

文章编号: 1672-8785(2016)02-0036-06

广东省主要人工林树种光谱分类

李 丹^{1,2} 杨 龙^{1,2} 王重洋^{1,2} 陈水森^{1,2*}

(1. 广东省地理空间信息技术与应用公共实验室, 广东广州, 510070 ;

2. 广州地理研究所, 广东广州, 510070)

摘 要: 为了探索不同树种叶片光谱区分的最佳特征波段以及不同树种光谱分类的性能, 利用地物光谱仪对广东省 10 个主要人工林树种的叶片光谱数进行了采集。采用遗传算法 (Genetic Algorithm, GA) 和连续投影变换算法 (Successive Projections Algorithm, SPA) 进行了高光谱数据降维处理, 然后结合支持向量机 (Support Vector Machine, SVM) 和随机森林 (Random Forest, RF) 两种方法进行了树种分类。研究发现, 通过两种变量选择方式筛选的用于树种分类的光谱范围主要位于近红外波段。其中, 经 GA 算法筛选的变量建模和预测精度与基于全波段光谱数据的分类精度较为接近, 且比经 SPA 算法变量筛选的分类结果好。通过比较可知, RF 算法所建模型的性能比 SVM 算法更稳定, 且 GA-RF 算法在几种分类处理中的性能最好。结果表明, GA-RF 算法可用于基于光谱数据的树种分类研究。

关键词: 人工林; 树种分类; 叶片光谱

中图分类号: S771 **文献标志码:** A **DOI:** 10.3969/j.issn.1672-8785.2016.02.006

Classification of Main Tree Species of Plantation in Guangdong Province by Leaf Spectra

LI Dan^{1,2}, YANG Long^{1,2}, WANG Chong-yang^{1,2}, CHEN Shui-sen^{1,2*}

(1. Guangdong Open Laboratory of Geospatial Information Technology and Application, Guangzhou 510070, China; 2. Guangzhou Institute of Geography, Guangzhou 510070, China)

Abstract: To explore the optimal feature wavebands for spectral discrimination of different tree species and the ability of classifying different tree species by spectroscopy, the leaf spectra of 10 tree species in the plantation in Guangdong Province are collected by using a spectroradiometer. The Genetic Algorithm (GA) and Successive Projection Algorithm (SPA) are used to reduce the data dimension. Then, the Support Vector Machine (SVM) and Random Forest Algorithm are used to classify the tree species. The result shows that the wavebands selected by two data dimension reduction methods are mainly located in the near infrared region. GA performs better than SPA in variable selection. The model established by RF is more stable than that established by SVM in performance. The GA-RF algorithms can be used for the tree species classification based on spectral data.

Key words: plantation; classification of tree species; leaf spectra

收稿日期: 2016-01-07

基金项目: 广东省自然科学基金项目 (S2013010014983); 国家自然科学基金项目 (41301401)

作者简介: 李丹 (1985-), 女, 河南通许人, 助理研究员, 主要研究方向为农业遥感。

*** 通讯作者:** 陈水森 (1965-), 男, 江西高安人, 研究员, 主要研究方向为定量遥感。E-mail: css@gdas.ac.cn

0 引言

森林树种识别对于林业资源管理、环境保护和生物多样性保护具有重要作用。基于地面调查的森林树种识别方法往往需要耗费大量的时间。高光谱技术以其快速、高效等特点在树种识别方面得到了较多应用。高光谱数据波段间隔窄, 光谱数据间存在大量的冗余信息。这些冗余信息会在一定程度上导致估计方差增大, 模型稳定性下降, 预测性能降低^[1]。对光谱数据进行有效处理是高光谱遥感研究的重要内容。近年来, 研究人员提出了诸多有关光谱数据处理^[2-4], 光谱特征选择、提取^[5-8]以及分类的方法^[9-10], 并对其在树种分类中的性能进行了研究^[1,11]。但目前关于广东省人工林树种的光谱可分性的研究还很少, 而且针对主要人工林树种的光谱特征分析研究也相对较少。因此, 本文从认识广东省主要人工林树种的光谱特征

入手, 通过采用不同的特征提取方法, 尝试分析 10 种主要人工林树种的光谱可分性以及用于主要人工林树种识别的主要光谱特征, 并评价基于光谱进行树种分类的可行性。

