



# 融合 $k$ -means聚类 and Hausdorff距离的散乱点云精简算法

李健<sup>1</sup> 曹垚<sup>2</sup> 王宗敏<sup>2,3</sup> 王广印<sup>2</sup>

1 郑州大学地球科学与技术学院,河南 郑州,450001

2 郑州大学水利与环境学院,河南 郑州,450001

3 中原工学院,河南 郑州,450007

**摘要:**针对点云精简算法在处理点云数据时特征保留不完整和对小曲率点云精简造成数据空洞的问题,提出了一种融合 $k$ -means聚类 and Hausdorff距离的点云精简算法。该算法在八叉树算法的基础上构建点云数据的拓扑关系,首先计算所有点云数据点的主曲率,然后计算点云数据点主曲率的 Hausdorff距离,根据精简目标要求设定 Hausdorff距离阈值,实现点云特征提取,最后对非特征区域进行 $k$ -means聚类提取特征点,并将两次提取的特征点融合得到精简结果。实验结果表明,该算法能较完整地保留模型的特征信息,并能避免形成空洞现象。

**关键词:**点云精简;Hausdorff距离; $k$ -means聚类;曲面拟合

**中图分类号:**P237

**文献标志码:**A

为了提高点云数据处理和应用的效率,需要对海量点云数据进行精简<sup>[1-4]</sup>。近年来,国内外学者对点云精简进行了大量研究,并取得了大量的研究成果。经典的点云算法有包围盒法<sup>[5]</sup>、曲率采样<sup>[6]</sup>、保留边界法<sup>[7]</sup>、聚类法等<sup>[8]</sup>。Weir等<sup>[9]</sup>利用体包围盒法对点云进行精简,该算法具有体包围盒高效的特性,但由于其精简过程是以一个或者一部分点来代替区域点,所以容易造成点云显著特征丢失。Wang等<sup>[10]</sup>利用特征参数法提取特征,并利用均匀球面采样法对非特征点进行精简,该算法具有特征保留的特性,但易造成点云数据空洞现象。Lee等<sup>[11]</sup>利用非均匀网格精简点云数据,该方法较好地保留了点云模型的特征点,但在平坦区域保留点分布不均匀。Yuan等<sup>[12]</sup>为了改进传统算法经常容易丢失几何信息的问题,利用球体树构造和 $k$ -means法对点云进行精简,该算法不仅对局部细节有很好的效果,而且对模型完整性具有良好的保留效果,但其构型和聚类过程比较复杂,效率较低。杨秋翔等<sup>[13]</sup>利用 Hausdorff距离算法提取特征点,实现了点云精

简,优点是几何特征保留较好,精简效率高,但容易造成数据空洞。陈璋雯等<sup>[14]</sup>提出了基于模糊熵迭代的三维点云精简算法,该算法能很好地保留细节特征,但没有考虑非特征区域精简问题。目前,点云精简算法优点突出,精简效果较好,但也存在点云数据特征保留不完整和对小曲率点云精简易造成数据空洞等不足。

针对上述算法精简散乱点云时出现几何特征保留较少和过精简的问题,本文提出了一种融合 $k$ -means聚类<sup>[15]</sup>和 Hausdorff距离<sup>[16-17]</sup>的点云精简算法。为了提高点数据的索引效率,该算法首先在八叉树算法的基础上构建点云数据的拓扑关系;然后利用最小二乘法进行局部曲面拟合,分别计算目标点与其邻域点集内各点连线和目标点切平面的夹角,以夹角的平均值估算目标点曲率,通过局部二次曲面拟合方程分别计算点云数据点的主曲率;接着根据主曲率计算 Hausdorff距离,通过估计曲率和主曲率的 Hausdorff距离来确定每个八叉树小长方体块的特征点和特征区域,根据精简目标要求实现点云特征提取;最后

收稿日期:2019-06-14

项目资助:国家自然科学基金(51678536);测绘遥感信息工程国家重点实验室开放基金(15E01);河南省教育厅高等学校重点科研项目(14A420002)。

第一作者:李健,博士,副教授,主要从事 LiDAR 点云数据处理与三维重建研究。lijian5277@163.com

通讯作者:曹垚,硕士。772440651@qq.com

对非特征区域进行  $k$ -means 聚类提取特征点。本文算法利用 Hausdorff 距离具有相似性度量的特性提取特征点,并利用  $k$ -means 聚类算法对非特征区域进行再精简,能够保证点云数据整体的完整性,避免过精简造成数据空洞。

## 1 点云精简算法

### 1.1 局部点云曲面拟合

本文在最小二乘法曲面拟合的基础上对点云数据点的邻域点集进行拟合。针对实验点云散乱无序的特性,在八叉树算法的基础上构建点云数据的拓扑关系,以提高点云数据索引的效率。将目标点  $p_i$  到相邻点的欧氏距离存入数组中,并对数组数据进行降序排序,取数组中前  $k$  个点  $\{(x_i, y_i, z_i)\}_{i=1}^k$ ,邻域点集记为  $k(q_i)$ 。点云数据目标点的  $k$  邻域点集只与目标点  $p_i$  有关,与其他点及其邻域点不相关。许多学者对  $k$  值的选取做了大量实验<sup>[6]</sup>,证明  $k$  取 15~30 效果较好,本文  $k$  取 17。运用最小二乘二次多项式对经过坐标转换的目标点及其邻域点集  $k(q_i)$  拟合得到局部曲面,二次多项式曲面拟合方程  $f(x_i, y_i)$ <sup>[18]</sup>为:

$$f(x_i, y_i) = a_1 + a_2 x_i + a_3 x_i^2 + a_4 x_i y_i + a_5 y_i + a_6 y_i^2 \quad (1)$$

