

doi:10.6041/j.issn.1000-1298.2021.11.020

# 基于无人机高光谱遥感影像的防护林树种分类

赵庆展<sup>1,2</sup> 江萍<sup>2,3</sup> 王学文<sup>1,2</sup> 张丽红<sup>4</sup> 张建新<sup>4</sup>

(1. 石河子大学信息科学与技术学院, 石河子 832000; 2. 兵团空间信息技术研究中心, 石河子 832000;  
3. 石河子大学机械电气工程学院, 石河子 832000; 4. 150 团农业发展服务中心, 石河子 832000)

**摘要:** 树种信息对林业资源监测和管理具有重要意义, 及时准确地掌握树种及长势状况是防护林工程建设与效益评价的基础。为研究利用无人机高光谱数据进行防护林树种分类的效果, 选取典型区域使用 Matrice600 型六旋翼无人机搭载 Rikola 高光谱成像仪获取高光谱影像, 基于支持向量机-递归特征消除算法 (SVM-RFE) 选取原始波段最佳组合, 再结合纹理特征、植被指数和数理统计特征, 使用随机森林算法对所有特征进行重要性评估并与分类精度相结合进行特征优化, 进而构建高光谱影像全波段、原始波段最佳组合、全部特征变量、基于随机森林 (RF) 特征优化后特征变量 4 种分类方案, 分别采用最大似然法 (MLC)、支持向量机 (SVM)、随机森林对防护林优势树种进行分类。结果表明: 所提出的基于交叉验证的 SVM-RFE 算法选出的原始波段组合能更好地还原原始光谱特征; 通过 RF 算法的特征重要性分析与分类精度相结合的方法可以有效选出重要特征, 当使用全部特征的 85% (包括 17 个光谱特征、3 个纹理特征、5 个植被指数和 3 个数理统计特征) 进行分类时, 总体精度最高为 95.93% (Kappa 系数为 0.9475); 所有特征中植被指数特征最重要, 3 种分类方法中 RF 算法分类总体精度 (OA) 最高。

**关键词:** 树种分类; 无人机; 高光谱; 特征挖掘; 随机森林

中图分类号: S252<sup>+</sup>.9 文献标识码: A 文章编号: 1000-1298(2021)11-0190-10

OSID:



## Classification of Protection Forest Tree Species Based on UAV Hyperspectral Data

ZHAO Qingzhan<sup>1,2</sup> JIANG Ping<sup>2,3</sup> WANG Xuewen<sup>1,2</sup> ZHANG Lihong<sup>4</sup> ZHANG Jianxin<sup>4</sup>

(1. College of Information Science and Technology, Shihezi University, Shihezi 832000, China

2. Corps Space Information Technology Research Center, Shihezi 832000, China

3. School of Mechanical and Electrical Engineering, Shihezi University, Shihezi 832000, China

4. 150 Regiment Agricultural Development Service Center, Shihezi 832000, China)

**Abstract:** Tree species information is of great significance to forestry resource monitoring and management, timely and accurate control of tree species and growth status is the basis for protection forest project construction and benefit evaluation. In order to study the effect of using UAV hyperspectral images to classify protection forest tree species, the advantages of UAV hyperspectral images high-resolution, multi-band, and short-period were used, taking the "Three North" protection forest which on the northern edge of the 150th Regiment of the Eighth Division of Xinjiang Production and Construction Corps as the research area, typical areas were selected and Matrice600 hexarotor drones equipped with Rikola hyperspectral imaging sensor was used to obtain hyperspectral images. Firstly, spectral features, textural features, vegetation indices (VIs) and characteristics of mathematical statistics were extracted from the UAV hyperspectral image, the support vector machine - recursive feature elimination (SVM-RFE) algorithm were used selection bands. Through the random forest algorithm to evaluate the importance of all features and combination with the overall classification accuracies was employed for feature reduction, and then four classification schemes of hyperspectral image full band, the best combination of original band, all feature variables, and feature variables based on random forest (RF) feature reduction were constructed. The classification results showed that the original band combination selected by the SVM -

收稿日期: 2020-11-12 修回日期: 2021-01-15

基金项目: 新疆生产建设兵团科技计划项目(2017DB005)、兵团空间信息工程技术中心创建项目(2016BA001)和中央引导地方科技发展专项资金项目(201610011)

作者简介: 赵庆展(1972—),男,教授,主要从事农业信息化、空间信息系统集成与服务研究, E-mail: zqz\_inf@shzu.edu.cn

RFE algorithm based on cross-validation proposed can better restore the original spectral features; when considering the four features (spectral features, textural features, hyperspectral Vis and mathematical statistics features) and after feature reduction, the three classifiers used, random forest (RF), maximum likelihood classification (MLC) and support vector machine (SVM), the overall classification accuracies of RF was the highest, which were 95.93%, respectively. These results also suggested that vegetation indices were effective for discriminating tree species with similar spectral signatures. The overall results provided evidence for the effectiveness and potential of UAV hyperspectral data for protection forest tree species identification.

**Key words:** tree species classification; unmanned aerial vehicle; hyperspectral; feature mining; random forests

## 0 引言

建设“三北”(西北、华北和东北)防护林体系工程是改善地区生态环境、解决生态灾难的根本措施,对于巩固和发展我国绿色生态屏障具有重要意义<sup>[1]</sup>。树种多样性是评价防护林林分结构的指标之一,树种类别属性信息对于建设和监测防护林生态系统意义重大,可为防护林规划者提供监测、管理、评估防护林的重要依据,确保其发挥防风固沙、农田保护等作用<sup>[2]</sup>。

与常规的实地勘测相比,遥感技术可以快速高效地监测林业生态系统<sup>[3-4]</sup>。林业遥感的一个重要方向就是树种分类及识别技术,被动遥感和主动遥感的迅速发展,使得研究人员能在更精细的尺度上识别森林各项属性<sup>[5]</sup>。部分学者融合激光雷达和高光谱数据对森林树种进行分类研究并取得了积极进展,但激光雷达数据的获取成本昂贵且受飞行区域面积影响,应用前景受到限制<sup>[6]</sup>。多光谱卫星遥感受到空间分辨率和光谱分辨率的影响,对于地物精细分类同样具有局限性<sup>[7]</sup>。高光谱数据能发掘更多的植被信息,当前无人机高光谱数据获取便捷、成本低、周期短、空间分辨率高,给树种分类带来新的遥感数据源。

高光谱影像具有高维特性,在样本数量有限时,直接分类易导致维数灾难,不仅增加了数据处理成本,而且会降低地物分类精度。有学者研究发现,在不影响地物分类精度的前提下,高光谱的部分波段是冗余的<sup>[8]</sup>。因此对高光谱影像进行分类之前先进行降维处理。文献[9]运用最佳指数法(Optimum index factor, OIF)、自适应波段选择法(Adaptive band selection, ABS)、自动子空间划分(Automatic subspace partitioning, ASP)与自适应波段相结合(ASP + ABS)3种方法进行无人机高光谱数据波段选择,提取信息量较大且波段间相关性较低的原始最佳波段组合,基于支持向量机(SVM)分类器对薇甘菊进行分类。OIF、ASP、自适应波段选择法,或是

