

基于 k -means 聚类的点云精简方法

贺一波^{1*}, 陈冉丽², 吴侃³, 段志鑫³

¹大同煤炭职业技术学院建筑工程系, 山西 大同 037003;

²石家庄铁路职业技术学院测绘工程系, 河北 石家庄 050041;

³中国矿业大学环境与测绘学院, 江苏 徐州 221116

摘要 提出了一种基于 k 均值(k -means)聚类的点云精简方法。与包围盒法相比,在压缩率近似相同的条件下, k -means 聚类方法能较好地保留细节特征,与原始数据的稠密稀疏分布更加一致,所建模型表面更光滑。

关键词 图像处理; 点云精简; k 均值聚类; 曲面拟合; 均方根曲率; 压缩率

中图分类号 TP391

文献标识码 A

doi: 10.3788/LOP56.091002

Point Cloud Simplification Method Based on k -Means Clustering

He Yibo^{1*}, Chen Ranli², Wu Kan³, Duan Zhixin³

¹Department of Architecture and Civil Engineering, Datong Vocational and Technical College of Coal, Datong, Shanxi 037003, China;

²Department of Surveying Engineering, Shijiazhuang Institute of Railway Technology, Shijiazhuang, Hebei 050041, China;

³School of Environment Science and Spatial Informatics, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China

Abstract A point cloud simplification method is proposed based on k -means clustering. Compared with the bounding box method with a similar compression rate, the k -means clustering method can preserve the details better, and the result is more consistent with the dense and sparse distribution of the original data. Moreover, the surface of the constructed model is smoother.

Key words image processing; point cloud simplification; k -means clustering; surface fitting; root mean square curvature; compression rate

OCIS codes 100.2000; 280.3640; 140.3430

1 引言

三维激光扫描技术作为一种新型技术,主要采用非接触的方式获取目标物体表面数据点的三维坐标,具有高效、高精度等优点,广泛应用于各个领域^[1]。但是,海量的点云数据减缓了数据处理的速度^[1-3]。所以在保证目标物体模型精度的前提下,研究点云数据的精简问题很有必要。

国内外学者针对点云数据精简的问题进行了大量的研究,主要有 3 类方法:包围盒法、随机采样法和曲率采样法^[4]。这 3 种方法都有其适用条件和不

足:包围盒法多适用于曲率较小的数据,对于曲率起伏较大的数据,若压缩率较大,则精简后的点云数据不能保留较多的特征点,会影响数据的细节表达^[5];随机采样法没有考虑目标物的形状和特征,而是随机地保留点云数据,若压缩率较大,则可能导致数据失真^[6];曲率采样法适合于曲率较大的数据,对于平坦的数据,若压缩率较大,则可能造成点云数据的大量缺失^[7-8]。在利用上述 3 种方法进行点云精简时,精简后的点云数据量往往较多,达不到较好的精简效果。为了保证点云数据精度,确保精简后的点云数据能较好地表达细节特征,需要引进一种方法,使

收稿日期: 2018-10-17; 修回日期: 2018-11-10; 录用日期: 2018-11-30

* E-mail: 511149199@qq.com

精简后的点云数据量较少,且能较好地保留特征点,真正达到精简点云的效果。

鉴于此,本文提出一种新的点云精简方法。引入 k 均值(k -means)聚类方法(k 为聚类数目),对点云数据进行分类,将每一聚类中的点云数据拟合成曲面,通过比较每一点的均方根曲率值和所有点的均方根曲率的平均值,删减点云数据,达到精简点云的目的。不同于包围盒法, k -means 聚类方法充分考虑到点云整体的形状,根据距离直接进行点云分类,不需要建立网格,提高了程序运行的速度。