曲面的一般方程可以表示为:

$$z_i = f(x_i, y_i) \quad (2)$$

对目标点及其邻域点集的坐标  $(x, y, z)$  先进行坐标转换,即转换成局部坐标系,再将坐标代入到方程(2)中求解二次多项式曲面拟合系数。

由于方程个数超过了拟合曲面方程系数,因此将拟合曲面系数的求解过程转换成式(3),即根据拟合曲面在目标及其邻域点集处的值与目标及其邻域点集的真实值之差的平方和达到最小来求解拟合曲面系数。

$$D = \sum_{i=1}^n [f(x_i, y_i) - z_i]^2 \quad (3)$$

由最小二乘算法原理可知,求解方程(3)可转换为:

$$\frac{\partial D}{\partial a_i} = 0, i = 1, 2, \dots, n \quad (4)$$

### 1.2 曲率的估算

首先根据局部点云曲面拟合的曲面方程计算目标点  $p_i$  的切平面法向量  $\mathbf{n}$ ,然后利用目标点坐标和切平面法向量计算该点的切平面方程,最后通过邻域点与目标点切平面的偏离角度平均

值估算曲率<sup>[14]</sup>。如图 1 所示,估算目标点  $p_i$  曲率步骤如下:

1) 将目标点  $p_i$  分别和邻域内的其他  $k$  个点  $k(q_j)$  相连,得到  $k$  个向量  $\overrightarrow{p_i q_1}, \overrightarrow{p_i q_2}, \dots, \overrightarrow{p_i q_k}$ 。

2) 向量  $\overrightarrow{p_i q_j}$  与切平面的夹角为  $\theta_{ij}$ ,它的正弦值为向量  $\overrightarrow{p_i q_j}$  与切平面法向量  $\mathbf{n}$  的夹角  $\theta_n$  余弦的绝对值:

$$\sin \theta_{ij} = |\cos \theta_n| = \left| \frac{\overrightarrow{p_i q_j} \cdot \mathbf{n}}{|\overrightarrow{p_i q_j}| \times |\mathbf{n}|} \right| \quad (5)$$

由式(5)分别计算向量  $\overrightarrow{p_i q_1}, \overrightarrow{p_i q_2}, \dots, \overrightarrow{p_i q_k}$  与切平面夹角的正弦值  $\sin \theta_{i1}, \sin \theta_{i2}, \dots, \sin \theta_{ik}$ ,得到目标点的估计曲率:

$$Q = (\sin \theta_{i1} + \sin \theta_{i2} + \dots + \sin \theta_{ik}) / k \quad (6)$$

式中,  $Q$  为目标点的估计曲率;  $k$  为目标点的邻域点数。

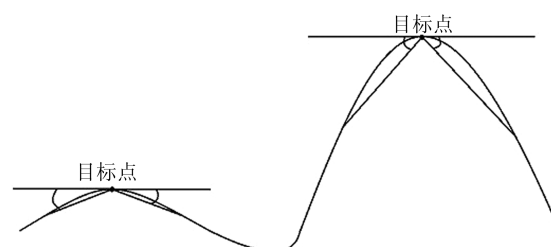


图 1 点云数据局部曲率

Fig.1 Local Curvature of Point Cloud Data

曲率代表几何物体表面高低起伏不平坦的一种衡量,曲率较大的区域一般弯曲程度较大,且包含较多的特征信息,曲率较小的区域较为平坦或者平缓。基于曲率的点云精简算法以曲率作为精简原则,保留曲率较大的点,在曲率较小的地方精简大量点易造成精简结果出现数据空洞现象;曲率较大的点与周围点共同构成物体特征信息,并且曲率存在渐变过程,而基于曲率的点云精简算法随着精简率的提高会去掉部分大曲率点周围的点,导致物体特征区域出现变形,降低了数据精度。因此,本文引入 Hausdorff 距离计算目标点与  $k$  邻域点的相似性来确定特征点。

### 1.3 基于 Hausdorff 距离的 $k$ -means 点云聚类

假设有两组点集  $A, B$ ,则两组点集之间的 Hausdorff 距离定义为<sup>[19]</sup>:

$$h(A, B) = \max \{d(A, B), d(B, A)\} \quad (7)$$

式中,  $d(A, B)$  和  $d(B, A)$  分别表示集合  $A$  到集合  $B$  和集合  $B$  到集合  $A$  的单向 Hausdorff 距离。本文通过 §1.1 中曲面拟合方程计算点云数据目标点