基于类别可分性的标准距离、离散度、B距离法、J-M(Jeffrey - Matusia)距离法等依赖于大量的数学计算,且选取的波段在光谱覆盖范围内分布不均匀<sup>[10]</sup>。对于荒漠区防风固沙林,林地各树种冠层原始光谱相似,仅靠原始光谱特征进行树种分类容易产生异物同谱现象,加入一些其他特征(纹理特征、植被指数、数理统计特征等)可有效提高分类精度<sup>[11-12]</sup>。支持向量机和随机森林(RF)两种分类器在面对样本数据少、维数高的遥感数据时都表现出了优良性能。而最大似然法(MLC)分类器在运行速度上具有明显优势,但易产生休斯现象(Hughes phenomenon, HP)<sup>[12]</sup>。一些学者对SVM、RF、MLC分类器对于树种分类精度的影响进行了研究,当面对不同遥感数据和不同特征变量时,这3种分类器表现出了不同分类性能<sup>[13-20]</sup>。在实际应用中,面对不同研究区和数据源,还无法确定树种分类哪种算法最为适宜。

针对以上问题,本文采用基于交叉验证的SVM-RFE(递归特征消除)算法,在RFE算法中引入交叉验证,对高光谱所有波段进行5折交叉验证,并选出最优波段的评分集合,根据评分对无人机高光谱数据进行原始最佳波段组合的选择,并与OIF进行比较;结合纹理特征、植被指数以及数理统计特征,采用RF特征重要性分析与分类精度相结合的方法进行特征优化,并评估不同特征对分类结果的影响;采用4种分类方案分别基于最大似然法、支持向量机和随机森林进行研究区树种分类,评价不同分类算法对于研究区树种的分类性能,以选择最适于研究区树种分类的分类算法。

## 1 研究区数据获取与预处理

### 1.1 研究区概况

研究区位于新疆天山以北,准噶尔盆地南缘,莫索湾垦区北端的150团三北防护林区域(45°10'N, 85°56'E,见图1)。研究区防护林属于防风固沙林,考虑到150团的地理条件和适生植物种类以及“三

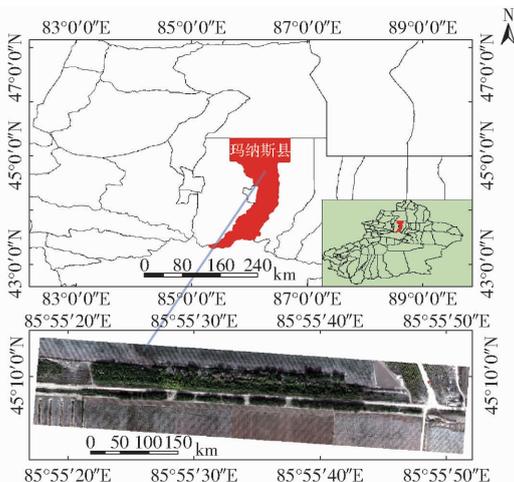


图1 研究区示意图

Fig.1 Schematic of study area

北”防护林工程树种组成配置原则<sup>[1]</sup>,林分形式为草类、灌木、乔木相结合,为固定沙地以及保护耕地,榆树林、混合阔叶林分布在道路两旁,其垂直结构为乔木层、灌木层、草本层。总面积约为451 km<sup>2</sup>。该地区为大陆性暖温带荒漠干旱气候,地势平坦,海拔300~500 m,主要乔木树种为榆树、新疆杨、胡杨、沙枣、梭梭等,林下植被以杂灌为主。本研究选择榆树、新疆杨、胡杨以及沙枣作为优势树种进行分类。

## 1.2 无人机平台

所用无人机平台为Matrice600型六旋翼无人机,最大承受风速可达8 m/s,最大平均飞行速度18 m/s,续航时间20 min左右,最大航程5 km,适配RONIN-MX型多功能云台,可搭载多种可见光与高光谱载荷设备。高光谱传感器采用芬兰SENOP公司生产的Rikola型高光谱成像仪,成像方式为框幅式成像,该成像光谱仪质量轻(720 g)、体积小,既可手持测量又可以搭载在小型无人机上,默认光谱

范围为500~900 nm,最多可达380个波段,可以根据数据采集要求调整光谱范围和波段个数。

## 1.3 高光谱数据获取

影像获取时间为2019年10月9日,无人机飞行高度200 m。为满足影像高分辨率要求以及飞行高度和相机参数需求,设置42个波段(光谱范围为502~903 nm,间隔约为10 nm),航向重叠率为82%,旁向重叠率为80%,空间分辨率为0.12 m,同时在地面设置4块地面辐射靶标,反射率分别为3%、22%、48%和64%,用于后期进行影像辐射校正处理。

## 1.4 地面调查数据

为了保证利用遥感影像进行样本数据标注过程的准确性,同期在研究区开展实地数据调查,根据研究区林分具体情况,选取了3个典型区域作为研究区进行采样。样地调查时间为2019年10月7日,采样过程中利用便携式GPS设备记录采样点位置信息,在各个样地中记录每棵单木树种情况。调查完成后,将原始数据记录整理至Excel表格,形成完整的调查记录表,采样共获取了40个样本点的树种类型数据,包括榆树、新疆杨、胡杨和沙枣等。

## 2 研究方法

首先对高光谱影像进行预处理以及分类特征集的构建,基于交叉验证的SVM-RFE(支持向量机-递归特征消除)算法选取原始波段最佳组合,再结合植被指数、纹理特征、数理统计特征等,构建新的影像分类特征。其次基于RF算法进行特征重要性分析,与分类精度相结合去除重要性得分排名靠后15%的特征。最后构建4种分类方案进行影像分类并作精度评价,比较不同分类器的分类效果。本研究技术路线如图2所示。

