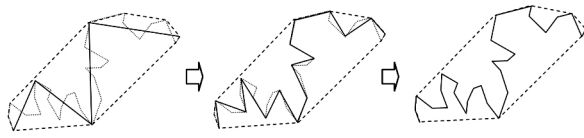


图1 渐进传输中的曲线图形

Fig. 1 Curve Shape Under Progressive Transmission Mode

关系变化的权重为 2, 同类之间拓扑关系变化的权重为 1, 可形成表 1 所示的曲线拓扑关系邻域差异矩阵。

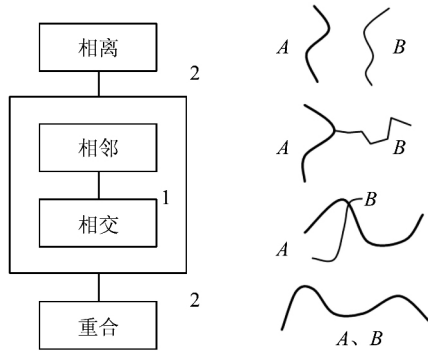


图2 曲线拓扑关系概念邻域图

Fig. 2 Conceptual Neighborhood of Topological Relations Between Curves

表1 曲线拓扑关系邻域差异矩阵

Tab. 1 Neighborhood Dissimilarity Matrix of Topological Relationships Between Curves

	相离	相邻	相交	重合
相离	0	2	3	5
相邻		0	1	3
相交			0	2
重合				0

记分别与曲线 L, L' 满足相邻、相交、重合的曲线集合为 A, B , 取 $A \cup B$ 为候选集, 则与曲线 L' 拓扑关系发生改变的曲线必然位于候选集内, 且候选集内的曲线是与 L' 关系密切的曲线集合, 可基于候选集来计算曲线 L' 的拓扑相似度。对候选集内的曲线逐条判断其与 L, L' 的拓扑关系, 按表 1 计算其拓扑关系权重 W_i , 记候选集内的曲线数为 n , 则曲线 L' 的拓扑相似度 Sim_t 可表示为:

$$Sim_t = 1 - \sum_{i=1}^n W_i / (5n) \quad (2)$$

单曲线的相似度取其图形相似度与拓扑相似度的加权平均值, 如式(3)所示, 其中 α, β 为常数, 且 $\alpha > 0, \beta > 0, \alpha + \beta = 1$ 。

$$Sim = \alpha * Sim_s + \beta * Sim_t \quad (3)$$

1.2 曲线集的相似性度量

曲线集在传输前和传输中的图形分别记为

M_p, M'_q (p, q 表示曲线个数, 且 $p \geq q$), 如图 3 所示。同一曲线在 M_p 与 M'_q 的差异具体有三种: ① 在 M'_q 中被舍弃, 如图 3 的曲线 d ; ② 在 M'_q 中出现拓扑异化, 如曲线 b, c ; ③ 在 M'_q 中曲线图形得以简化, 如曲线 a, b, c, e 。构建曲线最小凸包, 以代表曲线的图形范围, 凸包面积越大, 对应曲线在曲线集图形中所占的份量越大。曲线集的相似性可看成是曲线集所有曲线相似性的综合表现, 本文以曲线凸包面积为权, 将所有曲线的加权相似度作为曲线集的相似度。曲线集相似度的计算方法为: 首先遍历曲线集 M_p 中各曲线, 若曲线在 M'_q 被舍弃, 则其相似度为 0, 若有, 则按式(1)~式(3)计算相似度; 然后以各曲线的凸包面积为权, 按式(4)计算所有曲线的加权相似度, 其中, SIM 为曲线集相似度, Sim_i, S_i 为分别为曲线 i 的相似度和凸包面积。

$$SIM = \sum_{i=1}^n (Sim_i * S_i) / \sum_{i=1}^n S_i \quad (4)$$

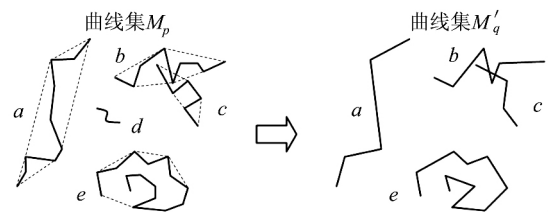


图3 渐进传输中的曲线集图形

Fig. 3 Curve Datasets Shape Under Progressive Transmission Mode

若将曲线集的相似度看作均值, 可得到各曲线相对于该均值的相似度误差, 据此计算相似中误差, 该中误差可代表曲线集的相似精度。如式(5)、式(6)所示, ΔSim_i 表示曲线 i 的相似度误差, σ_{SIM} 为曲线集的相似中误差。

$$\Delta Sim_i = SIM - Sim_i \quad (5)$$

$$\sigma_{SIM} = \sqrt{\sum_{i=1}^n (\Delta Sim_i * \Delta Sim_i) / n} \quad (6)$$