$p_i$  与该点邻域点集某一点  $q_j$  的主曲率  $k_1, k_2$  和  $k_1', k_2'$ , 分别记为集合  $\{k_1, k_2\}$  和  $\{k_1', k_2'\}$ 。则目标点  $p_i$  与其邻域内点曲率的 Hausdorff 距离  $h_{i1}, h_{i2} \dots h_{ik}$  的计算公式如下:

$$d(k, k') = \max_{i=1,2} \left( \min_{j=1,2} \left( \frac{\|k_i - k_j'\|}{\|k_i\| + \|k_j'\|} \right) \right) \quad (8)$$

$$d(k', k) = \max_{j=1,2} \left( \min_{i=1,2} \left( \frac{\|k_i - k_j'\|}{\|k_i\| + \|k_j'\|} \right) \right) \quad (9)$$

$$h(k, k') = \max\{d(k, k'), d(k', k)\} \quad (10)$$

因此, 目标点  $p_i$  曲率的 Hausdorff 距离  $H_{p_i}$  可以表示为:

$$H_{p_i} = \max\{h_{i1}, h_{i2} \dots h_{ik}\} \quad (11)$$

如果式(8)和式(9)中的分母  $\|k_i\| + \|k_j'\|$  无限接近零并且分子远大于分母时, 会造成计算结果出现较大偏差。为了避免分母过小所带来的误差, 本文引入限制变量  $\partial$ , 如果分母小于  $\partial$  时, 用  $\frac{\|k_i\| + \|k_j'\| + \partial}{2}$  来表示分母。

本文通过目标点与其邻域点主曲率的 Hausdorff 距离来确定目标点与其邻域点的相似性, 进而确定点云模型中点的曲率变化程度, 结合曲率对点云进行分类。图2为两种方法提取的点云特征点分布图, 图2(a)为原始点云; 图2(b)为估计曲率分布模型, 本文根据曲率大小将曲率分成大曲率、中曲率和小曲率3类, 其中红色表示曲率较大的点云, 青色表示曲率较小、区域平坦的点云, 黄色表示介于红色和青色之间的点云; 图2(c)为 Hausdorff 距离分布模型, 分层设色方法与曲率估计方法相同。由图2(b)可知, 在兔子特征较为明显的地方, 其曲率较大; 在兔子的腿部和耳朵等地方, 红色和黄色点分布较多。对比图2(b)和图2(c)可以发现, 图2(c)在兔子特征明显的耳朵和腿部的红色点和黄色点区域面积要大于图2(b)。

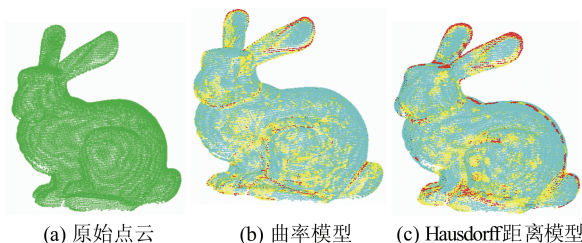


图2 兔子特征点模型

Fig.2 Models of Bunny Feature Points

点云  $k$ -means 聚类算法的核心是利用点云数据点与点之间的欧氏距离作为聚类指标, 即选择  $k$  个聚类中心, 记为  $J(j_1, j_2 \dots j_k)$ , 形成  $k$  个点簇, 记为  $C(c_1(p_{1i}), c_2(p_{2i}) \dots c_k(p_{ki}))$ 。首先分别计算点云中每个点到聚类中心点集  $J(p_1, p_2 \dots p_k)$  的欧氏距离, 比较计算所得的欧氏距离, 将每个点云数据点归为距离聚类中心最近的点; 然后根据每个聚类中心所包含的点集重新计算聚类中心  $J'(j_1, j_2 \dots j_k)$ , 比较此次聚类中心点集和前一次聚类中心点集, 如果相邻两次聚类中心点集距离之和小于预先设定值, 则停止聚类, 否则重复计算聚类中心; 最后以聚类中心代替每一类别中的点, 从而达到精简的目的<sup>[20]</sup>。

本文  $k$ -means 聚类中  $k$  值的选取方法如下:

(1) 设原始点云数量为  $S_0$ , 八叉树立方体数为  $S_c$ , 目标精简率为  $\delta$ , 则精简后的点云数量为  $(1 - \delta)S_0$ , 本文设定  $k$  为精简后的点云数量的 10%, 即  $(1 - \delta)S_0/10$ , 假设  $(1 - \delta)S_0/10 < 2S_c$ , 则  $k$  取值为  $2S_c$ ; (2) 分析点云主曲率的 Hausdorff 距离并确定初精简点集, 将初精简集中的点分别划分到八叉树的小长方体中, 根据  $k$  值大小和小长方体中点的个数来确定小长方体中聚类中心点的个数, 按照每个小长方体的聚类中心点的个数均匀地选取聚类中心点。

若本文选取第  $i$  个八叉树划分算法中的小长方体作为示例, 设该小长方体的非特征点集为  $D_i$ , 点数为  $n_i$ , 体积为  $V_i$ , 则它的密度  $\rho_i$  为:

$$\rho_i = n_i / V_i \quad (12)$$

$k$  值的计算公式为:

$$k_i = \left\lceil k \rho_i / \sum_{j=1}^n \rho_j \right\rceil \quad (13)$$

式中,  $\lceil \cdot \rceil$  为向上取整符号;  $n$  为小长方体的数量。随机选取  $k_i$  个数据点作为它的初始聚类中心。

点云精简算法在精简过程中得到特征点和非特征点后, 往往会直接删除非特征点。本文使用 Hausdorff 距离计算出特征点和非特征点后, 得到特征区域, 并利用  $k$ -means 聚类算法对非特征点进行处理。由于  $k$ -means 具有聚类的特性, 从而保证了点云模型非特征区域能够保留部分点且分布均匀。

融合  $k$ -means 聚类 and Hausdorff 距离的点云精简算法步骤如下:

1) 将 Matlab 读取到的点云数据点坐标存入矩阵  $O\_CloudPoints$  中。



2)在八叉树算法的基础上构建点云数据的空间拓扑关系。八叉树构建完成后,对八叉树划分结果建立索引并存入数组  $O\_OctTree$  中。

3)在八叉树算法的基础上构建每个数据点的  $k$ -邻域索引并存入  $K\_Neighbors$  中。

4)在最小二乘曲面拟合算法的基础上对点云数据每个点的邻域点集进行拟合,根据拟合方程计算主曲率,通过§1.3的计算方法计算目标点主曲率的 Hausdorff 距离。

5)利用步骤 4)中的方法依次计算点云数据所有点的 Hausdorff 距离。

6)按照由大到小的顺序对点云数据所有点的主曲率的 Hausdorff 距离进行排序,根据精简比例以及二次精简保留的  $k$  值确定第 1 次保留的点数  $(1-\delta)S_0-k$ ,以排序后的点云数据点序号为  $(1-\delta)S_0-k$  的点的 Hausdorff 距离作为阈值  $\epsilon$ 。根据构建的  $O\_OctTree$  遍历每个小长方体包含的点云数据点,并将非特征点加入  $D_i$  中。

7)根据式(12)和式(13)计算第  $i$  个小长方体的  $k$  值大小( $k_i$ ),随机选取小长方体内的  $k_i$  个数据点作为该小长方体的初始聚类中心,记为  $J_i(p_1, p_2 \dots p_{k_i})$ 。根据  $k$ -means 算法确定最终的聚类中心  $J'(p_1, p_2 \dots p_{k_i})$ ,以该聚类结果代替原始数据完成精简。

8)将 Hausdorff 距离和  $k$ -means 算法提取的结果进行融合,得到最终的精简数据。

融合  $k$ -means 聚类和 Hausdorff 距离的点云精简算法流程如图 3 所示。

## 2 实验结果与分析

将本文的融合算法分别与基于  $k$ -means 聚类的点云精简算法和基于 Hausdorff 的点云精简算法进行对比,以验证本文算法的准确性和有效性。采用经典的兔子数据和马数据(该数据来源于斯坦福大学)作为实验数据,点云数量分别为

31 606 和 48 485。兔子模型的特征较为集中,主要集中在兔子四肢和耳朵等处;马模型的特征主要集中在马的四肢和头部等地方。图 4 是 3 种算法对兔子点云精简后的结果,图 5 是 3 种算法对马点云精简后的结果,表 1 是 3 种精简算法对兔子和马点云精简后的精简率和点云数量对比。图 4(b)和图 5(b)表示  $k$ -means 聚类精简结果,其精简结果具有  $k$ -means 聚类的特性,精简后的点云分布均匀,并具有一定填充空洞的效果,但易精简大量特征点,造成特征区域变形,丢失大量特征信息。图 4(c)和图 5(c)表示基于 Hausdorff 距离的精简结果,由于 Hausdorff 距离能判断相似性,所以该算法能保留特征点周围大量的点,但其精简结果出现了大量的空洞现象。图 4(d)和图 5(d)表示本文融合算法的精简结果,通过图 4、图 5 和表 1 可以看出,在总的保留点数差不多的情况下,本文算法在兔子和马的特征区域保留了大量点,并且本文融合算法精简结果的轮廓保留完整,未造成数据空洞。

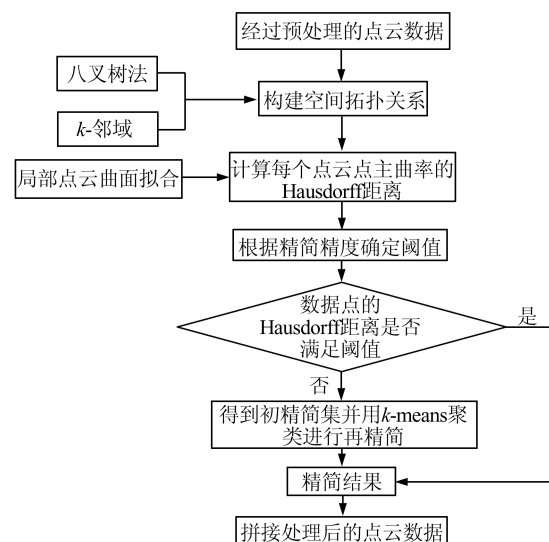


图 3 融合  $k$ -means 聚类和 Hausdorff 距离的点云精简算法流程

Fig.3 Flowchart of Point Cloud Simplification Algorithm Integrating  $k$ -means Clustering and Hausdorff Distance

