

DOI:10.19651/j.cnki.emt.2108293

基于二分 K-means 聚类的曲率分级点云数据 精简优化算法研究*

李佩佩 崔凤英

(青岛科技大学自动化与电子工程学院 青岛 266061)

摘要: 针对单一精简算法无法精确保留模型特征信息、易造成点云表面孔洞等问题,提出了一种基于二分 K-means 聚类的曲率分级优化精简算法。首先采用最小二乘法对邻域进行曲面拟合,计算曲率值,依据曲率值划分显著特征区与非显著特征区,其次采用二分 K-means 聚类划分非显著特征区,依据子簇的曲率阈值筛选保留具有特征重要性的亚特征点,最后合并更新显著特征区的数据集和亚特征点,得到简化结果。通过仿真实验,从算法运行速度和信息熵两方面与空间包围盒法、曲率精简算法进行对比分析,结果表明,该算法在精简质量上优于其他两种算法,在点云数据重建方面具有一定的应用价值。

关键词: 点云数据;二分 K-means 聚类;曲率简化

中图分类号: TP274 **文献标识码:** A **国家标准学科分类代码:** 510.4030

The optimization algorithm for curvature graded point cloud data based on dichotomous K-means clustering

Li Peipei Cui Fengying

(College of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao 266061, China)

Abstract: For the problems of losing the model feature information and causing easily the point cloud surface holes in single simplification algorithms, a streamlined algorithm for curvature classification optimization based on dichotomous K-means clustering is proposed. First, the least squares method was used to fit the neighborhood surface, calculate the curvature value, and divide the significant and non-significant feature regions based on the curvature value. Second, dichotomous K-means clustering was used to divide non-significant feature regions, select the subfeature points with feature importance retained according to the curvature threshold of subclusters, and finally the datasets and subfeature points were merged to obtain simplified results. The simplification algorithm is compared with the space surrounding box algorithm and the curvature reduction algorithm by the simulation experiments in terms of speed and information entropy. The results show that the proposed algorithm outperforms the other two algorithms in streamlining quality and has a certain application value in point cloud data reconstruction.

Keywords: point cloud data; dichotomous K-means clustering; curvature simplification

0 引言

由三维点云数据形成的残肢形状会影响假肢接受腔设计效果,准确高效地捕获残肢特征在制作假肢中起关键作用。目前,激光扫描被广泛用于快速收集物体表面的三维坐标^[1]。然而,高密度激光扫描点云极大地增加了数据量,在数据存储、处理、可视化和传输方面带来了挑战^[2]。处理所有的点可能不会在点云的特征提取、三维重建等方面提

供显著的改进,因此有必要简化残肢点云,以便在保留重要特性或结构的情况下选择更易于管理的点云数据集。

近年来,国内外学者在点云数据精简方面也开展了大量的研究。目前的点云精简方式大致包括两类:基于网格的精简方法和直接点的精简。基于网格的精简方法利用点云建立大量不规则网格,然后再根据既定的准则逐步消除多余网格,以实现精简的目的。Hu等^[3]通过网格压缩技术来简化点云,该策略对网格的简化能够更有效地保留点云

收稿日期:2021-11-08

* 基金项目:山东省研究生教育质量提升计划项目(SDYJD18029)资助

的基本结构,但是在计算机上建立网格的费用仍然是巨大的,尤其是在点云数据量很大时,计算成本将会非常高昂。而直接简化技术则是通过点的分布特性对点云加以精简,从而可以降低运算费用。王建强等^[7]提出了一种自适应简化的点云多参数 K-means 聚类,该方法可以较好地体现模型的细节特征,但是无法保证均匀区域的数据量,适用的点云模型形状单一。常俊飞等^[6]采用基于边界保留的 K-means 聚类来减少点云,但该方法对初始聚类中心的依赖性较大,容易陷入局部最小值。传统单一简化算法主要包括空间包围盒简化法^[6,7]和曲率简化方法^[8,9],文献[6-7]通过每个小包围盒的重心来近似表达整个体素中的所有数据点,虽然可以直观有效地简化点云,但该精简策略不区分特征点,易丢失细节特征。文献[8-9]通过曲率对点云数据量进行压缩。该精简算法保留了大量能反映模型几何信息的特征点,但时间开销大,此外该方法在曲率小的平坦区域会因去除过多的数据点而造成局部空洞现象。

