

基于高光谱激光雷达的林木病虫害样本分类研究

卢京^{1,2}, 陈玖英^{1,2*}, 李伟¹, 周梅¹, 胡坚¹, 田汶鑫¹, 李传荣¹

¹中国科学院空天信息创新研究院中国科学院定量遥感信息技术重点实验室, 北京 100094;

²中国科学院大学光电学院, 北京 100049

摘要 本研究搭建一套 91 通道、光谱分辨率为 5 nm、波长范围为 650~1100 nm、生物安全性高的可调谐高光谱激光雷达系统, 并完成臭椿、云南松和栎树等林木样本的探测实验。通过实验探测目标回波强度, 进而获得目标光谱反射率, 最后利用支持向量机分类器对不同种类的健康与染病样本进行分类与识别, 臭椿样本的分类精度可以达到 96.98%、云南松样本的分类精度可以达到 91.21%、栎树样本的分类精度可以达到 66.21%。该实验结果具有研究意义和参考价值, 为林业病虫害监测提供一种新的发展方向。

关键词 遥感; 高光谱激光雷达; 林木病虫害; 支持向量机; 参数选取; 信号处理

中图分类号 TP721.2

文献标志码 A

doi: 10.3788/LOP202158.1628004

Research on Classification of Pest and Disease Tree Samples Based on Hyperspectral Lidar

Lu Jing^{1,2}, Chen Jiuying^{1,2*}, Li Wei¹, Zhou Mei¹, Hu Jian¹, Tian Wenxin¹, Li Chuanrong¹

¹Key Laboratory of Quantitative Remote Sensing Information Technology of CAS, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China;

²College of Optoelectronics, Chinese Academy of Sciences, Beijing 100049, China

Abstract In this study, a set of tunable hyperspectral lidar system with 91 channels, spectral resolution of 5 nm, wavelength range of 650–1100 nm, and high biological safety is built, and the detection experiments of forest tree samples such as *Ailanthus altissima*, *Pinus yunnanensis*, and *Koelreuteria paniculata* are completed. The target echo intensity is detected through experiments, and the target spectral reflectance is obtained. Finally, the support vector machine classifier is used to classify and identify different types of healthy and diseased samples. The classification accuracy of *Ailanthus altissima* samples can reach 96.98%. The classification accuracy of *Pinus yunnanensis* samples can reach 91.21%, and the classification accuracy of *Koelreuteria paniculata* samples can reach 66.21%. The experimental results have research significance and reference value, and provide a new development direction for forestry pest monitoring.

Key words remote sensing; hyperspectral lidar; tree pests and diseases; support vector machine; parameter selection; signal processing

OCIS codes 280.3640; 010.3640; 280.3420; 010.0280

1 引言

为了改善空气质量、防风固沙、减少水土流失等环境问题, 中国政府进行了大规模的植树造林活动。

臭椿、云南松和栎树作为中国常见的绿化树种, 自然而然地成为植树造林的首选。由于我国绿化造林面积巨大, 林业病虫害几乎是不可避免的问题, 而解决这一问题要花费巨大的人力物力。高光谱激光雷达

收稿日期: 2020-11-25; 修回日期: 2020-12-03; 录用日期: 2020-12-17

基金项目: 北京市科技计划(Z181100001018036)

通信作者: *chenjy@aircas.ac.cn

探测系统可以快速同步获得地物目标的测距信息和光谱信息,可应用于目标的探测与识别等。这为研究一种精准高效的林业病虫害监测方法提供了保障,因此开展基于高光谱激光雷达的林业病虫害监测方法的研究具有重要的意义。

自 20 世纪 60 年代第一台激光器出现以来,以测距功能为主的激光雷达便应运而生^[1],其可以进行无接触扫描获得目标的三维信息,目前已经在数字城市、生态环境监测、地质学、农业、考古、电力巡检和国防建设等诸多领域得到了广泛的应用与发展^[2-3]。基于激光雷达三维成像能力的目标分类研究也是不断深入,包括基于三维激光雷达点云的目标分割、空间物体三维建模与体积计算、树种分类等^[4-6]。高光谱激光雷达是在激光雷达的基础上发展而来的,不但具备无接触扫描获得目标的三维信息的能力,还有获得目标的光谱信息的能力^[7]。目前国内外基于高光谱激光雷达的林业病虫害光谱信息的分类研究还处于起步阶段,并未进行深入研究与发展。2007 年,芬兰地球空间研究所曾开展了关于高光谱激光雷达技术的研究,利用其进行基于测距、光谱、点云分类的林业方向研究^[8]。2010 年,武汉大学利用 4 波长多光谱激光雷达地面演示系统获取了大量地物的多光谱激光雷达地面实测数据,发现利用多光谱对地观测激光雷达系统所获取的激光光谱信息可实现对植被冠层进行有效监测^[9]。2016 年,中国科学院遥感与数字地球研究所利用研