2 k -means 聚类方法

k -means 聚类算法是硬聚类算法的一种,根据目标函数进行聚类^[9]。本文中,目标函数即为点云数据到初始聚类中心的欧氏距离。具体过程如下^[10]。

1) 原始点云数据 $P=(p_1, p_2, \dots, p_n)$ (其中, n 为点云个数), 给定分类组数 $B(B \leq n)$, 根据 B 值的大小, 点云数据会随机产生 B 个聚类中心 $O=(O_1, O_2, \dots, O_B)$ 。

2) 计算所有点云数据 P 与每个聚类中心 O 的欧氏距离, 根据距离的最小值, 选择最近的聚类, 从而得到 B 个聚类。

3) 分别计算每个聚类中点云数据的重心值, 作为新的聚类中心。

4) 所有点云数据按照新的聚类中心重新进行聚类。

5) 重复步骤 3)、4), 直到聚类中心不再变化, 以此得到的 B 个聚类即为聚类结果。

3 点云曲率计算和精简过程

3.1 曲率计算

根据每个聚类中的点云数据, 利用最小二乘法拟合曲面, 并计算每个三维点的均方根曲率值。设曲面方程为

$$z = a_0 + a_1x + a_2y + a_3x^2 + a_4y^2 + a_5xy, \quad (1)$$

式中: x, y 为某一点云数据的横、纵坐标; $a_0, a_1, a_2, a_3, a_4, a_5$ 为系数。

根据文献[11], 令

$$u = \frac{\partial z}{\partial x}, \quad (2)$$

$$v = \frac{\partial z}{\partial y}, \quad (3)$$

式中: u, v 为曲面方程函数对某一点云数据横、纵坐

标的偏导。

为简化公式, 令:

$$E = 1 + u^2, \quad (4)$$

$$F = uv, \quad (5)$$

$$G = 1 + v^2, \quad (6)$$

$$L = \frac{\frac{\partial u}{\partial x}}{\sqrt{1 + u^2 + v^2}}, \quad (7)$$

$$M = \frac{\frac{\partial u}{\partial y}}{\sqrt{1 + u^2 + v^2}} = \frac{\frac{\partial v}{\partial x}}{\sqrt{1 + u^2 + v^2}}, \quad (8)$$

$$N = \frac{\frac{\partial v}{\partial y}}{\sqrt{1 + u^2 + v^2}}. \quad (9)$$

则高斯曲率可表示为

$$K = k_1 k_2 = \frac{LN - M^2}{EG - F^2}, \quad (10)$$

式中: k_1, k_2 为曲面上某一点云数据的主曲率。

平均曲率可表示为

$$H = \frac{k_1 + k_2}{2} = \frac{EN - 2FM + GL}{2(EG - F^2)}. \quad (11)$$

均方根曲率可表示为

$$C = \sqrt{\frac{k_1^2 + k_2^2}{2}} = \sqrt{2H^2 - K}. \quad (12)$$

3.2 点云精简

以均方根曲率作为点云精简的依据。以第 i ($i=1, 2, \dots, B$) 个聚类为例, 根据聚类中各点的均方根曲率 C_{ij} (表示第 i 个聚类中第 j 个点云数据的均方根曲率), 计算出均方根曲率的平均值 \bar{C}_i , 比较 C_{ij} 和 \bar{C}_i 的大小: 若 $C_{ij} < \bar{C}_i$, 认为此点曲率过小, 所包含的特征信息较少, 删除该点; 若 $C_{ij} \geq \bar{C}_i$, 认为此点曲率过大, 所包含的特征信息较多, 保留该点。

4 实验与结果分析

4.1 分类组数对点云精简的影响

利用 k -means 聚类方法, 需要先确定分类组数。本文先分析分类组数对点云精简的影响。以斯坦福大学三维扫描库中的 bunny 为原数据, 点云个数为 161032, 在 Matlab 中编写相应的程序, 分别设置分类组数为 1000、2500、5000、10000、20000, 得到精简后的点云数据, 导入到 Geomagic 中, 建立相应的模型, 分别计算其与原始点云数据模型的标准偏差, 计算结果如表 1 所示(表中压缩率是剩余点云个数与原数据点云数据的比值)。

