引文格式:刘泽邦,杨岸然,马梦宇,等.可视驱动的大规模地理矢量点数据实时热力图生成方法[J].武汉大学学报(信息科学版),2026,51(1):114-125.DOI:10.13203/j.whugis20250031

Citation: LIU Zebang, YANG Anran, MA Mengyu, et al. Visualization-Driven Real-Time Heat Map Generation Method for Large-Scale Geographic Vector Point Data[J]. Geomatics and Information Science of Wuhan University, 2026, 51(1): 114-125. DOI:10.13203/j.whugis20250031

可视驱动的大规模地理矢量点数据实时热力图生成方法

刘泽邦[1] 杨岸然[1] 马梦宇[1] 陈 荦[1,2] 周嘉里[1] 景 宁[1,2]

1 国防科技大学电子科学学院,湖南 长沙,410073

2 自然资源部南方丘陵区自然资源监测监管重点实验室,湖南 长沙,410118

摘 要:热力图作为一种流行的可视分析方法,有助于用户直观地浏览地理矢量点数据的空间分布和数据密度。针对当前热力图生成方法计算效率随点数据规模增长而大幅下降的问题,提出可视驱动的实时热力图生成方法。该方法从直接生成最终热力可视结果的角度出发,将像素点作为独立的计算单元,直接计算像素热力值来生成最终的热力可视效果。首先,基于瓦片金字塔结构对点数据进行分层组织,构建用于支持基于像素点进行计算的可视驱动型空间索引。然后,基于可视驱动型空间索引设计像素热力值生成算法,采用邻域像素叠加的方式计算像素热力值,大幅提升计算效率且保持了数据的空间分布特性。最后,设计并行热力图可视计算框架,实现了交互式热力可视化。实验结果表明,所提方法大幅提升了热力可视化效率,为千万级规模地理点数据集生成热力图的耗时仅为现有方法的13.5%,并可在0.75 s内快速完成热力可视化交互,从而支撑大规模地理矢量点数据的交互式热力分析。

关键词:地理矢量点数据;空间大数据;热力图;可视驱动;实时计算

中图分类号:P208 **文献标识码**:A **收稿日期**:2025-05-16

DOI:10.13203/j.whugis20250031 **文章编号**:1671-8860(2026)01-0114-12

Visualization-Driven Real-Time Heat Map Generation Method for Large-Scale Geographic Vector Point Data

LIU Zebang[1] *YANG Anran*[1] *MA Mengyu*[1] *CHEN Luo*[1,2] *ZHOU Jiali*[1] *JING Ning*[1,2]

1 College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China

2 Key Laboratory of Monitoring and Supervision of Natural Resources in Southern Hilly Region, Ministry of Natural Resources, Changsha 410118, China

Abstract: **Objectives**: Heat map as a popular and mature visual analysis method at present, shows the distribution trend of geospatial data through the change of color shading. It helps users visualize the spatial distribution and data density of geographic vector point data. The rapid generation technology of geospatial vector data heat maps will significantly empower the improvement of the real-time performance and interactivity of spatial decision support. **Methods**: Aiming at the problem that the computational efficiency of current heat map generation method decreases greatly with the increase of point data scale, a visualization-driven real-time heat map generation method is proposed, which uses pixels as computing units to directly calculate the heat value to generate the final heat map effects. First, the point data is organized hierarchically based on the tile-pyramid to build a specialized spatial index that can support pixel-based computation. Then, the pixel heat generation algorithm is designed, and the neighborhood pixel superposition strategy is adopted to greatly improve the computational efficiency and maintain the spatial distribution characteristics. Finally, a parallel computing framework is designed to realize interactive heat map visualization. **Results**:

基金项目:国家自然科学基金(42101432,41971362);湖南省自然科学基金(2022JJ40546)。

第一作者:刘泽邦,博士生,主要研究领域为时空大数据可视分析。liuzebang19@nudt.edu.cn

通信作者:马梦宇,博士,副教授。mamengyu10@nudt.edu.cn

Experimental results show that the proposed method greatly improves the efficiency of heat map visualization, and the time required to generate heat maps for tens of millions of scale geographic point data sets is only 13.5% of the existing method, and the heat map visualization interaction can be quickly completed within 0.75 s, thus supporting the interactive heat map analysis of large-scale geographic vector point data. **Conclusions**: A large number of experimental results show that the proposed heat map generation method has the real-time performance and interactivity, and compared with the baseline method, the visualization efficiency has been significantly improved.

Key words: geographic vector point data; geospatial big data; heat map; visualization-driven; real-time computing

在大数据时代,随着基础测绘、位置传感、定位导航、智能交通以及物联网服务等技术的快速发展和广泛应用,带有空间位置信息的地理矢量数据爆炸式增长[1-2]。例如,在道路车流监控、社交媒体打卡、兴趣点推荐等应用背景下,产生了大量的地理矢量点数据,这些点数据体量庞大、数据繁杂,使得用户难以快速获取和挖掘其中蕴藏的规律和价值,故需要借助一些特殊的数据可视化手段进行表达[3-4]。热力图[5-8]也被称为热区图,作为当前一种流行和成熟的可视分析方法,其通过颜色着色深浅的变化来展现空间数据某一特征的分布态势,不仅可以反映空间维度上要素分布的疏密程度,而且可以反映属性维度上要素值的大小变化,在生物医学、生态环境以及地理信息等领域均具有广泛的应用[9-12]。因此,对大规模地理矢量点数据进行热力图可视分析,有助于解决直接进行大数据可视化的情形下由于数据分布不均产生的数据堆叠问题,便于用户更为直观地感受大规模点数据的分布态势和疏密程度,为下一步空间决策提供辅助支撑。

在探索式可视分析范式中,通常遵循“假设生成-可视化验证-策略修正”的动态循环,在此过程中,一方面要求可视化技术达到为每一次交互操作提供秒级反馈的实时响应性能,另一方面要求可视化技术具备支持快速生成多版本可视化结果,供用户进行对比验证的能力。基于此应用背景,地理矢量数据热力图的快速生成技术将显著提升空间决策支持的实时性与交互性。例如,在智能交通管理中,分析师通过缩放百万级移动轨迹位置数据的时空范围,要求完成从街区级至城市级尺度切换的道路压力热力图的秒级响应,以探索不同时段的拥堵热点演化规律;在城市规划动态模拟中,规划人员通过实时调整人口密度权重、交通流量阈值,以交互式验证不同参数下公共服务设施的空间覆盖均衡性;在商业智能分析中,分析师需基于兴趣点数据与用户签到记录实时生成消费密度热力图,并通过交互式调整权重参数与缓冲区半径,快速评估不同候选地块的客流聚集效应。综上所述,在探索式可视分析的应用场景中,热力图生成性能是直接决定人机协同决策的流畅度、维持空间认知连贯性的核心技术保障。