制的全波形高光谱激光雷达系统获取带有目标光谱信息的点云,用于对目标结构信息和生化信息提取的相关研究^[10]。然而这些研究大多聚焦于目标相关参数和结构信息的提取,并未展开对目标光谱信息的深度挖掘,尤其是针对基于目标光谱信息的分类与识别的研究还有待探索。

本文利用新型可调谐高光谱激光雷达系统对多种林木样本进行实验,采集样本的高光谱激光雷达回波强度信号,对其进行处理得到样本的光谱反射率,最后运用支持向量机分类器对光谱反射率进行分类与识别研究,本研究探索了一种基于高光谱激光雷达的林木病虫害样本的分类方法,并对该方法在不同支持向量机参数模型下的分类识别效果进行了分析,该方法能在有限的光谱通道条件下对不同种类林木样本的健康或者染病进行分类识别。

2 实验系统与分类方法

2.1 新型可调谐高光谱激光雷达系统

新型可调谐高光谱激光雷达系统主要由发射系统、接收探测系统、扫描控制系统三个主要部分组成。发射系统主要由激光器和声光可调谐滤波器(AOTF)组成;接收探测系统主要包括接收光学系统、光电探测器和数据采集系统;扫描控制系统负责对发射光束进行空间位置和激光出射方向的控制,实现预定的空间扫描。图 1 为高光谱激光雷达原理,表 1 为高光谱激光雷达系统的主要参数。

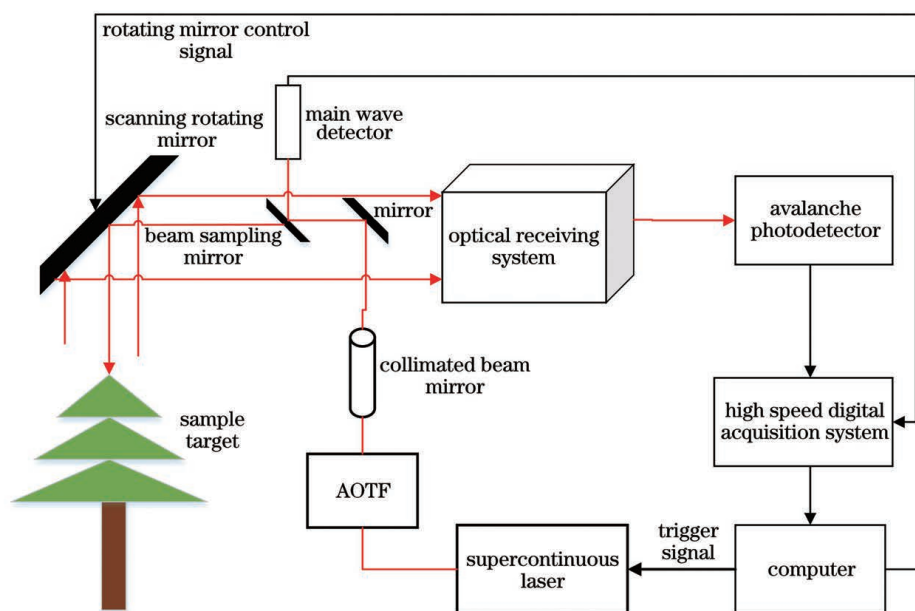


图 1 高光谱激光雷达原理示意图

Fig. 1 Schematic of hyperspectral lidar

表 1 高光谱激光雷达的主要参数

Table 1 Main parameters of hyperspectral lidar

Parameter	Value
Spectral range/nm	400-2400
Spectral resolution/nm	2-10
Beam divergence/mmrad	0.4
Beam diameter/mm	10