由于本方法可得到曲线集所有曲线的相似度, 可参照误差分布曲线^[15]的绘制方法, 根据一组相似度值来绘制相似度分布曲线, 该曲线能直观反映曲线集内各曲线相似度的大小及分布规律。

2 实验与分析

基于矢量等高线数据进行了网络多分辨率传输实验, 服务器端程序采用 Douglas-Peucker 算法实时生成 4 个分辨率等高线数据。最低分辨率 a 最早到达客户端, 然后是 b, c , 最后是完整的等

高线数据集 d 。实验中未系统地对数据进行拓扑一致性维护,多分辨率等高线仍存在着拓扑异化,客户端数据视图如图 4 所示。

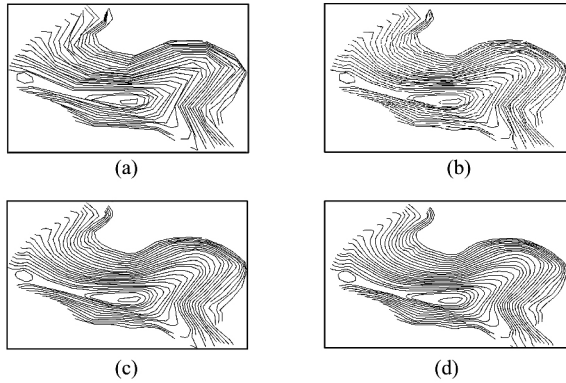


图 4 渐进传输模式下客户端的多分辨率数据视图
Fig. 4 Multi-resolution Data View of the Client Under Progressive Transmission Mode

取 $\alpha=0.6, \beta=0.4$,按本文方法计算图 4 的曲线集 a, b, c 相对于 d 的相似度分别为 0.796、0.882 和 0.957;相似中误差分别为 0.085、0.061 和 0.039。结果表明,最早到达的粗糙图形也具有较好的相似度,这是由于低分辨率图形是筛选后的典型特征点,能代表图形轮廓特征。随着传输的继续,曲线集相似度逐步增大,相似中误差则逐步减小,表明多分辨率传输中曲线集的相似度、相似精度都在逐步提高。

对于曲线集 a, b, c ,可得到各曲线集内所有曲线的相似度,根据各组相似度值可绘制相似度分布曲线。绘制方法如下:将相似度域 $[S_1, S_2]$ 等间距分为 k 个区间,记区间大小为 $dSIM$,曲线个数为 n ,统计相似度值落在各区间内的曲线个数 V_i 。图 5 为曲线集 a, b, c 的相似度分布曲线,其中, $S_1=0.55, S_2=1, k=9$ 。

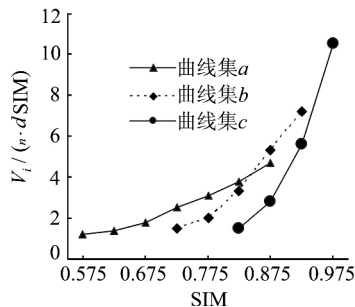


图 5 不同分辨率曲线集的相似度分布曲线
Fig. 5 Similarity Distribution Curve for Multi-resolution Curve Datasets

从图 5 可看出,曲线集 c 的相似度分布曲线最陡峭,图形陡峭表明位于较大相似度区间的曲

线数目较多,而位于较小区间的曲线数目较少; b 的次之; a 的最平缓。从相似度分布曲线对应的 X 轴区间看,曲线集 c 的 X 轴区间最小,且位于最右边; b 的次之; a 的最大,且靠近左边。区间越接近右边,表明其曲线相似度越接近于 1;区间越窄,表明相似度值越集中。从相似度分布曲线的形状可直观得出结论:曲线集 c 的相似性优于 b ,而 b 的相似性优于 a 。通过曲线集的相似度、相似中误差、相似度分布曲线可系统评估曲线集的相似性。

3 结 语

1) 曲线与其多分辨率表达形式间的差异主要体现在空间关系和几何图形上。本文通过拓扑相似度与图形相似度来计算单曲线的相似度,其中拓扑相似度的计算基于曲线拓扑关系邻域图,图形相似度的计算根据两曲线所围面积与曲线最小凸包面积的比值。单曲线的相似度能客观反映曲线在多分辨率传输中的相似程度。

2) 基于单曲线的相似度与凸包面积计算曲线集的加权相似度和相似精度,其中加权相似度表示曲线集经过选取、概括、拓扑关系变化后的整体相似程度,而相似精度指标即相似中误差可表示曲线集内各曲线间的相似差异大小。

3) 本文提出的相似度分布曲线可直观反映曲线集内各曲线的相似度大小及分布规律,利用相似度、相似中误差、相似度分布曲线可综合评估曲线集的相似性。

参 考 文 献

[1] An Xiaoya, Sun Qun, Xiao Qiang, et al. A Shape Multilevel Description Method and Application in Measuring Geometry Similarity of Multi-scale Spatial Data [J]. *Acta Geodaetica et Cartographica Sinica*, 2011,40(8):495-501(安晓亚,孙群,肖强,等.一种形状多级描述方法及在多尺度空间数据几何相似性度量中的应用[J].测绘学报,2011,40(8):495-501)