表1 不同分类组数下的压缩率和标准偏差

Table 1 Compression rates and standard deviations for different numbers of clusters

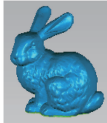
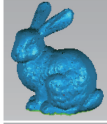
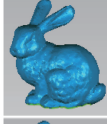
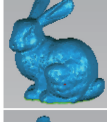
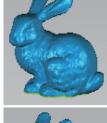
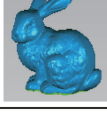
Number of clusters	Remaining number of point clouds	Compression rate	Standard deviation	Model of entire picture
1(original data)	161032			
1000	42045	0.2611	0.00013	
2500	25753	0.1599	0.000141	
5000	21793	0.1353	0.00013	
10000	24675	0.1532	0.000121	
20000	37542	0.2331	0.000093	

表1采用此方法精简数据,压缩率相对较低,但是却能很好地保存特征点,模型表面也较光滑。随着聚类组数的增加,标准偏差整体上呈下降趋势。当聚类组数为1000~5000时,压缩率会下降;当聚类组数为5000~20000时,压缩率会上升。分析原因:当聚类组数初始增加时,每组中的点云个数减少,且点云曲率差异较小,所以保留的点云数据会减少;当聚类组数继续增加时,每一聚类中至少保留一个点云数据,因此保留的点云数据会随着聚类组数的增加而增加。

4.2 k -means 与现有方法的对比分析

由于本文选用的数据曲率变化较小,相比于随机采样法和曲率采样法,包围盒法进行点云精简处理效果较好。因此,为证明 k -means 聚类方法的可行性,将其与包围盒算法进行对比分析。


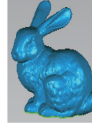
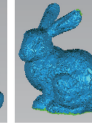
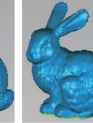
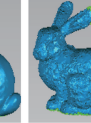
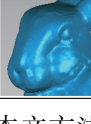
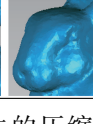
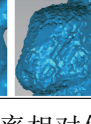
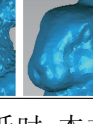
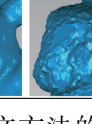
4.2.1 三维模型比较

当压缩率分别近似等于0.235和0.154时,将点云数据导入 Geomagic 软件中,利用同一种方法建模,比较两种方法的标准偏差、全图和头部细节展示图,如表2所示。

从表2可以看出,当两种方法的压缩率近似相

表2 两种方法的标准偏差和建模结果

Table 2 Standard deviations and modeling results for two methods

Index	Original data	Compression rate of 0.235		Compression rate of 0.154	
		Proposed method	Bounding box method	Proposed method	Bounding box method
Remaining number of point clouds	160132	0.2331	0.2371	0.1532	0.1555
Standard deviation		0.000093	0.000111	0.000121	0.00069
Model of entire picture					
Head detail display					

等,甚至本文方法的压缩率相对低时,本文方法的标准偏差更小,而且压缩率越小,两种方法的标准偏差差值越大,说明本文精简点云数据的方法越有优势。对比两种方法的模型,本文方法构建的模型更加光滑,特征保留效果也更好,而包围盒法构建的模型表面较粗糙,在细节展示图中很难分辨出兔子的眼睛和嘴,精简效果差。

4.2.2 精度评定

精度评定方法主要有3种:表面积法、体积法和切面法。为了进一步说明本文方法在保留细节上的优势,利用切面法处理点云数据。切面法是取任一平面切割原来的三角格网,将所形成的交线和交点投影到平面上进行比较,一般选择交点^[12]。选取两个平面截取点云数据,将两平面之间的点云数据投影到同一平面上进行比较。在细节展示图中可以看出,脸部数据变化较明显,所以选取脸部的数据进行比较,得到的数据如表3所示。