高光谱激光雷达系统首先将超连续激光光束注入 AOTF 中,经 AOTF 调谐滤波后按照设置要求在不同时刻发射不同波长的单色光,激光经准直扩散镜和反射镜出射,出射光束有小部分投射到光束采样镜后进入主波探测器中并传递到数据采集系统,大部分光束通过扫描转镜后投射到目标上,目标的回波信号由光学接收系统接收,再经光电探测器转化为电信号,由数据采集系统保存至上位机。对于采集到的数据首先进行预处理,即将采集到的样本回波强度信息转化为样本光谱反射率信息。为了减少外部环境因素的影响,保证同一次实验的数据之间可以横向比较,使用回波反射率对每个测量目标进行归一化。回波反射率由同一光谱通道的目标回波强度与参考板回波强度的归一化处理得到。待测目标的反射率可表示为

$$R_{\text{target}} = \frac{I_{\text{target}}}{I_{\text{reference}}} \times R_{\text{reference}}, \quad (1)$$

式中: I_{target} 和 $I_{\text{reference}}$ 分别为待测目标和参考靶标的激光回波强度; $R_{\text{reference}}$ 为参考靶标的反射率,参考靶标的反射率可通过地物光谱仪测量得到。起始波长为 650 nm,终止波长为 1100 nm,波长间隔为 5 nm,共 91 个光谱通道,参考靶标可以选用不同反射率的白色参考板。

2.2 分类方法

支持向量机是一种二类分类模型,其基本模型是定义在特征空间上间隔最大化的线性分类器,在分类问题中给定输入数据和学习目标 $x = \{x_1, x_2, x_3, \dots, x_n\}$ 和 $y = \{y_1, y_2, y_3, \dots, y_n\}$,其中输入数

据的每个样本都包含多个特征并由此构成特征空间,而学习目标为二元变量并分别表示负类和正类。若输入数据所在的特征空间存在作为决策边界的超平面,则将学习目标按正类和负类分开,并使任意样本的点 \mathbf{X} 到平面距离不小于最小距离,则称该分类问题具有线性可分性,超平面的法向量 ω 和截距 b 可表示为

$$\omega^T \mathbf{X} + b = 0, \quad (2)$$

$$y_i (\omega^T \mathbf{X} + b) \geq 1, \quad (i = 1, 2, 3, \dots, n). \quad (3)$$

但是支持向量机还是局限于线性可分的数据,当引入核函数之后,则使其成为实质上的非线性分类器^[11]。最早 Cortes 等^[12]提出了线性支持向量机,之后 Stephan^[13]通过引入核技巧提出了非线性支持向量机。支持向量机是一种常用的数据分类和回归的机器学习方法,该方法已成功应用于遥感监测领域。支持向量机的优点是鲁棒性强,对维数不敏感,对训练样本数量要求较低,而且支持向量机具有良好的泛化性能,具有较强的统计学习理论基础。

3 分类实验与数据分析

3.1 实验样本与样本光谱反射率

本实验采用的样本包括标准反射率白板、健康的臭椿样本、感染沟眶象的臭椿样本、健康的云南松样本、感染云南松切梢小蠹的云南松样本、健康的栎树样本、感染洁长棒长蠹的栎树样本等。植被、土壤和水体等物质是地球表面的主要组成部分,具有显著不同的光谱特征^[14]。色素吸收决定着可见光波段的光谱反射率,细胞结构决定近红外波段的光谱反射率^[15]。一般情况下,植被在 350~2500 nm 波长范围内具有典型的反射光谱特征,而且植被分析常使用红边区域和近红外区域的工作波段。因此,本实验的工作波长选定为 650~1100 nm,光谱通道数量为 91 个,光谱分辨率为 5 nm,在吸收率为 100% 的黑色参考板背景下对上述目标在固定距离处分别采集回波强度信号。实验采集的部分样本如图 2 所示。

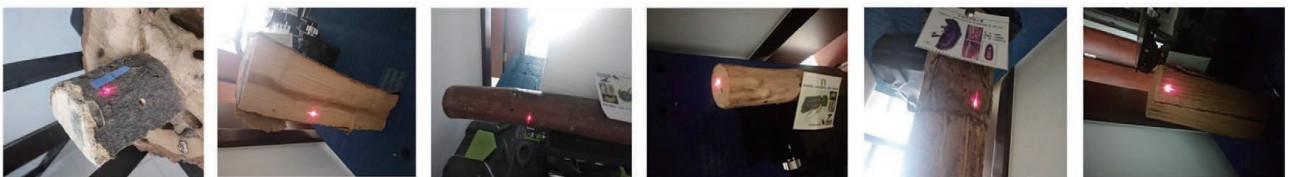


图 2 高光谱激光雷达数据采集的样本实例

Fig. 2 Sample example of hyperspectral lidar data acquisition

由于林木样本的生长环境和保存状况一直处于变化之中,而且木质样本易受大气介质等因素的影响,所以林木样本表面的成分随着时间推移而不断变化,因此激光回波强度受林木样本表面特性的影响很大。首先对健康林木样本表面与感染林木样本