为了能够克服单一简化方法中点云特征保留差和表面空洞的现象,最大程度地还原残肢表面形状的真实性,提高残肢患者佩戴假肢的舒适度,本文提出一种基于二分 K-means 聚类的曲率分级优化精简方法。该优化算法在三维点云精简中常用的曲率简化方法基础上进行改进,结合二分 K-means 聚类方法^[10]对非显著特征区残肢点云进行分割,并使用曲率简化算法对分割的点云区域进行精简处理,最后更新特征显著区和非显著特征区的残肢特征点,完成简化。

1 二分 K-means 聚类的曲率分级精简优化算法

针对患者残肢部位点云分布不均匀且存在冗余信息,传统单一简化算法存在特征细节缺失、空洞等问题,本文提出一种基于二分 K-means 聚类的曲率分级优化精简算法。该优化算法是在曲率简化算法的基础上引入二分 K-means 聚类方法,并多次使用曲率简化算法,具体的工作步骤如下所述。

1) 利用 k 邻域索引查找与残肢点云源数据相邻的 k 个点,建立残肢点云邻域关系,并采用最小二乘法拟合残肢轮廓表面。

2) 根据拟合出的曲面估算每个残肢点的曲率值,并计算所有点的平均曲率。

3) 比较每个点云曲率值与平均曲率值的大小,保留大于平均曲率的残肢点云,其所在局部区域定义为显著特征区,剩余残肢点云所在区域被命名为非显著特征区。

4) 采用二分 K-means 聚类将非显著特征区划分成 K 个子簇,每簇的残肢点云分别采用曲率简化算法即步骤 1)~3),进行精简,直至 K 个子簇全部处理完毕,并保留满足各子簇点云平均曲率的点云数据。

5) 更新满足特征条件的所有点云,完成残肢点云简化。

1.1 最小二乘法曲面拟合求曲率

任取残肢点云中的一个数据点 N_j , 利用 KD-Tree^[11] 进行 k 邻域搜索(本文 k 取 20), 通过查找与其相邻的 20 个数据点构建邻域。使用最小二乘法^[12-13] 对构建的空间邻域进行二次曲面拟合, 将曲面方程设为 $Z(x, y) = ax^2 + bxy + cy^2$, 计算系数, 然后根据空间曲面曲线的性质计算目标点 N_j 的高斯曲率、平均曲率和主曲率。

根据最小二乘原理应该使式(1)取最小值。

$$Q^2 = \sum_i (ax_i^2 + bx_i y_i + cy_i^2 - z_i)^2, i \in (0, 20) \quad (1)$$

式中: x_i, y_i, z_i 为数据点 N_j 邻域内的点。将式(1)分别对系数 a, b, c 求偏导, 使其为 0, 得出式(2):

$$\begin{cases} \frac{\partial^2 Q}{\partial a} = \sum_i 2x_i^2 (ax_i^2 + bx_i y_i + cy_i^2 - z_i) = 0 \\ \frac{\partial^2 Q}{\partial b} = \sum_i 2x_i y_i (ax_i^2 + bx_i y_i + cy_i^2 - z_i) = 0 \\ \frac{\partial^2 Q}{\partial c} = \sum_i 2y_i^2 (ax_i^2 + bx_i y_i + cy_i^2 - z_i) = 0 \\ i \in (0, 20) \end{cases} \quad (2)$$