热力图是大数据可视分析的常用方法,现有关于热力图的研究主要集中于场景应用和算法设计两个方面。在场景应用方面,主流的GIS软件如ArcGIS、QGIS以及主要的前端可视化库百度地图、OpenLayers、Mapbox等均包含热力图可视化功能,均可交互地为地理矢量数据生成相应的地图热力可视效果。文献[8]测试了部分前端地图可视化库的热力图生成功能,以衡量它们的性能和处理大数据的能力。文献[13]对热力图泛化过程进行定量分析,比较了不同泛化程度下的热力图的可用性。许多研究直接将热力图可视方法应用于城市分析[14-16]、灾害应急调度[17]、交通热点分析[18-19]等各类场景;同时,部分研究关注约束情况下热力图生成方法的实现[20-22]。在算法设计方面,热力图的主要原理是基于核密度估计实现的,考虑到对大规模地理点数据进行核密度估计的计算量巨大,为提升热力图生成效率,现有研究主要可分为4类:(1)采用预计算和索引加速的方式来提升核密度计算中的查询速率。文献[23]采用预聚类减少计算量的方式,实现了百万级数据热力图的快速生成;文献[24]提出一种高效的基于四叉树的自适应核密度估计方法,在保证较高计算精度的同时提高了计算效率和视觉效果;文献[25]提出改进核密度估计的空间点密度算法,可较好地保持离散点的属性与整体形状。(2)采用采样的方式选取用于近似表达整个数据集热力效果的子集,从而减少用于核密度计算的点数据规模。文献[26]提出了一种基于空间填充曲线的采样方法,可以有效选取用于核密度估计的核心子集。(3)采用渐进式计算的方式,确保

用户持续接受核密度估计结果。文献[27]提出了一种基于二次函数剪枝的核密度可视化框架QUAD,并进一步设计了一种渐进式核密度可视化探索工具LibKDV[28],其可以在单机环境下近实时地产生合理的热力可视化结果。(4)通过高性能计算方式来直接提升计算速率。文献[29]基于Spark为大数据热力图设计了并行瓦片生成算法;文献[30]基于Spark计算框架设计了路网核密度估计并行算法,有效提升了计算效率并具备良好的可扩展性;文献[31]提出了一个基于Spark的分布式空间数据管理与可视化框架,拓展了高效的热力图生成性能,可为大规模空间点数据快速生成热力图。同时,有许多研究利用图形处理单元的图形处理优势来提升核密度估计算法的效率,以加速热力图生成过程,显著提升了针对空间大数据的计算效率[32-36]。

现有的热力图生成方法基于数据驱动的计算模式,其核心思想是以地理矢量要素为计算单元,先根据屏幕分辨率逐个计算每个地理矢量要素在目标栅格点的核密度值,再将其进行叠加,这一过程的计算复杂度为$O(nm)$,其中n为地理矢量要素数量,m为栅格化的像素数量。采用该计算模式,随着数据规模的急剧增长和屏幕分辨率的细化,串行算法的时间成本会呈现近似平方级增长,这将使得生成热力图的计算量大幅增加,无法实现随地图缩放而实时生成不同分辨率下的热力图可视效果。在当前研究中,利用高性能并行计算技术对上述计算模式进行并行化,能够缓解数据规模增长带来的热力图计算压力,然而,在地理矢量要素规模较大的情况下,其仍很难支持地理矢量的实时热力可视化,原因在于现有并行技术侧重于地理矢量数据在多计算节点中的任务划分及绘制结果的合并策略,而数据分片策略难以同时匹配数据负载均衡和核密度计算对于空间局部邻近的需求,这导致每个计算任务在跨区访问数据间的通信开销急剧增长,且任务负载的空间异质性将进一步加剧计算资源的低效利用,即使采用高性能并行计算技术仍难以满足实时热力可视的需求。例如,在一个带有64个进程的Spark集群中,仅为一个包含500万点要素的兴趣点数据集生成一张500×500像素分辨率的热力图就需要46 s,这对于交互式分析的响应需求是无法接受的。因此,本文从可视角度出发,提出可视驱动的实时热力图生成方法,首先,针对点数据进行分层组织,构建定制的可视驱动型索引,按照瓦片金字塔的划分方式得到多层分辨率下的像素区域内的统计值;然后,设计像素热力生成算法,以像素点为计算单元,将热力值计算优化为基于索引加速的像素查询,再对匹配所得的统计值计算热力值,由此大幅提升计算效率;最后,为上述流程构建了并行热力图可视计算框架。该方法可根据当前屏幕范围和分辨率直接计算像素热力值,为地图放缩平移以及热力半径更改等交互操作实时渲染生成热力可视效果。本文在6个数据集上进行了大量实验,其数据规模逐渐增长至千万级,实验结果表明,本文方法生成千万规模数据集的耗时仅为现有方法的13.5%,并且可在0.75 s内完成热力图生成交互,可高效地支撑实时热力可视化。

1 问题定义

目前主要的热力图生成方法中,最典型的是基于核密度的热力图生成方法,主要采用数据驱动的计算方式,如图1所示,其核心思想是将矢量点要素作为计算单元,在用户交互过程中实现热力图的可视化生成,该计算流程包括4个主要步骤:

1)获取屏幕范围内的矢量点要素;

2)逐一以矢量点要素坐标为圆心,根据设定的搜索半径计算热力渐变圆,通过核函数来表达热力值的渐变,在圆心处的热力值最高,随着与圆心距离的增加,热力值逐渐减小,在与圆心点距离等于搜索半径的位置处热力值为零;

3)依据当前的屏幕分辨率,将所得热力渐变圆进行叠加,生成数据集的热力值合并结果;

4)选取渐变色带条,对热力值合并结果进行颜色映射,生成最终的热力图可视效果。

采用上述数据驱动的计算方式,主要存在两点不足:(1)当数据集包含的矢量点要素规模较大时,其分析计算量较大,需要耗费大量时间计算得到热力图可视化结果,难以满足交互式可视化的性能响应;(2)可视结果的分辨率有限,在缩放过程中存在失真问题,无法实现随地图缩放而实时生成不同分辨率下的热力图可视效果。综上,当前主要热力图可视方法难以应对大规模地理矢量点数据的实时热力图可视化。

考虑到对空间矢量点数据生成热力图的最终目的是面向屏幕可视化显示的需求,其本质即是计算生成栅格图像,故本文提出可视驱动的热力图生成方法,从栅格图像可视显示的最小单元

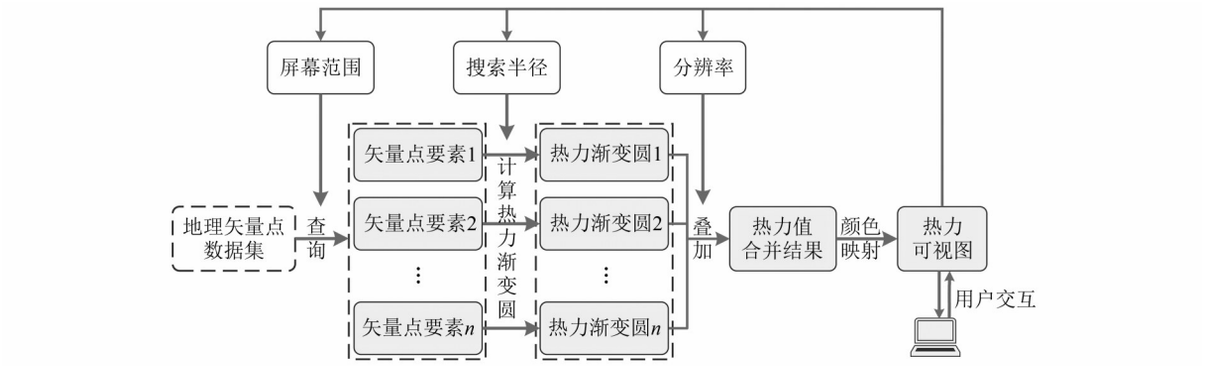


图 1　数据驱动的热力图生成流程

Fig. 1　Generation Process of Data-Driven Heat Map

即像素点的角度出发设计新的热力图生成流程。如图 2 所示,该方法直接以像素点为计算单元,通过计算数据集在像素点处的热力值,进而快速生成热力图可视效果。针对用户的交互热力图可视请求,处理流程如下:首先查询当前可视分辨率下屏幕范围内的像素点,并将原始点数据集转换为特定的可视驱动型空间索引结构,以支撑像素热力值的快速计算;然后,合并所有像素热力值;最后,进行颜色映射实时生成最终的热力图可视效果。

