

空间点群目标相似度计算

刘 涛^{1,2} 杜清运¹ 闫浩文²

(1 武汉大学资源与环境科学学院,武汉市珞喻路 129 号,430079)
(2 兰州交通大学数理与软件工程学院,兰州市安宁西路 88 号,730070)

摘 要:顾及到视觉认知的 Gestalt 原则,提出了一种综合考虑到点群的空间关系、空间分布和几何特征,注重空间点群的整体描述的相似度计算模型。实验结果表明,相似度计算结果与直观认知比较一致。
关键词:空间相似性;似度;空间点群;空间关系
中图法分类号:P208

相似性在认知过程中起着关键性的作用,是进行分类、归纳推理和类比推理的基础。空间相似性在空间认知中起着同样的作用^[1]。对空间目标相似性进行研究,横向上可以为同一尺度空间下的不同目标之间提供空间数据的查询和检索^[2]、数据集成和互操作^[3]、数据挖掘等方面的应用,纵向上可以为不同尺度空间下的同一目标提供自动综合的终止条件判定和综合质量评价^[4]。

相似关系在心理学和认知科学中研究得较多,但作为相似关系在地理空间中的应用——空间相似关系,一直以来研究得较少。一方面由于空间相似关系的可计算性差^[5],另一方面,由于地理空间的复杂性,导致空间相似关系的计算要考虑的因素很多,如空间关系、空间分布、几何特征、语义特征等^[1]。与空间相似关系研究比较密切的应用研究包括模式识别和图像检索。模式识别中对相似性的研究侧重于单个图形对象^[6],图像检索中对图像相似性的研究针对于栅格图像^[7]。Bruns 等引入心理学中的变换模型(transformation model)^[8]。Li 等在此基础上进行了完善,考虑到空间拓扑关系、方向关系和距离关系的优先等级,提出了 TDD(topology-direction-distance)模型^[1]。

点群目标是空间分布分析的重要对象^[9],本文考虑到视觉识别的 Gestalt 原则,注重于空间点群目标的整体描述,考虑到空间点群目标的空间关系和空间分布及几何特征,对空间点群目标相似度进行整体度量,并以深圳市医疗点群分布图及其综合

后的图,和深圳市采石场点群分布图进行了实验。实验表明,结果与直观认知比较符合。

1 空间相似性的定义

1.1 空间相似性的定义

Holt 综合考虑了空间尺度和空间上下文对空间相似性定义如下:在某一特定的粒度(比例尺)和内容(专题属性)上被认为是相似的两个区域。但该定义过于笼统和模糊。文献[4]用集合学的方法对空间相似性进行了定义,具有清晰的数学描述。

1.2 空间点群特征集合

关于空间目标相似性研究,Bruns 等提到一个重要原则,即空间目标之间的拓扑、方向和距离关系是最关键的,因为这几种关系抓住了空间目标集合体-空间场景的本质^[8]。关于空间点群的空间分布和几何特征,文献[9-13]等提出了如下参数用以描述点群的分布特征,包括点数(点群中包含点的数目)、点群分布范围、点群密度、分布中心、分布轴线等。由于点数比较简单,且论文在点群拓扑相似度中隐含了对点数的描述,分布中心和分布轴线在点群的方向和距离关系中也会有所涉及,故主要讨论点群分布范围和点群分布密度。

综上所述,定义空间点群目标的特征集合 C 如下。

定义 1 空间点群目标的特征几何 $C = \{ \text{空}$

间拓扑关系,空间方向关系,空间距离关系,点群分布范围,点群分布密度}。

2 空间点群目标相似度计算

2.1 点群拓扑相似度

对于点群之间的拓扑关系判定比较简单,可以用点的邻居来描述。点的邻居可以是 Voronoi 邻居、 k 个最近邻居或定长距离邻居等。选择 Voronoi 邻居作为点群拓扑信息的描述参数,其优势在于:① 和其他两类邻居相比,其生成没有参数约束;② 和 k 个最近邻居相比,其邻居个数不是人为固定的;③ 和定长邻居相比,它对于尺度变化和点群密度变化具有适应性;④ 它的邻居关系具有对称性^[13]。

