

# 基于多特征约束的露天采场道路点云提取

毛亚纯<sup>1</sup>, 杨哲玺<sup>1</sup>, 曹旺<sup>1</sup>, 齐迹<sup>2</sup>

(1. 东北大学 资源与土木工程学院, 辽宁 沈阳 110819; 2. 辽宁工程技术大学 测绘与地理科学学院, 辽宁 阜新 123000)

**摘要:** 针对露天采场道路点云数据通过法向量、路缘石等点云特征难以准确提取的问题, 提出了一种多特征约束的露天采场道路点云提取方法. 以辽阳市千山石灰石矿露天采场激光点云为数据源, 首先对原始数据进行降采样; 然后基于单点RGB信息、邻域RGB信息、邻域高差、邻域粗糙度、反射强度5类点云特征, 制作并划分了训练集和验证集, 利用随机森林算法构建了道路点云提取模型并进行了优化, 进一步引入欧式聚类算法改进了道路点云提取模型结果, 最后评估了露天采场道路点云提取结果. 结果表明, 本文方法可以实时有效准确地提取露天采场道路点云数据.

**关键词:** 露天采场; 道路点云; 点云特征信息; 随机森林算法; 欧式聚类算法

中图分类号: P 231 文献标志码: A 文章编号: 1005-3026(2024)09-1326-08

## Extraction of Road Point Cloud in Open Pit Based on Multi-feature Constraints

MAO Ya-chun<sup>1</sup>, YANG Zhe-xi<sup>1</sup>, CAO Wang<sup>1</sup>, QI Ji<sup>2</sup>

(1. School of Resources & Civil Engineering, Northeastern University, Shenyang 110819, China; 2. School of Geomatics & Geographic Sciences, Liaoning Technical University, Fuxin 123000, China. Corresponding author: YANG Zhe-xi, E-mail: yangzhexi\_neu@163.com)

**Abstract:** Aiming at the problem that road point cloud data in open pit is difficult to be accurately extracted through point cloud features such as normal vector and kerb, a method of road point cloud extraction in open pit with multi-feature constraints was proposed. Taking the laser point cloud in the open pit of Qianshan limestone mine in Liaoyang City as the data source, the original data was downsampled firstly, and then the training set and verification set were made and divided based on the five kinds of point cloud features including single point RGB information, neighborhood RGB information, neighborhood height difference, neighborhood roughness, and reflection intensity. The road point cloud extraction model was constructed and optimized using the random forest algorithm. Furthermore, European clustering algorithm was introduced to improve the road point cloud extraction model. Finally, the road point cloud extraction results were evaluated in open pit. The results show that the proposed method can effectively and accurately extract the road point cloud data in open pit in real time.

**Key words:** open pit; road point cloud; point cloud characteristic information; random forest algorithm; European clustering algorithm

道路是露天采场现状图的重要组成部分, 其实时更新是采场运输效率、经济效益和安全的重要保障, 更可为无人驾驶提供精确的道路及属性信息. 随着露天采场资源的不断开发, 采剥区域也随之变化, 导致了道路位置发生改变. 因此实

时有效准确地更新露天采场道路数据, 对于提升露天采场运输效率、经济效益和安全具有重要的现实意义.

随着微型导航定位系统及传感器的迅速发展, 基于无人机开展的非接触式扫描测量技术因

其便捷高效的优点,被广泛应用于地理数据采集及三维建模<sup>[1]</sup>.激光点云数据可清晰呈现道路的空间特征,因此国内外学者基于点云数据深入研究了不同方面的道路点云提取方法<sup>[2]</sup>.现有的道路点云提取方法主要有以下几类:

1) 通过数学方法提取道路点云, Huang等<sup>[3]</sup>提出了1种基于马尔可夫随机场(Markov random field, MRF)的快速点云地面分割方法. Mi等<sup>[4]</sup>提出了1种超体素提取道路边界及矢量化方法,通过对路缘石信息的候选学习实现了具有鲁棒性的各种道路形状和点密度特征结合的道路提取. Li等<sup>[5]</sup>提出了1种基于激光点云的道路边界检测多特征窗口检索方法,该方法通过主次检索窗口设置及双向查找方法,并结合二次曲线模型拟合道路边界实现校园道路提取. Lu等<sup>[6]</sup>提出了1种基于LiDAR数据的实时道路边界检测跟踪方法,根据空间距离特征和角度特征,通过检测得到的地面点云上方反馈提取道路边界点. Kumar等<sup>[7]</sup>提出了1种Snake模型用于道路边缘提取. Xiong等<sup>[8]</sup>提出了1种基于多回波LiDAR数据构建道路模型并自动提取边界线的方法,针对道路模型拟合地面点云后降噪处理,再根据多项点云特征约束提取道路边界线并完成边界信息优化. 程效军等<sup>[9]</sup>提出了1种基于地面点云强度的分类方法,利用颜色对强度进行补充,实现了对激光强度纠正结果的容错机制. Yadav<sup>[10]</sup>提出了1种基于强度滤波结合象限的区域增长提取道路网.

2) 在基于机器学习算法提取道路的研究中, Grilli等<sup>[11]</sup>研究了机器学习在点云分类中的应用,将特定数据集上训练的随机森林模型应用于未知场景. Bicić等<sup>[12]</sup>将RGB、表面法线、曲率、全方差、平面度、线性度等点云特征相结合,通过不同训练样本对机器学习模型优化后提取无路缘石路面,对露天采场道路提取具有指导意义. Ferraz等<sup>[13]</sup>提出了1种提取大范围路网的方法,基于图推理和道路点云特征的弱先验知识进行预处理,结合2种机器学习算法提取道路点云. Ferraz等<sup>[14]</sup>通过二维窗口化提取点云特征并采用随机森林分类,再经由OBIA(object-based image analysis)实现路网分割.

3) 在深度学习方面, Deng等<sup>[15]</sup>提出了1种基于超点的池化方法来融合2种不同模式的特征以进行联合学习. 通过联立点云及图像特征,经由

超点连接2D与3D特征,实现大规模道路场景点云分割. Ma等<sup>[16]</sup>提出了1种基于BoundaryNet的深度学习框架,通过移动激光扫描点云和高分影像来提取并细化道路边界. 张爱武等<sup>[17]</sup>提出了1种基于卷积神经网络并结合激光点云多特征的端到端语义分割方法,提高了道路及障碍物分割精度. 李健等<sup>[18]</sup>提出了1种基于多尺度球形邻域特征的深度神经网络算法,通过点云邻域特征组合选取实现地面点云语义分割.

上述研究多以城区及山区道路为研究对象,利用法向量、路缘石等信息提取道路点云,而对露天采场道路而言,法向量、路缘石等信息并不明显,利用上述方法难以取得较好的结果. 因此,本文以辽阳市千山石灰石矿点云数据为数据源,以单点RGB信息、邻域RGB信息、邻域高差、邻域粗糙度、反射强度作为点云约束特征,建立随机森林模型,实现多特征约束的露天采场道路点云实时有效准确提取.

## 1 研究区域与方法

### 1.1 研究区域概况

本研究以辽宁省辽阳市千山石灰石矿为研究区域,如图1所示. 利用大疆M300无人机搭载禅思L1三维激光扫描仪获取露天采场点云数据,如图2所示. 无人机飞行时航线高度为100 m,激光旁向重叠度为20%,可见光的旁向重叠度和航向重叠度分别为37%与70%,扫描获取的点云数据总点数为94 142 496.

为解决点云运算过程中出现的内存限制及效率较低问题,在满足道路点云提取精度要求下,利用降采样方法将点云数目降为2 000 000. 数据集制作环境为Microsoft Visual Studio 2019、PCL 1.12.0和CLOUD COMPARE 2.13 alpha,模型训练环境为PyCharm 2020.3和Python 3.9.

### 1.2 露天采场道路提取方法

本文结合露天采场道路的实际特征情况,基于无人机激光点云数据,提出多特征约束随机森林模型,以实现露天采场道路点云的有效提取,本文方法流程如图3所示.

## 2 数据集及模型训练

### 2.1 点云特征

针对露天采场道路点云色彩、形态及具备一

定噪点的特性,选用以下 5 类特征作为道路点云 提取的特征约束.