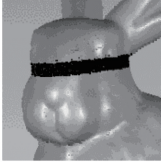





表3中点云数据是通过投影得到的,所以存在点云密集和稀疏的地方。当压缩率近似相等时,本文算法得到的点云数据与原始数据的稠密稀疏分布更加一致,而包围盒法得到的点云数据稠密稀疏比较均匀,与原始数据分布不一致。在原始数据中存在一些弯曲部分,本文算法得到的点云数据也在相应位置存在弯曲部分,而包围盒法得到的点云数据较平滑,点云弯曲部分不明显。随着压缩率的降低,本文算法依然可以保留细节特征,而包围盒法的点云数据平缓,未能很好地保留特征数据。

5 结 论

基于 k -means 聚类方法提出了一种新的点云

表3 切面法投影后的点云数据

Table 3 Point cloud data after projection by section method

Index	Plane selected by section method	Original data	Compression rate of 0.235		Compression rate of 0.154	
			Proposed method	Bounding box method	Proposed method	Bounding box method
Point cloud data after projection by section method						

精简方法。利用 k -means 聚类方法对点云数据进行分类,在每类中利用曲率来精简点云数据。同时,分析了不同分类情况下点云精简的效果,并在压缩率近似相等的前提下,与包围盒法进行比较。结果发现,本文算法能较多地删减点云中的冗余数据,且较好地保留点云特征信息。

参 考 文 献

- [1] Axelsson P. Processing of laser scanner data: algorithms and applications [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 1999, 54(2/3): 138-147.
- [2] Zheng D H. The data reduction of point cloud and analysis of reduction effect [J]. Engineering of Surveying and Mapping, 2006, 15(4): 27-30.
郑德华. 点云数据直接缩减方法及缩减效果研究 [J]. 测绘工程, 2006, 15(4): 27-30.
- [3] Venkataramani R, Bresler Y. Optimal sub-Nyquist nonuniform sampling and reconstruction for multiband signals [J]. IEEE Transactions on Signal Processing, 2001, 49(10): 2301-2313.
- [4] Sun P F. Streamlined compression analysis and practice of point cloud data based on coordinate increment [D]. Xi'an: Xi'an University of Science and Technology, 2014.
孙鹏飞. 基于坐标增量的点云数据精简压缩分析与实践 [D]. 西安: 西安科技大学, 2014.
- [5] Fleishman S, Drori I, Cohen-Or D. Bilateral mesh denoising [J]. ACM Transactions on Graphics, 2003, 22(3): 950-953.
- [6] Pauly M, Gross M, Kobbelt L P. Efficient simplification of point-sampled surfaces [C] // IEEE Visualization, 2002: 163-170.
- [7] Li Q M. Research on processing technology of point cloud from laser scanning system [D]. Dalian: Dalian Maritime University, 2013.
李青蒙. 激光扫描点云处理技术研究 [D]. 大连: 大连海事大学, 2013.
- [8] Lee K H, Woo H, Suk T. Point data reduction using 3D grids [J]. The International Journal of Advanced Manufacturing Technology, 2001, 18(3): 201-210.
- [9] Hua Y M. Auto iterative repairing of surface point cloud sequences with missing [D]. Dalian: Dalian University of Technology, 2015.
滑亚妹. 带缺失的曲面点云序列的自动迭代修复 [D]. 大连: 大连理工大学, 2015.
- [10] Hartigan J A, Wong M A. Algorithm AS 136: a K -means clustering algorithm [J]. Applied Statistics, 1979, 28(1): 100-108.
- [11] Zhou Y, Zhang W B, Du F R, *et al.* Algorithm for reduction of scattered point cloud data based on curvature [J]. Transactions of Beijing Institute of Technology, 2010, 30(7): 785-789.
周煜, 张万兵, 杜发荣, 等. 散乱点云数据的曲率精简算法 [J]. 北京理工大学学报, 2010, 30(7): 785-789.
- [12] Xi W F. Reduction of laser point cloud data compression [D]. Kunming: Kunming University of Science and Technology, 2011.
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.