

基于树木激光点云的有效特征抽取与识别方法

卢晓艺¹, 云挺^{1,2}, 薛联凤^{1*}, 徐强法¹, 曹林²

¹南京林业大学信息科学技术学院, 江苏 南京 210037;

²南京林业大学南方现代林业协同创新中心, 江苏 南京 210037

摘要 采用地面激光扫描获取树木的光探测和测距数据, 并将其作为遥感数据源, 选取水杉、棕榈、无患子、竹子和橡胶树为研究对象, 提出了三类有效特征: 树木相对聚类特征、点云分布特征和树木表观特征, 列举了 68 个特征参数。采用支持向量机在交叉验证中对训练数据集进行检验计算, 确定最优的特征参数组, 最终在测试数据集中进行树种分类。研究表明: 基于树木相对聚类特征的最优特征参数组进行树种分类的平均分类精度较低 (45%); 基于点云分布特征的最优特征参数组进行树种分类的平均分类精度有所增加 (58.8%); 基于树木表观特征的最优特征参数组进行树种分类的平均分类精度较高 (63.8%); 基于三类特征的 13 个最优特征参数进行树种分类的平均分类精度最高 (87.5%)。此外, 由于水杉与其他树种形态差异较为明显, 在分类中表现突出, 错判率最低 (6.5%)。所提方法具有较高的可行性, 为获得更准确的森林树种分布提供了强有力的工具。

关键词 遥感; 激光雷达; 树种识别; 支持向量机(SVM); 交叉验证; 组合特征参数

中图分类号 S771

文献标识码 A

doi: 10.3788/CJL201946.0510002

Effective Feature Extraction and Identification Method Based on Tree Laser Point Cloud

Lu Xiaoyi¹, Yun Ting¹, Xue Lianfeng^{1*}, Xu Qiangfa¹, Cao Lin²

¹College of Information Science and Technology, Nanjing Forestry University, Nanjing, Jiangsu 210037, China;

²Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing, Jiangsu 210037, China

Abstract Herein, light detection and ranging data were collected as remoting data sources by terrestrial laser scanning (TLS). Metasequoia, palm, sapindus, bamboo, and rubber trees were selected as research objects. Three effective features are proposed, which are relative clustering features of trees, features of point cloud distribution of trees, and apparent features of trees. 68 feature parameters are listed. A support vector machine (SVM) classifier was then used to verify and calculate the training dataset and to determine the optimal feature parameters in cross-validation. Finally, the tree species is classified in the test dataset. The research results show that the average classification accuracy of tree classification based on the optimal parameters of relative clustering features of trees is low (45%), that based on the optimal feature parameters of point cloud distribution slightly increases (58.8%), that based on the optimal parameters of tree appearance features is relatively high (63.8%), and that based on the 13 optimal parameters of three types of features is the highest (87.5%). In addition, due to the difference between metasequoia and other tree species is obvious, the metasequoia is outstanding in classification and its misjudgement rate is the lowest (6.5%). The proposed method has high feasibility and provides a powerful tool for obtaining a more accurate distribution of forest species.

Key words remote sensing; light detection and ranging (LiDAR); tree species classification; support vector machine (SVM); cross-validation; combination characteristic parameter

OCIS codes 280.3420; 280.3640

收稿日期: 2018-12-04; 修回日期: 2019-01-20; 录用日期: 2019-02-18

基金项目: 国家重点研发计划 (2017YFD0600900)、国家自然科学基金项目 (31770591)、中国博士后面上基金 (2016M601823)

* E-mail: xuelianfeng@njfu.edu.cn

1 引言

树种信息为森林可持续管理中的生态系统评估^[1]和生物多样性监测^[2]提供了基础数据,正确识别森林树种是利用和保护森林资源的前提和依据。传统的树种识别方法主要通过观察、比较和分析各树种的主要特征,然后由宏观特征到微观特征逐步对树种进行识别^[3]。现阶段,国内的树种识别技术大多依赖于一些高成本、费时和费力的野外实体调查或根据精度不高的航片进行人工判读^[4]。

遥感技术的进步使得树种识别在多种传感器捕获的数据中得以实现^[5]。目前,树种识别的数据源主要来自两个类别:光谱和激光雷达(LiDAR),由于数据形式的多样性、数据存储容量的庞大以及难以直观表示树种信息,面向光谱数据和点云数据的树种识别算法便应运而生。早期的树种识别研究使用的分类算法主要有最大似然监督分类算法(MLC)^[6]和无监督聚类算法(K-means、迭代自组织数据分析算法 ISODATA)^[7]。随着树种识别研究的不断演变和进步,基于无参数决策树的分类器和神经网络逐渐成为分类器的主流^[8]。由于硬件和软件计算能力的快速提高,现阶段的树种识别研究优选无参数的机器学习方法,如随机森林算法^[9]和支持向量机(SVM)算法^[10]。此外,适用于树种识别研究的分类算法还有参数判别分析方法^[11]、波谱角识别方法^[12]、贝叶斯分类算法^[13]、广义线性模型^[14]、光谱混合分析^[15]、模糊逻辑^[16]、Thresholding 算法^[17]。

在过去的40年中,虽然有关树种识别的研究还在不断增加,但由于树木分布的密集性、树种种植的混合性以及林木竞争造成的树冠交叉现象,导致树木结构特征的描述模糊不清,所以有关树种识别的研究仍然是一项具有挑战性的任务^[18]。一般而言,抽取的特征参数越丰富,就越容易表达一个树种的本质,但应用对象的全部特征进行分类又会造成信息冗余及误判。因此,最优特征参数集合的构建是树种识别的关键。然而,目前有关最优特征抽取的研究尚未成熟。此外,单株分离作为树种分类的第一步,对树种分类精度也具有影响。不可置否的是,现有算法的单木分割结果会受到树顶检测精度和树顶凸度的影响。在郁闭度较高的针阔混交林中单木树冠的大小不同,树冠之间交叠,也会对结果产生影响。因此,使用面向LiDAR数据进行精准的单木分割仍然具有挑战性。然而,本文的目标是基于激

光点云的树种分类研究,从单个遥感数据中获取树种信息,以期对激光测绘在林业中的应用提供参考。

面向上述问题,本文基于LiDAR点云数据的树种识别研究,提出了一个有效的基于地面激光扫描(TLS)的LiDAR数据进行树种识别的方法框架。具体而言,本文的主要研究目标如下:1)对基于TLS的LiDAR数据进行特征分析,实现多类别树种特征的初步抽取;2)采用SVM分类器在交叉验证中根据初步抽取的特征参数进行树种分类,对分类结果进行统计、分析和比较,从中抽取贡献率较大的特征参数构成最优特征参数集合;3)基于TLS的LiDAR数据通过SVM分类器在交叉验证中评估树种识别的准确性。

2 研究区概况及数据预处理

2.1 研究区概况

研究区选择江苏省南京市南京林业大学新庄校园的部分林业用地和中国海南省儋州市的橡胶实验基地。南京林业大学地理上属于华东地区,长江下游中部,气候类型属于亚热带季风气候,冬季温和少雨,夏季高温多雨。校园内的植物有91个科,245个属,603个种,其中正种546种,变型种12种,栽培变型种15种。优势树种主要有水杉、杂交马褂木、桂花、意杨、香樟、悬铃木、喜树、龙柏、垂柳等。选取校园内的数棵水杉、无患子、棕榈和竹子为实验对象。儋州市地理上属于海南岛西北部,濒临北部湾,气候类型属于热带湿润季风气候,处于东亚大陆季风气候的南缘,夏无酷暑,冬无严寒,阳光充足,雨量充沛。选取该地区橡胶实验基地的数棵橡胶树作为实验对象。