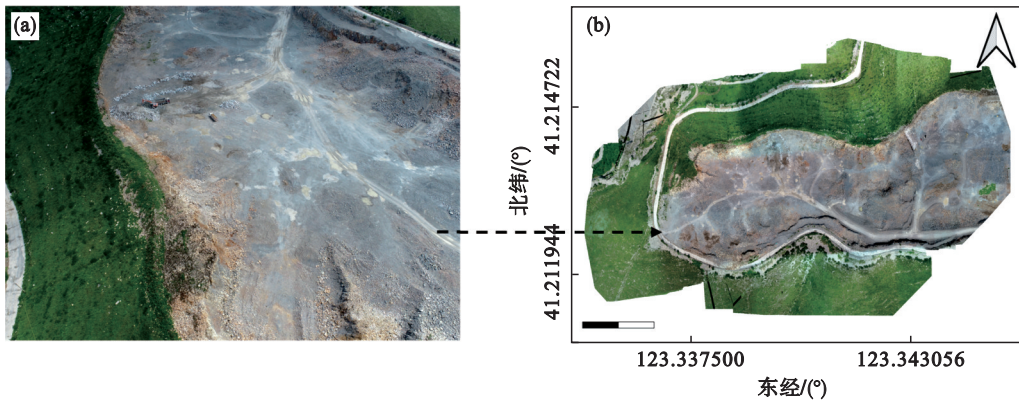


图 1 辽阳市千山石灰石矿  
Fig. 1 Qianshan limestone mine, Liaoyang City  
(a)—影像数据; (b)—点云数据.



图 2 大疆 M300 无人机及禅思 L1 三维激光扫描仪  
Fig. 2 DJI M300 drone and Zenith L1 three-dimensional laser scanner

1) 单点 RGB 信息. RGB 信息反映了目标的表面纹理信息,在露天采场中,车辆反复碾压导致道路与非道路的显著颜色差异,通过目视即可观察得到,因此采用单点 RGB 信息作为约束条件,用于随机森林模型的训练与验证.

2) 邻域粗糙度.道路点所在小范围邻域内点分布较平坦,极少出现凹凸不平现象,而非道路点邻域内点分布分散导致邻域粗糙度数值较大.计算邻域粗糙度时,首先利用 K 近邻法(K-nearest neighbor, KNN)获取检索点周围 20 个点,以此作为检索点邻域点集,再针对该点集采用随机抽样一致性算法(random sample consensus, RANSAC)拟合邻域点集平面,拟合后平面方程为

$$Ax + By + Cz + D = 0. \quad (1)$$

式中  $A, B, C, D$  均为平面参数且  $A, B, C$  三者不同时为 0.平面拟合后邻域内点距平面距离  $d_i$  为

$$d_i = \frac{Ax_i + By_i + Cz_i + D}{\sqrt{A^2 + B^2 + C^2}}. \quad (2)$$

式中  $x_i, y_i, z_i$  为检索点邻域内点的三维坐标.

