果 树 学 报 2025,42(12):2898-2909
Journal of Fruit Science

DOI:10.13925/j.cnki.gsxb.20250015

结合光谱变换与特征选择的火龙果株数高光谱遥感提取

郭 松[1],舒 田[1],赵泽英[1*],许元红[1],陈智虎[1],蒋丹垚[2]

([1]贵州省农业科技信息研究所,贵阳 550006;[2]西北农林科技大学资源环境学院,陕西杨凌 712100)

摘 要:【目的】火龙果植株空间信息及株数的快速、无损获取是长势精准监测及区域种植结构调整的重要前提。构建精度较高的火龙果株数提取模型,进一步为贵州省山地特色智慧农业发展提供科研理论支持。**【方法】**使用低空无人机搭载Pika XC2传感器采集贵州省关岭县上官镇火龙果种植区高光谱遥感影像,通过该影像对地表主要地物光谱曲线进行光谱变换以挖掘高光谱影像数据潜力并使用特征选择剔除冗余变量,基于多种机器学习分类模型精确划分研究区不同地物,同时结合实测的火龙果植株投影面积计算其株数。**【结果】**(1)火龙果植株原始高光谱曲线在可见光波长区间反射率较低,近红外波长区间反射率较高;不同类型光谱下火龙果植株与其余地物光谱反射率差异较大的波段不同;(2)以特征距离定义筛选策略的特征选择算法降维效果较好,各区域不同类型光谱下火龙果植株特征波段数量介于2~9个,降维比均在97%以上;(3)地物分类精度以及株数提取精度与地表复杂度成反比。所有分类模型中连续统去除光谱下的随机森林模型精度最高,其总体分类精度与*Kappa*系数分别在84%及0.87以上。该模型下火龙果株数提取效果最好,不同区域提取精度在83%以上。**【结论】**连续统去除变换与随机森林算法的结合可较为准确地识别火龙果植株信息,该思路可为大尺度下喀斯特地区火龙果植株空间信息获取提供技术参考。

关键词:火龙果;无人机高光谱;光谱变换;特征选择;机器学习;株数提取

中图分类号:S667.9 **文献标志码:**A **文章编号:**1009-9980(2025)12-2898-12

Hyperspectral remote sensing extraction of pitaya number by combining spectral transformation and feature selection

GUO Song[1], SHU Tian[1], ZHAO Zeying[1*], XU Yuanhong[1], CHEN Zhihu[1], JIANG Danyao[2]

([1]Guizhou Agricultural Science and Technology Information Institute, Guiyang 550006, Guizhou, China; [2]College of Natural Resources and Environment, Northwest A & F University, Yangling 712100, Shaanxi, China)

Abstract:【Objective】Rapid and non-destructive acquisition of plant spatial information and plant number of pitaya is an important prerequisite for accurate monitoring its growth and adjusting regional planting structure. Traditional field measurement is costly and inefficient, however, hyperspectral remote sensing is simple to operate and the data is more sensitive to vegetation, so it has become an effective means for non-contact acquisition of vegetation spatial information at a large scale.【Methods】The DJI M600 low-altitude UAV equipped with Pika XC2 sensor was used to collect hyperspectral remote sensing images of pitaya growing areas in Shangguan town, Guanling county, Guizhou province. Different regions were divided according to surface complexity and the spectral curves of major surface objects were calculated using Envi 5.3. After Savitzky-Golay second-order smoothing, first derivative spectrum (FDS) and continuum removal spectrum (CRS) were derived to explore the potential of hyperspectral image data, and a feature selection method was proposed to eliminate redundant variables by defining dimension reduction strategy from feature distance. Based on artificial neural network (ANN),

收稿日期:2025-01-21 **接受日期:**2025-09-02

基金项目:科研机构创新能力建设专项(黔科合服企(2021)15号)

作者简介:郭松,男,研究实习员,硕士,研究方向为农业遥感。E-mail:1185716519@qq.com

***通信作者** Author for correspondence. E-mail:605538133@qq.com

support vector machine (SVM) and random forest (RF) machine learning models, different ground objects in the study area were divided, and the plant number was calculated by combining the projected area of pitaya measured on the surface.【Results】The results were as follows: (1) The reflectance of the primary hyperspectral curve of pitaya was lower in the visible wavelength region and higher in the near infrared wavelength region, and the reflectance between them was connected by red edge; The spectral reflectance of pitaya and other ground objects were different in different spectral types. The primary spectrum was located in the "red valley" and "high reflective platform", the first derivative spectrum was located in the "red edge" and "green peak", and the continuum removal spectrum was located in the "red valley" and "green peak". (2) The feature selection algorithm defined by the feature distance had a better dimensionality reduction effect, and the number of feature bands was proportional to the surface complexity. The number of feature bands of pitaya ranged from 2 to 9 under different spectral types in each region, and the dimensionality reduction ratio was all above 97%. The spectral transformation could effectively reduce the number of feature bands and the distance between features. The characteristic bands of different spectral types in each region were mainly concentrated in the "red valley", "red edge" and "near infrared" regions. (3) The classification accuracy of ground objects and the extraction accuracy of plant number were inversely proportional to the surface complexity. Among all classification models, the accuracy of CRS-RF models was the best, and the overall classification accuracy and Kappa coefficient were above 84% and 0.87, respectively, indicating that the training set and the result set were completely consistent. CRS-RF models had the best effect on the number of pitaya, and the accuracy in different regions was above 83.33%.【Conclusion】The combination of continuum removal transformation and random forest algorithm can accurately identify pitaya plant information, which can provide technical reference for obtaining the spatial information of pitaya plants in karst area at a large scale. In practical application, it is only necessary to input hyperspectral remote sensing image of the study area into the trained CRS-RF model, and then the space position and plant number of pitaya in the corresponding region can be output.

Key words: Pitaya; UAV hyperspectral remote sensing; Spectral transformation; Feature selection; Machine learning; Extraction number of pitaya plant

火龙果又名三角柱，是仙人掌科量天尺属灌木，其茎和花均可入药，二者分别具备活血化瘀和清热解毒功效，其果实不仅鲜甜多汁，还可润肠通便、养阴润燥[1]。火龙果原产于中美洲沙漠地区，在19世纪80年代引入中国台湾省、广东省等东南沿海地区[2]，而后逐渐向西南地区扩大种植[3]。火龙果植株空间数据获取是监测长势状况的必要前提，传统的人工实地定位效率低下、成本颇高且难以应用于大尺度地块。随着地区种植规模的不断扩大，如何快速、大面积提取植株信息已成为火龙果种植业急需解决的热点问题。

近年来，遥感技术不断推陈出新，为快速、非接触式植被识别提供了新途径，其原理是从地物反射或发射的电磁辐射信息中提取出区分目标与背景的重要特征，最后基于该特征将二者精准分离[4]。在此方面，国内外学者进行了大量研究。何艺等[5]采用局域最大值法和多尺度分割算法从乐昌林场无人机数字正射影像中获取树木密度，两种算法的分割精度均在90%以上；Al等[6]使用不同尺度下的卷积算法从Wroldview-3卫星遥感影像中提取枣椰树的单位面积株数，发现7 m的卷积尺度识别效果最佳，准确率在88%以上；刘帅兵等[7]对玉米苗期无人机可见光影像进行HSV色彩变换，利用Harris角点检测算法计算株数，最高计算精度达99%；Fareed等[8]联合雷达遥感点云和DSM数据提取人工林木种植行，与单一遥感数据相比，引入DSM可提高单株林木识别模型的鲁棒性；郑晓岚等[9]为获取棉花苗期株数，结合Hough变换与支持向量机回归构建光谱信息与实测株数估算模型，模型决定系数高至0.95；束美艳等[10]借助分水岭算法精准分割无人机RGB影像中