表面的光谱反射率差异进行初步分析。图 3 为健康样本表面与染病样本表面的反射率趋势。从图 3 可以看到,除了健康臭椿与感染沟眶象臭椿的反射率在部分波长范围内反射率差异较大,其他几种林木样本健康与染病的反射率都十分接近。

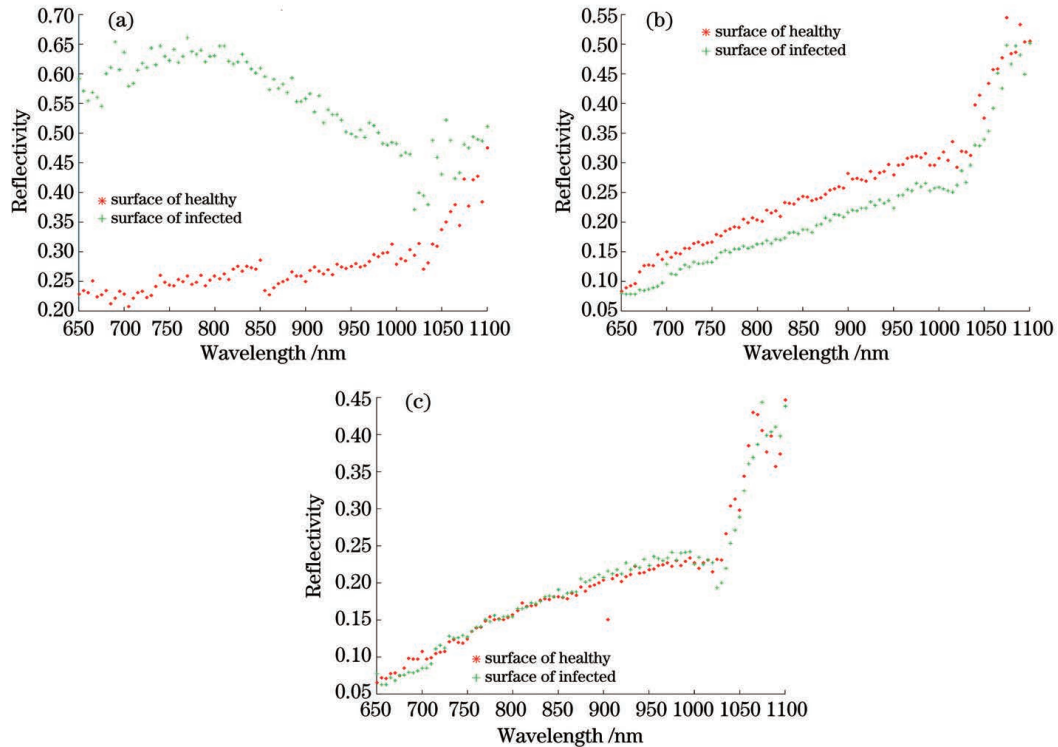


图 3 健康样本表面与染病样本表面的反射率。(a)臭椿;(b)云南松;(c)栾树

Fig. 3 Reflectivity of surface of healthy and infected samples. (a) *Ailanthus altissima*; (b) *Pinus yunnanensis*; (c) *Koelreuteria paniculata paniculata*

3.2 支持向量机分类器模型的构建及参数选择

由于林木样本的波长与反射率是非线性的,为此本文采用了非线性支持向量机分类器。支持向量机可以配置不同的核函数来处理各种非线性决策边界,目前常用的核函数有多项式核函数、高斯径向基核函数、线性核函数。因为高斯径向基核函数具有平滑、稳定、同步的特性,所以本研究采用的是高斯径向基核函数,即

$$K(\mathbf{x}, \mathbf{z}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{z}\|_2}{2\sigma^2}\right), \quad (4)$$

式中: \mathbf{x} 和 \mathbf{z} 为空间内不同的样本点; σ 为核半径。

关于实验数据处理,本文使用软件对回波信号进行处理、划分样本集、建模以及评价。

通过数据预处理得到了样本的光谱信息,之后将健康的臭椿样本与感染沟眶象的臭椿样本、健康的云南松样本与感染云南松切梢小蠹的云南松样本、健康的栾树样本与感染洁长棒长蠹的栾树样本