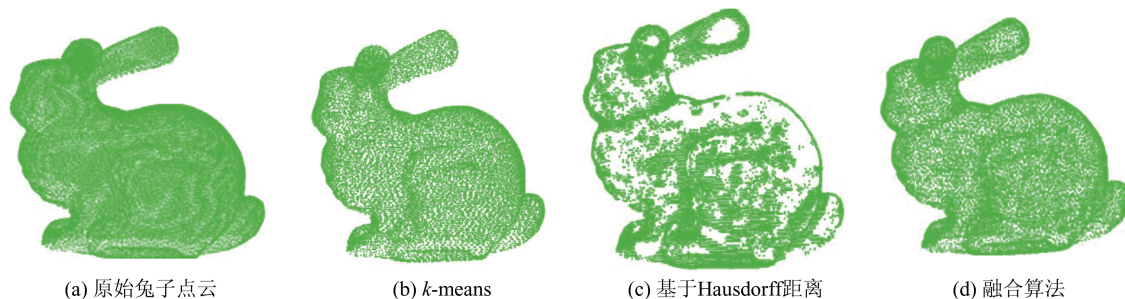


图 4 兔子点云数据 3 种算法精简结果

Fig.4 Simplification Results of the Rabbit Point Cloud of Three Algorithms



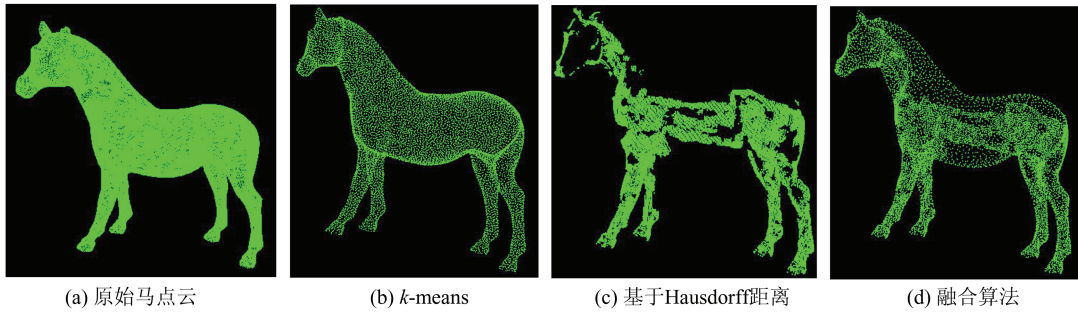


图 5 马点云数据 3 种算法精简结果

Fig.5 Simplification Results of the Horse Point Cloud of Three Algorithms

表 1 3 种算法精简前后数据对比

Tab.1 Data Comparison of Three Algorithms Before and After Simplification

点云类型	原始点数	<i>k</i> -means		基于 Hausdorff 距离		融合算法	
		精简后点数	精简率/%	精简后点数	精简率/%	精简后点数	精简率/%
兔子点云	31 606	15 000	52.5	12 924	59.1	16 887	46.6
马点云	48 485	10 471	78.4	12 512	74.2	10 787	77.8

本文利用 Geomagic 软件对原始点云数据以及 3 种算法精简后的点云数据进行三维曲面重构,分别对实验数据中特征明显的区域(兔子耳朵和马四肢)进行了对比,并对兔子耳朵和马四肢的三角面片数量进行了统计分析。为了保证截取的细节部位形状、大小和位置相同,本文在裁剪原始点云重构模型和 3 种算法精简后的点云数据重构模型时,裁剪面的坐标和旋转角度完全相同。图 6 和图 7 分别是兔子耳朵和马四肢原始点云重构模型和 3 种算法精简后的点云数据重构模型,表 2 是兔子耳朵和马四肢的三角面片数量。由表 2 可知,基于 Hausdorff 距离的精简结果重构网格切割后,兔子耳朵处三角面片数量为 5 698,

马四肢处为 8 490;基于 *k*-means 的精简算法在兔子耳朵处的三角面片数量为 3 726,在马四肢处为 4 650。由此可见,基于 Hausdorff 距离的精简结果在特征区域保留的信息明显多于基于 *k*-means 的精简算法,但基于 Hausdorff 距离的精简结果出现了大量空洞。本文融合算法融合了 *k*-means 聚类 and Hausdorff 距离,其重构网格在兔子耳朵处的三角面片数量为 4 942,在马四肢的三角面片数量为 6 908,与基于 Hausdorff 距离算法的结果相差不大,并且未造成数据空洞。因此,本文算法在保留更多特征信息的同时,也弥补了基于 Hausdorff 距离易造成数据空洞的不足,提升了算法的适用范围。

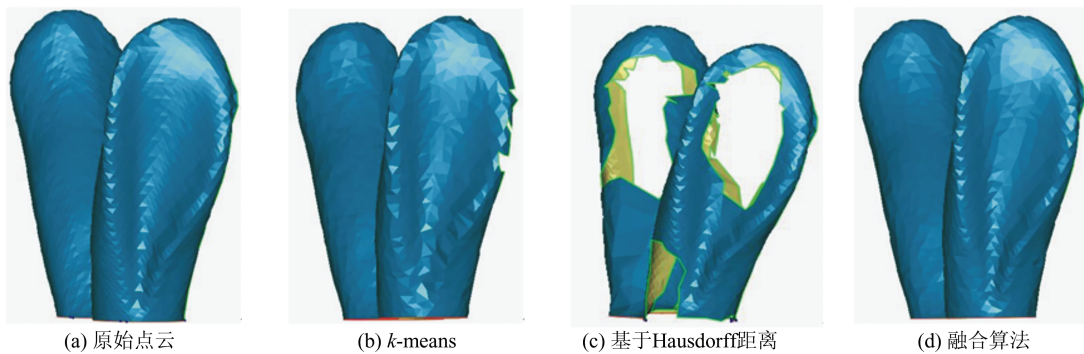


图 6 兔子耳朵部位 3 种算法重构网格

Fig.6 Reconstruction Grids of Three Algorithms in Rabbit Ears

为了分析不同精简率的精简结果,利用本文算法分别对精简率目标约 19.5%、34.9%、46.6%、66.2% 和 86.3% 的兔子点云进行精简,其结果如图 8 所示。其中,图 8(a)为原始兔子点云,