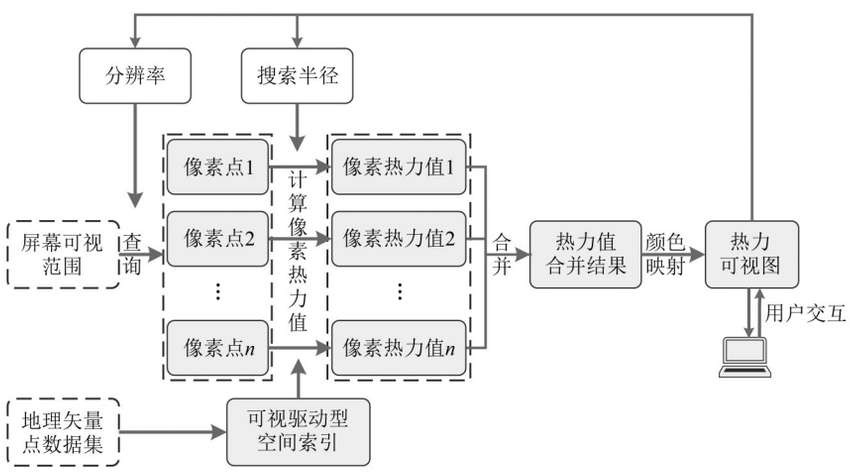


图 2　可视驱动的热力图生成流程

Fig. 2　Generation Process of Visualization-Driven Heat Map

2　可视驱动的热力图生成方法

可视驱动的热力图生成方法的核心思想是以像素点为计算单元,快速计算像素热力值生成最终的热力图可视效果。针对可视驱动的计算方式,可将热力图可视化处理转换为如何计算点数据集在像素点处的热力影响的问题。一个直观的方式是计算所有点要素在像素点处的热力值并进行合并得到像素热力值,然而,并不是所有点要素在像素点处都会产生热力影响,而且当影响像素点热力值的点要素数量较多时,计算热力值也非常耗时。

考虑到矢量点数据分布不均的空间特性,在可视化时许多点要素会落在同一个像素点表示的空间范围内,故选取该像素点代表这些点要素来计算对其他像素点的热力影响,这样不仅可大幅减少计算量,而且可保持空间数据的分布特性。同时,为实现多层级的显示效果,采用瓦片金字塔模型来呈现热力图可视结果,因此还需要根据瓦片金字塔的可视显示方式来对点要素进行特定的数据组织。

综上所述,可视驱动的热力图生成方法的设计流程可概括为如图 3 所示。首先,将大规模地理点数据的热力可视问题转换为如何基于像素点计算热力值的问题;然后,针对像素热力值计算问题,基于瓦片金字塔模型将矢量点数据集转换为可视驱动型空间索引;最后,基于像素热力值计算问题和所设计的空间索引,设计像素热力值生

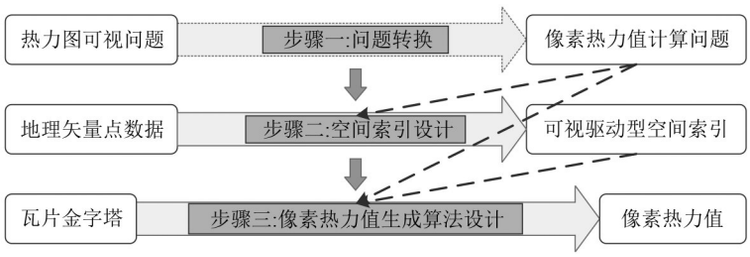


图3 可视驱动的热力图生成方法设计流程

Fig. 3 Design Flow for Visualization-Driven Heat Map Generation Method

成算法,以像素点为计算单元,计算瓦片金字塔中像素点的像素热力值,快速生成热力图可视效果。

2.1 可视驱动型空间索引

面向可视驱动的地理矢量点要素组织是该方法的基础,针对瓦片金字塔的显示方式为矢量点数据集设计可视驱动型空间索引[37]。如图4所示,瓦片金字塔是一种多层级、多分辨率图片组织模型,在瓦片金字塔的第0层级,通过一张瓦片来呈现全球空间范围内所有要素的可视化结果,随着显示层级的增加,瓦片数量按照1∶4的比例增加,可视分辨率也随之升高一倍。在瓦片金字塔中,所有的瓦片均为规则的256×256像素。采用该方式可实现热力可视效果的动态无缝浏览。

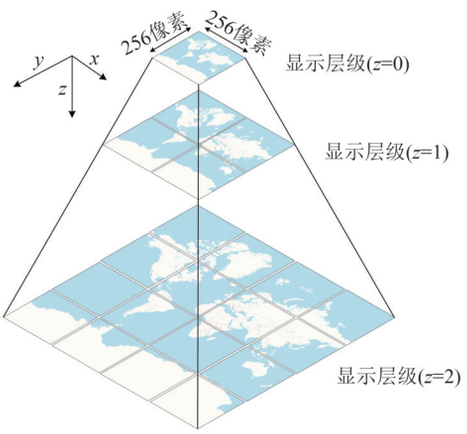


图4 瓦片金字塔模型

Fig. 4 Tile-Pyramid Model

因为最终的热力可视结果通过瓦片金字塔进行展示,为了能够直接基于瓦片像素点来计算热力影响,该索引基于瓦片金字塔的划分方式对点要素进行分层组织,将数据集转换为像素粒度的组织结构。索引生成流程如图5所示,L为赤道周长的1/2,索引从墨卡托投影下[38]的全球空间范围开始向下递归四分,按照Z曲线对划分区域进行Geohash编码,分别用0~3表示一次划分所得的左上(left up,LU)、右上(right up,RU)、左下(left bottom,LB)、右下(right bottom,RB)区域。在构建索引时,索引初始为单个根节点,表示全球空间范围的相关信息,从根节点向下,对落入相同划分区域内的点要素创建索引节点,在索引节点中记录该划分区域的编码、包含矢量点要素数量、空间区域(用LB和RU角点表示)、4个子节点指针,当划分区域内不包含矢量点要素时,则不创建索引节点。当继续向下递归划分时,若该划分区域的子区域创建了索引节点,则将相应的子节点指针指向创建的子节点,反之指针设为空。依照上述递归流程,生成最终的可视驱动型空间索引。