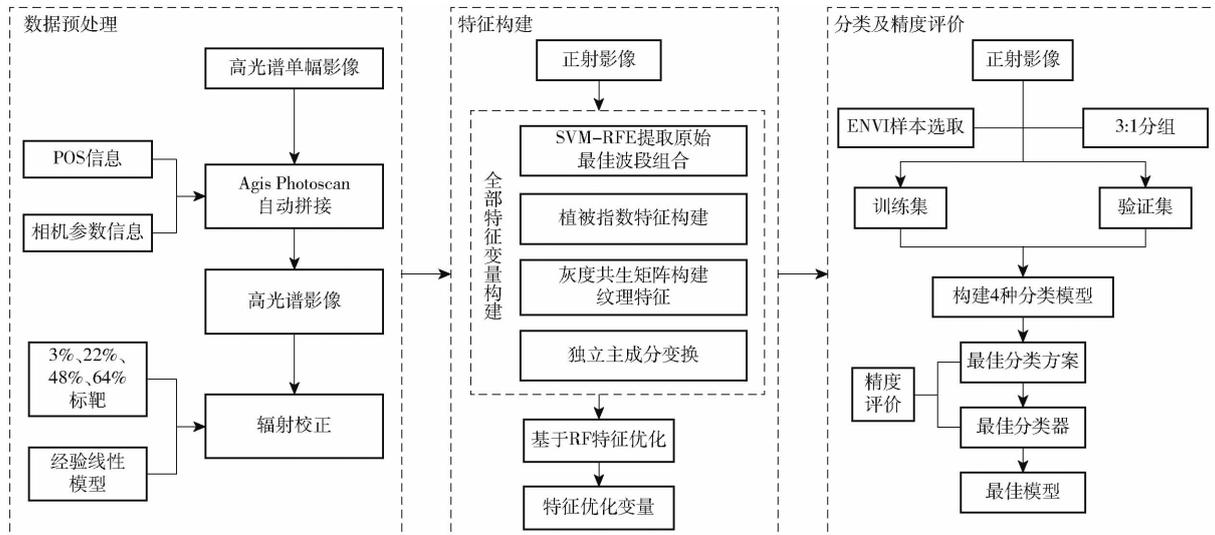


图2 高光谱影像树种分类技术路线图

Fig.2 Technical route diagram of hyperspectral image tree species classification

## 2.1 高光谱数据预处理及样本数据获取

将获取到的原始高光谱影像导出,使用 Agis Photoscan 软件完成系统校正工作(包括暗电流校正、镜头渐晕校正、辐射定标和图像格式转换),使用 ENVI 软件将其转换成标准 TIFF 格式的栅格影像数据。将 POS 数据(影像的经纬度信息)与影像一一对应,在 Agis Photoscan 中进行拼接处理,得到 0.12 m 空间分辨率的高光谱影像数据。对拼接完成的正射影像进行辐射校正,机载高光谱影像的实际数字量化值  $D_N$  与地面反射率  $R_{ef}$  的关系表达式为

$$R_{ef} = D_N a + b \quad (1)$$

式中  $a$ ——定标增益系数  $b$ ——偏移值

根据定标方程,以绘制感兴趣区域的方式分别统计 4 块靶标影像的 DN 值(实际数字量化值),并与 4 块靶标的标准反射率相对应,采用最小二乘法进行拟合,从而获得机载 Rikola 型高光谱仪辐射定标系数  $a$  和偏移值  $b$ <sup>[21]</sup>。

根据采样获取的地面样本点信息,对研究区主要树种进行记录,排除了数量较少或被其他冠层遮挡的树种,并根据实际分类效果选取地物样本;通过 ENVI 的创建 ROI 工具直接在影像上选取感兴趣区作为样本点。根据各地物在研究区的分布,最终共选择了 33 689 个像元作为样本数据,取 3/4 作为训练样本,其余作为验证样本。

## 2.2 原始最佳波段组合

训练样本较少且各树种样本数量不均匀时会降低分类器的分类性能。为了避免此类问题,对高光谱影像进行特征挖掘<sup>[22]</sup>非常必要,特征挖掘包括特征提取、波段选择和其他用于特征空间重构的运算。

特征提取是通过对原始光谱数据进行数学变换来降维和增强光谱特征,例如基于信息量正交线性变换的主成分分析法(Principal component analysis, PCA)、基于图像数据相关性的最小噪声分离(Maximum noise fraction, MNF)等,通过 PCA 对数据进行预处理一般需要假设数据服从高斯分布,由于高光谱数据非线性特点,此类算法进行降维后会丢失大量原始数据中的非线性信息<sup>[23]</sup>。波段选择是按照一定标准选择一个原始光谱数据的最优波段子集,该波段子集保留了地物的光谱特征或是提高了地物类别可分性。波段组合就是一个组合优化问题,其组合标准称为目标函数<sup>[24]</sup>。目前应用较为广泛的为最佳指数因子选择法(OIF)<sup>[25]</sup>、波段相关性比较、波段指数法等。

最佳指数因子选择法(OIF)综合考虑单波段影像信息量以及各波段间相关性,更接近于波段选择原则,且计算简单易于实现。OIF 计算公式为

$$R_{OIF} = \frac{\sum_{i=1}^3 S_i}{\sum_{i=1}^3 \sum_{\substack{j=1 \\ i \neq j}}^3 |R_{ij}|} \quad (2)$$

式中  $S_i$ ——第  $i$  个波段标准差

$R_{ij}$ ——波段间相关系数

$R_{OIF}$ ——最佳指数因子

OIF 越大,说明波段组合质量越好。

支持向量机算法也可被用于波段选择,与递归特征消除算法结合应用效果良好。SVM-RFE 算法是一种嵌入式的特征选择方法,与包装法不同,嵌入法不将分类器视为黑盒,而是使用训练分类器获得的信息来选择特征。因此本研究采用 SVM-RFE 算法选择原始最佳波段组合,其主要思想是构建一个模型进行多次训练,每轮训练移除若干权值系数较低的特征,再基于新的特征集进行下一步训练,直至选出满意的特征(利用 SVM 的分类性能进行波段选择,利用 SVM 的结果评价降维性能)<sup>[26]</sup>。在进行高光谱数据降维时如何确定降维后特征数目非常困难,传统的 RFE 算法需要在训练前确定降维后特征个数,本文将交叉验证的思想引入 RFE 算法,提出了基于交叉验证的 SVM-RFE 算法,在 REF 的基础上对不同的波段组合进行交叉验证,学习器 SVM 本身不变,通过计算其决策系数之和,最终得到不同波段对于分类的重要程度,然后保留最佳的波段组合。

## 2.3 植被指数特征构建

植被指数特征利用波段间数学变换对影像进行指数特征提取,根据本研究区的地物种类以及高光谱数据可选波段,构建了 10 个相关的植被指数进行研究。在进行指数计算后,对所有植被指数进行归一化处理,基于 CART 决策树的特征重要性评估对各个植被指数进行重要性评分,最终选取了 5 个重要性得分最高的植被指数特征,如表 1 所示。

高光谱数据不仅包含地物连续的光谱信息,也包含丰富的地物空间分布信息,仅考虑光谱信息,难以准确地对高光谱图像进行分析<sup>[27]</sup>。可以通过灰度共生矩阵构建纹理特征引入空间信息,纹理特征是表达高光谱影像空间特征的有效方法之一<sup>[28]</sup>。

对影像数据通过主成分分析进行降维处理,将影像信息压缩至几个主成分中,提取主成分包含信息量最高的第一主成分。通过对第一主成分进行灰度共生矩阵计算,空间域选用二阶概率统计滤波提取纹理特征。分别计算  $3 \times 3$ 、 $5 \times 5$ 、 $7 \times 7$  滤波窗口纹理特征,发现  $7 \times 7$  窗口的纹理特征标准差最大,因此将滤波窗口设置为  $7 \times 7$ ,获得 8 项纹理特征,分别为均值、方差、同质性、对比度、非相似性、熵、二