根据式(2)求解二次曲面系数 a, b, c 值。

将曲线方程写成曲面参数方程形式:

$$g(x, y) = \begin{cases} X(x, y) = x \\ Y(x, y) = y \\ Z(x, y) = ax^2 + bxy + cy^2 \end{cases} \quad (3)$$

若曲面上存在一条曲线 g , 则 g 的表达式为:

$$g = g(x(t), y(t)) \quad (4)$$

若以 l 表示曲线 g 的弧长, 那么根据复合函数求导公式可求得弧长微分公式:

$$(dl)^2 = (dg)^2 = (g_x dx + g_y dy)^2 = g_x^2 (dx)^2 + 2g_x g_y dx dy + g_y^2 (dy)^2 \quad (5)$$

式中: g_x, g_y 表示分别对 x, y 求偏导。

由曲面的第一基本公式可得:

$$(dl)^2 = I = E(dx)^2 + 2F dx dy + G(dy)^2 \quad (6)$$

式中: $E = g_x \cdot g_x, F = g_x \cdot g_y, G = g_y \cdot g_y$ 。

假如 P 是曲线 g 上一点, 用 t 和 n 分别表示点 P 的单位切向量和单位法向量, 则曲率向量可分解为:

$$k = \frac{dt}{dl} = k_n n + k_{g'} (n \times t) \quad (7)$$

曲线的单位法向量 n 表示为:

$$n = \frac{g_x \times g_y}{|g_x \times g_y|} \quad (8)$$

可得曲面的第二基本公式:

$$II = -dg \cdot dn = L(dx)^2 + 2M dx dy + N(dy)^2 \quad (9)$$

式中: $L = g_{xx} \cdot n, M = g_{xy} \cdot n, N = g_{yy} \cdot n; g_{xx}, g_{yy}, g_{zz}$ 指曲面的微分。

法曲率可表示为:

$$k = \frac{II}{I} = \frac{L + 2M\lambda + N\lambda^2}{E + 2F\lambda + G\lambda^2} \quad (10)$$

$$\lambda = \frac{dy}{dx}$$

处理变量 λ 可得到 k_n 的两个主曲率 k_1, k_2 , 由曲率特性可推导出:

高斯曲率:

$$K = k_1 k_2 = \frac{LN - M^2}{EG - F^2} \quad (11)$$

平均曲率:

$$H = \frac{1}{2}(k_1 + k_2) = \frac{EN - 2FM + GL}{2(EG - F^2)} \quad (12)$$

主曲率:

$$k_1, k_2 = H \pm \sqrt{H^2 - K} \quad (13)$$

比较邻域的平均曲率值和平均曲率值, 如果邻域平均曲率高于平均曲率, 则说明该邻域点云分布较为陡峭, 特征信息重要等级高, 为显著特征区; 若邻域平均曲率低于平均曲率, 则说明该区域点分布较为平坦, 特征信息重要性较低, 为非显著特征区^[14]。保留显著特征区的点云能够表征残肢表面特征。

1.2 二分 K-means 聚类划分点云特征

二分 K-means 聚类分析是一种具有弱化聚类中心功能的方法, 是对传统 K-means 聚类算法^[15] 的优化。其工作原理如下。

1) 将非显著特征区的残肢点云作为整体放到一个队列中。

2) 从队列中选择一个簇, 采用 K-means 算法进行划分 (相当于将所有样本进行 $K=2$ 的划分), 划分为两个子簇, 对所有子簇计算误差平方和 SSE。

$$SSE = \sum_{i=1}^n \omega_i (p_i - \hat{p}_i)^2 \quad (14)$$

其中, ω_i 为每簇中聚类中心的权重, p_i 为子簇的点云数据, \hat{p}_i 为聚类中心点。该值越小, 代表数据点距离它们的质心越近, 聚类效果越好。