通过定制化设计上述索引生成流程,可视驱动型空间索引可以很好地适应基于像素生成热力值的计算需求,其原因在于:当递归至第8次时,可将全球空间范围划分为$2^8\times2^8$(256×256)个子区域,这恰好与一张瓦片的像素规模相等,故索引中第8层的节点与瓦片金字塔中第0层级中一张瓦片中的像素点是唯一匹配的。照此规律,从索引第8层开始,可直接根据索引节点信息来计算对应的瓦片像素点的热力影响。

2.2 像素热力值生成算法

基于可视驱动型空间索引设计了像素热力值生成算法,其核心思想是直接计算像素热力值以快速生成热力图可视效果。由于对像素点产生热力影响的点要素在空间上必然是分布于其邻域像素点所表示的空间区域内,为减小像素热力值的计算量,采用邻域像素点叠加的方式,直接将以待计算像素点为圆心、热力半径为半径的圆内的像素点在该点处的热力值进行叠加,像素热力值的计算公式为:

$$H_p=\frac{1}{nR^2}\sum_{i=1}^{n}(S(p_i)/s_{\max})K(D(p_i,p)/R) \quad (1)$$

式中,p为待计算像素点;H_p为像素热力值;n为热力范围内的像素数量;R为热力搜索半径;p_i为p相邻邻域内的相距小于R的像素点;$s_{\max}$为像素点包含点要素数的最大值;S函数用于获取p_i内的点要素数;D函数用于计算像素点间的距离;K函数表示核函数。

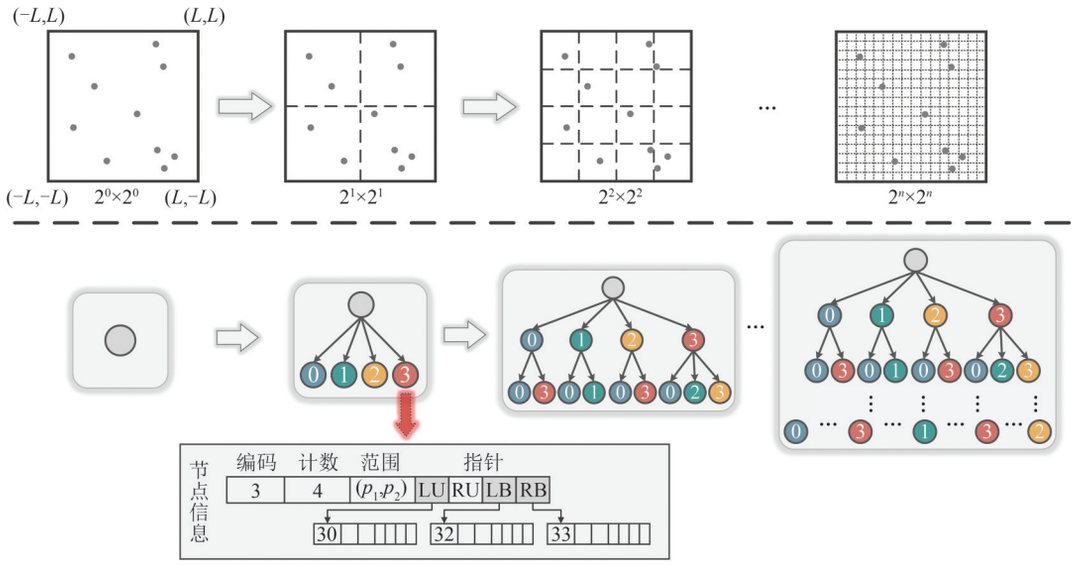


图 5　可视驱动型空间索引生成流程

Fig. 5　Generating Process of Visualization-Driven Spatial Index

式(1)的关键是快速获取像素点内的点要素数,可通过在可视驱动型空间索引中查找像素点所对应的索引节点信息来实现,图 6 为索引节点查找流程。在查找前,首先根据像素点计算其空间范围,并得到其地理哈希编码。在查找时,从索引的根节点开始遍历其子节点,通过子节点指针是否为空来判断子节点是否存在,若存在,则获取子节点编码,与像素编码进行前缀匹配。若匹配失败,则说明索引节点不存在,像素点内的点要素数为 0;反之,继续判断两者是否相等,若不相等,则以该子节点作为当前节点,继续按照上述步骤进行查找,直至最后两个编码相等时输出得到像素点对应的索引节点,获取节点的计数信息即是像素点内的点要素数。

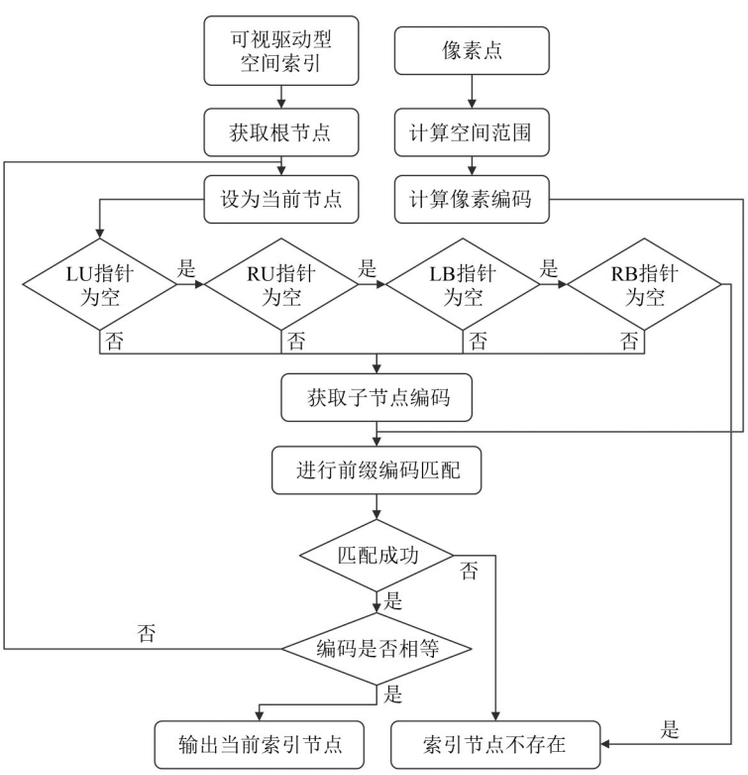


图 6　索引节点查找流程

Fig. 6　Search Process of Index Node

同时,在计算邻域像素点的热力影响时,邻域像素点需要满足距离小于热力半径的条件。基于像素点规律排布的特性,某一像素点的邻域像素点易通过瓦片分辨率计算得到,且越邻近该

像素点的像素点满足距离条件的几率越大,所以以待计算像素点为中心,由内向外逐层判定邻域像素点是否小于热力距离。

为进一步评估方法的时间复杂度,假设矢量点要素数量为N,考虑到热力半径和屏幕分辨率是固定的,故每个点要素热力半径内的像素数量均相同,设为M,则数据驱动的计算方式的时间复杂度为$O(NM)$。而在像素热力生成算法中,假设屏幕内的像素总数为C,在遍历计算像素值时所需的邻域像素点数量也为M,则该算法的时间复杂度为$O(CM)$。在常用的可视化设备中,C一般不超过百万量级,而N可轻易达到千万量级,由此分析可知本文算法可大幅降低热力图生成的计算复杂度。