2.2 数据获取

将地面激光雷达(Leica C10,徕卡测量系统公司,瑞士)放置在距离目标树3m的不同侧面,并设置相同的扫描精度。扫描后,在Leica Cyclone软件中将不同角度下得到的扫描结果配准到同一坐标系中,继而得到目标树的全覆盖点云扫描数据。最终选取240株林木进行树种分类研究,这些树种包括水杉、棕榈、无患子、橡胶树和竹子。

2.3 数据预处理

点云数据滤波是数据处理过程中至关重要的一步,为了最大限度地保留点云数据的原始信息和提高滤波的可靠性,采用基于多尺度虚拟网格的改进滤波算法^[19]进行数据降噪,以减小树种分类结果的误差累计。为了能保持良好的树种特

征,继而达到理想的实验结果,本研究采用迭代 H-Minima 改进分水岭算法^[20]辅以人工修正进行单株分离,尽量减少分割过程中因林木郁闭度过高而造成的过分割和合并等问题,在南京林业大学

校园内样地进行单株分离操作的效果如图 1 所示。图 2 为水杉、棕榈、无患子、橡胶树和竹子这 5 种样本树的点云表示。表 1 列出了上述样本树在统计学中的基本结构特征。

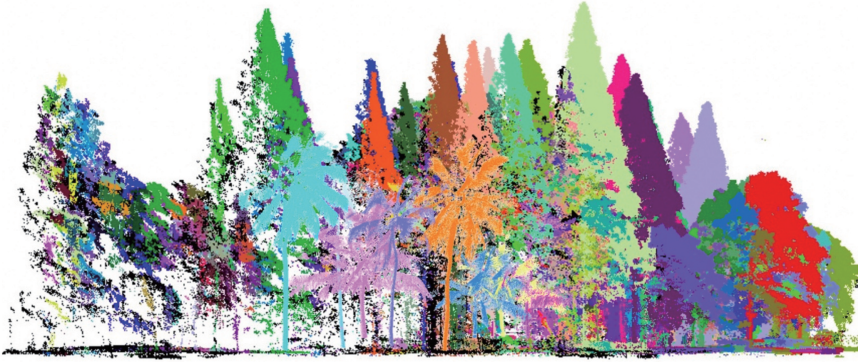


图 1 南京林业大学校园内样地单株分离效果图

Fig. 1 Delineation effect of individual tree in an example area of Nanjing Forestry University

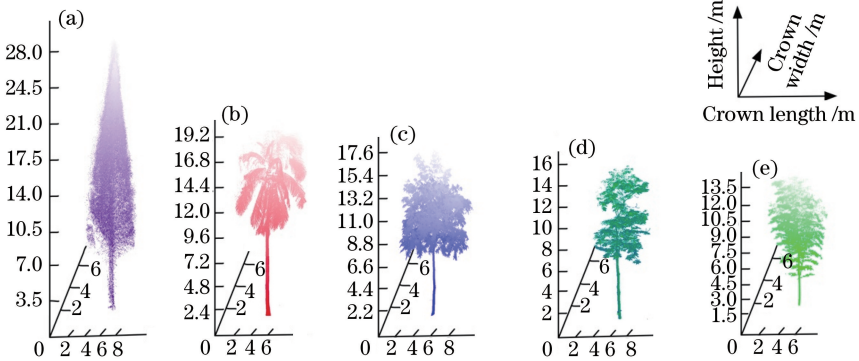


图 2 样例树的点云表示。(a)水杉;(b)棕榈;(c)无患子;(d)橡胶树;(e)竹子

Fig. 2 Scanned point clouds of sample trees. (a) Metasequoia; (b) palm; (c) spaindus; (d) rubber tree; (e) bamboo

表 1 样例树在统计学中的基本结构特征描述

Table 1 Characterization of basic structures of sample trees in statistics

Attribute	Metasequoia	Palm	Sapindus	Rubber tree	Bamboo
Tree height /m	28.8±4.1	18.4±1.8	16.2±2.3	15.6±2.9	13.7±3.6
Crown breadth /m	6.3±1.8	5.1±1.1	4.2±2.3	5.2±0.5	2.81±0.7
Crown volume /m ³	463.8±48.4	163.9±23.4	243.2±28.6	186.3±18.2	106.7±13.8
Number of points cloud	187556±9773	105539±4184	142317±6251	120843±5735	88496±3857
Number of trees	48	54	42	51	45

3 方法介绍

3.1 方法框架

为了能够提出更好地反映每种树木的私有特征,从而进行更有效的树种分类,首先从分析树木的特征开始。本研究主要提出了三种类别的树木特征:1)树木相对聚类特征;2)点云分布特征;3)树木表观特征。接下来,针对每种类别,通过分类器将其

特征参数的所有组合方式作为输入变量进行分类,比较其分类精度。在寻找树种分类最佳精度的同时保持参与分类的特征参数尽可能少的过程中,可以获得相关的特征参数,这些特征参数被定义为最优特征参数。然后,组合三种类别的最优特征参数,并将其作为变量输入到分类器中,以获得更高的分类精度。最后,评估树种分类的整体性能。具体的工作流程框架如图 3 所示。