的表面反射率数据一分为二,一半作为支持向量机的训练集并用于构建支持向量机模型,另一半作为测试集并用来验证支持向量机分类器模型的效果。

得到三组不同的支持向量机模型之后,调节模型参数使模型性能更加优越。使用基于高斯径向基核函数的 SVM 模型有两个非常重要的参数 C 与 γ 。 C 为惩罚系数,即对误差的宽容度,若 C 值越大,则说明越不能容忍出现误差,容易过拟合;若 C 值越小,则容易欠拟合,综上 C 值过大或过小,都会影响模型的泛化能力。 γ 是选择高斯径向基函数作为核函数后,该函数自带的一个参数,其隐含地决定了数据映射到新的特征空间后的分布, γ 值越大,则支持向量越少,这会造成 SVM 只作用于支持向量样本附近,则训练集上的准确率很高,而测试集上的准确率较低,存在过拟合的风险; γ 值越小,则支持向量越多,则会造成平滑效应太大,无

法在训练集上得到特别高的准确率,也会影响测试集的准确率^[16]。此外高斯径向基核函数公式里面的 σ 和 γ 的关系为

$$K(\mathbf{x}, \mathbf{z}) = \exp \left[-\frac{d(\mathbf{x}, \mathbf{z})^2}{2\sigma^2} \right] = \exp[-\gamma \cdot d(\mathbf{x}, \mathbf{z})^2], \quad (5)$$

$$\gamma = \frac{1}{2\sigma^2}. \quad (6)$$

γ 与高斯径向基核函数的幅宽有关,其会影响每个支持向量所对应的高斯作用范围,从而影响泛化性能。根据(6)式已知 γ 与 σ 的关系,可以通过改

变 σ 值来影响 γ , 先根据经验对其惩罚因子 C 和核函数的固有参数进行初始化赋值,然后通过实验测试,再根据测试数据反复调整参数值,直到得到满意的结果为止。因为实验选用正类与负类的的数据数量一样多,所以对惩罚系数在 $0 \sim 64$ 之间进行选择。根据台湾大学林智仁教授的研究表明 σ 值通常可以在 $2^{-5} \sim 2^5$ 之间^[17], 所以对 σ 在 $2^{-5} \sim 2^5$ 之间进行选择。图 4 为不同 σ 值条件下健康林木样本与染病林木样本的分类精度趋势,图 5 为不同 C 值条件下健康林木样本与染病林木样本的分类精度趋势。