对于点数为 N_1 ,拓扑邻居总数为 B_1 的空间点群 P_1 和点数为 N_2 ,拓扑邻居总数为 B_2 的点群 P_2 ,其拓扑相似性判定可用式(1)来进行判定:

$$\text{SIM_topo} = 1 - \frac{|H_2 - H_1|}{\max(H_2, H_1)} \quad (1)$$

式中, $H_1 = \frac{B_1}{N_1}$, $H_2 = \frac{B_2}{N_2}$, SIM_topo 即为点群目标 P_1 和 P_2 之间的拓扑相似度。

2.2 点群方向相似度和距离相似度

现实空间的点群分布,通常是描述带有一定

$$\tan\theta = \frac{(\sum_{i=1}^n x_i'^2 - \sum_{i=1}^n y_i'^2) \pm \sqrt{(\sum_{i=1}^n x_i'^2 - \sum_{i=1}^n y_i'^2)^2 + 4(\sum_{i=1}^n x_i' y_i')^2}}{2 \sum_{i=1}^n x_i' y_i'}$$

4) 计算沿长轴方向的标准差 δ_x 和短轴方向的标准差 δ_y :

$$\delta_x = \sqrt{\frac{\sum_{i=1}^n (x_i' \cos\theta - y_i' \sin\theta)^2}{n}}$$
$$\delta_y = \sqrt{\frac{\sum_{i=1}^n (x_i' \sin\theta + y_i' \cos\theta)^2}{n}}$$

有了椭圆中心、转角、长、短轴,即可确定标准差椭圆。需要说明的是,前面的计算中均默认所有点的权重为 1。实际上各点具有不同权重的情况是很常见的,此时,只需将权重值代入算式中,与第 i 个位置的坐标相关的项乘以对应的权重 w_i 即可,这样亦可考虑到点群的专题属性。

2.2.2 点群方向相似度

图 1 中可以看出,椭圆的长轴方向即为点群目标的主要分布方向,所有点到该方向的标准差

的方向偏离的地理现象。如某段高速公路上发生的交通事故,其发生位置通常会表现为由这段高速公路的形状决定的线状模式^[15]。标准差椭圆是用来描述点群分布的方向偏离的,该方法作为点群的分布分析方法已经嵌入在 ArcGIS9 软件中并得到广泛应用。采用点群的标准差椭圆来定义空间点群方向相似度和距离相似度,既考虑到实际空间点群目标的地理分布,又是对点群目标作整体考虑,符合视觉的 Gestalt 认知原则。

本文对目标点群生成标准差椭圆,计算其方向偏离程度来度量点群目标方向相似度,并在此基础上,根据目标点群的标准差椭圆的长、短轴之间的关系来进行点群目标距离相似度的度量。

2.2.1 标准差椭圆

标准差椭圆的中心为点群的分布中心。描述和定义标准差椭圆的三要素为:转角、沿长轴的标准差和沿短轴的标准差。其计算方法简述如下。

1) 对一点群数目为 n 的点群来说,计算其平均中心:

$$(\bar{x}_{mc}, \bar{y}_{mc}) = \left(\frac{\sum_{i=1}^n x_i}{n}, \frac{\sum_{i=1}^n y_i}{n} \right)$$

2) 对分布在研究区域内的每个点 (x_i, y_i) 进行坐标变换: $x_i' = x_i - \bar{x}_{mc}$, $y_i' = y_i - \bar{y}_{mc}$ 。

3) 计算转角 θ :

距离最小。该方向与 x 轴的夹角 θ 就是点群的主要分布方向的定量描述参数, θ 取值范围为 $[0, \pi]$, 则对标准差椭圆夹角分别为 θ_1 和 θ_2 的两个点群来说,定义其方向相似度如下:

$$\text{SIM_dire} = |\cos(\theta_1 - \theta_2)| \quad (2)$$

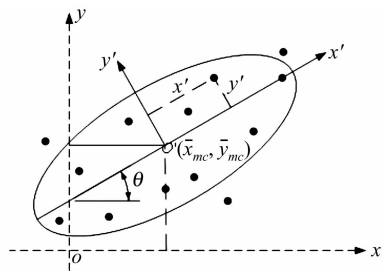


图 1 点群的标准差椭圆
Fig. 1 Standard Deviation Ellipse of the Point Cluster

当两点群的标准差椭圆夹角成 90° 时,两点群之间的方向相似度为 0;当两点群的标准差椭圆夹角相等或者相差 180° 时,两点群之间的方向

相似度为 1。同时,该式具有旋转不变性,比较符合直观认知。

2.2.3 点群距离相似度

标准差椭圆夹角描述了点群分布的方向偏离特征,但无法描述标准差椭圆的形态特征所表示的点群分布集中程度。对标准差椭圆长、短轴分别为 a_1 、 b_1 和 a_2 、 b_2 的两个点群来说,定义其距离相似度如下:

$$\text{SIM_dist} = 1 - \frac{\left| \frac{b_1}{a_1} - \frac{b_2}{a_2} \right|}{\max\left(\frac{b_1}{a_1}, \frac{b_2}{a_2}\right)} \quad (3)$$

式中,以点群的标准差椭圆的长、短轴的距离之比作为点群距离相似度的度量,实际上是描述了点群分布的集中程度,或者说描述了点群之间的距离集中程度。

2.3 分布范围相似度

点群的分布范围是一个不确定的问题。艾廷华^[9]利用 Delaunay 三角网进行“剥皮”操作从而得到点群的分布范围。该算法充分利用了 Delaunay 三角网的三角形边连接具有“最邻近性”特点,顾及了点群要素化简的 Gestalt 邻近性原则。本文采用该方法对点群构造边界并形成点群的“范围多边形”,点群的分布范围相似度的判定就转化为范围多边形形状相似度的判定。

考虑到分布范围相似度的度量需要具有旋转、平移和缩放不变性,郝燕玲^[14]提出了一种形状描述函数来计算面目标之间的形状相似度:

$$\text{SIM_scope} = 1 - \frac{\sqrt{\sum_{i=0}^n (f_1(l_i) - f_2(l_i))^2}}{\max\left(\sum_{i=0}^n f_1(l_i), \sum_{i=0}^n f_2(l_i)\right)} \quad (4)$$

2.4 分布密度相似度

生成原始点群的加权 Voronoi 图,将点的密度定义为该点所生成的 V 图的面积(a_i)的倒数 $1/a_i$ 。建立点密度 $1/a_i$ 与灰度的线性关系,生成点密度分布的灰度图像,计算其灰度直方图,把点群密度从空间域变换到频率域,利用灰度图像相似度判定算法进行点群分布密度相似性判断。

2.4.1 加权 Voronoi 图

由于 Voronoi 格网中的每一点到该格网中心的距离都小于到其他格网中心的距离,所以 Voronoi 图在空间剖分得到了广泛的应用。根据点的权重生成点集的加权 Voronoi 图更能准确地表达点群的分布情况。朱渭宁等^[16]将 Voronoi 的加权方法分为比例模型、引力模型和高斯模型

等 3 种模型。研究的对象为现实地理空间中的目标实体,故选择运用引力模型生成原始点群的加权 Voronoi 图。引力模型用公式描述如下。

设点 p_1 和 p_2 ,其权重(属性值)分别为 w_1 和 w_2 ,受该两点引力平衡点设为 p_e , $d(p_1, p_2)$ 代表两点之间的距离,根据引力模型,则有:

$$w_1/d(p_1, p_e)^2 = w_2/d(p_e, p_2)^2$$

当 $w_1 \neq w_2$ 时,可求得:

$$d(p_1, p_e) = d(p_1, p_2) \cdot (w_1 \pm \sqrt{w_1 w_2}) / (w_1 - w_2)$$

两个解中,应该去除位于 $p_1 p_2$ 延长线的一点,剩下的点即为所要找的引力平衡点 p_e 。根据引力模型生成点群的加权 Voronoi 图,计算各点 Voronoi 图的面积的倒数,与灰度级建立线性关系,生成点群密度灰度图像,从而将点群分布密度相似度的计算转化为灰度图像相似度的计算。

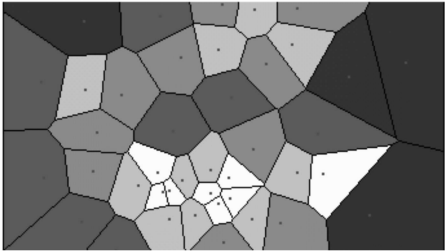


图 2 Voronoi 图及灰度级图像
Fig. 2 Voronoi Diagram and Its Corresponding Gray Image

2.4.2 图像相似度计算

由于图像的统计直方图具有旋转、平移不变性,在图像的检索中得到了重要应用。不同图像之间的相似性或差异性可以通过计算图像直方图之间的距离来进行度量。选用何种距离进行度量是一个关键的问题。此处采用传统的用到比较多的 L_1 距离^[17]进行相似度度量。假设有两幅图像 H 和 C ,其灰度直方图分别为 h 和 c ,则 L_1 距离定义如下: $d_{L_1}(H, C) = \sum_{i=1}^n |h_i - c_i|$ 。