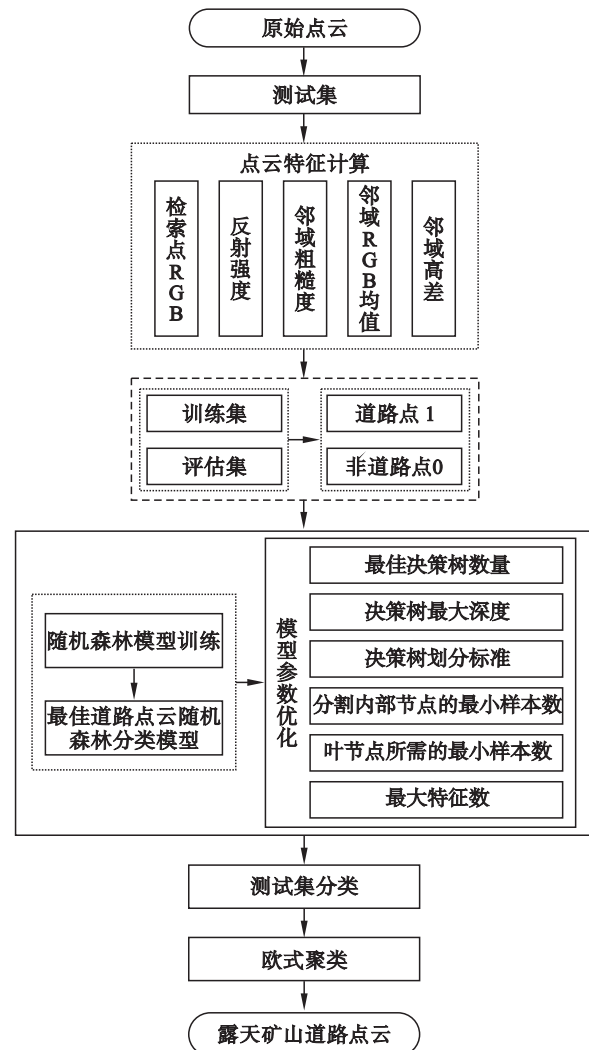


图 3 露天采场道路点云提取流程图  
Fig. 3 Flow chart for point cloud extraction of open-pit mine roadways

当前检索点粗糙度  $R_a$  即为

$$R_a = \frac{1}{20} \sum_{i=1}^{20} d_i. \quad (3)$$

3) 反射强度.在一定范围内,回波目标特性反映激光雷达反射强度大小,对于具有差异性的物体,反射强度相差较大.因此反射强度可作为点云目标分类的一项重要指数.

4) 邻域RGB信息.点云中的噪点致使误分类情况增多.道路点相比于非道路点,其纹理信息更加集中,以邻域RGB信息作为模型分类的主要特征,能有效降低因噪点带来的训练错误率,并提升道路与非道路间的识别度.通过邻域RGB信息对模型训练的约束,最终减弱模型的误分类情况.邻域RGB信息采取式(4)~(6)计算:

$$r_{\text{ave}} = \frac{1}{20} \sum_{i=1}^{20} R_i, \quad (4)$$

$$g_{\text{ave}} = \frac{1}{20} \sum_{i=1}^{20} G_i, \quad (5)$$

$$b_{\text{ave}} = \frac{1}{20} \sum_{i=1}^{20} B_i. \quad (6)$$

式中: $r_{\text{ave}}$ 、 $g_{\text{ave}}$ 、 $b_{\text{ave}}$ 分别为检索点邻域内点RGB信息均值; $R_i$ 、 $G_i$ 、 $B_i$ 为邻域内点RGB三通道数值.

5) 邻域高差.道路点邻域高差变化较小,而非道路点邻域高差由于露天采场地物起伏影响相对较大,因此将邻域高差作为道路分类提取的一项辅助特征.

## 2.2 数据集制作

首先手动标记露天采场点云数据,将样本标签分为道路点与非道路点,形成露天采场点云数据集,其中非道路点标签为0,道路点标签为1.然后将数据集划分为训练集和验证集.训练集用于模型训练及模型参数确定,训练集点数共计703 436,其中道路点点数为36 512,非道路点点数为666 924;验证集用于模型选择,不参与模型参数选取,当模型训练时,采取梯度下降法最小化损失函数,从而降低损失函数的值,而验证集不参与梯度下降过程.验证集点数为7 032,其中道路点点数为1 389,非道路点点数为5 643,训练集与验证集整体比例接近于99:1.数据集划分情况如图4所示,图中蓝色区域为道路点,红色区域为非道路点.

当训练集与验证集点云选取后对其赋予点云多特征时,为便于训练集内部道路点与非道路点间点云特征数值差异直观表示,采取特征图形式对多项特征进行可视化;针对道路点与非道路点点数不一致问题,对两类标签点进行点数归一化,归一化后横坐标即为排序后当前特征下点数占总点数的比例;纵坐标即为当前点数下的特

征值,如图5所示.

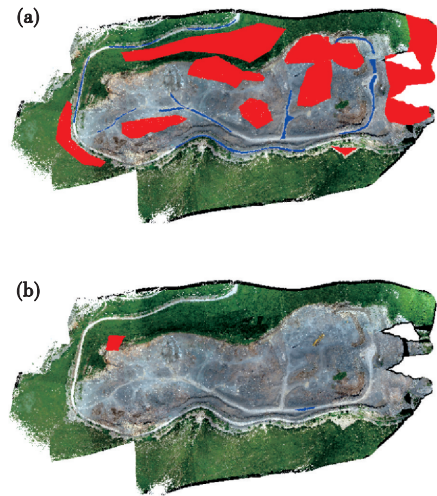


图4 数据集划分

Fig. 4 Data set partitioning

(a)—训练集;(b)—验证集.