1 试验部分

1.1 试验区

人工林叶片采集的野外试验区位于广东省鹤山市鹤山森林生态系统国家野外研究站内(112°50'E, 22°34'N), 其海拔高度为 60.7 m。站内包括针叶树混交林、乡土阔叶树混交林、豆科树种混交林、尾叶桉林、马占相思林和大叶相思林等。选择 10 种优势树种作为研究对象(见表 1), 其中每个树种随机选择 15 棵成年个体。然后用高枝剪摘下一个冠层树枝, 其中每个树枝选择 3 片成熟全展叶作为研究对象。将采集好的叶片保存在样品袋中, 并将其带回实验室进行光谱测量。

表 1 10 种优势树种的特征数据

中文名	拉丁名	物种源地	叶形
马尾松	<i>Pinus massoniana</i>	本土	针叶
湿地松	<i>Pinus elliottii</i>	外来	针叶
杉木	<i>Cunninghamia lanceolata</i>	本土	针叶
尾叶桉	<i>Eucalyptus urophylla</i>	外来	阔叶
大叶相思	<i>Acacia auriculaeformis</i>	外来	阔叶
马占相思	<i>Acacia mangium</i>	外来	阔叶
荷木	<i>Schima superba</i>	本土	阔叶
红椎	<i>Castanopsis hystrix</i>	本土	阔叶
阴香	<i>Cinnamomum burmannii</i>	本土	阔叶
火力楠	<i>Michelia macclurei</i>	本土	阔叶

1.2 数据采集

在室内, 采用美国 ASD 公司生产的 Field-Spec Pro F 型便携式分光辐射光谱仪测定了每个叶片的反射光谱。反射光谱的范围为 350 ~ 2500 nm, 光谱间隔为 1 nm。光源采用光谱仪自带的 50 W 卤化灯, 光源的方位角为 60°; 光谱仪的视场角为 25°, 探头垂直向下, 距叶片 0.10 m; 叶片正面平整置于反射率近似为 0 的黑色橡胶

垫上。探头分别正对待测叶片的前、中、后部进行测定。每次测量记录 5 个采样光谱, 共计测定 15 个采样光谱, 并以其平均值作为该叶片的光谱反射率值。在测定过程中, 每隔 15 min 进行一次标准白板校正^[12]。

1.3 数据处理与分类

为了降低高光谱数据的信息冗余和提高光谱处理效率, 我们对原始光谱数据进行了 10 波

段平均处理,同时去除了光谱噪声较大的波段,而仅保留了 415 ~ 2475 nm 光谱范围内共 207 个光谱数据。由于在室内光谱测量中,室内光源的照明强度和叶片测量位置等因素对原始反射光谱具有一定影响,因此对原始光谱数据进行了一阶导数处理,以增强光谱特征^[13]。图 1 和图 2 所示分别为原始光谱和经一阶导数处理后的 10 种叶片的光谱。

分别利用在特征提取方面具有较好性能的 GA 算法^[1,14]和 SPA 算法^[15-16]对一阶导数光谱数据进行了特征选择,筛选出了不同树种分类的光谱特征波段。然后利用 SVM 算法^[17-18]和 RF 算法^[18-19]进行了分类识别。10 个树种中,每个树种均取 15 个样本,其中随机选择 10 个样本用来建模,另外 5 个样本用来验证。这样就组成了具有 100 个样本的建模样本集和具有 50 个样本的验证数据集。

2 结果与讨论

2.1 反射光谱数据

从图 1 中可以看出,不同树种叶片的反射光谱具有绿色植物叶片光谱在可见光-近红外波段的普遍特征:在 400 ~ 760 nm 可见光波段,反射率较低;在 550 nm 附近,存在叶绿素的强反射峰,反射率在 20% 左右;在 680 nm 附近,存在

低的叶绿素的吸收谷;在 690 ~ 760 nm 波段,反射率急剧抬升,达到 45% 以上;在 760 ~ 900 nm 波段,反射率曲线较为平缓,反射率保持在 45% 以上;在 900 ~ 2400 nm 波段以及在 1020 nm、1350 nm、1430 nm、1950 nm 附近,由于存在水分吸收,反射率呈现出吸收谷特征。但不同树种叶片的反射率光谱之间差异较大,而且这种差异在近红外波段愈加明显。在 750 ~ 2400 nm 波段,荷木的反射率最高,其次是红椎。在 400 ~ 750 nm 波段,马占相思在 550 nm 左右的反射率最高。