表 1 植被指数计算及基于 CART 决策树特征重要性评估

指数名称	计算公式	重要性得分
NDVI(归一化植被指数)	$NDVI = \frac{\rho_{NIR} - \rho_R}{\rho_{NIR} + \rho_R}$	0.477 4
NDGI(归一化绿度差值指数)	$NDGI = \frac{\rho_G - \rho_R}{\rho_G + \rho_R}$	0.298 5
CRI(类胡萝卜素指数)	$CRI = \frac{1}{\rho_B} - \frac{1}{\rho_G}$	0.167 7
GRVI(绿色比值植被指数)	$GRVI = \frac{\rho_{NIR}}{\rho_G} - 1$	0.033 3
MSAI(修正型土壤调整植被指数)	$MSAI = \frac{2\rho_{NIR} - \sqrt{(2\rho_{NIR} + 1)^2 - 8(\rho_{NIR} - \rho_R)}}{2}$	0.016 5
PSRI(光谱结构不敏感植被指数)	$PSRI = \frac{\rho_{NIR}^2 - \rho_R}{\rho_{NIR}^2 + \rho_R}$	
SR(比值植被指数)	$SR = \frac{\rho_{NIR}}{\rho_R}$	
REP(红边位置指数)		
DVI(差值植被敏感指数)	$DVI = \rho_{NIR} - \rho_R$	
NDWI(归一化水指数)	$NDWI = \frac{\rho_G - \rho_{NIR}}{\rho_G + \rho_{NIR}}$	

注:其中 $\rho_{NIR}$ 表示近红外波段, $\rho_R$ 表示红波段, $\rho_G$ 表示绿波段, $\rho_B$ 表示蓝波段。

阶矩、相关性。

## 2.4 分类特征优化

特征优化选取了特征变量中重要性较高的特征变量。RF 可以对特征变量的重要性和贡献度进行评估。在构造决策树时,通过 bootstrap 抽样法从训练样本集中提取数据,对于决策树中的每个节点,先从所有特征中随机选取  $m_{try}$  个特征,根据基尼系数进行分裂测试并找到最佳特征。训练过程中未被抽取的样本被称为袋外(Out of bag, OOB)数据,利用 OOB 数据进行分类结果的精度评价以及不同特征变量的重要性计算<sup>[29]</sup>。本文通过改变森林中树的数量进行多次实验,实验中记录特征重要性得分,最后与分类总体精度相结合,去除冗余的特征变量。

## 2.5 分类模型构建

RF 通过集成学习的思想将多棵决策树集成,不仅能够有效地运行在大数据集,而且可以在无需降维的情况下处理具有高维特征的输入样本<sup>[28]</sup>。随机森林分类需要设置两个参数,随机森林树的个数( $ntree$ )以及树节点预选的变量个数( $m_{try}$ )。本研究通过多次实验发现, $ntree$ 在 550 以后分类总体精

度几乎不变,因此将  $ntree$  设置为 550,  $m_{try}$  默认使用“Square Root”方法,即输入分类器特征变量数的平方根。SVM 可以自动寻找那些对分类有较大区分能力的支持向量,由此构造出分类器,将类与类间隔最大化<sup>[11]</sup>。支持向量机所选核函数为 RBF 函数(径向基函数),最优惩罚系数  $C$  为 0.036。

分类方案为:第 1 组选择原始高光谱全部波段( $n=42$ ,  $n$  表示波段数)作为分类器的输入;第 2 组选择原始最佳波段组合( $n=17$ )作为分类器的输入;第 3 组选择全部特征变量( $n=33$ )作为分类器的输入;第 4 组选择优化特征变量( $n=28$ )作为分类器的输入;分别基于 SVM 和 RF 以及 MLC 这 3 种分类器进行分类。为了评估分类结果的准确性以及无人机高光谱数据在防护林物种分类应用中的有效性,利用采样点建立地面真实的感兴趣区并构建混淆矩阵。对于每个分类结果,混淆矩阵提供了总体精度(Overall accuracy, OA)、Kappa 系数、用户精度(User accuracy, UA)和生产者精度(Producer accuracy, PA)来评价分类精度。

## 2.6 影像分类后处理

运用监督分类以及随机森林基于像素的分类时,分类结果中不可避免地会产生椒盐现象和一些小的图斑,为了增强分类效果,对初步的分类结果通过中值滤波对图像进行平滑,消除椒盐噪声和小的斑点(卷积核尺寸为  $5 \times 5$ )。

对于分类类别中产生的错分像元,采用聚类分析工具(MajorityMinority)进行归类(变换核尺寸为  $5 \times 5$ ,中心像元权重 1)。对于因缺少空间连续性而产生的斑点和洞,采用聚类处理工具(Clump)解决。对图像中出现的孤岛点,运用过滤处理(Sieve)工具消除。

## 3 结果与分析

### 3.1 原始最佳波段组合

以研究区优势树种为研究对象,分别使用 OIF 法(表 2)和本文提出的基于交叉验证的 SVM-RFE 算法进行原始最佳波段组合选择。

预处理后的无人机高光谱数据有 42 个波段,计算所有波段组合的 OIF 指数,选取指数排序前 20 的波段组合,如表 2 所示。本研究从 OIF 值前 20 的最佳波段组合中选取了 11 个波段进行原始波段组合,所选波段序号为 1、2、3、4、5、6、35、36、37、38、42。通过 OIF 算法选取原始波段最佳组合计算复杂度、高、耗费时间长,由表 2 可知,选择的波段呈现两极分化现象,分别集中在蓝光波段和近红外波段,绿色和红色波段未被选择,丢失了大量原始信息。

表 2 OIF 指数排序  
Tab.2 OIF index ranking

序号	波段(序号)组合	OIF 值
1	1,2,36	432.517 1
2	1,2,42	432.505 5
3	2,4,35	432.505 5
4	1,2,37	432.488 9
5	2,3,35	432.461 9
6	1,4,36	432.422 9
7	1,4,42	432.411 6
8	2,5,35	432.407 3
9	1,3,36	432.396 0
10	2,4,36	432.392 4
11	1,4,37	429.262 4
12	1,2,38	429.260 3
13	1,3,42	429.255 3
14	2,6,35	429.255 2
15	2,4,42	429.251 4
16	3,4,35	429.238 3
17	2,3,36	429.236 4
18	1,3,37	429.235 9
19	2,4,37	429.232 1
20	2,4,42	429.224 9

而基于本文提出的方法所选的原始波段为 17 个,对选取的波段进行波段组合,所选波段序号为 1、2、3、8、11、13、17、20、21、23、26、28、35、36、40、42。基于本文提出的方法所选的波段覆盖了整个范围,波段选择均匀且计算量相对较小。将两种波段组合选择方法的榆树光谱响应曲线分别与原始的 42 个波段榆树光谱响应曲线进行对比,如图 3 所示,由图 3 可知,基于交叉验证的 SVM-RFE 算法选出的波段特征组合,榆树光谱响应曲线与原始 42 个波段榆树光