道路点与非道路点间单点RGB信息及邻域RGB信息在不同点数量区间范围内差别较大;反射强度、邻域粗糙度及邻域高差在点数比例较小时差异较小,在点数比例较大时差异逐渐扩大.因此将单点RGB信息、邻域RGB信息、反射强度、邻域粗糙度、邻域高差这5类点云特征信息相结合,通过邻域RGB进行主要判别并利用单点RGB辅助约束,再结合反射强度、邻域粗糙度、邻域高差作为露天采场道路提取的重要特征.

## 2.3 模型构建及模型参数优化

随机森林算法是典型的集成学习Bagging算法<sup>[19]</sup>.随机森林算法的基本组成单元为决策树,每棵决策树代表1个分类器.针对分类问题,同一输入样本, $N$ 棵决策树将得到 $N$ 个分类结果.随机森林通过集成所有的分类结果,将获得票数最多的类别作为最终输出,随机森林Bagging思想如图6所示.

随机森林算法中的各决策树间没有相关性,因此可以通过并行计算提升算法运行效率.并行计算时由于特征维度过多、模型假设过于复杂等原因,导致自由生长的决策树易产生过拟合现象,但是由于Bagging思想使得随机森林对异常值不敏感,通过强大数定理即可证明随机森林的强抗干扰能力.随机森林能够有效运行在大数据集上,在处理具有高维特征的输入样本时无须对数据进行降维处理.