经一阶导数处理后,反射光谱在色素和水分等方面的吸收特征增强。从图 2 中可以看出,在可见光波段内的 510 nm、530 nm、560 nm、580 nm、690 nm、740 nm 附近,一阶导数曲线呈现峰、谷特征。其中,红椎的光谱曲线在 510 nm、530 nm、560 nm、690 nm 附近的波谷和波峰深度最深;荷木的光谱曲线在 580 nm 和 740 nm 附近的波峰和波谷深度最深。在 1375 nm 和 1415 nm 附近,反射光谱呈现出波谷和波峰形态。其中,荷木的光谱曲线具有最大的谷/峰深度;马尾松和湿地松的光谱曲线的谷/峰深度最浅。在 1635 nm 和 1660 nm 附近,反射光谱呈现出谷/峰形态。其中,荷木的光谱曲线具有最大的谷/峰深度;其次是红椎,马尾松深度最浅。在 1865 nm

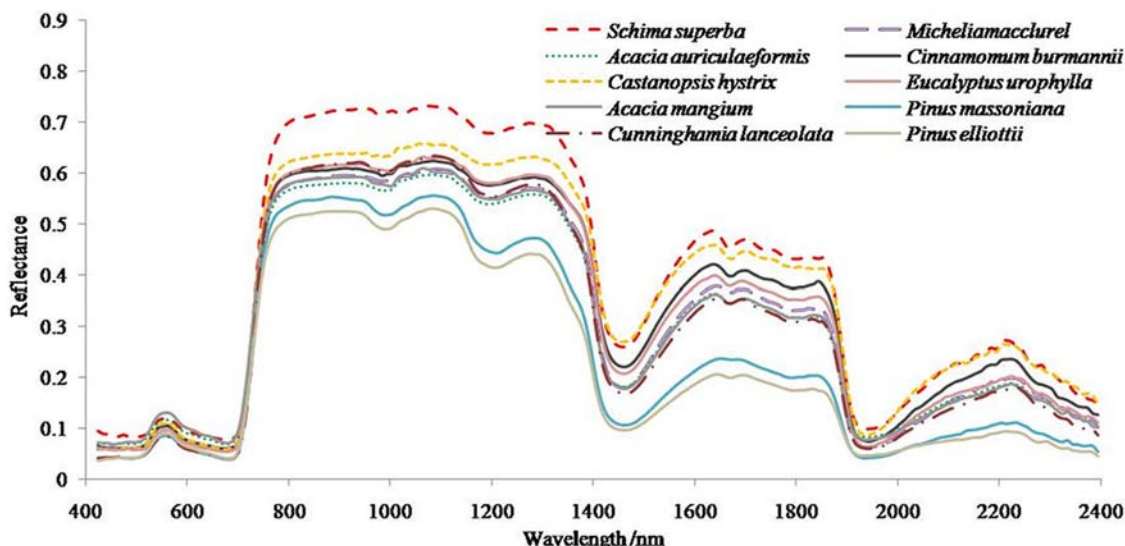


图 1 10 种树种叶片的反射光谱

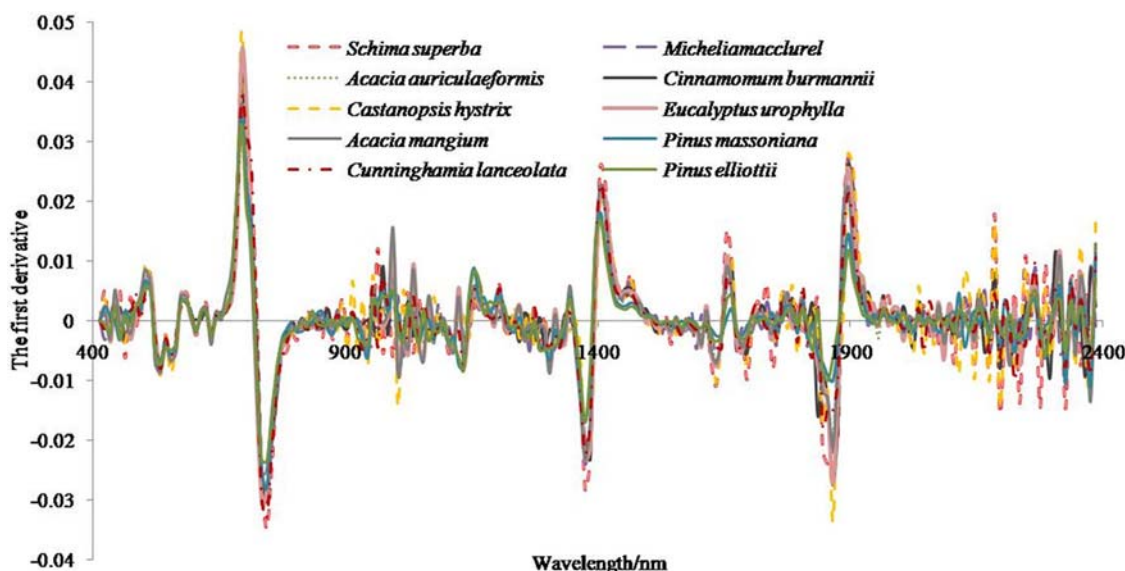


图 2 10 种树种叶片的一阶导数曲线

表 2 采用 GA 算法和 SPA 算法对一阶导数光谱数据所选择的特征波长

方法	波段 (nm)
GA 算法	455、935 ~ 965、1145 ~ 1205、1245 ~ 1265、1285 ~ 1305、1455 ~ 1515、1575 ~ 1625、1805
SPA 算法	475、515、545、875、995、1045、1195、1465、2055、2145、2165 ~ 2175、2195、 2225、2245 ~ 2255、2285、2325、2355、2485

和 1900 nm 附近, 反射光谱呈现出谷/峰形态。其中, 红椎的光谱曲线具有最大的谷/峰深度; 湿地松的光谱曲线的谷/峰深度最浅。

2.2 特征波长选择

表 2 列出了采用 GA 算法和 SPA 算法对一阶导数光谱数据所选择的特征波长。其中, GA 算法的参数设置如下: 群体数目为 30, 杂交概率为 50%, 变异概率为 1%, 迭代次数和运行轮数设置为 100; 当 SECV 值最小时, 确定入选的变量数目。SPA 算法的运行步骤与其他文献设置的相同^[20], 其中设定的最小特征变量数为 20, 最大特征变量数为 40。