3) 选择 SSE 最大的聚簇采用 K-means 算法进行再次二分划分操作, 直到聚簇数目等于设定的数目 K 为止。经交叉验证发现, $K=6$ 时聚类效果最好, 因此, 设定簇数 $K=6$ 。

二分 K-means 算法不受初值问题的影响, 解决了 K-means 算法对初始聚类质心比较敏感的问题, 降低了聚类过程中多次迭代的概率, 提高聚类效率。

2 实例结果与分析

2.1 数据精简结果

本文通过车载环式激光扫描仪采集获取残肢端点云数据, 为了说明提出简化算法的普适性, 选取了残肢点云包含

显著特征点较少的 1 号患者和残肢点云包含特征点较多 (如膝盖、腘窝) 的 2 号患者为研究对象, 分别采用空间包围盒简化法^[6]、曲率精简算法^[8] 以及本文的精简算法对两位患者残肢进行简化处理, 其中 1 号患者残肢端点云设定简化率约为 80%, 2 号患者残肢端点云设定简化率约为 75%。

1 号患者残肢原始点云散点图如图 1(a) 所示, 各算法简化后的点云分布如图 1(b)~(d) 所示。

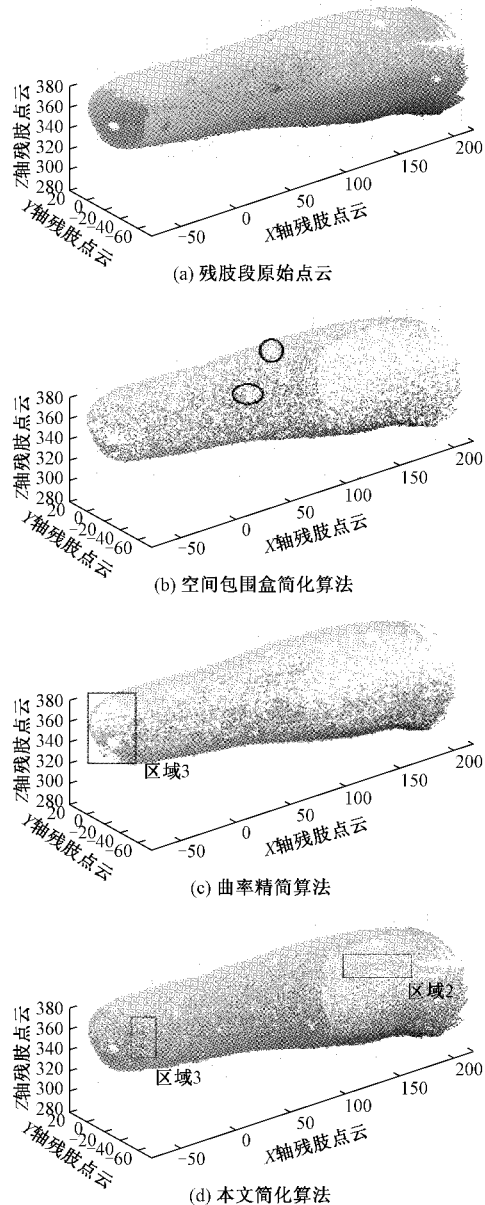


图 1 1 号患者残肢端点云数据分布图

如图 1 所示, 空间包围盒简化算法、基于曲率的精简算法以及本文提出的简化算法都能实现点云数据的精简, 文献[6]的空间包围盒法只是进行单一的网格划分, 再选择网格中的代表点, 实现点云简化的目的, 所保留的特征信息较少; 文献[8]的曲率精简算法主要依据曲率保留残肢特征信息。如图 1(b) 和 (d) 所示, 空间包围盒简化算法改善孔洞