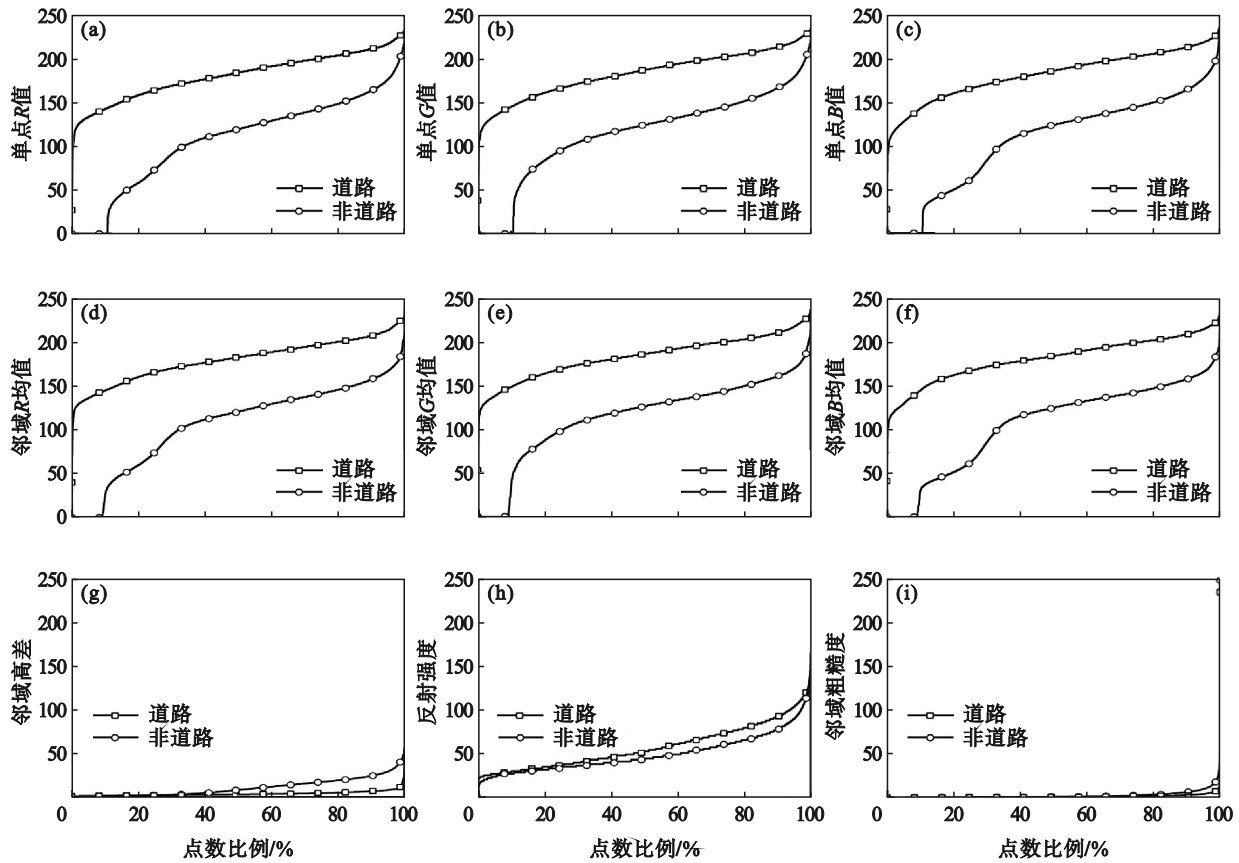


图5 两种类型点云特征值

Fig. 5 Two types of point cloud characteristic values

(a)—单点R值;(b)—单点G值;(c)—单点B值;(d)—邻域R均值;(e)—邻域G均值;(f)—邻域B均值;  
 (g)—邻域高差;(h)—反射强度;(i)—邻域粗糙度.

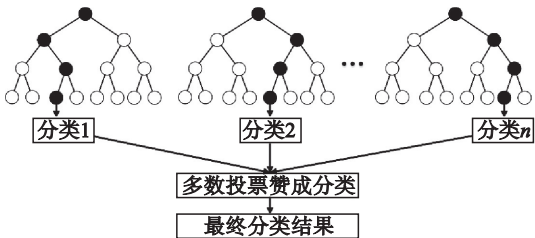


图6 随机森林 Bagging 思想

Fig. 6 Random forest Bagging idea

露天采场点云数据降采样后点数量以百万计,因此对露天采场道路点云提取而言,采用随机森林构建模型过程中,可以获取内部生成误差的一种无偏估计,保证输入特征在分类问题上的重要性以及处理缺省值的有效性.

随机森林算法中任意2棵树的相关性越大,错误率越高;每棵树的分类能力越强,整个森林的错误率也就越低,因此在构建随机森林模型时,模型各项参数会影响分类结果.在模型训练时需要优化参数,参数优化过程如图7所示.

随机森林模型参数优化主要包括以下内容:

1) 最佳决策树数量( $n_{estimators}$ ).决策树数

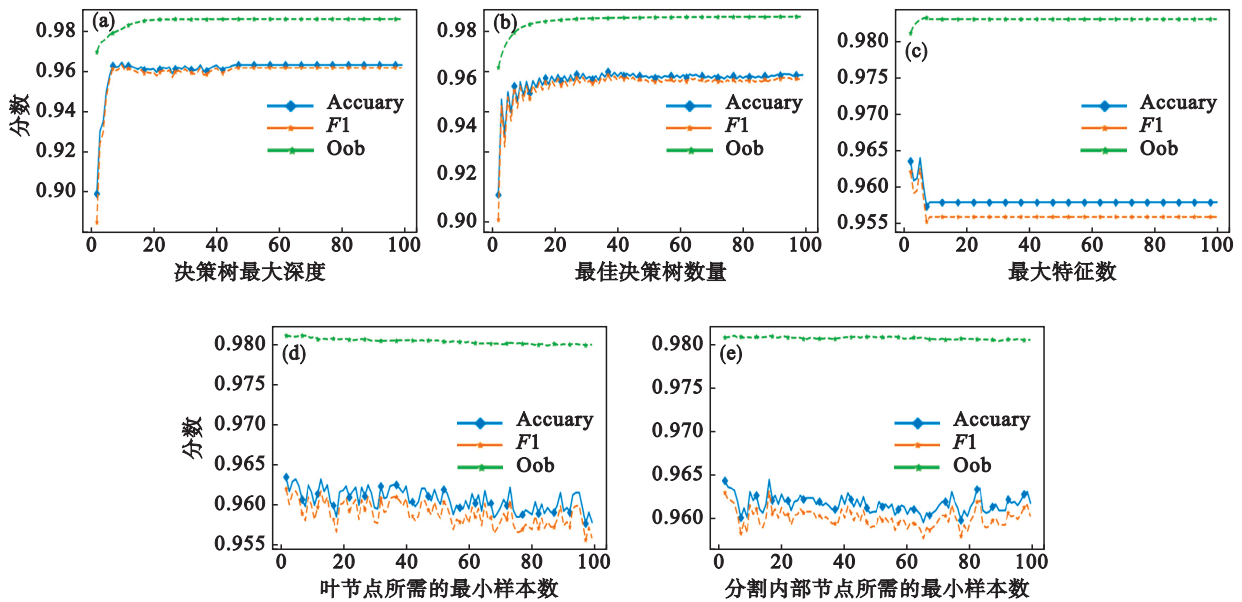
量决定了弱分类器的个数.决策树的数量越大,其精确度越高,但当其数量大于特定值时,所带来的提升效果非常有限,决策树数量上的冗余反倒降低模型训练效率.