GA 算法总共选择了 29 个波长的信息, 其中有些波段是连续的。这些波长之间具有较强的相关性。SPA 算法总共选择了 20 个波长的信息, 它们大多不连续。但两种特征选择算法均在近红外波段选择了较多的光谱变量, 说明近红外光谱对于树种类型识别可贡献出较多信息。这一结果与徐光彩等人^[21]以及王志辉和丁丽霞^[7,22]的研究结论之间存在相似

之处, 即用于区分不同树种的有效波段大都位于近红外波段。

两种算法均选择了蓝光波段以及 1195 nm、1465 nm 附近的光谱信息。其中, 蓝光波段与色素吸收等有关; 近红外波段内的 1195 nm 波长位于 C-H 的一阶倍频吸收范围内, 这也是水分的主要吸收区间^[23]; 1465 nm 接近 1450 nm 附近存在的由 O-H 官能团的一阶倍频引起的吸收区间, 这一区间被认为与淀粉、糖类、纤维素、水分等有关^[23]。GA 算法选择了 935 ~ 965 nm 和 1145 ~ 1205 nm 两个连续区间的光谱信息。这些光谱区间位于水分、纤维素等吸收范围内。文献^[16]指出, 1265 nm 和 1465 nm 附近的光谱数据可用于不同作物类型的区分^[16]。1575 ~ 1625 nm 波段与叶片中的氮^[24]、磷含量^[25]等有关; 1805 nm 接近 OH 和 C=O 官能团在 1820 nm 附近的吸收特征, 该现象被认为与纤维素吸收有关^[23]。SPA 算法选择的 515 nm 和 545 nm 附近的光谱位于绿光波段内, 这与叶片中的胡萝卜素吸收密

表 3 不同分类模型的对比情况

数据模型	方法	建模精度 (%)	预测精度 (%)
全波段光谱	SVM	99.8±0.5	84.4±2.2
	RF	95.0±0.7	94.8±2.3
SPA	SVM	95.6±4.8	62.4±3.8
	RF	83.0±2.2	81.6±5.2
GA	SVM	94.8±1.3	88.4±2.6
	RF	91.8±2.4	91.0±4.1

切相关。在 875 nm、995 nm、1045 nm 附近的反射光谱除了与水分等吸收有关,而且还与叶片结构等差异密切相关。更多的波段则位于 2000 nm 以后,这些波段大多在蛋白质、纤维素和木质素等的吸收特征附近^[23-24]。