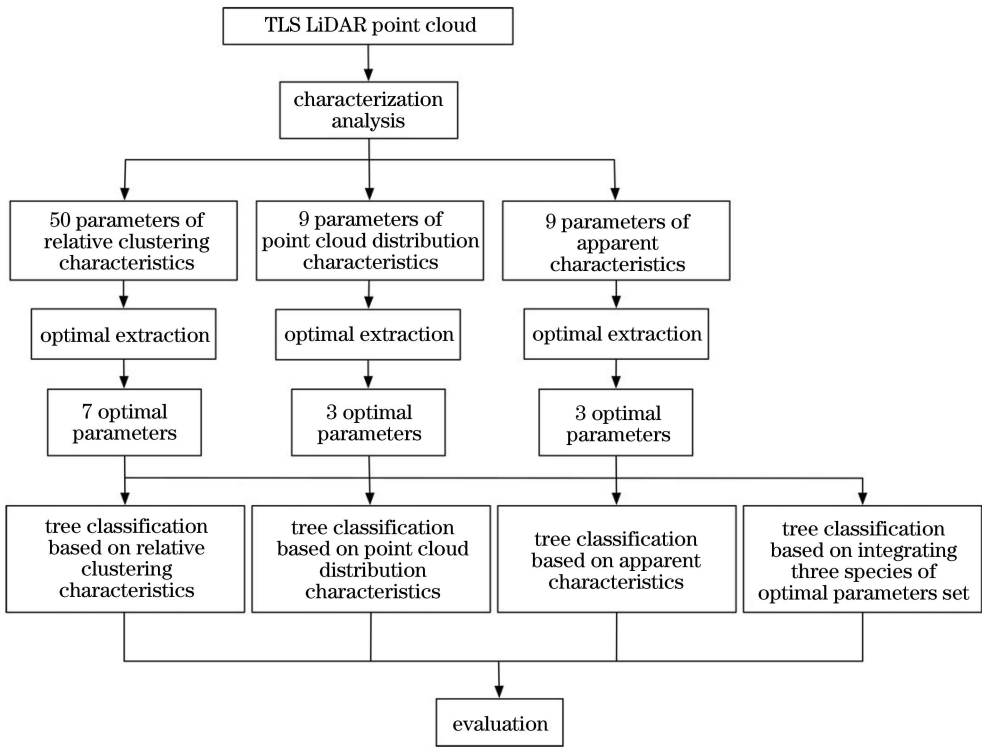


图 3 树种识别工作流程概述

Fig. 3 Overview of work flow of tree species identification

3.2 树木相对聚类特征

为了描述单株树的聚集离散结构,提出树木相对聚类这一特征。对于一株给定的树,采用文献[21]的方法对其点云进行计算,实现枝叶分离。将获取到的只有树叶点云的树冠自顶向下水平分割成10个相同高度的矩形层 R_k ,其中 k 为序号, $k=1, 2, \dots, 10$ 。取树高与冠底高度之差作为树冠高度 H_{crown} ,则 R_k 的高度为 $H_{\text{crown}}/10$ 。 R_k 的长度和宽度由所有扫描点在三维(3-D)坐标中 X 轴和 Y 轴的最大值与最小值确定,设坐标系下的任意扫描点坐标为 $p_i^j(x_i^j, y_i^j, z_i^j)$,其中 j 为样本树中的第 j 株树, i 为第 j 株树内的第 i 个扫描点,则 R_k 的长度和宽度可以分别表示为 $|\max(x_i^j) - \min(x_i^j)|$ 和 $|\max(y_i^j) - \min(y_i^j)|$ 。接下来,将 R_k 分割成 m 个边长为 $H_{\text{crown}}/10$ 的正方形区域样方,当 R_k 的长度和宽度不能整除 $H_{\text{crown}}/10$ 时,将长度和宽度扩充以满足整除要求。设 n_l^k 为 R_k 内第 l 个样方中的扫描点数量。在每个 R_k 内用特征值 V_k 描述对应 k 层点云的聚集离散结构, V_k 定义为

$$V_k = \frac{\left\{ m \sum_{l=1}^m (n_l^k)^2 - \left[\left(\sum_{l=1}^m n_l^k \right)^2 / m \right] \right\} / (m-1)}{\left(\sum_{l=1}^m n_l^k \right) / m}, \quad k = 1, 2, \dots, 10. \quad (1)$$

通过定量比较可以确定 R_k 的结构特征:如果 $V_k < 1$,则 R_k 的结构特征为分散;如果 $V_k > 1$,则 R_k 的结构特征为聚集;如果 $V_k = 1$,则 R_k 的结构特征是随机的。一般来说, V_k 值越大,点聚类越多。树冠自顶向下的每个矩形层的 V_k 特征值可以分别表示为 V_1, V_2, \dots, V_{10} (见表2)。由于不同树木之间矩形层的高度 $H_{\text{crown}}/10$ 不同,故而不同树木之间 V_k 特征值的比较存在尺度上的差异。因此,在具体的分类过程中,针对每株树,通过(2)式将 V_k 特征值作0~1之间的归一化处理,这一做法进一步解释了“相对”的概念,使得不同树木之间 V_k 特征值的比较是可行的。

$$V_{\text{new}_k} = \frac{V_k - \min(V)}{\max(V) - \min(V)}, \quad (2)$$

式中: $\min(V)$ 和 $\max(V)$ 分别为给定树的10个矩形层中对应的最小值与最大 V_k 值。通过 V_k 特征值的大小可以得知样本树中任意树冠在某一高度层的聚集离散结构。为了进一步探讨该层最大化聚集离散的相对尺度,提出了 L_k 函数这一概念,参见(5)式,具体过程如下:先定义 $P_k(r)$ 为关于距离度量 r 的函数,即

$$P_k(r) = A_k \sum_{i_2=1}^{n_k} \sum_{i_1=1}^{n_k} \frac{I_r(p_{i_1}^j, p_{i_2}^j)}{n_k^2}, \quad (3)$$

其中,

$$I_r(p_{i_1}^j, p_{i_2}^j) = \begin{cases} 1, & d_{p_{i_1}^j, p_{i_2}^j} \leq r \\ 0, & d_{p_{i_1}^j, p_{i_2}^j} > r \end{cases}, \quad (4)$$

式中: A_k 为 R_k 垂直于地面方向上的投影面积; n_k 为 R_k 内的扫描点个数; $I_r(p_{i_1}^j, p_{i_2}^j)$ 为指标函数; $d_{p_{i_1}^j, p_{i_2}^j}$ 为任意两点 $p_{i_1}^j$ 和 $p_{i_2}^j$ 之间的距离, 若 $d_{p_{i_1}^j, p_{i_2}^j} \leq r$, 则 $I_r(p_{i_1}^j, p_{i_2}^j) = 1$, 否则为 0。然后求取 $P_k(r)$ 函数的偏差:

$$L_k(r) = r - \sqrt{\frac{P_k(r)}{\pi}}, \quad (5)$$

根据对扫描点水平间距和样本树平均冠幅大小的估计, 将 r 的取值设置为在 0.2~6 m 区间内的 30 个数值(间隔为 0.2 m), 则单株树的每个矩形层包含 30 个 $L_k(r)$ 函数值, 可视为具有正峰值或负峰值的函数曲线。 $L_k(r)$ 的最大正值表示 R_k 在刻度 r 上存在最大程度的分散, $L_k(r)$ 的最小负值表示 R_k 在刻度 r 上存在最大程度的聚集。根据 $L_k(r)$ 函数计算出 4 个代表性数值: 1) $L_k(r)$ 函数的最大正值, 用 L_{\max_k} 表示; 2) $L_k(r)$ 函数的最小负值, 用 L_{\min_k} 表示; 3) $L_k(r)$ 函数最大正值处对应的距离度量 r , 用 r_{\max_k} 表示; 4) $L_k(r)$ 函数最小负值处对应的距离度量 r , 用 r_{\min_k} 表示。由于任意单株树都包含 10 个矩形层, 所以针对 $L_k(r)$ 函数的 4 个代表性数值, 任意单株树都将包含 40 个特征值, 分别表示为: $L_{\max_1}, L_{\max_2}, \dots, L_{\max_{10}}; L_{\min_1}, L_{\min_2}, \dots, L_{\min_{10}}; r_{\max_1}, r_{\max_2}, \dots, r_{\max_{10}}; r_{\min_1}, r_{\min_2}, \dots, r_{\min_{10}}$ (见表 2)。

表 2 树木相对聚类特征的相关参数列表

Table 2 Parameters associated with relative clustering characteristics of trees

Feature name	Denotation	Number of features
V	V_1, \dots, V_{10}	10
L_{\max}	$L_{\max_1}, \dots, L_{\max_{10}}$	10
L_{\min}	$L_{\min_1}, \dots, L_{\min_{10}}$	10
r_{\max}	$r_{\max_1}, \dots, r_{\max_{10}}$	10
r_{\min}	$r_{\min_1}, \dots, r_{\min_{10}}$	10

3.3 点云分布特征

虽然使用点云数据不能重建树木的详细结构, 但在统计意义上它可以反映出不同树种的结构特征。例如: TLS 点云数据的空间分布率较高, 因此其能有效反映树冠枝叶在不同层次的分布情况, 然而这种分布差异主要取决于树木种类的不同。本研究不仅提出了点高度类型的特征参数之外, 还提出了将位于不同高度分位点处的扫描点数量作为特征

参数。点云分布特征参数具体包括: 树高平均值, 其表达式为 $H_{\text{mean}} = \left(\sum_{i=1}^{n_{\text{tp}}} z_i^j \right) / n_{\text{tp}}$; 树高标准差, 其表达式为 $D_H = \left\{ \left[\sum_{i=1}^{n_{\text{tp}}} (z_i^j - H_{\text{mean}})^2 \right] / (n_{\text{tp}} - 1) \right\}^{1/2}$; 树高偏度, 其表达式为 $O_H = \left[\sum_{i=1}^{n_{\text{tp}}} \left(\frac{z_i^j - H_{\text{mean}}}{D_H} \right)^3 \right] / n_{\text{tp}}$; 树高峰度, 其表达式为 $P_H = \left\{ \left[\sum_{i=1}^{n_{\text{tp}}} \left(\frac{z_i^j - H_{\text{mean}}}{D_H} \right)^4 \right] / n_{\text{tp}} \right\} - 3$; 树高各五分之一分位点处的扫描点数量与目标树的扫描点总数之间的比率, 其表达式为 $R_{H_{\frac{h}{5} \sim \frac{h+1}{5}}} = n_{H_{\text{tree}} \times \frac{h}{5} \leq z_i^j \leq H_{\text{tree}} \times \frac{h+1}{5}} / n_{\text{tp}}, h = 0, 1, \dots, 4$ 。其中, n_{tp} 代表目标树的扫描点总数, H_{tree} 代表目标树的树高, $n_{H_{\text{tree}} \times \frac{h}{5} \leq z_i^j \leq H_{\text{tree}} \times \frac{h+1}{5}}$ 代表分布在树高 $\frac{h}{5}$ 与 $\frac{h+1}{5}$ 之间的扫描点数量。