柑橘单株，识别准确率在93%以上；Zheng等[11]以城市人工林为研究对象，提出一种基于YOLOv4-Lite的单树检测方法并通过MobileNetv3卷积神经网络作为主要特征提取器，与分水岭等传统算法相比，模型识别精度高出26%以上；饶雄飞等[12]根据烟草形态中的关键节点构建烟草植株多策略识别算法，该算法在无人机多光谱影像中烟草株数定位精度在99%以上；Wu等[13]建立光谱纹理敏感指数（SPSI）以规避冬小麦光谱饱和以及背景效应问题，发现850 nm、730 nm、675 nm计算的SPSI反演单位面积冬小麦株数效果最佳；Lai等[14]以木瓜园大疆精灵4影像为基础，计算多个卷积窗口下的不同植被指数，发现当卷积窗口为23像素时，Otsu方法识别木瓜株数效果最佳，精度在95.54%。综上所述，基于遥感的植株识别研究已较为全面，数据源包括RGB可见光、多光谱以及雷达点云，采样设备多为卫星和无人机平台，采用的方法由传统阈值分割向机器学习识别扩展，研究对象有田间农作物和森林乔木等。但仍有可补之处：在数据源方面，无人机RGB影像虽然数据采集速度快，但是其数据维度单一；雷达遥感影像可体现植被形态特点，然而其点云仅包含了空间坐标，缺少植被生理特征信息。在株数提取方法方面，卷积、分割、植被指数以及机器学习等算法能挖掘影像内在规律，但数字影像本身的数学变换却尚未利用；高光谱遥感技术波段数量多、可变换形式丰富以及对植株响应强的特点可有效规避上述短板。尽管其庞大的数据量对计算机软硬件存在较高要求，但该技术对目标的各项细节标记较为全面且满足多种光谱变换，在细节提取方面优势明显[15]。

贵州省作为西南地区火龙果主要种植区之一，火龙果已成为当地农业经济的重要组成部分[16]，其株数的快速、无损获取是区域种植结构调整、农户水肥资源快速配置以及产量和产值估算的重要前提。因此，笔者以贵州省关岭县火龙果作为研究对象，基于无人机搭载高光谱传感器采集种植区影像，采用多种方法获取火龙果植株空间信息，经由地面实测的植株面积分割目标对象以提取株数，以期建立一个精度较高的火龙果株数提取方法，进一步为贵州省山地特色智慧农业发展提供科研理论支持。

1 材料和方法

1.1 研究区概况

研究区位于贵州省安顺市关岭县上官镇（105.741111° E，25.747222° N）（图1），平均海拔950 m，年均降水量1300 mm，年均气温16 ℃，全年无霜期超过200 d，气候类型主要为中亚热带季风湿润气候，雨热同期，四季分明，类似“天然温室”，地区

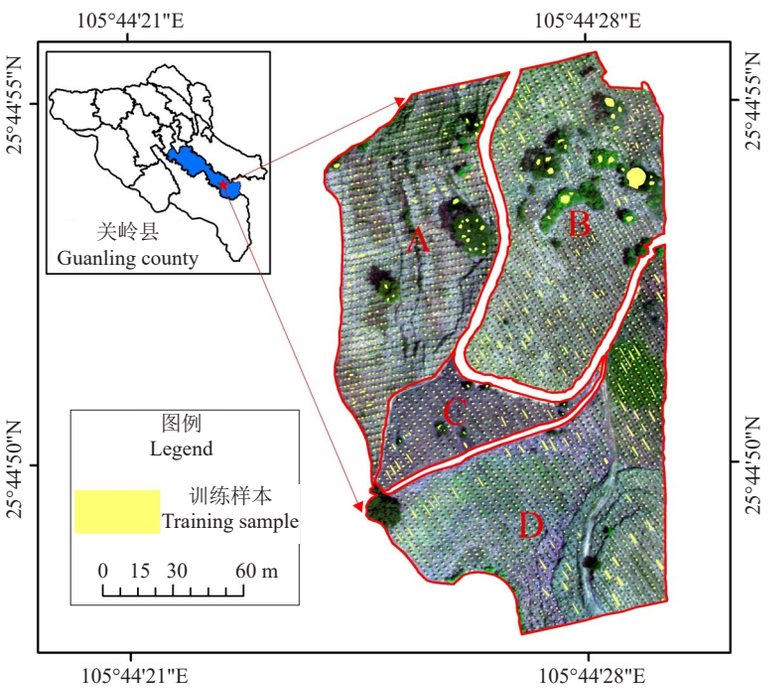

图1 研究区地理位置

Fig. 1 Geographical location of the study area

土壤类型为黄壤，pH呈弱酸性，且富含有机质，适宜火龙果种植。

1.2　项目测定与方法

1.2.1　高光谱影像获取　研究区的火龙果高光谱影像通过大疆M600 Pro云台（中国，深圳市大疆创新科技有限公司）搭载Pika XC2高光谱成像仪（美国，Resonon公司）获取。Pika XC2采用推扫式实时成像，数据采集时镜头自动垂直向下，明暗自动调焦的成像特点可有效避免人为调焦误差，该传感器光谱范围介于400~1000 nm，光谱分辨率1.9 nm，空间通道数1600个，横向视场角与瞬时视场角分别为30.8°、0.041°。数据采集时间为2019年7月25日正午11:00—13:00，天气晴朗无风。火龙果品种为紫红龙，树龄5 a（年），生育期为果实膨大期。无人机起飞前对镜头进行白板校正，并于地表适当位置放置灰阶靶标与空间校正点，飞行高度设定为50 m，飞行速度为4 m·s^{-1}，航向与旁向重叠度分别为70%、80%，飞行1个架次共采集1258张研究区高光谱遥感影像[17]。

1.2.2　火龙果植株投影面积获取　以单株火龙果冠层枝条伸展的最大水平距离为准，通过重力垂线于地面划定整个冠层投影区域并计算其面积。将此面积作为该株火龙果的地表投影面积，整个研究区共测定长势具有代表性的423株火龙果投影面积。

1.3　数据处理

1.3.1　影像数据处理　数据处理包括辐射定标、影像拼接、地理配准、影像降噪。在Pika XC2高光谱影像专业处理软件Spectronon Pro中进行辐射定标，以地表放置的靶标为准，对采集的1258张影像进行辐射校正，赋予影像中每个像元正确反射率；在Pix 4D Mapper软件中进行影像拼接，采用流程化操作将所有影像合并为覆盖整个研究区的单景影像；在ENVI 5.3软件中进行地理配准和降噪，基于地表布置的空间校正点赋予该单景影像正确的地理坐标；同时采用Savitzky-Golay二阶平滑滤波弱化数据采集过程中由设备暗电流以及环境等造成的影像噪声。