图 8(b)~8(f)分别对应上述精简率精简后的点数。由图 8 可以看出,本文融合算法在兔子特征信息多的区域保留了大量点,比如兔子四肢和耳朵等起伏较大的不平坦地方;在兔子较为平坦的

地方则点比较稀疏,但对兔子的轮廓保留较好, 并且在小曲率点区域的点分布均匀。

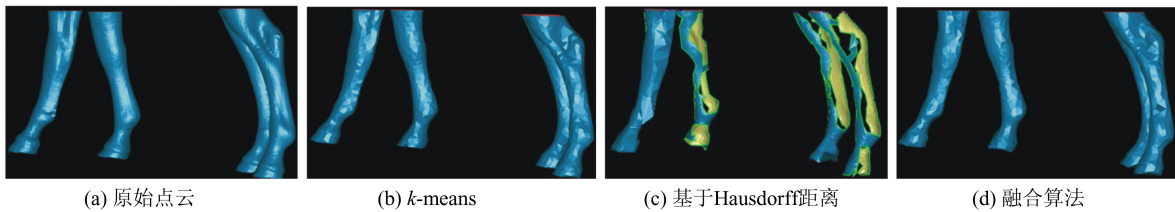


图 7 马四肢部位 3 种算法重构网格

Fig.7 Reconstruction Grids of Three Algorithms for Horse Limbs

表 2 3 种算法重构后三角面片数据对比

Tab.2 Data Comparison of Triangle Faces of Three Algorithms After Reconstruction

部位	三角面片数量			
	原始数据	$k$ -means	基于 Hausdorff 距离	融合算法
兔子耳朵	7 831	3 726	5 698	4 942
马四肢	21 147	4 650	8 490	6 908

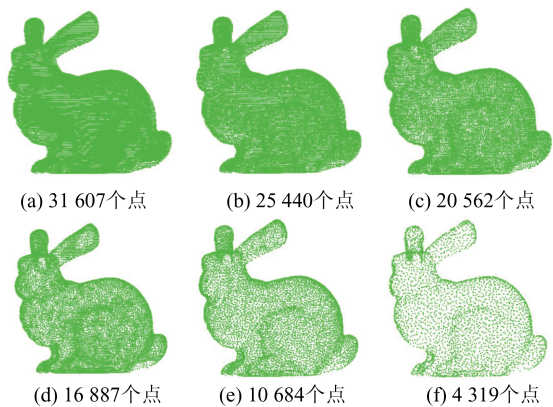


图 8 本文算法不同精简率的精减结果

Fig.8 Streamline Results of Different Simplification Rates of the Proposed Algorithm

为了对精简结果质量进行评价<sup>[21]</sup>,本文选取精简后点云重构网格的表面积和精简后兔子点云数据与原始兔子点云数据误差  $\Delta(S, S^*) = d(p, S^*)$  的标准差作为量化指标,其中  $d(p, S^*)$  表示原始点云数据重构网格曲面  $S$  上点  $p$  到精简后点云重构网格曲面  $S^*$  上投影点  $p^*$  的欧氏距离<sup>[15,22]</sup>。本文分别计算了 3 种精简算法精简结果的质量量化指标,如图 9 所示。

由图 9(a)可知,随着精简率的增大,精简结果的点云数量占比降低,3 种算法精简结果的误差标准差都呈现增大的趋势,本文融合算法和基于 Hausdorff 距离的精简算法的误差标准差明显小于基于  $k$ -means 聚类算法。由图 9(b)可知,基于 Hausdorff 距离算法的精简结果重构网格表面积随着精简结果点云数量占比的减少而减少,所

以本文算法具备精简结果误差标准差较小和重构网格表面积与原模型表面积相差不大的优点。但由于本文算法先运用主曲率的 Hausdorff 距离确定初精简集,再运用  $k$ -means 聚类,所以运算时间要高于另外两种算法。

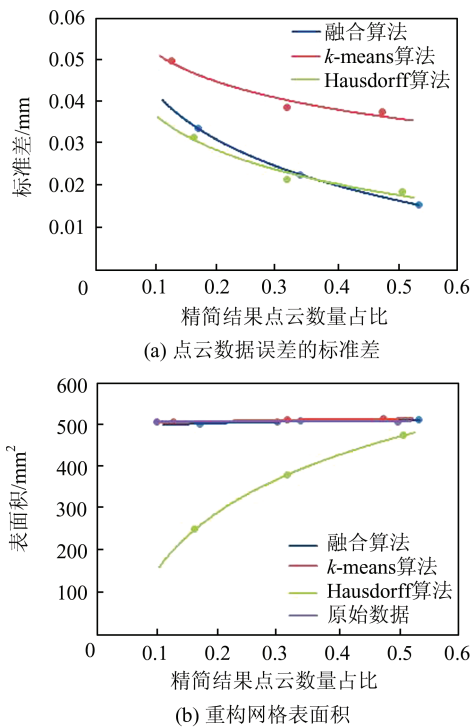


图 9 精简质量评价指标

Fig.9 Evaluation Indexes of Simplification Quality

综合分析图 4 至图 8 可以得出,基于  $k$ -means 的精简算法具有高效精简的特性,且精简结果分布均匀,但其在兔子和马的特征区域丢失了大量特征信息;基于 Hausdorff 距离的精简算法能较完整地保留兔子和马的特征信息,但其精简结果具有大量的空洞现象;本文算法能较完整地保留兔子和马的特征信息,并且能避免形成空洞现象。