3.4 树木表观特征

常见的树木表观特征包括树高 (H_{tree})、树冠高度 (H_{crown})、冠中最高分支高度 (H_{maxb})、冠中最低分支高度 (H_{minb})、冠盖最长分布 (L_{crown}) 以及与 L_{crown} 垂直的冠盖最长分布 (L_{crown}^v), 如图 4 所示。

除了常见的特征参数外, 还可以通过寻求更有效的特征参数, 或是组合两个通用特征参数的方法提出其他特征参数。为了寻求更有效的特征参数, 本课题组提出了如图 4(b) 所示的具有与冠盖相等面积的等效最小外接圆直径 D_{EA} , 其在一定程度上克服了冠形变化的影响。对于组合两个通用特征参数的策略, 通常用常见特征参数与树高的比率作为新的特征参数, 该度量可以减小树龄对树种分类的影响。树木表观特征参数包括: 等效冠直径, 其表达式为 $D_{\text{EA}} = 2 \times \text{sqrt}(A_{\text{base}})$; 等效冠直径与树高的比率, 其表达式为 $R_{D/H} = D_{\text{EA}} / H_{\text{tree}}$; 冠长与树高的比率, 其表达式 $R_{C/T} = H_{\text{crown}} / H_{\text{tree}}$; 冠长与等效冠直径的比率, 其表达式为 $R_{C/D} = H_{\text{crown}} / D_{\text{EA}}$; 树高和冠中最高分支高度之差与树高和冠中最低分支高度之差的比率, 其表达式为 $R_{T-\text{minb}/T-\text{maxb}} = (H_{\text{tree}} - H_{\text{minb}}) / (H_{\text{tree}} - H_{\text{maxb}})$; 冠盖的最长分布和与其垂直的冠盖最长分布的比率, 其表达式为 $R_{L_c/L_c^v} = L_{\text{crown}} / L_{\text{crown}}^v$; 树冠椭球模型的拟合参数, 其表达式为 $\eta_{\text{crown}} = \frac{a \times b}{c} \left| \left[\frac{(x_i^j)^2}{a^2} + \frac{(y_i^j)^2}{b^2} + \frac{(z_i^j)^2}{c^2} = 1 \right] \right|$; 网格的平均高度, 其表达式为 $H_{G_{\text{mean}}} =$

$$\sum_{t=1}^8 \left[\frac{\sum_{b_2=1}^g \sum_{b_1=1}^g \text{sum}(z_i^t)}{n_{C_{b_1, b_2}}} \right] / (8\delta_t^{\text{cover}}), i \in C_{b_1, b_2};$$

中心垂直剖面处的网格参数,其表达式为 $Q_L = \sum_{t=1}^8 (\delta_t^{\text{cover}} \times A_t^{\text{cover}}) / (8A_t^{\text{convex}})$ 。其中, A_{base} 表示树冠投影面积, a, b, c 分别表示椭球沿 X、Y、Z 轴的半径, $n_{C_{b_1, b_2}}$ 表

示在第 b_1 行第 b_2 列网格内扫描点的数量, δ_t^{cover} 表示在第 t 个中心垂直剖面内扫描点投影覆盖到的网格数量, A_t^{cover} 表示在第 t 个中心垂直剖面内扫描点投影覆盖到的网格面积, A_t^{convex} 表示在第 t 个中心垂直剖面内扫描点投影的凸包面积。

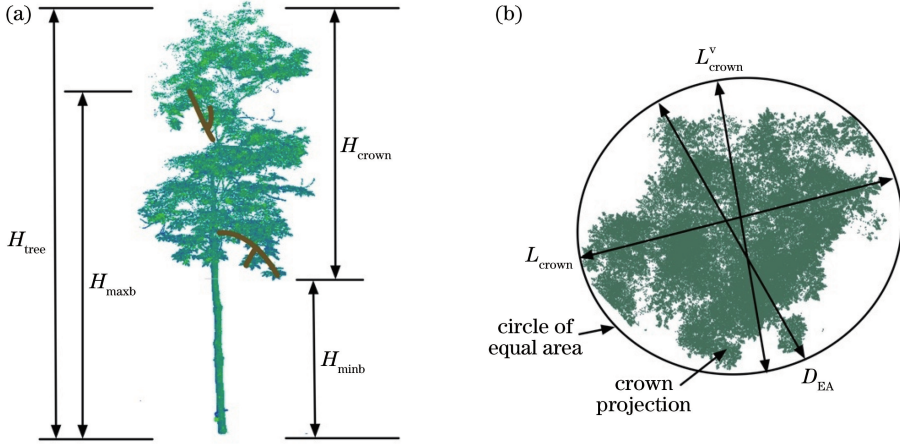


图 4 树木表观特征部分参数定义示意图。(a) 样例树(橡胶树)侧视图; (b) 样例树(橡胶树)俯视图
Fig. 4 Schematic of definition of partial parameters associated with apparent characteristics of trees.

(a) Side view of sample tree (rubber tree); (b) top view of sample tree (rubber tree)

在树木表观特征参数中, D_{EA} 的推导涉及到处理树冠投影到地面的二维散射点, 从最小外接圆的面积可以推导出相关的 D_{EA} 值。根据树冠叶子点云的正射投影, 可以从扫描点中获取 L_{crown} 、 L_{crown}^v 、 H_{maxb} 和 H_{minb} 。

影平面的网格划分, 划分过程具体如下: 首先以树干为中心垂直于地面方向切割树冠, 将其平均分成 8 段, 接下来将每个分段内的点云投影到各个分段的中心垂直剖面上, 如图 5(a) 所示, 最后将中心垂直剖面划分为 $g \times g$ 个网格, 用 C_{b_1, b_2} 代表第 b_1 行第 b_2 列的网格, 如图 5(b) 所示。

其他树木表观特征参数的获取主要来自点云投

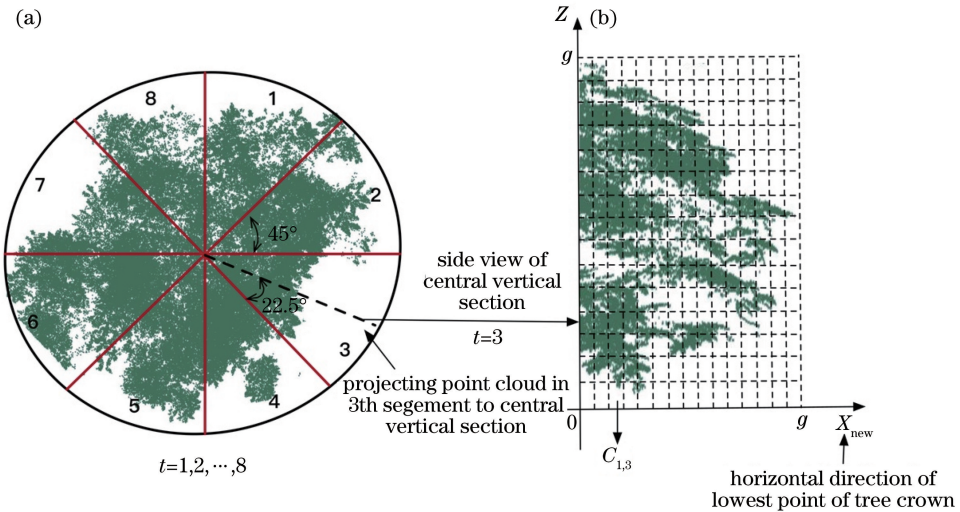


图 5 树木表观特征中与网格相关特征的求取示意图。(a) 样例树(橡胶树)8 分段划分与各分段的中心垂直剖面投影俯视图; (b) 中心垂直剖面网格化示意图
Fig. 5 Schematic of acquisition of features related to grids in apparent characteristics of trees. (a) Top view of sample tree (rubber tree) divided into 8 segments and central vertical section of each segment; (b) schematic for meshing central vertical section