1.3.2　地物高光谱反射率数据处理　以研究区高光谱影像为基础数据，在ENVI 5.3软件中基于红（660 nm）、绿（560 nm）、蓝（460 nm）3波段合成可见光影像。根据可见光影像中火龙果种植纹理及地物复杂度等区别划分出A、B、C、D共4个区域，并基于Region of Interest功能在每个区域中选择多个感兴趣区（ROI）作为地物分类训练样本。样本包括目标（火龙果植株）与背景（乔木、裸地等）。统计各ROI中像元的平均反射率，以此作为该ROI所覆盖地物的代表反射率，并据此绘制相应地物的高光谱曲线。光谱变换是放大地物曲线特征的有效手段之一[18]，目前已报道的光谱变换手段，其核心皆是围绕光谱曲线斜率和积分特征。为进一步挖掘高光谱影像地物识别潜力并兼顾变换手段的代表性，笔者采用一阶导数变换量化原始光谱曲线斜率特征，通过连续统去除变换体现原始光谱曲线反射峰和吸收谷积分特征。

贵州省地表复杂且地势起伏较大的特点在一定程度上会掣肘地物遥感识别模型在该地区的普适性[19]。经实地调查，笔者所选区域在地物复杂度、高差以及纹理方面可代表关岭县火龙果种植区基本情况。因此，依据地形及地物分布特征将研究区划分为4个区域，划分结果及ROI分布见图1。各区域特点分别为：A，横向纹理，地类复杂度最小，主要地物4类，高差3~6 m；B，纵向纹理，地类复杂度较高，主要地物10类，无明显高差；C，纵向纹理，地类复杂度较低，主要地物5类，无明显高差；D，纵向纹理，地类复杂度一般，主要地物6类，高差3~15 m。

1.4　模型建立

采用人工神经网络（artificial neural network，ANN）、支持向量机（support vector machine，SVM）以及随机森林（random forest，RF）构建分类模型。其中人工神经网络是对人类神经元活动的抽象处理，主要由输入层、隐含层以及输出层组成，该算法可将复杂的分类问题分解到不同神经元中，经由隐含层与激活函数，映射不同的特征输入到相应的类别输出；支持向量机是将特征集投影到高维空间中，将分类问题转换为“超平面”决策，经由“超平面”的隔绝划分目标对象不同类别；随机森林采用“集思广益”策略，将不同弱分类模型分化到不同“决策树”中，各“决策树”联合形成强分类模型，不同“决策树”可承担目标对象的不同特征，目标对象类别由所有“决策树”共同决定[20]。所有模型的建立均在ENVI 5.3软件中完成。

地物分类精度由*Kappa*系数和总体分类精度*K*进行评价，火龙果株数提取精度通过株数准确率*P*进行评价，三者计算方法见公式1、2和3。其中 p_0

为总体分类精度，是每1类正确分类的像元数之和除以总像元数，p_e为各类别预测像元数和实际像元数乘积之和除以训练像元总数的平方；I为正确分类的像元数目，M为总像元数目；N_0为提取火龙果株数，N为实地调查株数。当*Kappa*系数位于[0~0.2]、(0.2~0.4]、(0.4~0.6]、(0.6~0.8]、(0.8~1.0]时，则分别表示训练集与结果集具有极低一致性、一般一致性、中等一致性、高度一致性和完全一致性。

$$Kappa = \frac{p_0 - p_e}{1 - p_e}, \quad (1)$$

$$K = \frac{I}{M} \times 100\%, \quad (2)$$

$$P = \frac{N_0}{N} \times 100\%。 \quad (3)$$

2 结果与分析

2.1 主要地物高光谱特征

2.1.1 *原始光谱* 计算A、B、C、D各区域中不同地物ROI的平均反射率，获得研究区地物原始光谱特征，如图2。对比不同区域，B区地物种类最多，主要包含火龙果植株在内的10种地物，其次D区6种、C区5种，最少为A区4种。整体上看，不同区域相同地物高光谱曲线特点一致。以B区为例，不同类型乔木光谱曲线为典型植被光谱曲线，曲线形似根号，具有明显“两谷一峰一平台”特点，即蓝谷(420 nm)、红谷(680 nm)、绿峰(560 nm)以及高反射平台(780~880 nm)；裸地带有极少杂草光谱曲线可近似

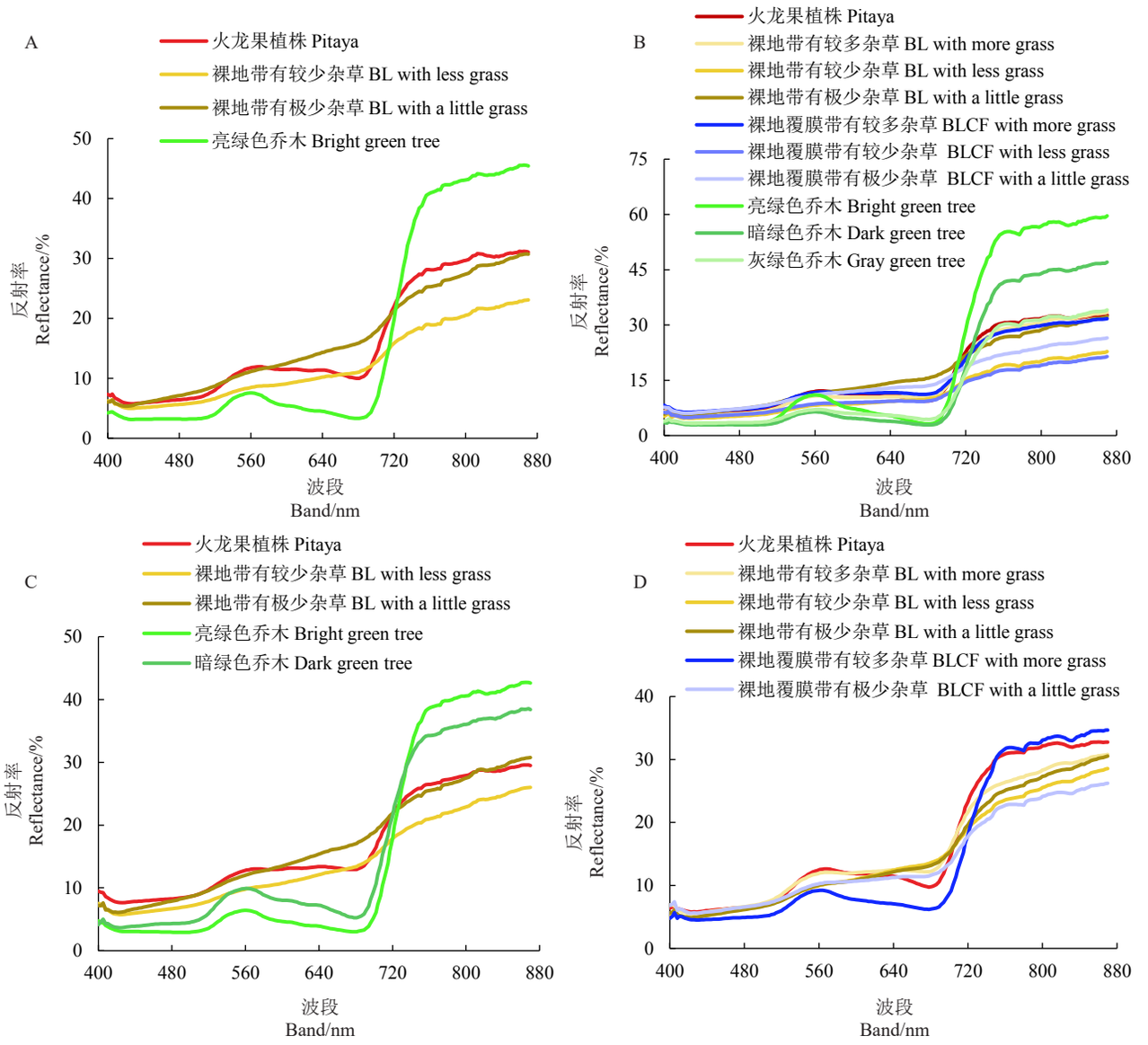

图2 研究区主要地物原始光谱特征

Fig. 2 Primary spectral characteristics of main ground objects in the experiment area