像素热力生成算法输入瓦片编号T、可视驱动型空间索引I和设定热力半径W,输出得到该瓦片的热力可视图,主要步骤为:(1)创建规格为256×256像素的空白瓦片画布;(2)获取索引第$T+8$层中节点计数最大值s_{max};(3)遍历瓦片像素点p,按照Z曲线计算瓦片的地理哈希编码,按照图6流程查找索引节点;(4)当节点存在时,获取节点计数信息s_{pixel},计算像素点自身的热力值,累加至最终的像素热力值;(5)当节点不存在时,遍历像素点的邻域像素点p_{near},重复步骤(3)~(4)查找索引节点并计算在p点的热力值,累加至最终的像素热力值;(6)将像素热力值进行颜色映射,将所得颜色样式绘制在画布上;(7)所有像素遍历完毕,将瓦片画布写出存为瓦片图像。

3 并行热力图可视计算框架

可视驱动的热力图生成方法的并行计算框架如图7所示,该框架基于浏览器/服务器模式进行搭建,在浏览器端设计了基于OpenLayers的可视框架,用于支撑热力图可视效果的多层级交互浏览,具体包括处理用户的交互请求并发送至服务器端,同时将服务器输出的瓦片图组织为最终的热力可视效果。在服务器端,主要包括热力可视地图瓦片服务和并行热力计算引擎两个部分。热力可视地图瓦片服务一方面接受浏览器的瓦片任务请求,并进行并行任务划分,另一方面将计算所得热力值结果封装为热力瓦片图,并传输至浏览器端进行显示。并行热力计算引擎是框架的核心,其为像素热力生成算法设计了基于消息传递接口(message passing interface,MPI)与开放式多处理混合模型(open multi-processing,OpenMP)的多进程与多线程结合的并行计算流程,因为瓦片每个像素热力值的计算相互独立,所以将每个瓦片任务请求被视为独立任务分配至MPI计算进程,并且为提升瓦片生成速率,在每个进程内,启动多个OpenMP线程用于像素热力值计算。在计算前,空间矢量点数据被转换为空间索引为并行热力计算引擎提供支撑。

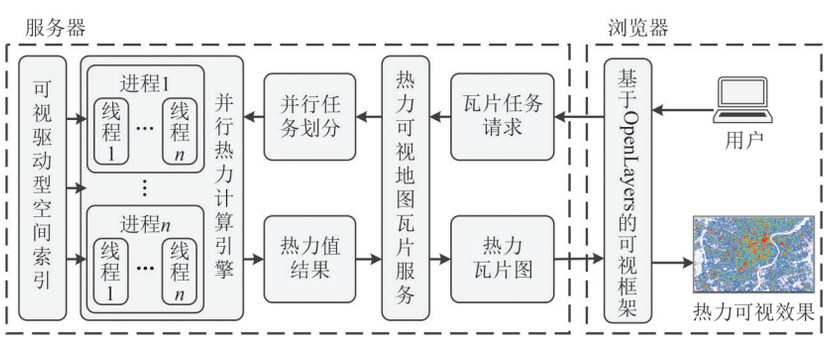


图7 并行热力图计算框架

Fig. 7 Parallel Heat Map Computing Framework

4 热力图生成实验

实验数据集如表1所示,均来源于开源数据。在数据集选取中,首先,从P_1至P_6数据规模逐渐增加,一是可说明本文方法在不同规模数据集上的算法提升性能,二是可测试本文方法热力生成性能对于不同数据规模的敏感性。同时,P_6数据集的规模达到了千万级别,可验证本文方法具备支持千万级规模地理矢量点数据热力图的秒级响应。最后,也可说明本文方法具备支持现实应用需求的能力,例如,对于P_3数据集的交互式热力图生成,可以量化分析不同类型兴趣点设施的覆盖率和集聚度,辅助政府进行民生项目选址与资源调配;对于P_5数据集的热力图生成,可以反

映不同国家和地区间的经济发展差异,辅助基础设施盲区识别与补建决策。共设计了3组实验:实验1将本文方法与现有主要方法进行性能对比,验证本文方法的优异性能;实验2验证本文方法能够支撑实时交互热力可视化的性能;实验3分析了热力半径的变化对本文方法性能的影响。在实验1中,GeoSparkViz的实验环境为:配置为128 GB,CPU为32核2.3 GHz的3台机器构成的Spark集群,其余的实验环境配置为一个内存为16 GB,CPU为8核2.60 GHz的单机。在测试本文方法时,实验中所有数据集均在内存中被转换为可视驱动型索引。在3个实验中,本文方法均启动4个MPI进程,且每个进程中包含4个OpenMP线程。

表1　实验数据集

Table 1　Experimental Datasets

数据集	数据规模/个
P_1:中国大陆自然地理位置点	48 148
P_2:中国大陆地名位置点	324 326
P_3:北京市兴趣点	679 414
P_4:中国点要素	1 304 980
P_5:全球4G基站位置点	2 350 953
P_6:全球地名位置点	12 365 365

4.1　实验1:与现有方法的性能对比

实验1将本文方法与数据驱动的热力图生成方法进行对比,验证本文方法相较现有方法的性能优势。将本文方法与前端热力可视化库(百度地图、OpenLayers、Mapbox)、主流GIS软件(ArcGIS、QGIS)和热力图生成方法(Seaborn、LibKDV[28]、GeoSparkViz[31])进行对比。Seaborn是一个支持探索性数据分析的Python数据可视化库;LibKDV是一个用于地理可视分析的多功能核密度可视化工具,其针对核密度计算的复杂度进行了优化处理;GeoSparkViz是一个面向大规模地理空间数据处理和分析的分布式地理空间可视化框架,可以并行化地生成热力图效果。在进行方法比较时,热力半径是影响方法效率的一个主要因素,为排除热力半径对于热力生成方法的影响,将所有方法的热力半径设置为相同的宽度,并统计生成热力图的时间。

所有方法生成热力图可视效果的时间如表2所示。从结果可知,在当前热力图生成方法中:首先,3个前端热力可视化库百度地图、OpenLayers和Mapbox在规模较小的数据集上具备较快的生成效率,但其受限于网页端的计算性能和数据传输瓶颈,无法实现大数据集的热力可视化。其次,热力图作为计算密集型应用,在没有进行优化的情况下,QGIS、ArcGIS与Seaborn 3种工具的计算非常耗时。最后,LibKDV和GeoSparkViz两个方法通过算法设计和并行优化大幅提升了计算效率,GeoSparkViz的性能明显优于其他方法,然而,其采用数据驱动的计算模型,使得热力图生成的计算效率随数据规模的增长而急剧下降,对于规模较大的P_5和P_6数据集仍需要耗费较长的计算时间,无法实现实时热力可视化。作为比较,本文方法具备高效的热力图生成性能,其生成P_5和P_6数据集热力图的耗时仅约为GeoSparkViz的9.9%(4.6 s÷46.3 s)和13.5%(27.2 s÷201.6 s)。实验1结果体现了本文方法具有计算复杂度低的优势。相较数据驱动的计算模式需要计算所有点要素在目标点的核密度值,本文方法只需计算查询范围内的像素点在目标点的热力值,其规模远远小于点要素数量规模,由此大幅降低了计算复杂度。