2) 决策树最大深度(max\_depth).决策树最大深度保障了单个决策树在分类过程中工作的完整性,在深度到达指定阈值后样本便会停止分裂.

3) 决策树划分标准(criterion).决策树划分标准决定划分子树评估的标准.随着划分过程加深,分支节点所包含的样本应尽可能属于同一类别,为此采用基尼(Gini)参数防止出现过拟合现象及提升运算效率.

4) 叶节点所需的最小样本数(min\_samples\_leaf)、分割内部节点所需的最小样本数(min\_samples\_split).两项参数都从寻找节点所需包含最小样本数出发,但 min\_samples\_leaf 作用于叶子节点, min\_samples\_split 作用于内部节点.

5) 最大特征数(max\_features).最大特征数用于限制决策树分枝时考虑的特征个数,超过限制个数的特征都会被舍弃.



注: Oob(out of bag)是袋外样本.

图7 参数优化

Fig. 7 Parameter optimization

(a)—决策树最大深度; (b)—最佳决策树数量; (c)—最大特征数;  
 (d)—叶节点所需的最小样本数; (e)—分割内部节点所需的最小样本数.

在随机森林模型训练过程中,引入 Accuracy\_score 函数进行模型参数得分情况评估,通过统计类别标签被正确分类和错误分类的个数计算得分,评估随机森林模型精度.但

Accuracy\_score 函数仅用于衡量分类正确百分比,为了进一步评估召回率及模型的泛化能力,引入 F1\_score、Oob\_score 两种类型评估函数进行辅助评估.各个最优参数总体得分情况如表 1 所示.

表 1 最优参数总体得分  
 Table 1 Overall score of optimal parameters

参数名	最优参数	Accuracy	F1	Oob
决策树划分标准	Gini	0.962 773	0.961 037	0.981 782
最佳决策树数量	37	0.963 026	0.961 533	0.985 605
决策树最大深度	10	0.964 306	0.962 922	0.980 794
分割内部节点所需的最小样本数	16	0.964 448	0.963 075	0.980 963
叶节点所需的最小样本数	2	0.963 453	0.961 997	0.981 033
最大特征数	5	0.963 879	0.962 459	0.982 851

经优化后,模型最优参数如表 1 所示,利用上述参数对测试集进行分类,分类结果如图 8 所示,中央白色区域为道路点云,黑色区域为非道路点云.分类后的道路点云点数为 102 506,非道路点云点数为 1 897 494.由于道路点云两边存在矿车碾压路段以及部分颜色明亮区域,在分类结果中存在部分散乱点.为此引入欧式聚类算法剔除散乱点,去除后道路点云点数为 97 647.道路点云提取结果如图 9 所示.通过图 1 可以看出实验区道路上存在岩土堆叠,因此提取后道路在图 9a 中方

框范围所示处存在截断情况.