为缓慢上升的曲线，其随着杂草的增多而逐渐逼近植被光谱曲线，该特点与裸地覆膜地物光谱特征一致，但二者在近红外区域（780~880 nm）有所不同，该区域裸地覆膜光谱曲线平滑度整体上优于裸地光谱曲线。火龙果植株光谱曲线与乔木光谱曲线类似，然而由于火龙果植株冠层葱郁度远低于乔木，影像上像元纯度较低。因此与乔木相比，火龙果植株在整个可见光区域（400~650 nm）反射率高于乔木，而近红外区域反射率低于乔木。综合来看，火龙果植株在红谷和高反射平台区域与其余地物的原始光谱反射率差异较大。

2.1.2　一阶导数光谱　一阶导数变换可放大原始光谱曲线中反射率突变的位置，一阶导数光谱曲线由原始光谱曲线斜率值组成，相较于原始光谱，其物理意义已发生变化。以 B 区火龙果植株光谱为例（图 3），一阶导数光谱上 520 nm、580 nm 以及 720 nm 处的极值点分别对应绿峰两端以及红谷向高反射平台过渡的红边。总体上，经一阶导数变换后，不同区域

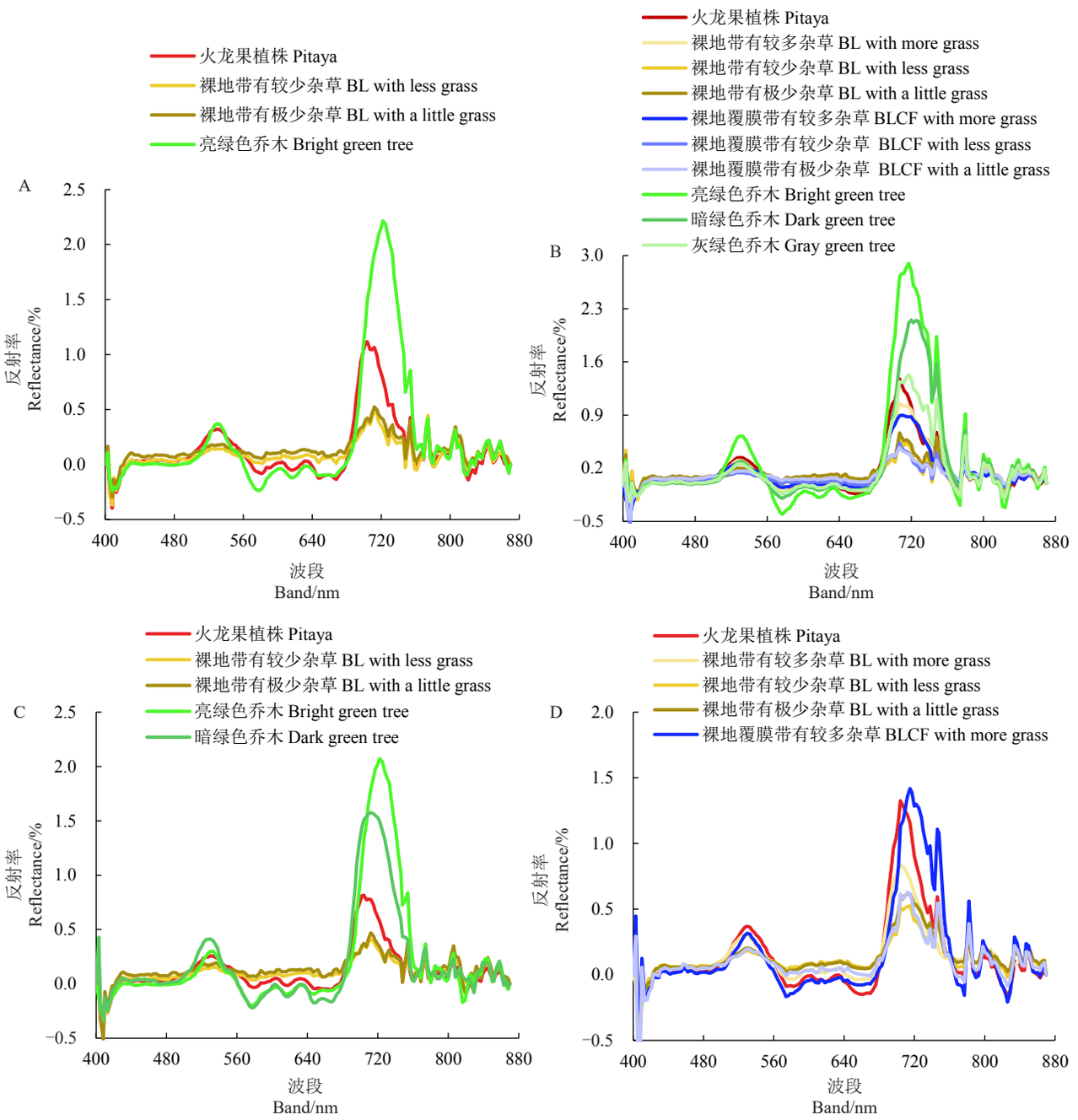

图 3　研究区主要地物一阶导数光谱特征

Fig. 3　First derivative spectral characteristics of main ground objects in the experiment area

火龙果植株与其余地物曲线值差异较大的区域主要集中于红边和绿峰。

2.1.3 连续统去除光谱 连续统去除变换原理为以直线方式连接原始光谱曲线上各极值点，连接时要保证连接得到的折线在各极值点处的外角大于180°，连接完成后将原始光谱反射率除以折线值即得到连续统值。连续统去除变换重点放大原始光谱曲线中的反射与吸收特征，如图4。与原始光谱相比，不同区域中各地物反射峰与吸收谷被放大。同时，不同区域火龙果植株连续统去除光谱在峰与谷的位置与其余地物的差异也被拉大，相应代表波段为红谷和绿峰。

2.2 火龙果植株特征波段选择

将火龙果植株光谱曲线与其余地物光谱曲线之间垂直距离最大的波段定义为特征波段（nm），其对应的垂直距离为特征距离。表1为各研究区火龙果植株特征波段和相应特征距离。从特征波段数量看，不同区域火龙果植株特征波段数量介于2~9个，