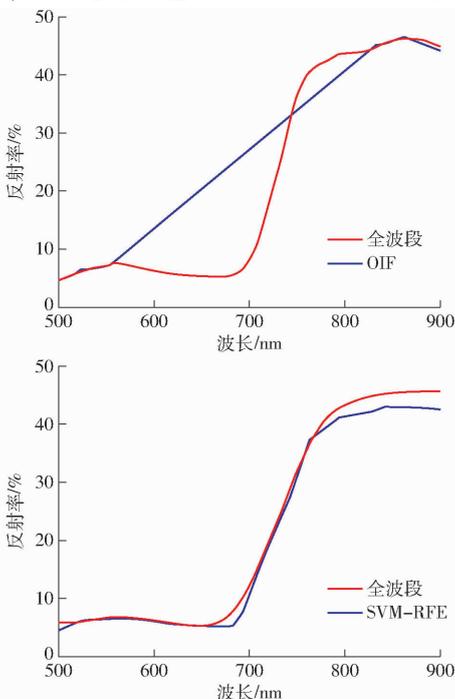


图 3 榆树光谱响应曲线

Fig.3 Spectral response curves of *Ulmus pumila* L

谱响应曲线拟合良好,较好地保留了原始光谱特征。基于两种方法选取的原始最佳波段组合得到的分类结果如表 3 所示,基于 OIF 法选取原始波段组合总体分类精度与 Kappa 系数与原始全波段相比都降低了,而基于本文提出的方法总体精度和 Kappa 系数与原始全波段相比几乎不变。相比较于 OIF 法,本文提出的方法更适合用作高光谱数据的原始最佳波段组合选择。

表 3 原始波段选择分类结果比较

Tab.3 Comparison of original band selection and classification results

类别	全波段 + RF		OIF + RF		SVM - RFE + RF	
	UA/%	PA/%	UA/%	PA/%	UA/%	PA/%
榆树	79.87	95.41	74.29	90.51	80.03	95.73
新疆杨	92.42	69.57	97.19	75.27	93.73	73.10
白杨	99.69	94.79	99.54	95.98	99.69	95.09
沙枣	95.24	76.92	85.33	61.54	95.37	74.28
OA/%	94.35		92.78		94.34	
Kappa 系数	0.924 8		0.903 8		0.924 8	

### 3.2 特征变量重要性分析与特征选择

通过随机森林的 OOB 误差分析得到特征变量重要性分布(图 4,图中 ica1 ~ ica3 表示独立主成分分析后 3 个主成分,bi(i = 1, 2, 3...)表示原始波段,glcm1 ~ glcm8 分别表示均值、方差、同质性、对比度、非相似性、熵、相关性、二阶矩),重要性得分越高,该变量对分类结果的影响以及贡献越大。通过改变决策树的数量发现,当决策树数量达到 2 000,特征重要性排序以及删除的特征变量基本不变。因此选择 RF 决策树数量为 2 000 时实验结果作为最终的特征重要性分析结果,如图 4 所示,分析可知:植被指数以及数理统计特征重要性得分较高,绿色比值植被指数重要性排名第一,这是由于研究区榆树占大多数,且长势良好,叶绿素含量较高,因此在近红外波段反射率较高。其他植被指数特征和数理统计特征均排在前 20%,说明本文构建的植被指数特征对于防护林树种分类有重要作用;而数理统计特征选取的是独立主成分分析的前 3 个主成分,其包含了原数据 95% 以上的信息量,因此重要性得分较高。

纹理特征重要性普遍较低,造成此现象的原因可能是因为经过主成分分析以后,排名靠后的纹理特征包含的原始数据信息量较少。

按照特征重要性排序和特征数量依次从 1 到 33 输入 RF 分类器,利用地面真实感兴趣区建立混淆矩阵,对分类结果进行评价,评价指标为总体分类精度(OA)和 Kappa 系数,特征变量数与分类精度关系如图 5 所示。分类精度在特征变量数 1 ~ 11 时明

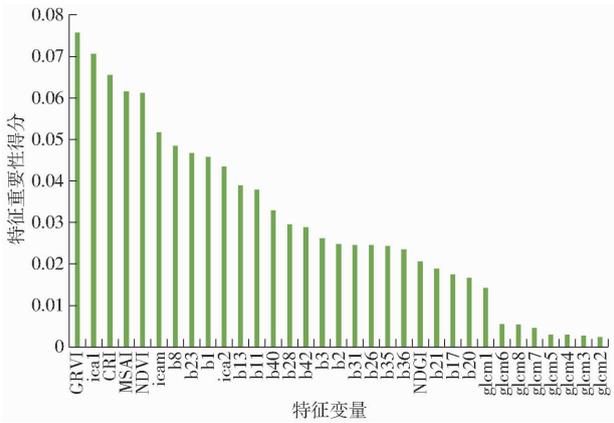


图4 特征变量重要性得分

Fig.4 Importance scores for characteristic variables

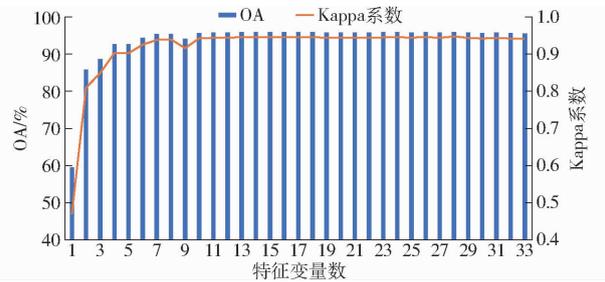


图5 特征变量数与分类精度关系

Fig.5 Relationship between number of features and classification accuracy

显提高,分类总体精度和 Kappa 系数分别为 94.73% 和 0.93。特征变量数从 11 开始,分类精度

呈上升趋势,但不明显。当  $n = 28$  时分类精度最高,总体精度和 Kappa 系数分别为 95.53% 和 0.9475。伴随着纹理特征的加入,分类精度呈略微下降趋势,说明特征变量过多易导致数据冗余和过拟合现象。最终确定重要性排名前 85% 的特征变量 ( $n = 28$ , 包括 17 个光谱特征、3 个纹理特征、5 个植被指数和 3 个数理统计特征) 作为分类器的输入。

### 3.3 不同分类器分类效果对比分析

本研究使用 28 个优选特征,基于同一训练样本,分别采用 MLC、SVM 和 RF 算法进行树种分类。从分类图(图 6)来看,3 种分类器对研究区的树种都有不同程度的混分与错分现象,沙枣分类精度最低,MLC 和 SVM 对于榆树和沙枣混分比较多,原因是研究区域内沙枣种植数量少,树冠大,侧斜枝较多,反映到图像上光谱信息较强,与区域中其他树种(榆树)混合种植,二者光谱特征相似,增加了分类难度,造成沙枣与榆树混分的现象。