3.5 分类方法

在本研究中,树种分类是基于 SVM 完成的,最优特征的抽取可以提高树种分类的效率和分类结果的准确性。因此,特征参数的抽取是树种分类的关键步骤。提出 68 个特征参数用于最优特征抽取,采用 SVM 分类器在交叉验证中进行样本树的分类。首先将每个类别的所有特征参数构成的组合作为输入变量进行分类,比较得到的分类精度,以确定基于该类别的最优特征参数组。然后整合三个类别的所有最优特征参数组,重复上述分类操作,获得整体意义上的最佳分类精度。

每种树木的样本树被随机分为两个子集,其中的一个子集作为训练数据集,用于特征抽取并训练个体,以获得最佳分类器;另一个子集作为测试数据集,用于评估最终树种分类的准确性。针对每个类别的样本树,随机抽取其样本数量的 1/3 作为训练

数据集,其余的样本树为测试数据集,训练数据集包含 16 株水杉、18 株棕榈、14 株无患子、17 株橡胶树以及 15 株竹子。

4 结果与分析

4.1 基于树木相对聚类特征的树种分类

图 6 显示了一株橡胶树和一棵水杉中两个矩形层的点云投影以及关于 V_k 特征值的坐标系描述。就单株树而言,橡胶树的 R_2 和 R_8 虽然含有相近的扫描点个数(R_2 内的扫描点数量为 28500, R_8 内的扫描点数量为 31900),但它们的点云分布却存在差异, R_8 相对 R_2 的点云分布更集中,其中 V_2 为 0.63 (<1),而 V_8 相对较高为 1.1(>1)。从图 6(b)可以明显地看出水杉的 R_2 相对 R_8 的点云分布更集中。通过图 6(c)进行验证,其中 V_2 为 4.07(>1), V_8 为 0.41(<1),验证结果与预估相符。

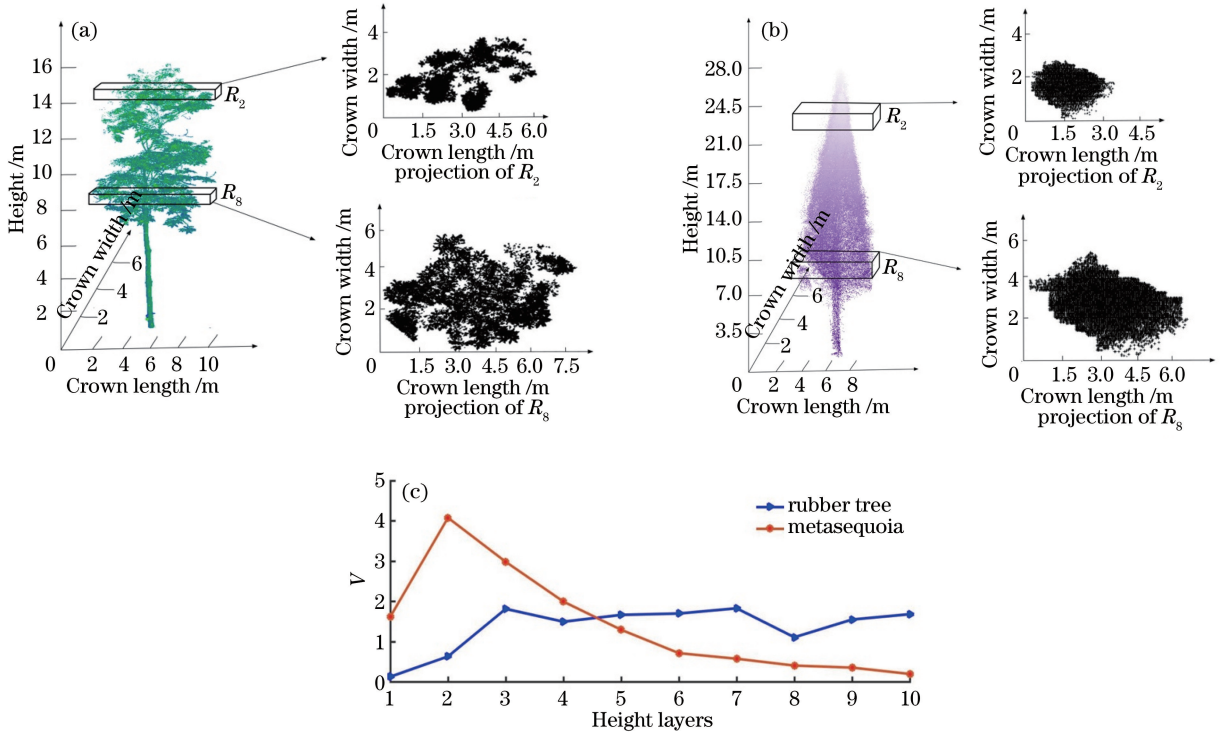


图 6 样例树(橡胶树和水杉)基于 V 特征的结果描述。(a)橡胶树的点云表示及其第 2 个和第 8 个矩形层的点云投影; (b)水杉的点云表示及其第 2 个和第 8 个矩形层的点云投影;(c)橡胶树和水杉各高度层的 V 特征折线图
Fig. 6 Sample trees (rubber tree and metasequoia) described based on results of V -feature. (a) Scanned point clouds of rubber tree and point cloud projections in the second and eighth rectangular layers; (b) scanned point clouds of metasequoia and point cloud projections in the second and eighth rectangular layers; (c) line chart indicating V -feature of each layer for rubber tree and metasequoia

图 7 显示了图 6 所示两株树内关于 L_2 函数和 L_8 函数(参见(5)式)的分布。从图 7(a)中可以看出:橡胶树的 R_2 在 $r=0.6$ m 时存在最大化的聚集,而水杉的 R_2 在 $r=1.2$ m 时存在最大化的聚

集,因此,在 R_2 中水杉相对橡胶树具有更大的点云聚集面积。从图 7(b)中可以看出:橡胶树的 R_8 在 $r=0.8$ m 时存在最大化的聚集,而水杉的 R_8 在 $r=2$ m 时存在最大化的聚集,因此,在 R_8 中水杉的点