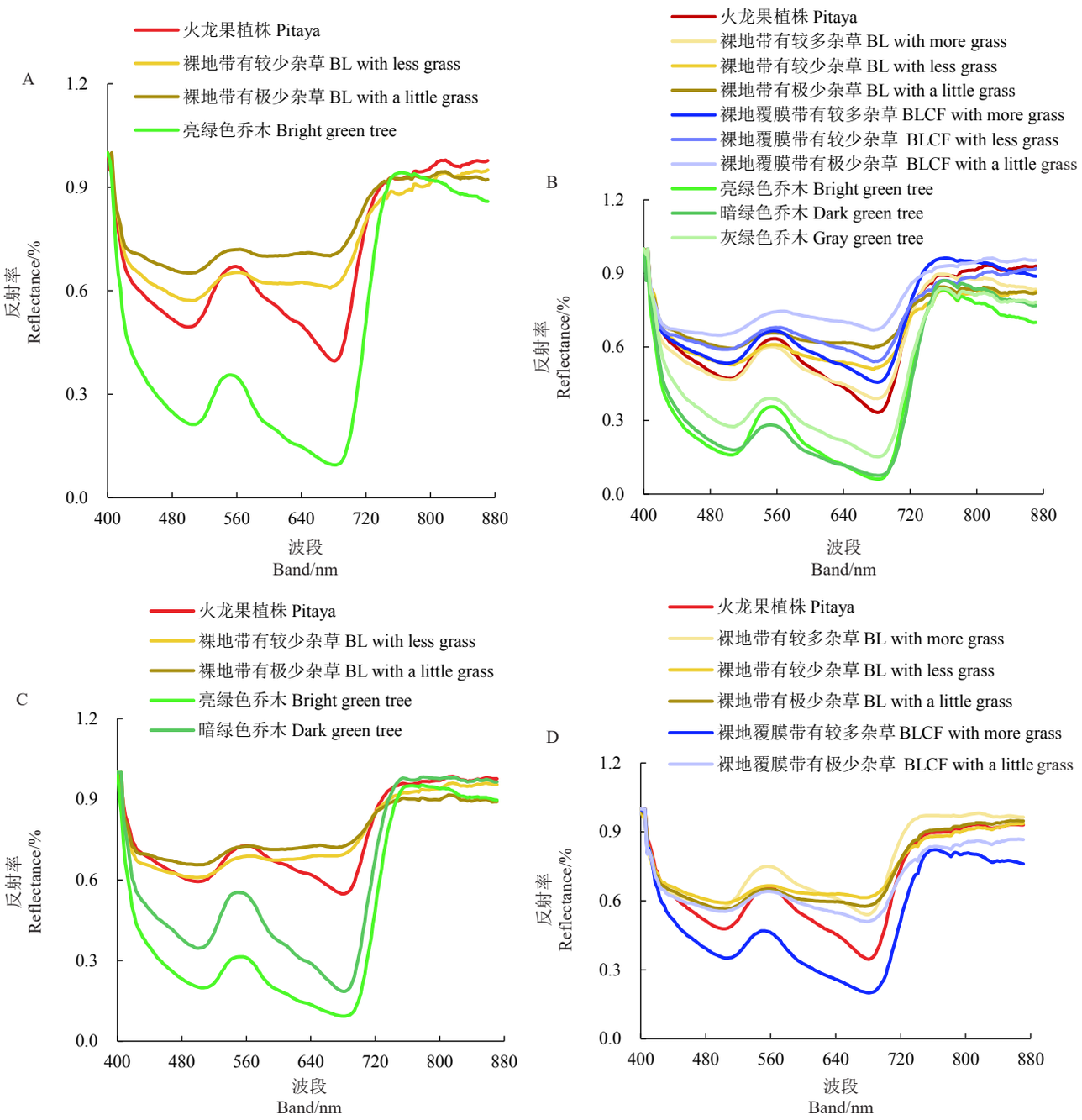

图4 研究区主要地物连续统去除光谱特征

Fig. 4 Continuum removal spectral characteristics of main ground objects in the experiment area

表1　不同类型光谱下火龙果识别特征波段

Table 1　Identification characteristic bands of pitaya under different spectral types

区域 Area	光谱类型 Spectrum type	特征波段（特征距离） Characteristic bands and characteristic distance	波段数量 Number of band	降维比 Ratio of dimension reduction/%
A	PS	483（9.26）、676（5.96）、556（14.40）	3	98.70
	FD	702（0.74）、702（0.70）、728（1.53）	2	99.13
	CR	683（0.22）、683（0.31）、704（0.36）	2	99.13
B	PS	722（1.81）、780（11.87）、683（6.33）、756（2.49）、782（12.94）、759（8.53）、868（26.89）、870（14.33）、707（6.65）	9	96.10
	FD	707（0.32）、704（0.84）、704、（0.76）、707（0.48）、707（0.84）、707（0.90）、720（1.78）、730（1.44）、738（0.45）	5	97.84
	CR	871（0.09）、683（0.18）、683（0.27）、681（0.12）、683（0.22）、683（0.34）、595（0.33）、574（0.36）、577（0.25）	6	97.40
C	PS	754（5.56）、686（4.24）、865（13.17）、868（8.98）	4	98.27
	FD	704（0.47）、704（0.44）、725（1.55）、722（0.92）	3	98.70
	CR	686（0.15）、683（0.18）、702（0.49）、683（0.36）	3	98.70
D	PS	761（4.51）、759（5.79）、709（6.27）、764（7.30）、780（8.37）	5	97.84
	FD	704（0.49）、704（0.75）、748（0.55）、704（0.82）、702（0.75）	3	98.70
	CR	689（0.20）、683（0.23）、712（0.24）、683（0.27）、683（0.17）	3	98.70

注：PS、FD、CR 分别指原始、一阶导数以及连续统去除光谱。下同。

Note: PS, FD, CR refer to primary spectrum, first derivative spectrum and continuum removal spectrum, respectively. The same below.

降维比均在97%以上，特征波段数量与地表复杂度成正比。各类型光谱下，特征波段数量均为B区最多，A区最少；同一区域内，原始光谱特征波段数量大于一阶导数光谱和连续统去除光谱，相应特征距离表现为原始光谱最大，其余二者较小，说明光谱变换在有效精简特征波段数量的同时也会缩小特征距离；从特征波段位置看，各区域不同类型光谱下特征波段主要集中于红谷、红边以及近红外区域，表明在原始以及变换光谱下，这些区域是区分火龙果植株与地表背景的重要区域。

2.3　地物分类模型构建及精度对比

将表1中特征波段作为输入变量，以图1中划定ROI作为训练样本，通过人工神经网络（ANN）、支持向量机（SVM）以及随机森林（RF）提取地表火龙果植株信息。由于ANN中关键参数为隐含层数目、随机森林为分类树棵数、支持向量机是核函数类型以及*Gamma*系数，因此本研究中以上参数均与特征数相关，其余参数采用默认值。其中ANN算法隐含层数目为特征波段数，训练贡献阈值与速率分别采用默认值0.9、0.2；RF算法中数目棵数为特征波段数，杂质函数选择gini；SVM算法中核函数为径向基函数，*Gamma*参数为特征波段数的倒数，惩罚系数*C*为默认值100。所有算法迭代次数统一设定为特征波段数的10倍。

2.3.1　*原始光谱*　表2为原始光谱下不同地物分类精度，整体上各区域总体精度和*Kappa*系数分别介于73.18%~89.03%、0.741~0.872。从不同区域看，A区分类精度最好，其次为C区和D区，最差为B区；对比不同分类模型，各区域分类效果均为RF＞SVC＞ANN。ANN模型中除A区外，其余区域总体精度均未超过80%；SVC模型中仅B区未超过80%；所有RF模型总体精度均大于81%，相应*Kappa*系数最低为0.812，表明经RF模型处理后，各区域训练集与结果集几乎完全一致，地物识别效果较好。