### 3 结 语

基于 Hausdorff 距离的点云精简算法能较好地保留点云形状特征,提高精简效率,然而直接

去除非特征区域点容易造成数据空洞;基于  $k$ -means 聚类的点云精简算法有较高的精简效率,但对细节特征保留效果较差。本文融合 Hausdorff 距离和  $k$ -means 聚类,利用 Hausdorff 距离具有相似性度量的特性提取特征区域,并对非特征区域进行  $k$ -means 聚类提取特征点,凸现了  $k$ -means 聚类算法保留点云数据整体轮廓的特点,避免了数据空洞问题,从而使精简结果更加完整,精简率高,且保留了较多的形状特征。实验表明,该算法不但能保持原始数据的整体轮廓且避免出现数据空洞,还能够较好地保留点云数据的局部细节形状特征。如何确定最优聚类中心以及提升融合算法的效率还有待进一步研究。

### 参 考 文 献

- [1] Gong Jianya, Cui Tingting, Shan Jie, et al. A Survey on Facade Modeling Using LiDAR Point Clouds and Image Sequences Collected by Mobile Mapping Systems[J]. *Geomatics and Information Science of Wuhan University*, 2015, 40(9): 1 137-1 143(龚健雅, 崔婷婷, 单杰, 等. 利用车载移动测量数据的建筑物立面建模方法[J]. 武汉大学学报·信息科学版, 2015, 40(9): 1 137-1 143)
- [2] Yan Li, Hu Xiaobin, Xie Hong. Data Management and Visualization of Mobile Laser Scanning Point Cloud [J]. *Geomatics and Information Science of Wuhan University*, 2017, 42(8): 1 131-1 136 (闫利, 胡晓斌, 谢洪. 车载 LiDAR 海量点云数据管理与可视化研究[J]. 武汉大学学报·信息科学版, 2017, 42(8): 1 131-1 136)
- [3] Zhang Shunlan, Mo Jianwen, Zou Lulu. Point Cloud Simplification Algorithm Based on  $k$  Neighbor and Normal Accuracy[J]. *Journal of Wuhan University of Technology (Transportation Science & Engineering)*, 2014, 38(3): 572-575(张顺岚, 莫建文, 邹路路. 基于  $k$  近邻和法向精度的点云精简算法[J]. 武汉理工大学学报(交通科学与工程版), 2014, 38(3): 572-575)
- [4] Li Jian, Wang Zongmin, Ma Yurong, et al. Automatic and Accurate Mosaicking of Point Clouds from Multi-station Laser Scanning [J]. *Geomatics and Information Science of Wuhan University*, 2014, 39(9): 1 114-1 120(李健, 王宗敏, 马玉荣, 等. 多站激光点云数据全自动高精度拼接方法研究[J]. 武汉大学学报·信息科学版, 2014, 39(9): 1 114-1 120)
- [5] Sun W, Bradley C, Zhang Y F, et al. Cloud Data Modelling Employing a Unified, Non-redundant Triangula Mesh [J]. *Computer - Aided Design*, 2001, 33(2): 183-193
- [6] Li Renzhong, Yang Man, Liu Yangyang, et al. An Uniform Simplification Algorithm for Scattered Point Cloud[J]. *Acta Optica Sinica*, 2017, 37(7): 97-105(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] Han H, Han X, Sun F, et al. Point Cloud Simplification with Preserved Edge Based on Normal Vector [J]. *Optik-International Journal for Light and Electron Optics*, 2015, 126(19): 2 157-2 162
- [8] Tonini M, Abellan A. Rockfall Detection from Terrestrial LiDAR Point Clouds: A Clustering Approach Using R[J]. *Journal of Spatial Information Science*, 2014(8): 95-110
- [9] Weir D J, Milroy M J, Bradley C, et al. Reverse Engineering Physical Models Employing Wrap-Around B-spline Surfaces and Quadrics [J]. *Proceedings of the Institution of Mechanical Engineers Part B Journal of Engineering Manufacture*, 1996, 210(22): 147-157
- [10] Wang L, Chen J, Yuan B. Simplified Representation for 3D Point Cloud Data[C]. IEEE International Conference on Signal Processing, Beijing, China, 2010
- [11] Lee K H, Woo H, Suk T. Point Data Reduction Using 3D Grids [J]. *International Journal of Advanced Manufacturing Technology*, 2001, 18(3): 201-210
- [12] Yuan S, Zhu S, Li D S, et al. Feature Preserving Multiresolution Subdivision and Simplification of Point Clouds: A Conformal Geometric Algebra Approach [J]. *Mathematical Methods in the Applied Sciences*, 2017, 41(10): 4 074-4 087
- [13] Yang Qiuxiang, Yang Xiaoqing, Du Jian. Point Cloud Simplification Algorithm Based on Hausdorff Distance and Segmentation[J]. *Computer Engineering and Design*, 2016, 37(8): 2 105-2 109(杨秋翔, 杨小青, 杜健. 基于 Hausdorff 距离的点云分片精简算法[J]. 计算机工程与设计, 2016, 37(8): 2 105-2 109)
- [14] Chen Zhangwen, Da Feipeng. 3D Point Cloud Simplification Algorithm Based on Fuzzy Entropy Iteration[J]. *Acta Optica Sinica*, 2013, 33(8): 161-167(陈璋雯, 达飞鹏. 基于模糊熵迭代的三维点云精简算法[J]. 光学学报, 2013, 33(8): 161-167)
- [15] Shi B Q, Liang J, Liu Q. Adaptive Simplification of Point Cloud Using  $k$ -means Clustering[J]. *Computer-Aided Design*, 2011, 43(8): 910-922
- [16] Lv Zhao, Kang Zhizhong. Change Detection of Buildings Based on Terrestrial Laser Scanning Data [J]. *Geomatics and Information Science of Wuhan University*, 2011, 36(11): 1 284-1 289(吕钊, 康志忠. 利用地面激光扫描数据进行建筑物变化检测[J]. 武汉大学学报·信息科学版, 2011, 36(11):