的效果不如综合简化算法如图 1(b) 圆圈处, 点云分布的有序性低, 这是因为该算法只是简单地根据设定的网格长度对整个数据集网格划分, 选取网格代表点没有考虑到点云数据的特征重要性, 所以造成特征区域特征信息的丢失以及相对平坦邻域的孔洞现象。如图 1(c) 和 (d) 所示, 曲率简化算法保留的特征信息少于本文算法, 精简效果较差, 如图 1(c) 区域 3 处, 残肢上下两端的边界不清晰。在点云分布较为平坦的邻域中, 数据点的曲率值相对较小, 如果只设定一个曲率阈值很容易导致邻域中大部分甚至全部的点云都被消除, 造成孔洞现象。本文提出的综合简化算法中二分 K-means 聚类方法弥补了以上缺陷, 利用欧氏距离度量数据间的相似性, 将相同特征重要性等级的数据点分布在同一区域内, 尽可能的保证特征信息保留, 降低非显著特征区出现孔洞的概率。

2 号患者残肢原始点云散点图如图 2(a) 所示, 各算法简化后的点云分布如图 2(b)~(d) 所示。

如图 2(b)~(d) 所示, 3 种方法都能够保留残肢的显著特征信息, 但基于曲率简化算法导致残肢局部出现空洞现象, 保留亚特征区域特征信息的能力较差, 点云的有序性较低; 空间包围盒法和本文综合算法充分保留了残肢膝盖等部位的特征信息, 其简化效果优于曲率精简算法, 如图 2(e)、(f) 所示, 本文综合简化算法的残肢点云分布均匀度明显高于空间包围盒, 有效解决了孔洞现象。

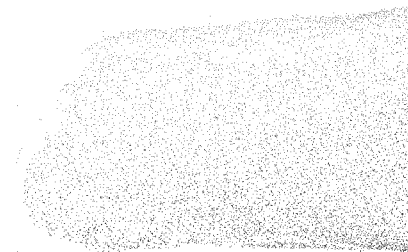
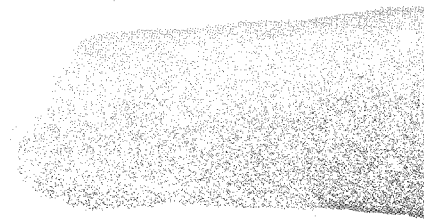
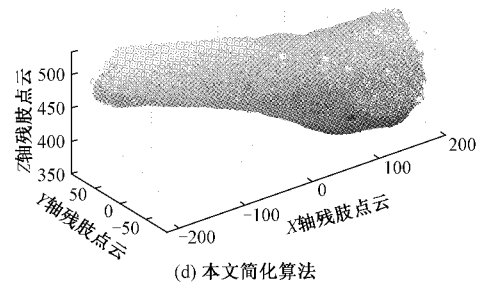
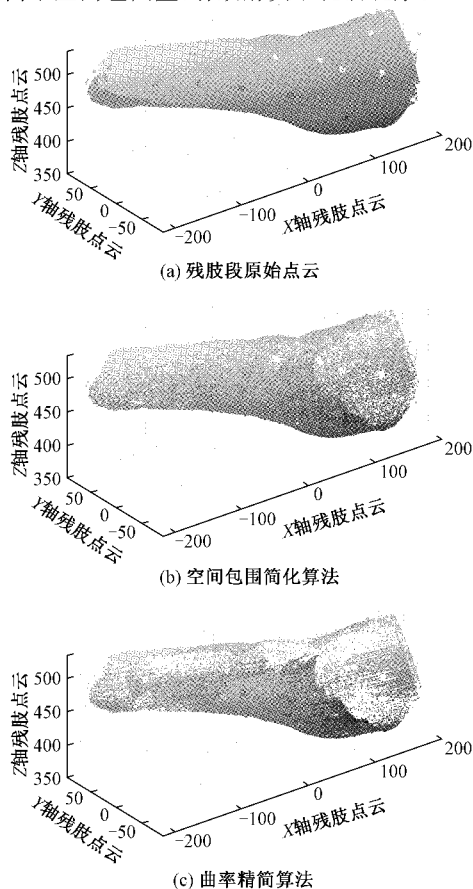