2.3.2　*变换光谱*　光谱变换可放大原始光谱细节特征并弱化背景噪声。表3为基于变换光谱的地物分类情况。在模型精度方面，相较于原始光谱（表2），一阶导数光谱下A区ANN模型、B区和D区SVC模型高于相应原始光谱下的各类模型，总体精度分别高出1.56、1.87以及1.20百分点，*Kappa*系数分别高出0.013、0.065和0.017；D区RF模型与原始光谱相比表现出总体精度降低0.28百分点，而*Kappa*系数提高0.016的特点；一阶导数光谱下的其余模型总体精度和*Kappa*系数均低于原始光谱。从光谱类型看，连续统去除光谱中各区域所有模型精度均优于原始光谱和一阶导数光谱。与原始光谱相比，其总体精度高出值介于1.59~3.93百分点，*Kappa*系数高出值介于0.023~0.079；与一阶导数光谱相比，其总

表2 原始光谱分类精度

Table 2 Classification accuracy of primary spectrum

区域 Area	分类模型 Classification model	总体精度 Total accuracy/%	*Kappa* 系数 *Kappa* coefficient
A	ANN	80.51	0.793
	SVC	85.14	0.857
	RF	89.03	0.872
B	ANN	73.18	0.744
	SVC	74.39	0.753
	RF	81.33	0.821
C	ANN	77.31	0.789
	SVC	83.42	0.827
	RF	85.19	0.862
D	ANN	75.03	0.741
	SVC	81.88	0.784
	RF	84.07	0.812

体精度和*Kappa*系数高出值分别位于0.88~5.56百分点、0.006~0.149区间。从建模方法看，与原始光谱相同，一阶导数及连续统去除光谱下同区域内均为RF优于SVC，而ANN最差。总体上变换光谱中各区域最优分类模型为RF模型，连续统去除光谱下A区、B区、C区以及D区相应RF模型精度分别为（92.01%，0.926）、（84.67%，0.871）、（86.78%，0.889）、（87.61%，0.875）。

2.4 火龙果株数提取

结合表2和表3可知，原始及变换光谱下均为RF模型表现较好。因此，选择不同类型光谱下RF模型分类结果作为火龙果株数提取自变量，提取结果如表4。从提取株数看，所有区域的火龙果估测株数均低于真实值；从株数准确率看，相同区域内CRS＞PS＞FDS，该规律与地物分类精度规律一致；同类型光谱下，株数准确率均为C、A区较好，D、B区较差，该特点与地物分类精度规律不同，可能是B区与D区地表高差较大和地物复杂度较高导致的影像垂直误差以及分类不准造成的。总体上，A、B、C、D各区域火龙果株数均为连续统去除光谱下估测准确率最高，分别为91.56%、83.33%、95.15%、86.64%。

3 讨 论

火龙果植株原始光谱曲线可近似为“根号”，在可见光波段反射率低、近红外波段反射率高，与乔木等背景地物的反射率差异是识别火龙果植株的数学基础[21]。有效放大并准确定位地物反射率差异是提高模型识别精度的关键，加之高光谱数据波段多而窄，数据量庞大的特性会导致“维度灾难”。因此，笔者从光谱变换及光谱降维两个角度进行研

表3 变换光谱分类精度

Table 3 Classification accuracy of transformation spectrum

光谱类型 Spectrum type	区域 Area	分类模型 Classification model	总体精度 Total accuracy/%	*Kappa* 系数 *Kappa* coefficient
FDS	A	ANN	82.07	0.806
		SVC	83.57	0.825
		RF	86.98	0.812
	B	ANN	72.06	0.705
		SVC	76.26	0.818
		RF	79.53	0.773
	C	ANN	75.15	0.713
		SVC	82.26	0.732
		RF	83.72	0.807
	D	ANN	73.34	0.686
		SVC	83.08	0.801
		RF	83.79	0.828
CRS	A	ANN	84.02	0.816
		SVC	86.85	0.923
		RF	92.01	0.926
	B	ANN	74.77	0.781
		SVC	77.14	0.824
		RF	84.67	0.871
	C	ANN	79.16	0.862
		SVC	87.31	0.872
		RF	86.78	0.889
	D	ANN	78.96	0.767
		SVC	84.77	0.863
		RF	87.61	0.875

表4 火龙果株数提取结果

Table 4 Extraction results of pitaya number

光谱类型 Spectrum type	区域 Area	实际株数 True number	提取株数 Extraction number	准确率 Accuracy/%
PS	A	1292	1126	87.15
	B	1296	1054	81.33
	C	309	279	90.29
	D	1175	981	83.49
FDS	A	1292	1120	86.69
	B	1296	963	74.31
	C	309	272	88.03
	D	1175	966	82.21
CRS	A	1292	1183	91.56
	B	1296	1080	83.33
	C	309	294	95.15
	D	1175	1018	86.64

究。

光谱变换是派生高光谱数据的常用手段之一，笔者采用一阶导数以及连续统去除变换。分类结果显示，基于连续统去除变换的火龙果识别以及株数提取精度优于原始光谱和一阶导数光谱，表明光谱变换在火龙果高光谱遥感识别中具有一定潜力，且火龙果识别的关键波段位于光谱曲线的波峰与波谷。该结论与前人[22-24]的研究结果类似，原因在于连续统去除变换可放大地物光谱曲线反射峰与吸收谷，一阶导数变换可量化光谱曲线的斜率，但火龙果植株与背景地物的光谱差异主要体现在反射峰与吸收谷而非曲线斜率，因此连续统去除变换优于原始光谱与一阶导数光谱。然而，笔者仅采取了两种较为传统的光谱变换方式，目前已有学者发现MTF、伪缨帽变换等新型变换方式可提高遥感影像地物识别精度[25-26]，未来应将对数变换、傅里叶变换等更多的变换方式应用到火龙果高光谱识别研究中，同时针对火龙果高光谱曲线波峰以及波谷位置的新型光谱变换算法开发也是可深入研究的领域之一。

当前高光谱降维方法主要包含特征提取以及特征选择，前者通过光谱空间旋转将高维特征投影到低维坐标系，如主成分分析、线性判别分析等；后者则采用特定策略从光谱空间中筛选目标变量，如连续投影算法、竞争性自适应重加权算法等。与特征选择相比，特征提取存在目标变量物理意义难以明确的短板，因此笔者提出一种以特征距离定义判别式的特征选择策略，使用该方法降维后的不同类型光谱下各区域降维比均在97%以上，对比刘子健等[27]和He等[28]的研究，虽降维效果较优，但传感器与研究区的不同也导致了特征波段位置不同。除高光谱特征外，火龙果植株在叶绿素含量、株高等植物性状方面与其他植被地类同样存在显著差异，后续火龙果植株识别研究可从定性与定量相结合的角度出发，在降低冗余特征的同时也应考虑其他有效特征的引入。不同类型光谱下各区域分类精度与地表复杂度成反比，相同区域内随机森林算法性能优于支持向量机和神经网络算法，该特点与徐祖平等[29]、李连发等[30]的研究一致，但由于气候、火龙果生育期、仪器以及数据采集时间等不同，本研究结果中的各分类模型以及数据处理方法是否适用于其他范式还有待通过大量基础研究进一步验证，同时，受制于研究区域类型和训练样本数量较少，地表复杂度与分类精度的定量关系并未明确，且在农业大数据背景下，机器学习算法运行的时间也是不可忽略的成本之一，因此模型普适性提升研究、参数寻优算法与分类算法的结合以及多区域类型和大量训练样本积累应是下一步研究的重点。