- 1 284-1 289)
- [17] Li T, Pan Q, Gao L, et al. A Novel Simplification Method of Point Cloud with Directed Hausdorff Distance [C]. IEEE 21st International Conference on Computer Supported Cooperative Work in Design (CSCWD), Wellington, New Zealand, 2017
- [18] Zhou Yulian. Feature Extruction of Point Cloud Based on Normal Information [D]. Harbin: Harbin Institute of Technology, 2012(周玉莲. 基于法矢信息的点云特征提取技术的研究 [D]. 哈尔滨: 哈尔滨工业大学, 2012)
- [19] Zhu Yu, Kang Baosheng, Li Hong'an, et al. Improved Algorithm for Point Cloud Data Simplification [J]. *Journal of Computer Applications*, 2012, 32 (2): 521-523(朱煜, 康宝生, 李洪安, 等. 一种改进的点云数据精简方法 [J]. 计算机应用, 2012, 32 (2): 521-523)
- [20] Yuan Xiaocui, Wu Lushen, Chen Huawei. Feature Preserving Point Cloud Simplification [J]. *Optics and Precision Engineering*, 2015, 23(9): 2 666-2 676 (袁小翠, 吴禄慎, 陈华伟. 特征保持点云数据精简 [J]. 光学精密工程, 2015, 23(9): 2 666-2 676)
- [21] Shi Baoquan, Liang Jin, Liu Qing, et al. Precision Inspection of Point Cloud & CAD Model Based on Constraint Search Sphere [J]. *Computer Integrated Manufacturing Systems*, 2010, 16(5): 929-934(史宝全, 梁晋, 刘青, 等. 基于约束搜索球的点云数据与 CAD 模型精确比对检测 [J]. 计算机集成制造系统, 2010, 16(5): 929-934)
- [22] Fu Siyong, Wu Lushen, Chen Huawei. Point Cloud Simplification Method Based on Space Grid Dynamic Partitioning [J]. *Acta Optica Sinica*, 2017, 37(11): 253-261(傅思勇, 吴禄慎, 陈华伟. 空间栅格动态划分的点云精简方法 [J]. 光学学报, 2017, 37(11): 253-261)

## Scattered Point Cloud Simplification Algorithm Integrating $k$ -means Clustering and Hausdorff Distance

LI Jian<sup>1</sup> CAO Yao<sup>2</sup> WANG Zongmin<sup>2,3</sup> WANG Guangyin<sup>2</sup>

<sup>1</sup> School of the Geo-Science & Technology, Zhengzhou University, Zhengzhou 450001, China

<sup>2</sup> School of Water Conservancy and Environment, Zhengzhou University, Zhengzhou 450001, China

<sup>3</sup> Zhongyuan University of Technology, Zhengzhou 450007, China

**Abstract:** Aiming at the incomplete retention of features during the point cloud data procession by point cloud simplification algorithm, and data holes caused by small-curvature point cloud simplification algorithm, this paper proposes a new point cloud simplification algorithm integrated  $k$ -means clustering and Hausdorff distance. The topological adjacency is established in the new simplification algorithm based on the OcTree algorithm. Then the principal curvatures of all point cloud is calculated and the Hausdorff distance of the principal curvatures is calculated, and the Hausdorff distance threshold set by the requirements of the reduced target is used to extracted the point cloud feature. Finally,  $k$ -means clustering is performed on non-feature regions to extract feature points, and the extracted feature points are merged to obtain reduced results. Results show that the proposed algorithm can retain the feature information of the model more completely and avoid the void phenomena.

**Key words:** point cloud simplification; Hausdorff distance;  $k$ -means clustering; curved surface fitting

**First author:** LI Jian, PhD, associate professor, specializes in LiDAR point cloud data processing and 3D reconstruction. E-mail: lijian5277@163.com

**Corresponding author:** CAO Yao, master. E-mail: 772440651@qq.com

**Foundation support:** The National Natural Science Foundation of China (51678536); the Open Research Fund of State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (15E01); Key Scientific Research Projects of Colleges and Universities of Henan Provincial Department of Education (14A420002).

**引文格式:** LI Jian, CAO Yao, WANG Zongmin, et al. Scattered Point Cloud Simplification Algorithm Integrating  $k$ -means Clustering and Hausdorff Distance [J]. *Geomatics and Information Science of Wuhan University*, 2020, 45(2): 250-257. DOI: 10.13203/j.whugis.20180204 (李健, 曹焱, 王宗敏, 等. 融合  $k$ -means 聚类 and Hausdorff 距离的散乱点云精简算法 [J]. 武汉大学学报·信息科学版, 2020, 45(2): 250-257. DOI: 10.13203/j.whugis.20180204)