灰度图像相似度也即点群分布密度相似度 SIM_densi 为:

$$\text{SIM_densi} = \frac{1}{N} \sum_{i=1}^N \left(1 - \frac{d_{L_1}(H, C)}{\max(h_i, c_i)} \right) \quad (5)$$

式中, N 为灰度空间样点数。

2.5 空间点群目标相似度计算

以空间拓扑、方向和距离相似度度量为基础,顾及点群的空间分布和几何特征,对空间点群相似度进行总体度量。在无显著侧重性的条件下,可取各点群相似度判定参数的平均数作为空间点

群的空间相似度,即有:

$$SIM = \sqrt[5]{SIM_topo \times SIM_dire \times SIM_dist \times SIM_scope \times SIM_densi}$$

该式为空间点群目标总体相似度计算模型。根据数理统计理论,相对于算术平均值,几何平均值在处理异常值方面表现得更加稳健、更不敏感。故对拓扑、方向、距离、分布范围和分布密度求取几何平均值,既是对几者综合考虑,同时也能减少其中某些项出现的异常值而造成相似度计算偏差。

3 实 验

为了验证算法的正确性,本文以深圳市为实验区,选择了深圳市 2008 年的两幅专题空间点群分布图,图 3(未严格按比例尺绘制)中,图 3(a)是

深圳市主要医疗点群分布图(150 个点,1:1 万),图 3(b)是该图的制图综合结果(67 个点,1:2 万),图 3(c)是同一地区重要采石场的点群分布(72 个点,1:1 万)。分别对 3 个空间点群两两求拓扑、方向、距离、分布范围和分布密度相似度及总体相似度,结果如表 1 所示。

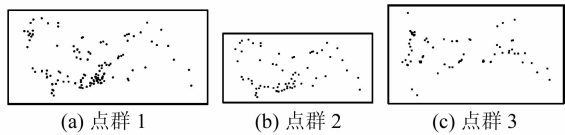


图 3 点群
Fig. 3 Point Cluster

表 1 3 个点群之间的相似度计算结果
Tab. 1 Results of Similarity Between the 3 Point Clusters

点群	SIM_topo	SIM_dire	SIM_dist	SIM_scope	Sim_densi	SIM
1,2	0.988	0.99	0.99	1	0.697	0.924
2,3	0.951	0.99	0.40	0.9	0.731	0.756
1,3	0.940	0.99	0.40	0.9	0.739	0.756

由表 1 中可以看出,点集 a 和 b 在各方面相似度都比较高,这说明地图综合结果比较理想。点集 a、b 分别和点集 c 之间的相似度度量值比较一致,这和点群 a、b 为同源目标点的事实比较一致;同时,点集 a、b 分别和点集 c 之间的方向相似度和范围相似度都比较高,这是因为同为同一地区的点集分布,分布范围和分布方向都受该地区地形及范围限制,导致两个专题的点集分布在方向和范围上比较相像;但在距离相似性上,由于为不同专题的点集集合,存在比较明显的差异。计算结果和直观感知比较符合。

参 考 文 献

[1] Li B, Fonseca F T. TDD-A Comprehensive Model for Qualitative Spatial Similarity Assessment. *Spatial Cognition and Computation*[J]. 2006, 6 (1): 31-62

[2] Goodchild M F, Egenhofer M J, Fegeas R. Interoperating GISs[R]. Report of a Specialist Meeting. http://www.ncgia.ucsb.edu/conf/interop97/interop_toc.html, 1998

[3] Frontiera P, Larson R, Radke J. A Comparison of Geometric Approaches to Assessing Spatial Similarity for GIR[J]. *International Journal of Geographical Information Science*, 2008, 22(3):337-360