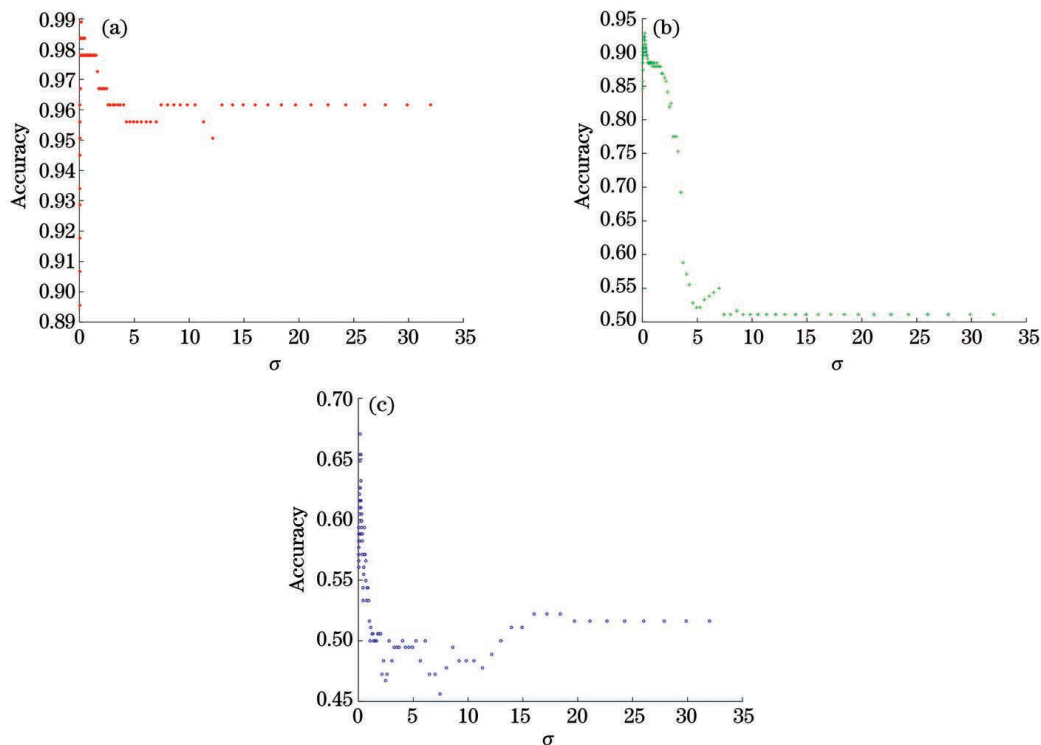


图 4 不同 σ 值条件下健康林木样本与染病林木样本的分类精度趋势。(a)臭椿;(b)云南松;(c)栾树

Fig. 4 Classification accuracy trend of healthy tree samples and infected tree samples under different σ values.

(a) *Ailanthus altissima*; (b) *Pinus yunnanensis*; (c) *Koelreuteria paniculata*

σ 的取值集中在 $0 \sim 1$ 之间, γ 的取值较大、支持向量减少,则展现出较高的分类精度;当 σ 增大时, γ 值减小、支持向量增多,则样本的分类精度呈现出明显的下降趋势。从图 4(a)可以看到,健康臭椿与感染沟眶象臭椿在 $\sigma = 2^{-2.7}$ 附近的分类精度可达 0.9890。从图 4(b)可以看到,健康云南松和感染切梢小蠹云南松在 $\sigma = 2^{-2.1}$ 附近的分类精度可达 0.9286。从图 4(c)可以看到,健康栾树与感染洁长棒长蠹栾树在 $\sigma = 2^{-2.5}$ 附近的分类精度可达 0.6703。对于惩罚系数 C 来说,随着 C 值的增加,分类精度迅速增加,当前取值超过一定范围之后分类器模型的性能将不会有较大提升。通过分析发

现,刚开始 C 值增加,模型的复杂度增加,支持向量减少;而当 C 值足够大之后,模型中边界支持向量的数量为 0, C 的变化就不会对模型的性能产生较大的影响。从图 5(a)可以看到,健康臭椿与感染沟眶象臭椿在参数 $C = 47$ 附近的分类精度可达 0.9945。从图 5(b)可以看到,健康云南松和感染切梢小蠹云南松在 $C = 18$ 附近的分类精度可达 0.9286。从图 5(c)可以看到健康栾树与感染洁长棒长蠹栾树在 $C = 48$ 附近的分类精度可达 0.6593。