2.3 分类精度分析

分别利用 SVM 算法和 RF 算法进行了运算。在 RF 分类过程中,存在两个关键参数: n_{tree} 为 RF 中决策树的个数(默认值为 500); m_{try} 为随机特征的数量以及输入变量的个数(这里设置为 10)^[26]。SVM 算法选择径向基函数作为核函数,并利用留一交叉验证法确定了模型的超参数和径向基函数的核参数^[26]。

在分类中,随机选择了分类建模样本($n=100$, 每次每个树种选择 10 个样本)和预测样本($n=50$, 每次每个树种选择 5 个样本),所以每次运算得到的预测结果和精度有所差异。表 3 列出了 10 次随机运行的结果,并给出了每次运行模型建模和预测精度的变化情况。通过对比两种分类方法的精度可以发现,RF 算法在建模和预测中的性能比 SVM 算法更为稳定,且建模精度和预测精度较为接近。其中,基于 GA 变量筛选结果建立的 RF 分类模型的建模和预测精度分别为 $91.8\%\pm 2.4\%$ 和 $91.0\%\pm 4.1\%$ 。从表 3 中可以看出,特征波段选择方法对分类精度具有影响;总体来讲,基于 SPA 算法所选特征波段建立的模型(SVM 和 RF 模型)的建模和预测性能均比基于 GA 算法所选波段建立的模型(SVM 和 RF 模型)要好。另外,通过 SPA 算法进行特征选择,还可以压缩光谱数据量,且光谱特征间

的冗余信息少。利用 GA 算法进行特征选择后所建立的模型与全波段光谱建模性能接近。通过利用 29 个变量信息也可大大减少光谱数据量。但从建模性能来看,SPA 算法的特征选择性能更好;通过结合 RF 算法与 SPA 算法,仅利用较少的光谱信息,就可获得与全波段建模相近的精度。

3 结束语

本文采用 GA 算法和 SPA 算法对广东省 10 种主要人工林树种叶片的可见光-近红外光谱变量进行了分析,并探讨了用于人工林树种区分的光谱特征。通过结合 SVM 算法和 RF 算法,对主要人工林树种叶片尺度的分类识别精度进行了分析,并为人工林光谱识别提供了依据。主要人工林树种叶片具有较高的可分性(模型精度在 85% 以上)。用于树种区分的光谱特征主要位于蓝绿光波段和近红外波段。基于 GA 算法选择的光谱特征变量能够较好地显示出不同树种之间的差异,且在分类中的性能较好。SVM 算法和 RF 算法的分类精度分析表明,RF 算法在建模和预测中具有更稳定的性能。通过结合 GA 算法和 RF 算法对 10 种人工林树种进行识别可以具有较好的精度,因此利用光谱数据对人工林树种叶片光谱进行分类是可行的。

本文主要利用 GA 算法和 SPA 算法对 10 种人工林树种叶片光谱进行了特征选择,并在 RF 与 SVM 回归方法的支撑下对树种分类的识别精度进行了评价。但本研究仅对 10 种人工林树种光谱识别的特征光谱区间进行了分析,所以我