[2] Zahn C T, Roskies R Z. Fourier Descriptors for Plane Closed Curves [J]. *IEEE Transactions on Computers*,1972,21(3):269-281

[3] Ai Tinghua, Shuai Yun, Li Jingzhong. A Spatial Query Based on Shape Similarity Cognition [J]. *Acta Geodaetica et Cartographica Sinica*, 2009,38(4):356-362(艾廷华,帅赞,李精忠.基于形状相似性识别的空间查询[J].测绘学报,2009,38(4):356-362)

[4] Liu Pengcheng, Luo Jing, Ai Tinghua, et al. Evalua-

- tion Model for Similarity Based on Curve Generalization[J]. *Geomatics and Information Science of Wuhan University*, 2012, 37(1):114-117(刘鹏程, 罗静, 艾廷华, 等. 基于线要素综合的形状相似性评价模型[J]. 武汉大学学报·信息科学版, 2012, 37(1):114-117)
- [5] Zhao Yu, Chen Yanqiu. Included Angle Chain: A Method for Curve Representation[J]. *Journal of Software*, 2004, 15(2):300-307(赵宇, 陈雁秋. 曲线描述的一种方法: 夹角链码[J]. 软件学报, 2004, 15(2):300-307)
- [6] Hao Yanling, Tang Wenjing, Zhao Yuxin, et al. Arcal Feature Matching Algorithm Based on Spatial Similarity[J]. *Acta Geodaetica et Cartographica Sinica*, 2008, 37(4):501-506(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)
- [7] Zhao Binbin, Deng Min, Xu Zhen, et al. Development of General Rules for Matching Multi-scale Area Objects[J]. *Geomatics and Information Science of Wuhan University*, 2011, 36(8):991-994(赵彬彬, 邓敏, 徐震, 等. 多尺度地图面目标匹配的统一规则研究[J]. 武汉大学学报·信息科学版, 2011, 36(8):991-994)
- [8] Tang Luliang, Yang Bisheng, Xu Kaiming. The Road Data Change Detection Based on Linear Shape Similarity[J]. *Geomatics and Information Science of Wuhan University*, 2008, 33(4):367-370(唐炉亮, 杨必胜, 徐开明. 基于线状图形相似性的道路数据变化检测[J]. 武汉大学学报·信息科学版, 2008, 33(4):367-370)
- [9] Egenhofer M J. A Model for Detailed Binary Topological Relationships[J]. *Geomatica*, 1993, 47(3):261-273
- [10] Bruns H T, Egenhofer M. Similarity of Spatial Scenes[C]. The 7th International Symposium on Spatial Data Handling, Netherlands, 1996
- [11] Li Bonan, Fonseca F T. TDD—A Comprehensive Model for Qualitative Spatial Similarity Assessment[J]. *Spatial Cognition and Computation*, 2006, 6(1):31-62
- [12] Liu Tao, Du Qingyun, Mao Haichen. Spatial Similarity Assessment Model and Its Application in Line Groups[J]. *Geomatics and Information Science of Wuhan University*, 2012, 37(8):992-995(刘涛, 杜清运, 毛海辰. 空间线群目标相似度计算模型研究[J]. 武汉大学学报·信息科学版, 2012, 37(8):992-995)
- [13] Bertolotto M, Egenhofer M J. Progressive Transmission of Vector Map Data over the World Wide Web[J]. *Geoinformatic*, 2001, 5(4):345-373
- [14] Tang Luliang, Li Qingquan, Yang Bisheng. Shape Similarity Measuring for Multi-resolution Transmission of Spatial Datasets over the Internet[J]. *Acta Geodaetica et Cartographica Sinica*, 2009, 38(4):336-340(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)
- [15] Surveying Adjustment Discipline Groups, School of Geodesy and Geomatics. Error Theory and Surveying Adjustment[M]. Wuhan: Wuhan University Press, 2003(武汉大学测绘学院测量平差学科组. 误差理论与测量平差基础[M]. 武汉: 武汉大学出版社, 2003)

A Similarity Measurement Model for Multi-resolution Transmission of Curve Datasets over the Internet

CAO Zhenzhou¹

¹ College of Geomatics Engineering, Nanjing University of Technology, Nanjing 211800, China

Abstract: Concerning the dissimilarity between original spatial data and progressively transmitted spatial data, this paper proposes a model for measuring the similarity of curve datasets. After curve similarity is measured based on differences of the topological and geometric characteristics, the weighted similarity and the similarity precision for curve datasets are computed. The model uses weighted similarity to assess the overall degree of similarity for curve datasets and uses similarity precision to assess the similarity differences between curves. Finally, the model was used in an experiment and verified for effectiveness.

Key words: similarity model; progressive transmission; multi-resolution; shape similarity; topology similarity

First author: CAO Zhenzhou, PhD candidate, lecturer, specializes in the multi-scale organization and progressive transmission of spatial data. E-mail: czz@njtech.edu.cn