长势旺盛的新疆杨与胡杨光谱、纹理特征相似,不易分辨,故两者有轻微混分现象。研究区周边农田棉花低矮,反映到图像上光谱信息较弱,与区域中灌木较为相似,棉花地和灌木也有少量混分现象。

由于考虑到防护林的总体效应,种植时会综合考虑特定的树木种类与灌木混合,防护林下分布了大量灌木,增加了树种精细分类的难度。此外,高光

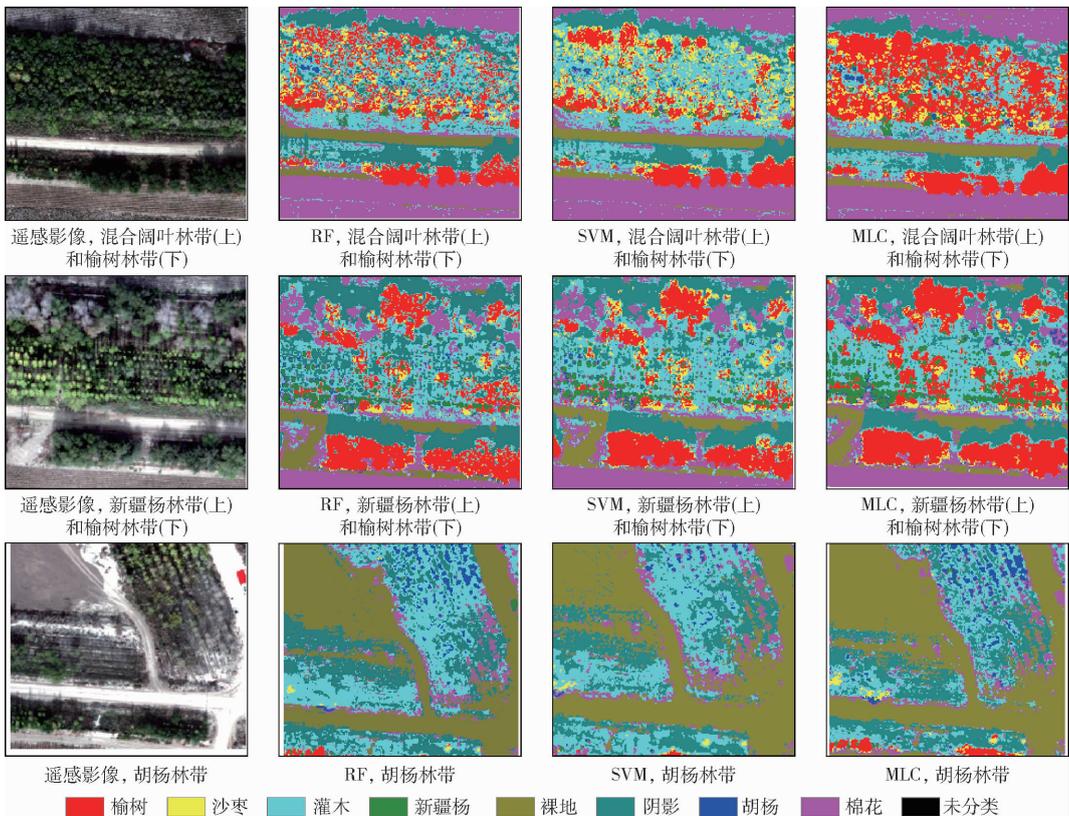


图6 特征优化后不同分类方法对树种分类的效果

Fig.6 Classification effects of tree species by different classification methods after feature optimization

谱数据采集时间在研究区的深秋季节,大部分植被已经开始枯萎,部分胡杨、沙枣已经枯萎或死亡而未被识别出来,影响了制图精度。3种算法都有不同程度的椒盐噪声,造成此现象的原因是本研究所使用的传感器空间分辨率较高。

表 4 汇总了基于特征优化后不同的分类器分类精度评估结果。实验结果表明,RF 分类精度最高,总体分类精度为 95.93% (Kappa 系数为 0.947 5);MLC 总体分类精度最低,为 88.70% (Kappa 系数为 0.850 8)。与 MLC 和 SVM 相比,RF 总体精度提高了 6.83 个百分点和 1.32 百分点,Kappa 系数提高了 0.096 7 和 0.024 1,大多数防护林树种的 PA 和 UA 都在 80% 以上。从单个树种分类精度来看,与 MLC 相比,随机森林分类器减少了白杨和榆树的漏分和错分;与 SVM 相比,减少了胡杨与沙枣的漏分与错分,PA 与 UA 较高证明了这一点。

表 4 不同分类器分类结果

Tab. 4 Results classified by different classifiers

类别	SVM		RF		MLC	
	UA/%	PA/%	UA/%	PA/%	UA/%	PA/%
榆树	84.26	93.99	93.85	100	82.39	94.78
新疆杨	86.18	86.41	94.41	84.67	93.61	94.78
白杨	97.70	88.69	100	92.86	0	0
沙枣	89.24	73.80	99.15	84.62	89.35	88.70
OA/%	94.21		95.53		88.70	
Kappa 系数	0.923 4		0.947 5		0.850 8	

从表 5 来看,分类方案 2 的结果表明,本文提出的高光谱数据波段选择方法是有效的。与仅使用光谱特征相比,植被指数和纹理特征以及数理统计特征的加入大大地提高了单个树种的分类精度,用户精度提高了 0.18 ~ 24.84 个百分点。特征优化后,分类总体精度提高了 0.39 个百分点,Kappa 系数提高了 0.065,虽然精度提高不明显,但数据处理效率提高,且挖掘出了适合于研究区树种分类的特征。RF 算法比 MLC、SVM 算法分类精度高,分类效果更好,对于研究区树种分类具有很好的适用性。

表 5 不同分类方案的 RF 分类结果

Tab. 5 RF classification results of different classification schemes

类别	方案 1 (n=42)		方案 2 (n=17)		方案 3 (n=33)		方案 4 (n=28)	
	UA/%	PA/%	UA/%	PA/%	UA/%	PA/%	UA/%	PA/%
榆树	79.87	94.51	80.05	95.25	92.72	86.94	93.85	100
新疆杨	99.69	95.41	99.84	95.09	100	94.05	100	92.86
胡杨	69.57	92.41	94.41	73.37	88.86	93.16	94.41	84.67
沙枣	76.92	96.54	99.84	95.09	94.76	87.02	99.15	84.62
OA/%	94.35		94.34		95.54		95.93	
Kappa 系数	0.924 8		0.924 8		0.941 0		0.947 5	

### 3.4 影像分类后处理

本研究最终使用 28 个优选特征,基于同一训练样本采用 RF 算法进行树种分类。对于初步分类结果,进行了图像分类后处理,采用中值滤波器去除椒盐噪声和较小图斑,效果良好;使用聚类分析、聚类处理和过滤处理,解决了图像中较大类别中的虚假像元问题,因缺少空间连续性而产生的斑点和洞以及图像中的孤岛问题,效果良好。从表 6 可以看到,处理后树种分类 OA 和 Kappa 系数分别提高了 1.01 个百分点和 0.012,本文的分类后处理方法可以有效增强分类效果。分类后处理效果如图 7 所示。