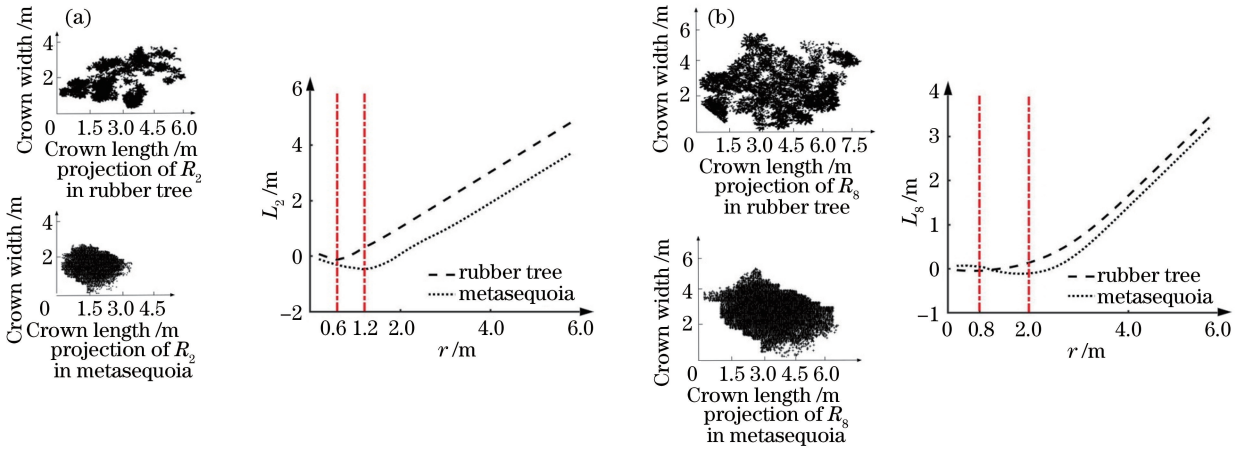


图 7 样例树(橡胶树和水杉)基于 L 特征的结果描述。(a)第 2 层的点云投影以及关于 $L_k(r)$ 函数的坐标系表示；(b)第 8 层的点云投影以及关于 $L_k(r)$ 函数的坐标系表示

Fig. 7 Sample trees (rubber tree and metasequoia) described based on results of L -feature. (a) Point cloud projections and coordinate system representation regarding to $L_k(r)$ function for rubber tree and metasequoia in the second rectangular layer; (b) point cloud projections and coordinate system representation regarding to $L_k(r)$ function for rubber tree and metasequoia in the eighth rectangular layer

云聚集面积仍相对较大。此外,当 r 取 6 m 时,橡胶树和水杉在 R_2 和 R_8 中都存在最大化的分散。

在分类过程中,为了降低程序复杂度和完善方法框架的模块化功能,将树木相对聚类特征(3.2 节)包含的 5 类特征参数(共 50 个特征值)按类别分别进行交叉组合,所有组合方式作为输入变量进行分类,分类结果的箱形图如图 8 所示,图 8 中所有的横坐标表示特征参数的组合数。

从图 8(a)可以看出,随着组合数增加,基于 V 特征参数组的分类性能得到了一定程度的改善,之后改善幅度逐渐趋于平稳,最终在组合数为 3、8、9 和 10 的条件下产生了较高的分类精度。因此,为了减少运算量,按照特征组合数为 3 选取 V_2 、 V_4 和 V_8 作为 V 特征参数组的最优特征参数。由图 8(b)~(c)可以看出,基于 L 特征参数组的分类一开始都表现出了较高的性能,出现这一现象的主要原因是

某一参数的显著性造成的。通过计算后发现在图 8(b)中仅使用 L_{\max_2} 参数就达到了 35.8% 的分类精度,在图 8(c)中仅使用 L_{\min_4} 参数同样达到了单个特征参数分类的最高精度 38.7%。因此,将 L_{\max_2} 和 L_{\min_4} 作为 L 特征参数组的最优特征参数。图 8(d)显示,由于 r_{\max} 特征参数组的平均分类精度较低,基本在 10%~20% 之间。因此,忽略 r_{\max} 特征参数组在树种分类中的影响。针对 r_{\min} 特征参数组,认为选取 2 个特征参数组合即可达到较高的分类精度 36.7%。经过对比,选取 r_{\min_2} 和 r_{\min_8} 作为 r_{\min} 特征参数组的最优特征参数。

综上所述,将 $\{V_2, V_4, V_8, L_{\max_2}, L_{\min_4}, r_{\min_2}, r_{\min_8}\}$ 作为基于树木相对聚类特征的最优特征参数组。表 3 列出了训练样本集(3.5 节)基于树木相对聚类最优特征参数组的交叉验证分类结果,可以看出,尽管 5 种树木的分类精度均为正值,但分类结果仍待改善。

表 3 基于树木相对聚类特征 $\{V_2, V_4, V_8, L_{\max_2}, L_{\min_4}, r_{\min_2}, r_{\min_8}\}$ 的交叉验证分类结果

Table 3 Classification results of cross-validation based on relative clustering features of $\{V_2, V_4, V_8, L_{\max_2}, L_{\min_4}, r_{\min_2}, r_{\min_8}\}$

Dataset	Metasequoia	Palm	Sapindus	Rubber tree	Bamboo	Total	Misjudgement rate / %	Leakage rate / %
Metasequoia	8	2	2	3	2	17	52.9	50
Palm	3	8	2	3	3	19	57.9	55.5
Sapindus	2	3	6	3	1	15	60	57.2
Rubber tree	2	3	3	7	1	16	56.3	58.8
Bamboo	1	2	1	1	7	12	41.7	53.3
Correct number	8	8	6	7	7	36		
Number of training samples	16	18	14	17	15	80		
Accuracy / %	50	44.5	42.8	41.2	46.7	45		

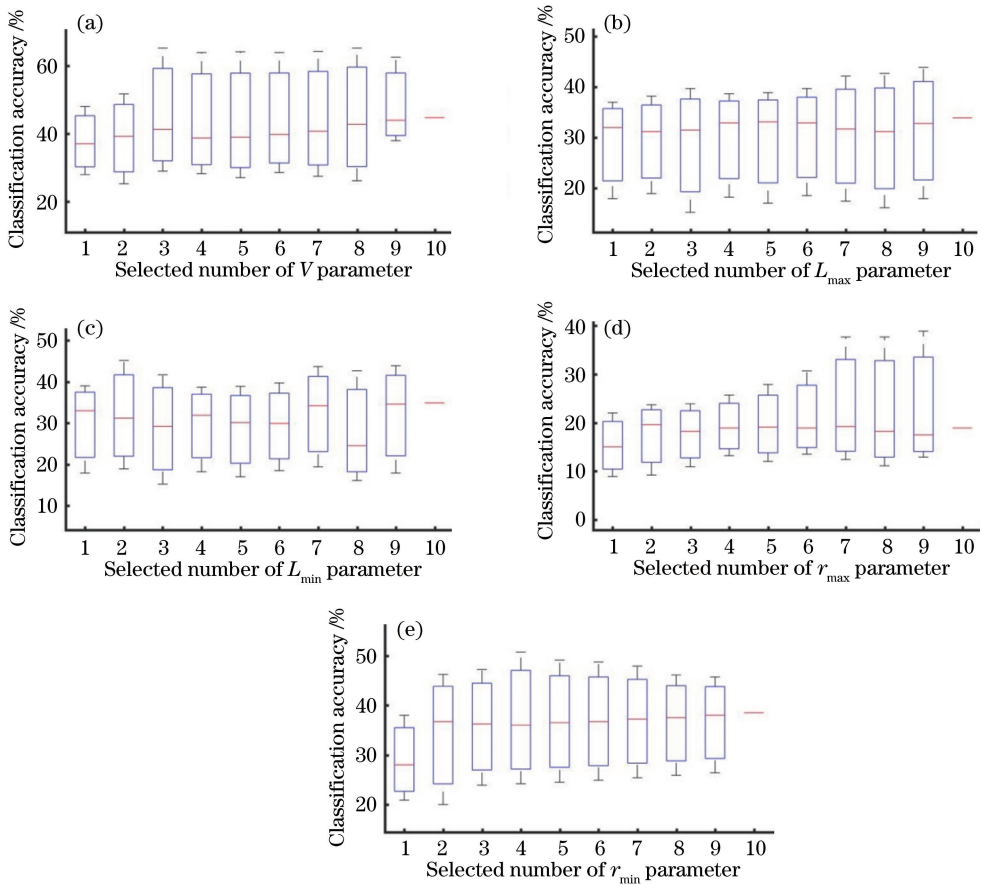


图 8 基于树木相对聚类特征的分类精度箱形图。(a)基于 V 特征；(b)基于 L_{max} 特征；
(c)基于 L_{min} 特征；(d)基于 r_{max} 特征；(e)基于 r_{min} 特征