图 2 2 号患者残肢端点云数据分布图

2.2 精简效果分析

1) 点云简化速度

一般地, 数据集的简化效果是根据精度、简度、速度 3 种指标进行分析评价的^[16]。精度指精简后的点云数据经曲面重建得到的模型与原始模型的差异程度。简度是指简化过程中点云减少的数量与原始点云数据集的比值。速度是通过简化点云所消耗的时间来度量。在本文中, 设定 1 号患者残肢点云简化度约为 80%, 2 号患者残肢点云简化率为 75%, 因此, 我们在简化时间上定量比较三种算法的运行效率, 结果如表 1 和 2 所示。

表 1 1 号患者 3 种简化算法的精简结果

简化算法	空间包围盒 算法	曲率精简 方法	本文简化 算法
原始点云	126 864	126 864	126 864
精简后点云	25 703	24 935	23 937
简化度/%	79.74	80.35	81.13
简化时间/s	7.598	11.216	8.479

表 2 2 号患者 3 种简化算法的精简结果

简化算法	空间包围盒 算法	曲率精简 方法	本文简化 算法
原始点云	233 938	233 938	233 938
精简后点云	57 869	58 236	58 402
简化度/%	74.88	75.11	75.04
简化时间/s	10.402	17.039	12.263

从表 1 和 2 可知,在简化率相近的条件下,本文提出的综合算法简化速度介于其他两种算法。空间包围盒算法在简化过程中直接对外接包围盒进行操作,不需要建立拓扑结构,因此算法速度较快;基于曲率的算法在简化过程中需要通过构建 K 邻域拓扑结构和拟合曲面来获得点云数据的曲率值,该计算过程较为复杂,花费的时间成本较高,算法运行速度慢;与曲率精简方法相比,本文提出的算法应用在两位患者的残肢上,运行时间分别缩短了 24% 和 28%,在简化过程中利用二分 K-means 聚类对数据集进行划分,相同特性的数据集聚在一起,加快了建立邻域以及计算曲率的速度,时间开销较小。

2) 信息熵理论

文献[17]通过有关信息熵的知识表征点云的特征信息,根据点云曲率的分布计算点的熵值,熵值越大,表示数据点所在区域无序程度越高,说明该点对于该区域所包含的信息量越大,对物体的细节特征表达越精确^[17]。所以,可以通过计算残肢点云数据的熵值,定量评估 3 种简化方式的精简质量。估计某一点熵值的标准公式为:

$$E_i = -p_i \log_2 p_i - \sum_{j=1}^k p_j \log_2 p_j \quad (15)$$

$$p_i = \frac{h_i}{h_i + \sum_{j=1}^k h_j} \quad (16)$$

$$p_j = \frac{h_j}{h_i + \sum_{j=1}^k h_j} \quad (17)$$

式中: h_i 代表点 P_i 的曲率; h_j 代表点 P_i 近邻点 P_j 的平均曲率; p_i 与 p_j 分别代表点 P_i 与 P_j 的曲率概率分布。所有点云数据的熵值为各个点云数据熵值的和:

$$E = \sum_{i=1}^n E_i \quad (18)$$

分别计算 3 种算法简化后数据集的熵值,定量评价残肢点云简化效果,如表 3 和 4 所示。

表 3 1 号患者残肢点云精简熵值

点云简化算法	熵值
空间包围盒法	4.8215×10^4
曲率精简方法	4.6560×10^4
本文简化算法	5.7677×10^4

表 4 2 号患者残肢点云精简熵值

点云简化算法	熵值
空间包围盒法	1.1229×10^5
曲率精简方法	9.3160×10^4
本文简化算法	2.5697×10^5

从表 3 和 4 可以看出,相较于其他两种点云简化结果的熵,本文方法精简结果的熵值最大,即残肢包含特征信息量最多,说明基于二分 K-means 聚类的曲率分级精简优化算法能够充分保留残肢点云细节特征。