3.3 测试集数据实验结果验证

完成支持向量机模型优化之后,就得到了适合识别不同种类林木样本健康或者染病的支持向量机

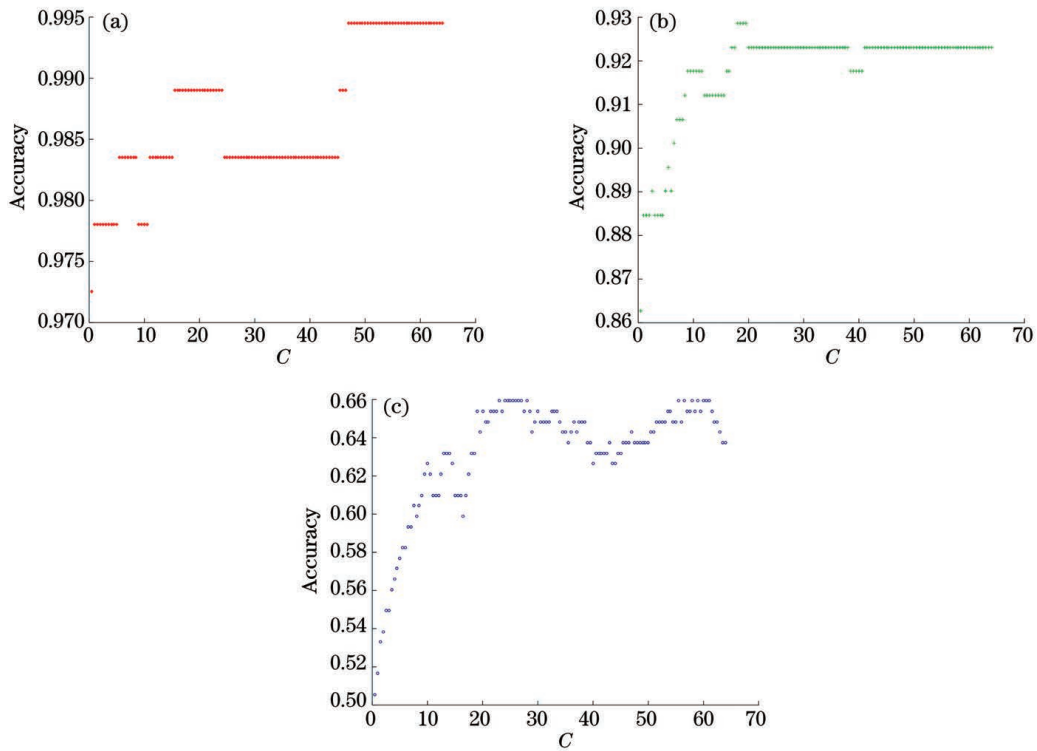


图 5 不同 C 值条件下健康林木样本与染病林木样本的分类精度趋势。(a)臭椿;(b)云南松;(c)栎树

Fig. 5 Classification accuracy trend of healthy tree samples and infected tree samples under different C values.

(a) *Ailanthus altissima*; (b) *Pinus yunnanensis*; (c) *Koelreuteria paniculata*

模型。将测试集数据输入支持向量机分类器模型中,用来验证其分类效果。表 2 为样本测试数据的分类精度。

表 2 测试集样本数据的分类精度

Table 2 Classification accuracy of test set sample data

Test set sample	Parameter	Classification accuracy
<i>Ailanthus altissima</i>	$\sigma = 2^{-2.7}, C = 47$	0.9698
<i>Pinus yunnanensis</i>	$\sigma = 2^{-2.1}, C = 18$	0.9121
<i>Koelreuteria paniculata</i>	$\sigma = 2^{-2.5}, C = 48$	0.6621

通过测试集数据的验证,经过参数优化之后的支持向量机模型可以很好地分类识别各种林木样本目标,而且都达到了预期效果,并未出现过拟合的现象。从表 2 可以看到,臭椿和云南松的分类精度较高,可以达到 90% 以上;栎树的分类精度也接近 70%。分类精度较高主要是因为其在 650 ~ 1100 nm 整个波长范围内的反射率差异显著,除了部分波段有混叠现象外,其他部分差异十分明显。而栎树在 650 ~ 1100 nm 整个波长范围内的反射率非常接近,这也是其分类精度较低的主要原因。综合来看支持向量机对于健康与染病的多种林木样本,分类效果明显,原因在于新型可调谐的高光谱激

光雷达的 91 通道光谱信息为分类提供了丰富的光谱信息,而且具有主动测量、占地面积小、抗干扰能力强等优点,这就使测量精度得到保证。

4 结 论

本研究利用自研的新型可调谐高光谱激光雷达系统,对固定距离下不同种类的健康林木样本和染病林木样本进行了分类研究,通过分析健康林木样本表面光谱与横截面光谱的关系证实表面光谱用于分类的可行性;运用支持向量机对健康和染病的林木样本的表面光谱反射率展开了关于核函数选取、模型参数选择、分类精度的相关研究;将实验获得的数据输入到优化之后的支持向量机模型中并进行验证,发现其对实验样本有良好的分类效果。其中健康臭椿和感染沟眶象臭椿在参数 $\sigma = 2^{-2.7}, C = 47$ 附近的分类精度可达 96.98%;健康云南松和感染切梢小蠹云南松在参数 $\sigma = 2^{-2.1}, C = 18$ 附近的分类精度可达 91.21%;健康栎树与感染洁长棒长蠹栎树在参数 $\sigma = 2^{-2.5}, C = 48$ 附近的分类精度 66.21%。虽然栎树的分类精度没有臭椿和云南松高,但在日后的研究中可以通过对模型的调整优化来进一步提升分类精度。本文在有限的光谱通道条

件下对健康和染病的林木样本表面光谱反射率进行分类识别,探索了一种适用于基于可调谐高光谱激光雷达光谱的病虫草害林木样本分类方法,目前只获得了初步结果,但具有一定的参考价值。本文的研究并未关注林木样本的染病程度,只是定性地分析了林木样本的健康染病与否,日后可以展开林木样本染病程度的相关研究,实现更加精确的监测。