表2　不同方法生成热力图时间对比/s

Table 2　Time to Generate Heat Map/s

方法	P_1	P_2	P_3	P_4	P_5	P_6
百度地图	0.4	5.8	10.1	24.9	138.3	
OpenLayers	0.3	4.4	8.4	21.2	120.9	
Mapbox	0.3	1.8	5.5	17.9		
ArcGIS	7.3	18.5	94.5	136.2	259.6	
QGIS	61.1	452.2	740.3			
Seaborn	20.5	216.6	460.0	957.9		
LibKDV	4.5	14.2	35.0	31.9	79.2	224.6
GeoSparkViz	5.6	12.5	15.9	21.7	46.3	201.6
本文方法	0.6	0.8	2.1	0.8	4.6	27.2

为进一步验证本文方法在各显示层级上对于大规模矢量点数据集的热力图生成性能,选取GeoSparkViz与本文方法进行比较,在规模最大的两个数据集P_5和P_6上进行实验,按照瓦片金字塔的格式分别绘制0、2、4、6、8、10显示层级上的全部可视瓦片,计算各显示层级上的瓦片绘制速率如图8所示。从结果可知,GeoSparkViz在显示层级较高时的性能稍好,但当显示层级较低时性能急剧下降,这是由于随着显示层级的降低,瓦片内的点要素规模较大,需要计算热力值的点要素也随之增加,使得计算非常耗时。本文方法相较GeoSparkViz在各个显示层级上的计算性能均具有明显的优势,在P_6数据集上的最低速率仍有42.6张/s,大幅优于现有方法性能,可很好地应用

于大规模地理矢量点数据的热力图生成。该实验结果进一步体现了本文方法具有交互响应快的优势,相较数据驱动计算模式的可视化效率随层级降低而急剧降低,本文方法在地图缩放中可根据显示层级和可视分辨率重新实时计算生成新的热力可视结果,实现对热力可视效果的无缝浏览,且不会出现放缩失真的情况。

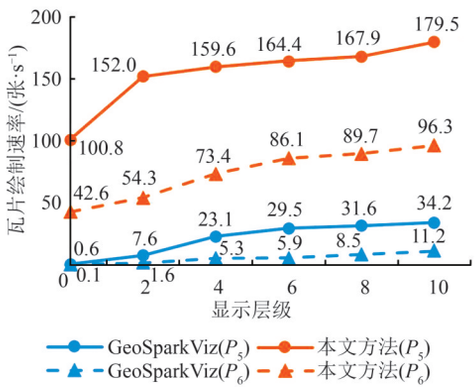


图8 各显示层级瓦片绘制速率

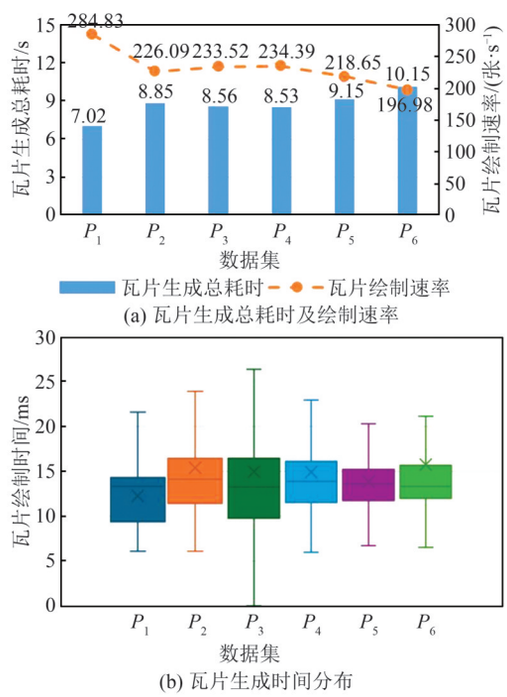
Fig. 8 Tile Rendering Speed at Each Zoom Level

4.2 实验2:实时交互性能的验证

在实验2中,对于每个数据集,设定热力半径为3个像素宽度,并在0~18的显示层级上随机选取2 000张非空的可视瓦片进行绘制,统计瓦片生成时间得到实验结果。图9展示了本文方法对于所有数据集的瓦片生成性能。图9(a)为瓦片生成总耗时及绘制速率,随着数据规模的增长,生成所有瓦片的总耗时并未明显增长,说明方法具备对数据规模不敏感的特性,在P_6上的瓦片绘制速率仍约有197张/s,在秒级交互响应的应用需求下,本文方法的瓦片生成速率远大于实际应用中屏幕视口范围内每秒通常不超过50张瓦片的数量要求。图9(b)为瓦片生成时间的箱图分布,对于所有数据集,瓦片最慢可在30 ms内完成绘制,假设当前用户浏览发送100个瓦片请求,根据当前启动的MPI进程数,通过25轮,最慢在0.75 s(25×30 ms)内完成所有瓦片绘制,可在1 s内完成交互响应。综上,该方法具备较好的实时交互性能,可用于大规模地理矢量点数据的交互式实时热力可视化。

4.3 实验3:热力半径对方法性能的影响

在实验3中,将输入的热力半径分别设置为2、4、6、8、10、15、20像素。对于每个数据集,随机选取0~18的显示层级上的2 000张可视瓦片,分别计算在上述热力半径下的瓦片绘制速率,结果如图10所示。从实验结果可知,对于所有数据集,随着热力半径的增加,瓦片生成速率呈现下降的趋势,这与实际相符,随着热力半径增加,需要用于叠加计算的邻域像素点增加,使得瓦片生成时间增加,必然会导致瓦片绘制速率下降。为测试方法性能下限,实验测试了较大热力半径下方法的性能,实验结果表明,当热力半径小于20像素时,本文方法具备较高的瓦片绘制速率;当热力半径达到20像素时,方法在P_3上的最慢瓦片绘制速率约为21张/s,仍基本具备支持交互式热力可视的性能。综上可得,热力半径越小,方法性能越好,当热力半径小于20像素时,上述进线程设置可以基本满足交互需求,当需要设置更大的热力半径时,若想提高交互性能,则可以通过增加计算进程和线程的方式来进一步提升瓦片绘制速率。



(a) 瓦片生成总耗时及绘制速率

(b) 瓦片生成时间分布

图9 瓦片生成性能

Fig. 9 Tile Rendering Performance Evaluation

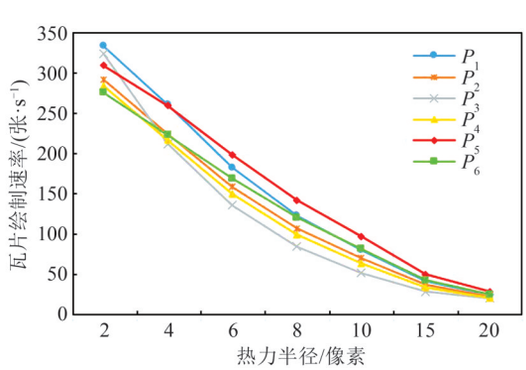


图10 不同热力半径下瓦片生成速率

Fig. 10 Tile Rendering Speed for Different Heat Radius

4.4　实验 4:热力生成案例展示

选取P_3数据集进行热力图生成展示,并从前端热力可视化库、主流GIS软件和热力图生成方法中各选取一个代表方法(Mapbox、QGIS、LibKDV)与本文方法进行比较,实验结果如图11所示。根据实验结果,不同方法进行热力值计算和渲染的方式有所区别,使得本文方法与其他方法在最终的热力图效果存在细微的差别,但所有方法对于数据密集区域的突出表现基本保持一致,验证了本文方法可以保证高质量的热力图可视化效果。