图8 露天采场点云分类结果

Fig. 8 Classification result of point cloud in open pit

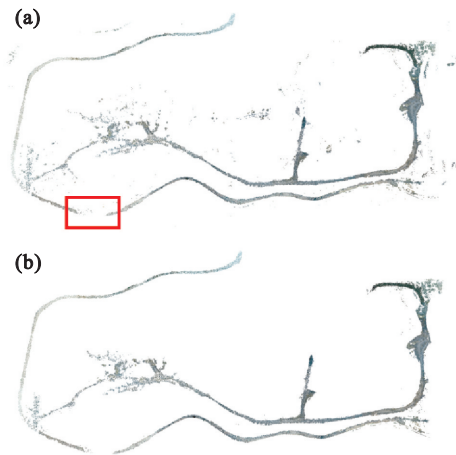


图9 测试集分类后道路点云

Fig. 9 Road point cloud after test set classification

(a)—聚类前;(b)—聚类后.

### 3 结果与分析

为验证露天采场道路点云提取准确性,本文引入国际摄影测量和遥感学会(ISPRS)提出的统计学误差评判标准<sup>[20]</sup>.本文将第一类误差,第二类误差,总误差定义为

$$\alpha = \frac{b}{a+b}, \quad (7)$$

$$\beta = \frac{c}{c+d}, \quad (8)$$

$$\theta = \frac{b+c}{a+b+c+d}. \quad (9)$$

式中: $a$ 为露天采场道路点正确分类个数; $b$ 为露天采场道路点被错误识别为非道路点个数; $c$ 为露天采场非道路点被识别为道路点个数; $d$ 为露天采场非道路点被正确分类的点数; $\alpha$ 为第一类误差; $\beta$ 为第二类误差; $\theta$ 为总误差.采用目视解译对露天采场道路点云进行提取,提取的道路点云点数为92 476,目视解译露天采场道路点云如图10所示.将提取结果与预测分类结果进行误差分析,总体提取分类误差如表2所示,道路点云判定误差如表3所示.



图10 目视解译露天采场道路点云

Fig. 10 Visual interpretation of road point cloud in open pit

表2 露天采场道路点云提取误差

Table 2 Extraction errors of road point cloud in open pit

类别	正确点个数	错误点个数	总数	误差/%
道路点	90 854	1 622	92 476	1.754
非道路点	1 902 353	6 793	1 909 146	0.356
总数	1 993 207	8 415	2 001 622	0.420

表3 露天采场道路点云判定误差

Table 3 Judgment errors of road point cloud in open pit

类别	道路判定误差/%
正确提取	93.043
错误提取	6.957

由表2及表3得出,通过本文提出的露天采场道路点云提取方法可以较好地提取道路点云数据.在道路点识别分类后引入欧式聚类算法,最终实现道路点云的提取误差及判定误差皆处于较小范围之内.

### 4 结 论

1) 提出了多特征约束下露天采场道路点云分类方法:将单点RGB信息、邻域RGB信息、邻域高差、邻域粗糙度、反射强度这5类点云特征相结合,可实时准确地提取露天采场道路.

2) 在随机森林模型6类参数优化后,基于该模型实现露天采场点云多特征约束下道路提取,并引入欧式聚类算法剔除非散乱点.经由ISPRS提出的统计学误差评判标准计算的第一类误差为1.754%,第二类误差为0.356%,总体误差为0.42%,露天采场道路点云提取后的三类误差皆处于较小范围之内.因此本文的研究成果对于提升露天采场的运输效率、经济效益和安全具有重要的现实意义.

#### 参考文献:

- [1] Tong X H, Liu X F, Chen P, et al. Integration of UAV-based photogrammetry and terrestrial laser scanning for the three-dimensional mapping and monitoring of open-pit mine areas[J]. *Remote Sensing*, 2015, 7(6): 6635-6662.
- [2] Ma L F, Li Y, Li J, et al. Mobile laser scanned point-clouds for road object detection and extraction: a review[J]. *Remote Sensing*, 2018, 10(10): 1531.
- [3] Huang W X, Liang H W, Lin L L, et al. A fast point cloud ground segmentation approach based on coarse-to-fine Markov random field[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(7): 7841-7854.
- [4] Mi X X, Yang B S, Dong Z, et al. Automated 3D road boundary extraction and vectorization using MLS point clouds[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(6): 5287-5297.