4　结　论

笔者以贵州省关岭县火龙果作为研究对象，基于特征选择与机器学习算法从火龙果种植区无人机高光谱影像中获取火龙果植株空间信息，经由地面实测的植株投影面积分割植株对象以提取火龙果株数，主要结论如下：

（1）火龙果植株原始高光谱曲线在可见光波长区间反射率较低，近红外波长区间反射率较高，二者间由红边过渡；不同类型光谱下火龙果植株与其余地物光谱反射率差异较大的波段不同，其中原始光谱位于红谷和高反射平台，一阶导数光谱位于红边和绿峰，连续统去除光谱位于红谷和绿峰。

（2）以特征距离定义筛选策略的特征选择算法降维效果较好，特征波段数量与地表复杂度成正比，各区域不同类型光谱下火龙果植株特征波段数量介于2~9个，降维比均在97%以上；光谱变换可有效精简特征波段数量并缩小特征距离；各区域不同类型光谱下特征波段主要集中于红谷、红边以及近红外区域。

（3）总体上地物分类精度以及株数提取精度与地表复杂度成反比。所有分类模型中，CRS-RF模型精度最好，总体分类精度与*Kappa*系数分别在84%及0.87以上，表明训练集与结果集具有完全一致性；CRS-RF模型下火龙果株数提取效果最佳，各区域精度分别为91.56%（A区）、83.33%（B区）、95.15%（C区）、86.64%（D区）。

由此可见，连续统去除变换与随机森林算法的结合可较为准确地识别火龙果植株信息，该思路可为大尺度下喀斯特地区火龙果植株空间信息获取提供技术参考。

参考文献 References：

[1] 谢国芳，周俊良，韩秀梅，张银，杨月 . 火龙果营养研究及开发现状[J]. 食品工业，2013，34(6)：171-174.
XIE Guofang，ZHOU Junliang，HAN Xiumei，ZHANG Yin，YANG Yue. Nutritional ingredient and exploitation of pitaya[J]. The Food Industry，2013，34(6)：171-174.

[2] 潘波，王冰洁，姜蕾，林勇 . 两种植保无人机对火龙果冠层的作业参数优化[J]. 植物保护学报，2021，48(3)：528-536.
PAN Bo，WANG Bingjie，JIANG Lei，LIN Yong. Optimization of the operational parameters of two types of plant protection unmanned aerial vehicles (UAVs) application to the pitaya canopy[J]. Journal of Plant Protection，2021，48(3)：528-536.

[3] 李莉婕，岳延滨，王延仓，赵泽英，黎瑞君，聂克艳，袁玲 . 高光谱定量反演火龙果茎枝叶绿素含量的研究[J]. 光谱学与光谱分析，2021，41(11)：3538-3544.
LI Lijie，YUE Yanbin，WANG Yancang，ZHAO Zeying，LI Ruijun，NIE Keyan，YUAN Ling. The quantitative study on chlorophyll content of *Hylocereus polyrhizus* based on hyperspectral analysis[J]. Spectroscopy and Spectral Analysis，2021，41(11)：3538-3544.

[4] 张裕，杨海涛，袁春慧 . 遥感图像分类方法综述[J]. 兵器装备工程学报，2018，39(8)：108-112.
ZHANG Yu，YANG Haitao，YUAN Chunhui. A survey of remote sensing image classification methods[J]. Journal of Ordnance Equipment Engineering，2018，39(8)：108-112.

[5] 何艺，周小成，黄洪宇，许雪琴 . 基于无人机遥感的亚热带森林林分株数提取[J]. 遥感技术与应用，2018，33(1)：168-176.
HE Yi，ZHOU Xiaocheng，HUANG Hongyu，XU Xueqin. Counting tree number in subtropical forest districts based on UAV remote sensing images[J]. Remote Sensing Technology and Application，2018，33(1)：168-176.

[6] AL SHIDI R H，KUMAR L，AL-KHATRI S A H，ALBAHRI M M，ALAUFI M S. Relationship of date palm tree density to dubas bug *Ommatissus lybicus* infestation in Omani orchards[J]. Agriculture，2018，8(5)：64.

[7] 刘帅兵，杨贵军，周成全，景海涛，冯海宽，徐波，杨浩 . 基于无人机遥感影像的玉米苗期株数信息提取[J]. 农业工程学报，2018，34(22)：69-77.
LIU Shuaibing，YANG Guijun，ZHOU Chengquan，JING Haitao，FENG Haikuan，XU Bo，YANG Hao. Extraction of maize seedling number information based on UAV imagery[J]. Transactions of the Chinese Society of Agricultural Engineering，2018，34(22)：69-77.

[8] FAREED N，REHMAN K. Integration of remote sensing and GIS to extract plantation rows from a drone-based image point cloud digital surface model[J]. ISPRS International Journal of Geo-Information，2020，9(3)：151.

[9] 郑晓岚，张显峰，程俊毅，任翔 . 利用无人机多光谱影像数据构建棉苗株数估算模型[J]. 中国图象图形学报，2020，25(3)：520-534.
ZHENG Xiaolan，ZHANG Xianfeng，CHENG Junyi，REN Xiang. Using the multispectral image data acquired by unmanned aerial vehicle to build an estimation model of the number of seedling stage cotton plants[J]. Journal of Image and Graphics，2020，25(3)：520-534.

[10] 束美艳，李世林，魏家玺，车荧璞，李保国，马韫韬 . 基于无人机平台的柑橘树冠信息提取[J]. 农业工程学报，2021，37(1)：68-76.
SHU Meiyan，LI Shilin，WEI Jiaxi，CHE Yingpu，LI Baoguo，MA Yuntao. Extraction of citrus crown parameters using UAV platform[J]. Transactions of the Chinese Society of Agricultural Engineering，2021，37(1)：68-76.

[11] ZHENG Y Y，WU G. YOLOv4-lite-based urban plantation tree detection and positioning with high-resolution remote sensing imagery[J]. Frontiers in Environmental Science，2022，9：756227.

[12] 饶雄飞，周龙宇，杨春雷，廖世鹏，李小坤，刘诗诗 . 基于无人机多光谱影像和关键点检测的雪茄烟株数提取[J]. 农业机械学报，2023，54(3)：266-273.
RAO Xiongfei，ZHOU Longyu，YANG Chunlei，LIAO Shipeng，LI Xiaokun，LIU Shishi. Counting cigar tobacco plants from UAV multispectral images *via* key points detection approach[J]. Transactions of the Chinese Society for Agricultural Machinery，2023，54(3)：266-273.

[13] WU Y P，WANG W H，GU Y Y，ZHENG H B，YAO X，ZHU Y，CAO W X，CHENG T. SPSI：A novel composite index for estimating panicle number in winter wheat before heading from UAV multispectral imagery[J]. Plant Phenomics，2023，5：0087.