Fig. 8 Boxplots of classification accuracies based on relative clustering characteristics. (a) Based on V -feature; (b) based on L_{max} -feature; (c) based on L_{min} -feature; (d) based on r_{max} -feature; (e) based on r_{min} -feature

4.2 基于点云分布特征的树种分类

将 9 个点云分布特征参数进行交叉组合后作为输入变量,分类结果的箱形图如图 9 所示。从图中可以看出,随着特征参数组合数增加,分类性能先是得到了一定程度的改善,而后改善情况逐渐趋于平稳,并且在组合数为 3 和 9 的情况下得到较高的分类精度(56.9%和 57.7%)。因此,通过比对后选取 P_H 、 $R_{H_{40\% \sim 60\%}}$ 和 $R_{H_{60\% \sim 80\%}}$ 这三个特征参数的组合作为基于点云分布特征的最优特征参数组。表 4 列出了训练样本集(3.5 节)基于点云分布最优特征参数组的交叉验证分类结果。由于水杉高大且树冠分布连续,竹子矮小且树冠分布不均,棕榈树高度适中但树冠分布密集,因此基于点云分布特征对水杉、竹子和棕榈的识别比较敏感,如表 4 所示。

4.3 基于树木表观特征的树种分类

将 9 个树木表观特征参数进行交叉组合后作为输入变量,分类结果的箱形图如图 10 所示。在这里,将 $R_{C/D}$ 、 η_{crown} 和 Q_L 这三个特征参数的组合确定

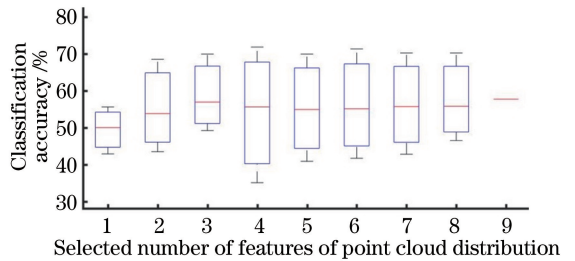


图 9 基于点云分布特征的分类精度箱形图
Fig. 9 Boxplots of classification accuracies based on features of point cloud distribution

为最优的特征参数组,其中, $R_{C/D}$ 和 η_{crown} 对冠形敏感, Q_L 可以反映枝叶是稠密还是稀疏。表 5 列出了训练样本集(3.5 节)基于树木表观最优特征参数组的交叉验证分类结果,由于水杉的树冠多为圆锥状,区别于其他树种的树冠形状,所以在本类特征分类中表现突出;而竹子的树冠和橡胶树的树冠相似度较高,并且林木枝叶密集程度相仿,因此在此特征分类中的区别较为模糊。

表 4 基于点云分布特征 $\{P_H, R_{H_{40\% \sim 60\%}}, R_{H_{60\% \sim 80\%}}\}$ 的交叉验证分类结果

Table 4 Classification results of cross-validation based on features of point cloud distribution of $\{P_H, R_{H_{40\% \sim 60\%}}, R_{H_{60\% \sim 80\%}}\}$

Dataset	Metasequoia	Palm	Sapindus	Rubber tree	Bamboo	Total	Misjudgement rate / %	Leakage rate / %
Metasequoia	10	1	1	1	0	13	23.1	37.5
Palm	1	11	1	2	1	16	31.3	38.9
Sapindus	2	2	8	4	2	18	55.6	42.9
Rubber tree	2	2	3	9	3	19	52.6	47.1
Bamboo	1	2	1	1	9	14	35.7	40
Correct number	10	11	8	9	9	47		
Number of training samples	16	18	14	17	15	80		
Accuracy / %	62.5	61.1	57.1	52.9	60	58.8		

表 5 基于树木表观特征 $\{R_{C/D}, \eta_{crown}, Q_L\}$ 的交叉验证分类结果

Table 5 Classification results of cross-validation based on apparent features of $\{R_{C/D}, \eta_{crown}, Q_L\}$

Dataset	Metasequoia	Palm	Sapindus	Rubber tree	Bamboo	Total	Misjudgement rate / %	Leakage rate / %
Metasequoia	12	0	1	1	1	15	20	25
Palm	1	11	3	1	1	17	35.3	38.9
Sapindus	1	2	9	2	1	15	40	42.9
Rubber tree	2	2	1	10	3	18	44.5	47.1
Bamboo	0	3	0	3	9	15	40	40
Correct number	12	11	9	10	9	51		
Number of training samples	16	18	14	17	15	80		
Accuracy / %	75	61.1	64.3	58.8	60	63.8		

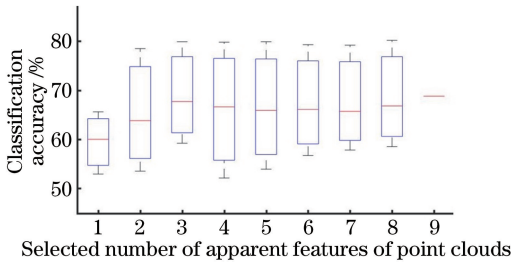


图 10 基于树木表观特征的分类精度箱形图

Fig. 10 Boxplots of classification accuracies based on apparent features

交叉验证分类结果,结果验证了整合三个类别最优特征参数组可以实现较高性能的树种分类。

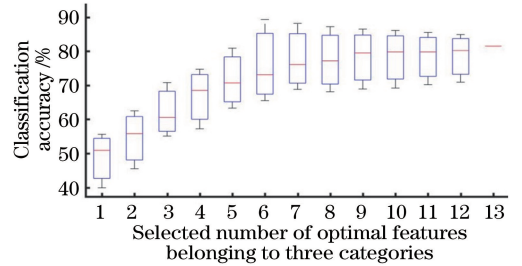


图 11 基于最优特征的分类精度箱形图

Fig. 11 Boxplots of classification accuracies based on optimal features

4.4 基于整合 3 个类别最优特征参数组的树种分类

将树木相对聚类特征、点云分布特征和树木表观特征的最优特征参数组共 13 个特征参数进行交叉组合后作为输入变量,分类结果的箱形图如图 11 所示。可以看出,随着组合最优特征参数数量增多,分类精度平均值呈平缓的上升趋势,当组合所有特征参数时,分类精度平均值达到最大值 87.5%。此外,以南京林业大学校园内样地为例,作基于整合三个类别最优特征参数组的树种分类,各树种的分类效果如图 12 所示。这里选取同一色系的四种颜色表示同种树木,以展现个体树的独立性。表 6 列出了测试样本集(3.5 节)基于整合三个类别最优特征参数组的

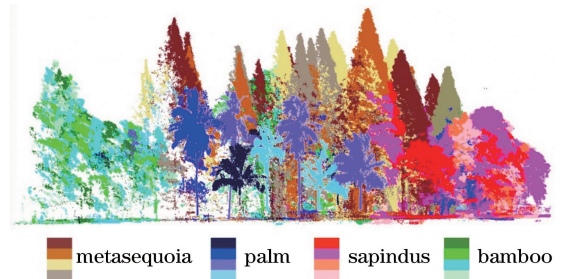


图 12 南京林业大学样地内各树种基于最优特征的分类图
Fig. 12 Classification map of individual tree species in an example area of Nanjing Forestry University based on optimal features