综上,本文所提出的综合简化优化算法能较好地实现残肢点云的简化,特别是对于显著特征信息有较为理想的简化效果;另外,对于非显著特征区域的点云数据,能有效去除子簇中的冗余点,避免产生空洞现象,是一种具有高可行性的点云简化方法。

3 结 论

本文提出的基于二分 K-means 聚类的曲率分级精简算法,采用基于曲率的简化策略有效保留了残肢特征点,利用二分 K-means 聚类的方法实现了对残肢非显著特征区的划分,在保证残肢表面完整无空洞的同时,尽可能地保留了残肢点云的亚特征信息。实验部分将本文方法与空间包围盒法、曲率简化算法进行对比,结果表明,该方法在两位患者的残肢点云数据中均有较好的精简效果,解决了特征信息丢失和空洞的问题,为后续重构高质量的假肢模型提供有利支撑。

参考文献

- [1] 梁严. 三维激光扫描点云数据精简及建模应用研究[D]. 徐州:中国矿业大学,2020.
- [2] 喜文飞. 激光点云数据压缩的精简研究[D]. 昆明:昆明理工大学,2011.
- [3] HU K, YAN D M, BOMME D, et al. Error-bounded and feature preserving surface remeshing with minimal angle improvement [J]. IEEE Transactions on Visualization and Computer Graphics, 2017, 23(12): 2560-2573.
- [4] 王建强,樊彦国,李国胜,等. 基于多参数 k-means 聚类的自适应点云精简[J]. 激光与光电子学进展, 2021, 58(6): 175-183.
- [5] 常俊飞,赵利民,王瀚斌. 基于边界保留的 k-means 聚类点云精简算法研究[J]. 测绘工程, 2018, 27(7): 60-65.
- [6] 李仁忠,杨曼,刘阳阳,等. 一种散乱点云的均匀精简算法[J]. 光学学报, 2017, 37(7): 97-105.
- [7] 娄凸. 三维激光扫描点云数据精简算法研究[D]. 昆明:昆明理工大学,2017.
- [8] 李雨欣. 基于激光雷达的点云数据处理算法研究[D].

- 长春:长春理工大学,2020.
- [9] 李国远,梁周雁,石信肖,等.基于曲率特征约束的激光点云精简方法研究[J].计算机与数字工程,2020,48(8):2034-2037,2063.
- [10] 刘广聪,黄婷婷,陈海南.改进的二分 K 均值聚类算法[J].计算机应用与软件,2015,32(2):261-263,277.
- [11] 王丽.基于 kd-tree 算法和法向量估计的点云数据精简方法[J].宿州学院学报,2019,34(12):65-68.
- [12] 段志鑫.三维激光扫描数据精简、表面重构方法及应用研究[D].徐州:中国矿业大学,2019.
- [13] 麻卫峰,王金亮,张建鹏,等.一种改进法向量估算的点云特征提取[J].测绘科学,2021,46(11):84-90,146.
- [14] 曹垚.三维激光点云数据精简算法及三角网格模型优化研究[D].郑州:郑州大学,2019.
- [15] 杨俊闯,赵超.K-Means 聚类算法研究综述[J].计算机工程与应用,2019,55(23):7-14,63.
- [16] 翟金磊.空间散乱点云去噪精简重建算法研究[D].西安:西安科技大学,2015.
- [17] 朱广堂,叶珉吕.基于曲率特征的点云去噪及定量评价方法研究[J].测绘通报,2019(6):105-108.

作者简介

崔凤英,硕士学位,副教授,主要研究方向为检测技术与自动化装置、智能控制和虚拟仪器设计。

E-mail:fcngying_cui@163.com

李佩佩(通信作者),研究生,主要研究方向为智能控制。

E-mail:2623321763@qq.com