参 考 文 献

- [1] Dai Y J. Principles of lidar [M]. Beijing: National Defense Industry Press, 2002.
戴永江. 激光雷达原理[M]. 北京: 国防工业出版社, 2002.
- [2] Du M. Developing lidar[J]. Modern Weaponry, 2001 (3): 32-34.
杜木. 发展中的激光雷达 [J]. 现代兵器, 2001(3): 32-34.
- [3] Sun M F. A Review of the development of lidar for earth observation [J]. Electro-optical Systems, 2009 (2): 1-6.
孙明甫. 对地观测激光雷达发展综述[J]. 光电系统, 2009 (2): 1-6.
- [4] Fan X H, Xu G L, Li W L, et al. Target segmentation method for three-dimensional LiDAR point cloud based on depth image[J]. Chinese Journal of Lasers, 2019, 46(7): 0710002.
范小辉, 许国良, 李万林, 等. 基于深度图的三维激光雷达点云目标分割方法[J]. 中国激光, 2019, 46 (7): 0710002.
- [5] Hu Y W, Wang J J, Fan Y Y, et al. LiDAR-based three-dimensional modeling and volume calculation for space objects [J]. Chinese Journal of Lasers, 2020, 47(5): 0510001.
胡燕威, 王建军, 范媛媛, 等. 基于激光雷达的空间物体三维建模与体积计算[J]. 中国激光, 2020, 47 (5): 0510001.
- [6] Chen X Y, Yun T, Xue L F, et al. Classification of tree species based on LiDAR point cloud data [J]. Laser & Optoelectronics Progress, 2019, 56(12): 122801.
陈向宇, 云挺, 薛联凤, 等. 基于激光雷达点云数据的树种分类[J]. 激光与光电子学进展, 2019, 56 (12): 122801.
- [7] Liu Q W, Li Z Y, Chen E X, et al. Waveform characteristics analysis of forest canopy detection lidar [J]. Scientia Sinica (Terrae), 2011, 41(11): 1670-1678.
刘清旺, 李增元, 陈尔学, 等. 森林冠层探测激光雷达的波形特征分析[J]. 中国科学: 地球科学, 2011, 41(11): 1670-1678.
- [8] Chen Y W, Räikkönen E, Kaasalainen S, et al. Two-channel hyperspectral LiDAR with a supercontinuum laser source[J]. Sensors, 2010, 10(7): 7057-7066.
- [9] Shi S, Gong W, Zhu B, et al. A novel multi-spectral LiDAR for earth observation and implementation of control system [J]. Geomatics and Information Science of Wuhan University, 2013, 38(11): 1294-1297.
史硕, 龚威, 祝波, 等. 新型对地观测多光谱激光雷达及其控制实现[J]. 武汉大学学报·信息科学版, 2013, 38(11): 1294-1297.
- [10] Wang Z, Chen Y, Li C, et al. A Hyperspectral LiDAR with Eight Channels Covering from VIS to SWIR [C] // IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, July 22-27, 2018, Valencia, Spain. New York: IEEE Press, 2018: 18244005.
- [11] Li H. Statistical learning methods [M]. Beijing: Tsinghua University Press, 2012.
李航. 统计学习方法[M]. 北京: 清华大学出版社, 2012.
- [12] Cortes C, Vapnik V. Support-vector networks [J]. Machine Learning, 1995, 20(3): 273-297.
- [13] Stephan R S. The nature of statistical learning theory [J]. Technometrics, 1997, 38(4): 409-409.
- [14] Li C R. Uav remote sensing load comprehensive verification system technology [M]. Beijing: Science Press, 2014.
李传荣. 无人机遥感载荷综合验证系统技术[M]. 北京: 科学出版社, 2014.
- [15] Feng W, Guo T C, Xie Y X, et al. Spectrum analytical technique and its applications for the crop growth detection [J]. Chinese Agricultural Science Bulletin, 2009(23): 192-198.
冯伟, 郭天财, 谢迎新, 等. 作物光谱分析技术及其在生长监测中的应用[J]. 中国农学通报, 2009(23): 192-198.
- [16] Fu Y, Li K L. A survey of model parameters selection method for support vector machines [J]. Computer Knowledge and Technology, 2010, 6(28): 8081-8082, 8085.
付阳, 李昆仑. 支持向量机模型参数选择方法综述 [J]. 电脑知识与技术, 2010, 6(28): 8081-8082, 8085.
- [17] Lin C J. Classification of multiclass data by support vector method (1/2) [M]. Taipei: Department of Information Engineering, National Taiwan University, 2002.
林智仁. 支撐向量法解多種類資料之分類(1/2) [M]. 臺北市: 國立臺灣大學資訊工程學系暨研究所, 2002.