[4] 闫浩文,褚衍东. 多尺度地图空间相似关系基本问题研究[J]. *地理与地理信息科学*, 2009, 25(4): 42-44

[5] 郭仁忠. 空间分析[M]. (第二版). 北京:高等教育出版社, 2001:240-244

[6] Arkin E M, Chew L P, Huttenlocher D P, et al. An Efficiently Computer Metric for Comparing Polygonal Shapes[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1991, 3 (13): 209-215

[7] Wang J Z, Li Jia, Wiederhold G. Simplicity: Semantics-Sensitive Integrated Matching for Picture Libraries[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, 9(23):947-963

[8] Bruns H T, Egenhofer M. Similarity of Spatial Scenes[C]. *Seventh, International Symposium on Spatial Data Handling, Delft*, 1996

[9] 艾廷华,刘耀林. 保持空间分布特征的群点化简方法[J]. *测绘学报*, 2002, 31(2):175-181

[10] 闫浩文,王家耀. 基于 Voronoi 图的点群目标普适综合算法[J]. *中国图像图形学报*, 2005, 10(5): 633-636

[11] 毋河海. 凸壳原理在点群目标综合中的应用[J]. *测绘工程*, 1997, 6(1):1-6

[12] 陆毅,翟京生,杜景海,等. 数字海图点群状特征的识别、测量与综合[J]. *武汉大学学报·信息科学版*, 2001, 26(2): 133-139

[13] 施建勇,闫浩文. 点群分布要素地图综合的质量评价方法[J]. *矿山测量*, 2008, 12(5):23-25

[14] 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

[15] David W S, Wong J L. ArcView GIS 与 ArcGIS 地理信息统计分析[M]. 张学良译. 北京:中国财政经济出版社,2008

[16] 朱渭宁,马劲松,黄杏元,等. GIS 中投影加权 Voronoi 图及竞争三角形生成算法研究[J]. 中国图

像图形学报, 2004, 9(3):334-338

[17] Stricker M, Orengo M. Similarity of color images [C]. SPIE95, San Jose, 1995

第一作者简介:刘涛,讲师,博士生,研究方向为空间关系。
E-mail:ltaochina@foxmail.com

Spatial Similarity Assessment of Point Clusters

LIU Tao^{1,2} DU Qingyun¹ YAN Haowen²

(1 School of Resource and Environmental Science, Wuhan University, 129 Luoyu Road, Wuhan 430072, China)
(2 School of Math, Physics & Software Engineering, Lanzhou Jiaotong University, 88 West Anning Road, Lanzhou 730070, China)

Abstract: With the consideration of Gestalt principles in visual cognition, this paper proposed a similarity assessment model which has a comprehensive consideration of point cluster’s spatial relationships, spatial distribution and geometry attributes. The topology relationship of point cluster is defined using point cluster’s topology neighbors, the direction relationship and distance relationship is defined using point cluster’s standard deviation ellipse, and then the computation method of point cluster’s topology, direction and distance similarity assessment is given out on this basis. Also the distribution range and the distribution density of point cluster and their similarity assessment method is put forward. Finally the formal five factors were combined to do a whole assessment of point clusters similarity. Experiment results show that the similarity assessment is consistent with intuitive cognition.

Key words: spatial similarity; similarity assessment; point cluster; spatial relationship

About the first author: LIU Tao, Ph.D candidate, with the research interests of spatial relationship.
E-mail: ltaochina@foxmail.com

(上接第 1143 页)

Three-Dimension Visualization Query Method Based on R-Tree

GONG Jun¹ XIE Xiao²

(1 School of Geography and Environment, Jiangxi Normal University, 99 Ziyang Road, Nanchang 330022, China)
(2 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

Abstract: Since k nearest neighbor (KNN) query which adopts relative conditions has more strict demands, this paper proposes a new KNN query algorithm, which adopts breadth traversal and optimal ordering principle and can fit different distributions of spatial features very well. Aiming at multi-scale phenomenon in 3D visualization, this paper presents a LOD feature query method. Experiments show that these spatial query methods have high efficiency and can retrieve focused features precisely. Moreover, LOD function can be realized to simplify 3d scene on the premise of guaranteeing rendering precision, which can satisfy current requirement of 3D GIS.

Key words: R-tree; spatial index; 3D spatial query; LOD

About the first author: GONG Jun, Ph.D, associate professor, majors in multi-dimension GIS and LiDAR.
E-mail: gongjunbox@gmail.com