表 6 RF 分类后处理结果

Tab. 6 RF (n=28) classification and post-processing

results

类别	RF 初步分类		分类后处理	
	UA/%	PA/%	UA/%	PA/%
榆树	93.85	100	89.69	96.36
新疆杨	100	92.86	100	93.01
胡杨	94.41	84.67	91.74	90.49
沙枣	99.15	84.62	97.73	93.03
OA/%	95.93		96.94	
Kappa 系数	0.947 5		0.959 5	

## 4 结论

(1) 提出了基于交叉验证的 SVM-RFE 算法提取原始波段最佳组合,相比较于 OIF 指数法,所提出的方法计算量相对较小且很好地保留了原始光谱特征信息,降低维度的同时增加了树种类别可分性,与最佳指数法相比,基于交叉验证的 SVM-RFE 算法更加适合于高光谱数据的波段选择。

(2) 纹理特征、植被指数以及数理统计特征的加入提高了单个树种的分类精度。绿色比值植被指数 (GRVI)、类胡萝卜素指数 (CRI)、修正型土壤调整植被指数 (MSAI)、归一化植被指数 (NDVI)、归一化绿度差值指数 (NDGI) 等植被指数特征对于提高研究区防护林树种的分类精度是有效的。基于 RF 的特征重要性分析与分类精度相结合的方式特征优化后,分类总体精度提高了 0.39 个百分点,Kappa 系数提高了 0.065,说明去除的纹理特征 (方差、同质性、对比度、非相似性、相关性以及二阶矩) 对于研究区的树种分类具有干扰性,去除以后提高了分类器的分类性能。

(3) 基于构建的 4 种分类方案进行树种分类,特征优化后的特征变量组合分类效果最好,RF 分类精度最高,总体精度可达 95.93% (Kappa 系数为

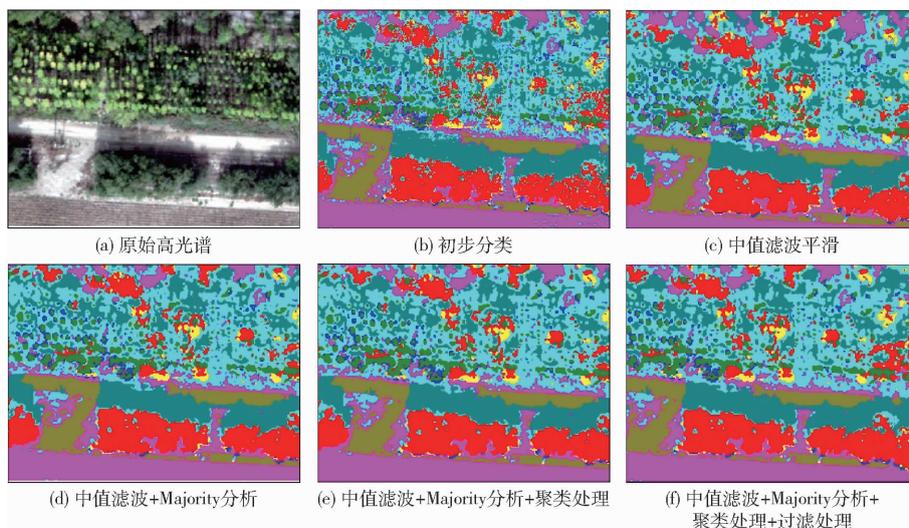


图7 影像分类后处理

Fig. 7 Image classification post-processing

0.947 5),MLC 分类精度最低,总体精度为 88.70% (Kappa 系数为 0.850 8),SVM 总体精度为 94.21% (Kappa 系数为 0.923 4),3 种分类方法比较,RF 是最适合于研究区的分类算法。

(4)对初步分类结果分别进行了中值滤波平滑、聚类分析、聚类处理和过滤处理,处理后分类总体精度和 Kappa 系数分别提高了 1.01 个百分点和 0.012,本研究的分类后处理方法可以增强分类效果。

## 参 考 文 献

- [1] 黄麟,祝萍,肖桐,等.近 35 年三北防护林体系建设工程的防风固沙效应[J].地理科学,2018,38(4):600-609.  
HUANG Lin,ZHU Ping,XIAO Tong,et al. The sand fixation effects of three-north shelter forest program in recent 35 years[J]. Scientia Geographica Sinica,2018,38(4):600-609. (in Chinese)
- [2] RYAN R J, PERRY J H, ANDREW J H, et al. Classification of urban tree species using hyperspectral imagery[J]. Geocarto International, 2012, 27(5):443-458.
- [3] 吴见,彭道黎.高光谱遥感林业信息提取技术研究进展[J].光谱学与光谱分析,2011,31(9):2305-2312.  
WU Jian,PENG Daoli. Advances in researches on hyperspectral remote sensing forestry information-extracting technology[J]. Spectroscopy and Spectral Analysis,2011,31(9):2305-2312. (in Chinese)
- [4] 吴楠,李增元,廖声熙,等.国内外林业遥感应用研究概况与展望[J].世界林业研究,2017,30(6):34-40.  
WU Nan,LI Zengyuan,LIAO Shengxi,et al. Current situation and prospect of research on application of remote sensing to forestry[J]. World Forestry Research, 2017,30(6):34-40. (in Chinese)
- [5] 黄华国.林业定量遥感研究进展和展望[J].北京林业大学学报,2019,41(12):1-14.  
HUANG Huaguo. Progress and perspective of quantitative remote sensing of forestry[J]. Journal of Beijing Forestry University, 2019, 41(12): 1-14. (in Chinese)
- [6] DAINI S W,STUART A,SHANNON J,et al. Improved classification of urban trees using a widespread multi-temporal aerial image dataset[J]. Remote Sensing,2020,12(15):2475.
- [7] 毕恺艺,牛铮,黄妮,等.基于 Sentinel-2A 时序数据和面向对象决策树方法的植被识别[J].地理与地理信息科学,2017,33(5):16-20,27.  
BI Kaiyi,NIU Zheng,HUANG Ni,et al. Identifying vegetation with decision tree model based on object-oriented method using multi-temporal Sentinel-2A images[J]. Geography and Geo-Information Science,2017,33(5):16-20,27. (in Chinese)
- [8] 刘舒,姜琦刚,马玥,等.基于多目标遗传随机森林特征选择的面向对象湿地分类[J/OL].农业机械学报,2017,48(1):119-127.  
LIU Shu,JIANG Qigang,MA Yue,et al. Object-oriented wetland classification based on hybrid feature selection method combining with relief F, multi-objective genetic algorithm and random forest[J/OL]. Transactions of the Chinese Society for Agricultural Machinery,2017,48(1):119-127. [http://www.j-csam.org/jcsam/ch/reader/view\\_abstract.aspx?file\\_no=20170116&flag=1](http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20170116&flag=1). DOI:10.6041/j.issn.1000-1298.2017.01.016. (in Chinese)
- [9] 刘彦君,张贵,王潇,等.基于无人机高光谱影像波段选择的薇甘菊分类[J].测绘通报,2020(4):34-39,52.  
LIU Yanjun,ZHANG Gui,WANG Xiao,et al. Classification study of *Mikania micrantha* Kunth from UAV hyperspectral image band selection[J]. Bulletin of Surveying and Mapping, 2020(4):34-39,52. (in Chinese)
- [10] 谭玉敏,夏玮.基于最佳波段组合的高光谱遥感影像分类[J].测绘与空间地理信息,2014,37(4):19-22.  
TAN Yumin,XIA Wei. Optimum band combination based hyperspectral remote sensing image classification[J]. Geomatics &