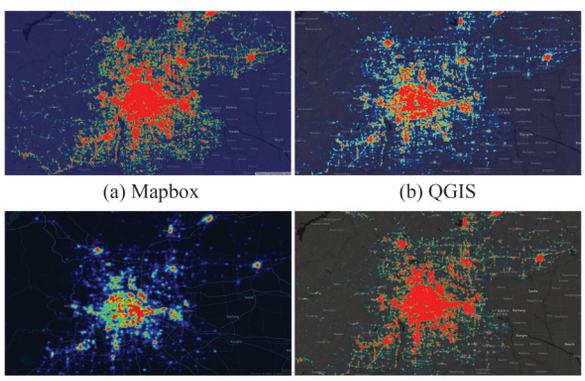


图11　不同方法生成热力图效果示例

Fig. 11　Demonstrations of Heat Map Effects Generated by Different Methods

5　结　语

热力图分析是一种重要的地理空间数据分析手段,有助于用户快速识别空间模式、热点区域和趋势。针对当前热力图生成方法的效率随数据规模增长而大幅下降的问题,本文提出了可视驱动的大规模地理矢量点数据实时热力生成方法。该方法以像素点为计算单元,设计匹配像素计算的可视驱动型空间索引,采用邻域像素点叠加的方式生成像素热力值,大幅提升了热力图生成效率。同时,该方法设计了并行计算框架,可根据显示层级和热力半径的变化实时生成新的热力可视结果。实验结果表明,该方法性能大幅优于当前热力图生成方法,可支撑千万级地理矢量点数据的实时热力可视化。本文方法主要面向静态的地理矢量点数据的热力图快速生成场景,故对数据类型和数据模式有一定的限制,对于几何结构更加复杂的地理矢量数据类型和动态更新的流式数据的支持度不足。在后续研究中,将进一步扩充矢量线数据和面数据类型的热力图生成算法,并研究流式地理矢量数据的热力图动态生成与增量更新算法。

参　考　文　献

[1] 李德仁,张过,蒋永华,等. 论大数据视角下的地球空间信息学的机遇与挑战[J]. 大数据,2022,8(2):3-14.
LI Deren, ZHANG Guo, JIANG Yonghua, et al. Opportunities and Challenges of Geospatial Information Science from the Perspective of Big Data[J]. *Big Data Research*, 2022, 8(2): 3-14.

[2] 白穆. 地理空间大数据建设面临的问题与对策[J]. 测绘与空间地理信息,2023,46(2):10-12.
BAI Mu. Problems and Countermeasures of Geospatial Big Data Construction[J]. *Geomatics & Spatial Information Technology*, 2023, 46(2): 10-12.

[3] GUO T, FENG K Y, CONG G, et al. Efficient Selection of Geospatial Data on Maps for Interactive and Visualized Exploration[C]//International Conference on Management of Data, Houston, TX, USA, 2018.

[4] 周志光,石晨,史林松,等. 地理空间数据可视分析综述[J]. 计算机辅助设计与图形学学报,2018,30(5):747-763.
ZHOU Zhiguang, SHI Chen, SHI Linsong, et al. A Survey on the Visual Analytics of Geospatial Data[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2018, 30(5): 747-763.

[5] DUCHOWSKI A T, PRICE M M, MEYER M, et al. Aggregate Gaze Visualization with Real-Time Heatmaps[C]//The Symposium on Eye Tracking Research and Applications, Santa Barbara, California, 2012.

[6] MAURUS M, HAMMER J H, BEYERER J. Realistic Heatmap Visualization for Interactive Analysis of 3D Gaze Data[C]//The Symposium on Eye Tracking Research and Applications, Safety Harbor, Florida, 2014.

[7] NETEK R, POUR T, SLEZAKOVA R. Implementation of Heat Maps in Geographical Information System - Exploratory Study on Traffic Accident Data[J]. *Open Geosciences*, 2018, 10(1): 367-384.

[8] NETEK R, BRUS J, TOMECKA O. Performance Testing on Marker Clustering and Heatmap Visualization Techniques: A Comparative Study on JavaScript Mapping Libraries[J]. *ISPRS International Journal of Geo-Information*, 2019, 8(8): 348.

[9] GALILI T, O'CALLAGHAN A, SIDI J, et al. Heatmaply: An R Package for Creating Interactive Cluster Heatmaps for Online Publishing[J]. *Bioinformatics*, 2018, 34(9): 1600-1602.

[10] RYAN M C, STUCKY M, WAKEFIELD C, et al. Interactive Clustered Heat Map Builder: An Easy Web-Based Tool for Creating Sophisticated Clustered Heat Maps[J]. *F1000Research*, 2020, 8: 1750.

[11] FENG D Y, TU L L, SUN Z W. Research on Population Spatiotemporal Aggregation Characteristics of a Small City: A Case Study on Shehong County Based on Baidu Heat Maps [J]. *Sustainability*, 2019, 11(22): 6276.

[12] REHMAN F, SAEED A, YASEEN M, et al. Genetic Evaluation and Characterization Using Cluster Heat Map to Assess NaCl Tolerance in Tomato Germplasm at the Seedling Stage[J]. *Chilean Journal of Agricultural Research*, 2019, 79(1): 56-65.

[13] SŁOMSKA-PRZECH K, PANECKI T, POKOJSKI W. Heat Maps: Perfect Maps for Quick Reading? Comparing Usability of Heat Maps with Different Levels of Generalization [J]. *ISPRS International Journal of Geo-Information*, 2021, 10(8): 562.

[14] 刘安琪. 基于百度热力图的城市公园访问特征及影响因素研究: 以南京市主城区为例[D]. 南京: 南京林业大学, 2021.
LIU Anqi. Features and Influencing Factors of Urban Park Visitation in Nanjing[D]. Nanjing: Nanjing Forestry University, 2021.

[15] 周雨霏, 杨家文, 周江评, 等. 基于热力图数据的轨道交通站点服务区活力测度研究: 以深圳市地铁为例[J]. 北京大学学报(自然科学版), 2020, 56(5): 875-883.
ZHOU Yufei, YANG Jiawen, ZHOU Jiangping, et al. Evaluating Vitality of Metro Station Service Area with Heat Map: A Case Study on Shenzhen Subway [J]. *Acta Scientiarum Naturalium Universitatis Pekinensis*, 2020, 56(5): 875-883.

[16] BAO W X, GONG A D, ZHANG T, et al. Mapping Population Distribution with High Spatiotemporal Resolution in Beijing Using Baidu Heat Map Data [J]. *Remote Sensing*, 2023, 15(2): 458.