[14] LAI S S，MING H L，HUANG Q Y，QIN Z H，DUAN L，CHENG F，HAN G P. Remote sensing extraction of crown planar area and plant number of papayas using UAV images with very high spatial resolution[J]. Agronomy，2024，14(3)：636.

[15] 李锋霞，黄勇，李强 . 光谱检测哈密瓜品质中异常样本的综合分析[J]. 中国瓜菜，2023，36(7)：18-23.
LI Fengxia，HUANG Yong，LI Qiang. Comprehensive analysis of abnormal samples in Hami melon quality based on spectral detection[J]. China Cucurbits and Vegetables，2023，36(7)：18-23.

[16] 张义鑫 . 西南地区数字经济与乡村振兴的耦合协调发展研究[D]. 贵阳：贵州大学，2023.
ZHANG Yixin. Research on the coupling and coordinated development of digital economy and rural revitalization in southwest China[D]. Guiyang：Guizhou University，2023.

[17] 杨红艳，杜健民，阮培英，朱相兵，刘浩，王圆 . 基于无人机遥感与随机森林的荒漠草原植被分类方法[J]. 农业机械学报，2021，52(6)：186-194.
YANG Hongyan，DU Jianmin，RUAN Peiying，ZHU Xiangbing，LIU Hao，WANG Yuan. Vegetation classification of desert steppe based on unmanned aerial vehicle remote sensing and random forest[J]. Transactions of the Chinese Society for Agricultural Machinery，2021，52(6)：186-194.

[18] ZHOU J C，FENG Z Y，LI Y P，WANG J L，MENG X R，LIU Y，QIU S B. Identification of dominant tree species based on Resource-1 02D hyperspectral image data[J]. Frontiers in Earth Science，2024，12：1418865.

[19] 何邦科，朱文泉，史培军，张慧，刘若杨，杨欣怡，赵涔良 . 复杂地形区植被覆盖度遥感精细估算方法：以青藏高原山地区为

例[J]. 生态学报，2024，44(20)：9039-9052.
HE Bangke，ZHU Wenquan，SHI Peijun，ZHANG Hui，LIU Ruoyang，YANG Xinyi，ZHAO Cenliang. A fine-scale remote sensing estimation method for fractional vegetation cover in complex terrain areas：A case study in the Qinghai-Tibet Plateau mountainous regions[J]. Acta Ecologica Sinica，2024，44(20)：9039-9052.

[20] 李文雄，靳瑰丽，刘文昊，马建，李嘉欣，王生菊，陈梦甜 . 基于不同机器学习算法伊犁绢蒿荒漠草地主要地物的高光谱分类[J]. 草业科学，2025，42(1)：35-43.
LI Wenxiong，JIN Guili，LIU Wenhao，MA Jian，LI Jiaxin，WANG Shengju，CHEN Mengtian. Hyperspectral classification of main features of *Seriphidium transiliense* desert grassland based on different machine learning algorithms[J]. Pratacultural Science，2025，42(1)：35-43.

[21] LOU C，AL-QANESS M A A，AL-ALIMI D，DAHOU A，ABD ELAZIZ M，ABUALIGAH L，EWEES A A. Land use/land cover (LULC) classification using hyperspectral images：A review[J/OL]. Geo-Spatial Information Science，2024：1-42(2024-04-15). https://doi.org/10.1080/10095020.2024.2332638.

[22] YANG H Y，DU J M. Classification of desert steppe species based on unmanned aerial vehicle hyperspectral remote sensing and continuum removal vegetation indices[J]. Optik，2021，247：167877.

[23] 刘炜，孙海霞，杨晓波，董建民 . 对数变换、导数变换的高寒草地反射光谱特征分析与识别：以那曲地区 HJ-1A/HSI 图像为例[J]. 光谱学与光谱分析，2020，40(7)：2200-2207.
LIU Wei，SUN Haixia，YANG Xiaobo，DONG Jianmin. Spectral reflectance characteristics of alpine grassland based on derivative and logarithmic transform spectra：Take HJ-1A/HSI images of Naqu prefecture as an example[J]. Spectroscopy and Spectral Analysis，2020，40(7)：2200-2207.

[24] 徐念旭，田庆久，申怀飞，徐凯健 . 基于微分变换的高光谱马尾松和杉木识别[J]. 国土资源遥感，2018，30(4)：28-32.
XU Nianxu，TIAN Qingjiu，SHEN Huaifei，XU Kaijian. Classification of *Pinus massoniana* and *Cunninghamia lanceolata* using hyperspectral image based on differential transformation[J]. Remote Sensing for Land & Resources，2018，30(4)：28-32.

[25] DENG J Q，DONG W Z，GUO Y W，CHEN X Y，ZHOU R H，LIU W Y. A novel remote sensing image enhancement method，the pseudo-tasseled cap transformation：Taking buildings and roads in GF-2 as an example[J]. Applied Sciences，2023，13(11)：6585.

[26] 张蔚，冯巍巍，蔡宗岐，王焕卿，闫奇，王清 . 基于 MTF 变换的拉曼光谱和卷积神经网络的海水微塑料识别方法研究[J]. 光谱学与光谱分析，2024，44(9)：2420-2427.
ZHANG Wei，FENG Weiwei，CAI Zongqi，WANG Huanqing，YAN Qi，WANG Qing. Study on recognition of marine microplastics using Raman spectra combined with MTF and CNN[J]. Spectroscopy and Spectral Analysis，2024，44(9)：2420-2427.

[27] 刘子健，顾佳盛，周聪，王游游，杨健，黄俊，王宏鹏，白瑞斌 . 基于高光谱成像技术的山楂产地判别研究[J]. 食品工业科技，2024，45(10)：282-291.
LIU Zijian，GU Jiasheng，ZHOU Cong，WANG Youyou，YANG Jian，HUANG Jun，WANG Hongpeng，BAI Ruibin. Identification of geographical origin for hawthorn based on hyperspectral imaging technology[J]. Science and Technology of Food Industry，2024，45(10)：282-291.

[28] HE G，YANG S B，WANG Y Z. A rapid method for identification of *Lanxangia tsaoko* origin and fruit shape：FT-NIR combined with chemometrics and image recognition[J]. Journal of Food Science，2024，89(4)：2316-2331.

[29] 徐祖平，舒朗朗，吴文桂，王子芝，程鑫萌，廖声熙 . 基于 GEE 的香格里拉草地分类及其生物量遥感估算[J]. 草业科学，2024，41(10)：2250-2262.
XU Zuping，SHU Langlang，WU Wengui，WANG Zizhi，CHENG Xinmeng，LIAO Shengxi. Classification of Shangri-La grasslands based on Google Earth Engine and remote sensing estimation of their biomass[J]. Pratacultural Science，2024，41(10)：2250-2262.

[30] 李连发，高茜琳，贺伟，陈淼淼，杨晓梅，王志华，张俊瑶，刘晓亮 . 自然资源遥感智能解译中复杂场景样本抽样方法研究[J]. 地球信息科学学报，2025，27(2)：331-349.
LI Lianfa，GAO Xilin，HE Wei，CHEN Miaomiao，YANG Xiaomei，WANG Zhihua，ZHANG Junyao，LIU Xiaoliang. Sampling method for complex scene samples in the intelligent interpretation of natural resources remote sensing[J]. Journal of Geo-Information Science，2025，27(2)：331-349.