- Spatial Information Technology, 2014, 37(4): 19–22. (in Chinese)
- [11] 王修信, 秦丽梅, 罗玲, 等. 遥感图像森林林型 SVM 分类的多特征选择[J]. 计算机工程与应用, 2013, 49(20): 259–262.  
WANG Xiuxin, QIN Limei, LUO Ling, et al. Multi-feature selection in remote sensing forest species classification with SVM [J]. Computer Engineering and Applications, 2013, 49(20): 259–262. (in Chinese)
- [12] CAO Jingjing, LENG Wanchun, LIU Kai, et al. Object-based mangrove species classification using unmanned aerial vehicle hyperspectral images and digital surface models[J]. Remote Sensing, 2018, 11(10): 89.
- [13] 刘怀鹏, 安慧君. Worldview-2 树种分类 T&N 波段优势分析[J]. 西部林业科学, 2019, 48(3): 41–46.  
LIU Huaipeng, AN Huijun. Analysis of T&N band advantage in tree species classification based on Worldview-2[J]. Journal of West China Forestry Science, 2019, 48(3): 41–46. (in Chinese)
- [14] PAL M. Random forest classifier for remote sensing classification[J]. International Journal of Remote Sensing, 2005, 26(1): 217–222.
- [15] 温小乐, 钟奥, 胡秀娟. 基于随机森林特征选择的绿化乔木树种分类[J]. 地球信息科学学报, 2018, 20(12): 1777–1786.  
WEN Xiaole, ZHONG Ao, HU Xiujian, et al. The classification of urban greening tree species based on feature selection of random forest[J]. Journal of Geo-information Science, 2018, 20(12): 1777–1786. (in Chinese)
- [16] 李哲, 张沁雨, 邱新彩, 等. 基于高分二号遥感影像树种分类的时相及方法选择[J]. 应用生态学报, 2019, 30(12): 4059–4070.  
LI Zhe, ZHANG Qinyu, QIU Xincan, et al. Temporal stage and method selection of tree species classification based on GF-2 remote sensing image[J]. Chinese Journal of Applied Ecology, 2019, 30(12): 4059–4070. (in Chinese)
- [17] ZHANG C Y, XIE Z X. Object-based vegetation mapping in the Kissimmee River watershed using hyMap data and machine learning techniques[J]. Wetlands, 2013, 33(2): 233–244.
- [18] MAXWELL A E, WARNER T A, STRAGER M P, et al. Assessing machine-learning algorithms and image- and lidar-derived variables for GEOBIA classification of mining and mine reclamation[J]. International Journal of Remote Sensing, 2015, 36(4): 954–978.
- [19] MAXWELL A E, WARNER T A, STRAGER M P, et al. Comparison of naip orthophotography and RapidEye satellite imagery for mapping of mining and mine reclamation[J]. GIScience & Remote Sensing, 2014, 51(3): 301–320.
- [20] MAXWELL A E, WARNER T A, STRAGER M P, et al. Combining RapidEye satellite imagery and LiDAR for mapping of mining and mine reclamation[J]. Photogrammetric Engineering & Remote Sensing, 2014, 80(2): 179–189.
- [21] 田文忠, 赵庆展, 胡浩伟, 等. 无人机高光谱载荷性能交叉验证[J]. 中国测试, 2019, 45(11): 131–137.  
TIAN Wenzhong, ZHAO Qingzhan, HU Haowei, et al. Cross-validation of performance about UAV hyperspectral load[J]. China Measurement & Test, 2019, 45(11): 131–137. (in Chinese)
- [22] BAJCSY P, GROVES P. Methodology for hyperspectral band selection[J]. Photogrammetric Engineering and Remote Sensing, 2004, 70(7): 793–802.
- [23] 张梓童, 王恒, 高世臣, 等. 基于特征选择和梯度提升算法的高光谱遥感地物识别[J]. 数学的实践与认识, 2019, 49(22): 187–196.  
ZHANG Zitong, WANG Heng, GAO Shichen, et al. Hyperspectral image classification based on feature selection and gradient boosting[J]. Mathematics in Practice and Theory, 2019, 49(22): 187–196. (in Chinese)
- [24] 杜培军, 夏俊士, 薛朝辉, 等. 高光谱遥感影像分类研究进展[J]. 遥感学报, 2016, 20(2): 236–256.  
DU Peijun, XIA Junshi, XUE Zhaohui, et al. Review of hyperspectral remote sensing image classification[J]. Journal of Remote Sensing, 2016, 20(2): 236–256. (in Chinese)
- [25] 马娜, 胡云锋, 庄大方, 等. 基于最佳波段指数和 J-M 距离可分性的高光谱数据最佳波段组合选取研究: 以环境小卫星高光谱数据在东莞市的应用为例[J]. 遥感技术与运用, 2010, 25(3): 358–365.  
MA Na, HU Yunfeng, ZHUANG Dafang, et al. Determination on the optimum band combination of H1-1A hyperspectral data in the case region of Dongguan based on optimum index factor and J-M distance[J]. Remote Sensing Technology and Application, 2010, 25(3): 358–365. (in Chinese)
- [26] ANDERW H, KENNETH C, MEGAN L, et al. Hyperspectral classification of plants: a review of waveband selection generalizability[J]. Remote Sensing, 2020, 12(1): 113.
- [27] 李娜, 李咏洁, 赵慧洁, 等. 基于光谱与空间特征结合的改进高光谱数据分类算法[J]. 光谱学与光谱分析, 2014, 34(2): 526–531.  
LI Na, LI Yongjie, ZHAO Huijie, et al. An improved classification approach based on spatial and spectral features for hyperspectral data[J]. Spectroscopy and Spectral Analysis, 2014, 34(2): 526–531. (in Chinese)
- [28] TIAN Y. Monitoring invasion process of spartina alterniflora by seasonal Sentinel-2 imagery and an object-based random forest classification[J]. Remote Sensing, 2020, 12(9): 1383.
- [29] BELGIU M, LUCIAN D. Random forest in remote sensing: a review of applications and future directions[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2016, 114: 24–31.