表 6 基于最优特征 $\{V_2, V_4, V_8, L_{max_2}, L_{min_4}, r_{min_2}, r_{min_8}, P_H, R_{H_{40\% \sim 60\%}}, R_{H_{60\% \sim 80\%}}, R_{C/D}, \eta_{crown}, Q_L\}$ 的交叉验证分类结果

Table 6 Classification results of cross-validation based on optimal features of

$$\{V_2, V_4, V_8, L_{max_2}, L_{min_4}, r_{min_2}, r_{min_8}, P_H, R_{H_{40\% \sim 60\%}}, R_{H_{60\% \sim 80\%}}, R_{C/D}, \eta_{crown}, Q_L\}$$

Dataset	Metasequoia	Palm	Sapindus	Rubber tree	Bamboo	Total	Misjudgement rate /%	Leakage rate /%
Metasequoia	29	2	0	0	0	31	6.5	9.4
Palm	2	31	2	1	0	36	13.9	13.9
Sapindus	0	2	24	1	1	28	14.3	14.3
Rubber tree	1	1	2	29	2	35	17.1	14.7
Bamboo	0	0	0	3	27	30	10	10
Correct number	29	31	24	29	27	140		
Number of training samples	32	36	28	34	30	160		
Accuracy /%	90.6	86.1	85.7	85.3	90	87.5		

最终的分类精度达到了 87.5%，但离完全精确还存在改进空间，因为分类效果受很多因素的影响，例如：树木空间竞争关系对树冠的影响，扫描时由于遮挡问题造成的数据缺失，同种树木的个体性差异以及树木生长环境的各异性等。为了减小这些因素对树种分类结果的影响，后续研究应尽量选取林分自然稀疏的样地进行多站多角度扫描，实现 LiDAR 数据和样地人工数据的多源融合，辅以更精确的单株分离算法，探索更详细的能够反映不同树种局部和整体结构特征的参数。

5 结 论

森林树种的正确识别是利用和保护森林资源的基础和依据，现阶段的树种分类结果尚未令人满意。本课题组试图开发一个全面而有效的方法框架，以检验基于 TLS 的点云数据对树种分类的效果。针对基于 TLS 的点云数据对水杉、棕榈、无患子、橡胶树和竹子进行树种分类的问题，提出三类特征（树木相对聚类特征、点云分布特征、树木表观特征），列举出 68 个特征参数供最优特征抽取，采用 SVM 分类器在交叉验证中对训练数据集进行检验计算，以确定最优的特征参数组，最终抽取 13 个最优特征参数在测试数据集中进行树种分类，达到了 87.5% 的分类精度，基本满足测试要求，进一步证明了该方法可为可持续森林管理和生态环境保护提供有力的辅助手段。

后续研究内容将侧重于大数据样本的建立、地基与机载激光全方位扫描以消除遮挡，以及结合更优的单株分离算法提高分类精度，为激光测绘在林业树种分类识别中的实际应用提供更坚实的理论基础。

参 考 文 献

[1] Shi Y F, Skidmore A K, Wang T J, *et al.* Tree

species classification using plant functional traits from LiDAR and hyperspectral data [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2018, 73: 207-219.

[2] Skidmore A K, Pettorelli N, Coops N C, *et al.* Environmental science: agree on biodiversity metrics to track from space [J]. *Nature*, 2015, 523(7561): 403-405.

[3] Bruggisser M, Roncat A, Schaepman M E, *et al.* Retrieval of higher order statistical moments from full-waveform LiDAR data for tree species classification [J]. *Remote Sensing of Environment*, 2017, 196: 28-41.

[4] Harrison D, Rivard B, Sánchez-Azofeifa A. Classification of tree species based on longwave hyperspectral data from leaves, a case study for a tropical dry forest [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2018, 66: 93-105.

[5] Fassnacht F E, Latifi H, Stereńczak K, *et al.* Review of studies on tree species classification from remotely sensed data [J]. *Remote Sensing of Environment*, 2016, 186: 64-87.

[6] Moore M M, Bauer M E. Classification of forest vegetation in north-central Minnesota using Landsat Multispectral Scanner and Thematic Mapper data [J]. *Forest Science*, 1990, 36(2): 330-342.

[7] Walsh S J. Coniferous tree species mapping using LANDSAT data [J]. *Remote Sensing of Environment*, 1980, 9(1): 11-26.

[8] Zhang C Y, Qiu F. Mapping individual tree species in an urban forest using airborne lidar data and hyperspectral imagery [J]. *Photogrammetric Engineering & Remote Sensing*, 2012, 78(10): 1079-1087.

[9] Richter R, Reu B, Wirth C, *et al.* The use of airborne hyperspectral data for tree species

- classification in a species-rich Central European forest area [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2016, 52: 464-474.
- [10] Pant P, Heikkinen V, Hovi A, *et al.* Evaluation of simulated bands in airborne optical sensors for tree species identification [J]. *Remote Sensing of Environment*, 2013, 138: 27-37.
- [11] Liu H J, Wu C S. Crown-level tree species classification from AISA hyperspectral imagery using an innovative pixel-weighting approach [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2018, 68: 298-307.
- [12] Cho M A, Debba P, Mathieu R, *et al.* Improving discrimination of savanna tree species through a multiple-endmember spectral angle mapper approach: canopy-level analysis [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2010: 48 (11): 4133-4142.
- [13] Pinacho-Pinacho C D, García-Varela M, Sereno-Uribe A L, *et al.* A hyper-diverse genus of Acanthocephalans revealed by tree-based and non-tree-based species delimitation methods: ten cryptic species of *Neoechinorhynchus* in Middle American freshwater fishes [J]. *Molecular Phylogenetics and Evolution*, 2018, 127: 30-45.
- [14] Åkerblom M, Raunonen P, Mäkipää R, *et al.* Automatic tree species recognition with quantitative structure models [J]. *Remote Sensing of Environment*, 2017, 191: 1-12.
- [15] do Amaral C H, de Almeida T I R, de Souza Filho C R, *et al.* Characterization of indicator tree species in neotropical environments and implications for geological mapping [J]. *Remote Sensing of Environment*, 2018, 216: 385-400.
- [16] Strnad D, Kohek Š, Kolmanič S. Fuzzy modelling of growth potential in forest development simulation [J]. *Ecological Informatics*, 2018, 48: 80-88.
- [17] Piironen R, Fassnacht F E, Heiskanen J, *et al.* Invasive tree species detection in the Eastern Arc Mountains biodiversity hotspot using one class classification [J]. *Remote Sensing of Environment*, 2018, 218: 119-131.
- [18] Heinzl J, Koch B. Corrigendum to “investigating multiple data sources for tree species classification in temperate forest and use for single tree delineation” [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2013, 21: 101-110.
- [19] Huang Z W, Liu F, Hu G W. Improved method for LiDAR point cloud data filtering based on hierarchical pseudo-grid [J]. *Acta Optica Sinica*, 2017, 37(8): 0828004.
黄作维, 刘峰, 胡光伟. 基于多尺度虚拟格网的 LiDAR 点云数据滤波改进方法 [J]. *光学学报*, 2017, 37(8): 0828004.
- [20] Teng W X, Wen X R, Wang N, *et al.* Individual tree crown extraction in high resolution remote sensing image based on iterative H-minima improved watershed algorithm [J]. *Laser & Optoelectronics Progress*, 2018, 55(12): 122802.
滕文秀, 温小荣, 王妮, 等. 基于迭代 H-minima 改进分水岭算法的高分辨率遥感影像单木树冠提取 [J]. *激光与光电子学进展*, 2018, 55(12): 122802.
- [21] Yun T, An F, Li W Z, *et al.* A novel approach for retrieving tree leaf area from ground-based LiDAR [J]. *Remote Sensing*, 2016, 8(11): 942.