们将在下一步的工作中深入讨论不同人工林树种之间光谱可分的具体原因以及叶片结构相似树种之间的光谱可分性。

参考文献

- [1] 刘燕德, 张光伟, 蔡丽君. 基于高光谱的 GA 和 SPA 算法对赣南脐橙叶绿素定量分析 [J]. *光谱学与光谱分析*, 2012, **32**(12): 3377–3380.
- [2] 王志辉, 丁丽霞. 基于叶片高光谱特性分析的树种识别 [J]. *光谱学与光谱分析*, 2010, **30**(7): 1825–1829.
- [3] 邢东兴, 常庆瑞. 基于花期果树冠层光谱反射率的果树树种辨识研究 [J]. *红外与毫米波学报*, 2009, **28**(3): 207–211.
- [4] Vasat R, Kodesova R, Boruvka L, et al. Consideration of Peak Parameters Derived from Continuum-removed Spectra to Predict Extractable Nutrients in Soils with Visible and Near-infrared Diffuse Reflectance Spectroscopy (VNIR-DRS) [J]. *Geoderma*, 2014, **232**(12): 208–218.
- [5] 褚西鹏. 基于叶片非成像高光谱数据树种分类 [D]. 杭州: 浙江农林大学, 2012.
- [6] 臧卓, 林辉, 孙华, 等. 南方主要针叶树种高光谱数据降维分类研究 [J]. *中南林业科技大学学报*, 2010, **30**(11): 20–25.
- [7] Kokaly R F, Skidmore A K. Plant Phenolics and Absorption Features in Vegetation Reflectance Spectra Near 1.66 μm [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2015, **43**: 55–83.
- [8] Nikopensius M, Pisek J, Raabe K. Spectral Reflectance Patterns and Seasonal Dynamics of Common Understory Types in Three Mature Hemi-boreal Forests [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2015, **43**: 84–91.
- [9] Vasques G M, Dematte J A M, Viscarra R R A, et al. Soil Classification Using Visible/near-infrared Diffuse Reflectance Spectra from Multiple Depths [J]. *Geoderma*, 2014, **223–225**: 73–78.
- [10] George R, Padalia H, Kushwaha S P S. Forest Tree Species Discrimination in Western Himalaya Using EO-1 Hyperion [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2014, **42**: 140–149.
- [11] Kuusinen N, Kolari P, Levula J, et al. Seasonal Variation in Boreal Pine Forest Albedo and Effects of Canopy Snow on Forest Reflectance [J]. *Agricultural and Forest Meteorology*, 2012, **164**: 53–60.
- [12] 朱西存, 赵庚星, 董芳, 等. 基于高光谱的苹果花磷素含量监测模型 [J]. *应用生态学报*, 2009, **20**(10): 2424–2430.
- [13] 陆婉珍. *现代近红外光谱分析技术 (第二版)* [M]. 北京: 中国石化出版社, 2007.
- [14] 孙旭东, 郝勇, 蔡丽君, 等. 基于抽取和连续投影算法的可见近红外光谱变量筛选 [J]. *光谱学与光谱分析*, 2011, **31**(9): 2399–2402.
- [15] 刘明博, 唐延林, 李晓利, 等. 水稻叶片氮含量光谱监测中使用连续投影算法的可行性 [J]. *红外与激光工程*, 2014, **32**(4): 1265–1271.
- [16] 刘国海, 江辉, 梅从立. 基于 dbiPLS-SPA 变量筛选的固态发酵湿度近红外光谱检测 [J]. *农业工程学报*, 2013, **29**(S1): 218–222.
- [17] Turker M, Koc-San D. Building Extraction from High-resolution Optical Spaceborne Images Using the Integration of Support Vector Machine (SVM) Classification, Hough Transformation and Perceptual Grouping [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2015, **43**: 58–69.
- [18] Nitze Ingmar, Barrett Brian, Cawkwell Fiona. Temporal optimisation of image acquisition for land cover classification with Random Forest and MODIS time-series [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2015, **34**: 136–146.
- [19] Torres M, Qiu G. Automatic Habitat Classification Using Image Analysis and Random Forest [J]. *Ecological Informatics*, 2014, **23**(9): 126–136.
- [20] 成忠, 张立庆, 刘赫扬, 等. 连续投影算法及其在小麦近红外光谱波长选择中的应用 [J]. *光谱学与光谱分析*, 2010, **30**(4): 949–952.
- [21] 徐光彩, 庞勇, 李增元, 等. 小兴安岭主要树种冠层光谱季相变化研究 [J]. *光谱学与光谱分析*, 2013, **33**(12): 3303–3307.
- [22] 陈树人, 栗移新, 毛罕平, 等. 基于光谱分析技术的作物中杂草识别研究 [J]. *光谱学与光谱分析*, 2009, **29**(2): 463–466.
- [23] Fourty T, Baret F, Jacquemoud S, et al. Leaf Optical Properties with Explicit Description of Its Biochemical Composition: Direct and Inverse Problems [J]. *Remote Sensing of Environment*, 1996, **56**(2): 104–117.
- [24] 施润和, 牛铮, 庄大方. 叶片生化组分浓度对单叶光谱影响研究——以 2100 nm 吸收特征的碳氮比反演为例 [J]. *遥感学报*, 2005, **9**(1): 1–7.
- [25] 王磊, 白由路, 杨俐苹. 春玉米磷素营养的光谱响应及诊断 [J]. *植物营养与肥料学报*, 2007, **13**(5): 802–808.
- [26] 王小川, 史峰, 郁磊, 等. *Matlab 神经网络 43 个案例分析* [M]. 北京: 北京航空航天大学出版社, 2013.