[17] 常春慧. 基于百度热力图的城市洪灾应急物资需求估算与调度[D]. 武汉: 武汉理工大学, 2020.
CHANG Chunhui. Demand Estimation and Dispatch of Urban Flood Emergency Materials Based on Baidu Heatmap [D]. Wuhan: Wuhan University of Technology, 2020.

[18] 高峰. 车流量热力图绘制系统[D]. 杭州: 浙江大学, 2017.
GAO Feng. A Traffic Flow Heat Map Drawing System[D]. Hangzhou: Zhejiang University, 2017.

[19] 孔立斌. 基于热力图与加权特征的车辆目标检测算法的研究[D]. 广州: 广东工业大学, 2021.
KONG Libin. Research on Vehicle Object Detection Algorithm Based on Heat Map and Weighted [D]. Guangzhou: Guangdong University of Technology, 2021.

[20] 龙雪琴, 周萌, 赵欢, 等. 基于网络核密度的网约车上下客热点识别[J]. 交通运输系统工程与信息, 2021, 21(3): 86-93.
LONG Xueqin, ZHOU Meng, ZHAO Huan, et al. Passengers' Hot Spots Identification of Online Car-Hailing Based on Network Kernel Density[J]. *Journal of Transportation Systems Engineering and Information Technology*, 2021, 21(3): 86-93.

[21] 康磊, 刘海砚, 程维应, 等. 顾及属性特征的城市设施热点识别方法[J]. 测绘通报, 2022(1): 8-14.
KANG Lei, LIU Haiyan, CHENG Weiying, et al. A Method of Urban Facility Hot Spot Recognition Considering Attribute Characteristics[J]. *Bulletin of Surveying and Mapping*, 2022(1): 8-14.

[22] 吴瑞龙, 王明省, 何华贵, 等. 顾及约束的交互式热力图服务研究与应用[J]. 地理空间信息, 2023, 21(2): 85-88.
WU Ruilong, WANG Mingsheng, HE Huagui, et al. Research and Application of Interactive Heatmap Service Considering Constraints[J]. *Geospatial Information*, 2023, 21(2): 85-88.

[23] 杨振凯, 李响, 杨飞. 一种面向百万级数据的热力图生成算法[J]. 测绘科学, 2018, 43(8): 85-89.
YANG Zhenkai, LI Xiang, YANG Fei. A Thermal Map Generating Algorithm for Mega Level Data [J]. *Science of Surveying and Mapping*, 2018, 43(8): 85-89.

[24] YUAN K, CHENG X Q, GUI Z P, et al. A Quad-Tree-Based Fast and Adaptive Kernel Density Estimation Algorithm for Heat-Map Generation[J]. *International Journal of Geographical Information Science*, 2019, 33(12): 2455-2476.

[25] 饶加旺, 马荣华. 改进核密度估计的空间点密度算法[J]. 计算机工程与应用, 2021, 57(11): 260-265.
RAO Jiawang, MA Ronghua. Improved Kernel Density Estimator Based Spatial Point Density Algorithm[J]. *Computer Engineering and Applications*, 2021, 57(11): 260-265.

[26] ZHENG Y, OU Y, LEX A, et al. Visualization of Big Spatial Data Using Coresets for Kernel Density Estimates [J]. *IEEE Transactions on Big Data*, 2021, 7(3): 524-534.

[27] CHAN T N, CHENG R, YIU M L. QUAD: Quadratic-Bound-Based Kernel Density Visualization

[C]//The 2020 ACM SIGMOD International Conference on Management of Data, Portland, OR, USA, 2020.

[28] CHAN T N, IP P L, U L H, et al. KDV-Explorer: A near Real-Time Kernel Density Visualization System for Spatial Analysis[J]. *Proceedings of the VLDB Endowment*, 2021, 14(12): 2655-2658.

[29] 张繁, 袁兆康, 肖凡平, 等. 基于Spark的大数据热图可视化方法[J]. 计算机辅助设计与图形学学报, 2016, 28(11): 1881-1886.
ZHANG Fan, YUAN Zhaokang, XIAO Fanping, et al. Research on Heatmap for Big Data Based on Spark[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2016, 28(11): 1881-1886.

[30] 郭宇达, 朱欣焰, 禹维, 等. 基于Spark计算框架的路网核密度估计并行算法[J]. 武汉大学学报(信息科学版), 2020, 45(2): 289-295.
GUO Yuda, ZHU Xinyan, GUO Wei, et al. Parallel Algorithm for Road Network Kernel Density Estimation Based on Spark Computing Framework[J]. *Geomatics and Information Science of Wuhan University*, 2020, 45(2): 289-295.

[31] YU J, SARWAT M. GeoSparkViz: A Cluster Computing System for Visualizing Massive-Scale Geospatial Data[J]. *The VLDB Journal*, 2021, 30(2): 237-258.

[32] JEŽEK J, JEDLIČKA K, MILDORF T, et al. Design and Evaluation of WebGL-Based Heat Map Visualization for Big Point Data[C]//The Symposium GIS, Ostrava, Czech Republic, 2016.

[33] ZHANG G M, ZHU A X, HUANG Q Y. A GPU-Accelerated Adaptive Kernel Density Estimation Approach for Efficient Point Pattern Analysis on Spatial Big Data[J]. *International Journal of Geographical Information Science*, 2017, 31(10): 2068-2097.

[34] 朱峥嵘, 黄亚锋, 赵立营. 一种利用GPU加速的轨迹线热力图生成显示方法[J]. 武汉大学学报(信息科学版), 2022, 47(7): 1035-1042.
ZHU Zhengrong, HUANG Yafeng, ZHAO Liying. A Method of Generating and Displaying Trajectory Line Heat Map with GPU Acceleration[J]. *Geomatics and Information Science of Wuhan University*, 2022, 47(7): 1035-1042.

[35] ZHANG G M. Detecting and Visualizing Observation Hot-Spots in Massive Volunteer-Contributed Geographic Data Across Spatial Scales Using GPU-Accelerated Kernel Density Estimation[J]. *ISPRS International Journal of Geo-Information*, 2022, 11(1): 55.

[36] ZHANG G M, XU J. Multi-GPU-Parallel and Tile-Based Kernel Density Estimation for Large-Scale Spatial Point Pattern Analysis[J]. *ISPRS International Journal of Geo-Information*, 2023, 12(2): 31.

[37] 刘泽邦, 陈荦, 杨岸然, 等. 支持大规模地理矢量数据实时可视化的高效索引技术[J]. 武汉大学学报(信息科学版), 2023, 48(9): 1512-1521.
LIU Zebang, CHEN Luo, YANG Anran, et al. Efficient Indexing Technology for Real-Time Visualization of Large-Scale Geographic Vector Data[J]. *Geomatics and Information Science of Wuhan University*, 2023, 48(9): 1512-1521.

[38] 周炤, 李少梅, 杨佳. Web墨卡托投影及其性质分析[J]. 测绘科学技术学报, 2021, 38(1): 71-74.
ZHOU Zhao, LI Shaomei, YANG Jia. Web Mercator Projection and Analysis of Its Properties[J]. *Journal of Geomatics Science and Technology*, 2021, 38(1): 71-74.