



武汉大学学报(信息科学版)

*Geomatics and Information Science of Wuhan University*

ISSN 1671-8860, CN 42-1676/TN

## 《武汉大学学报(信息科学版)》网络首发论文

题目：融合颜色特征的随机森林特征优选的黄河三角洲植被信息分类  
作者：刘逸娴，刘庆生，张新，黄翀，李贺，车纯广，陈一  
DOI：10.13203/j.whugis20240203  
收稿日期：2024-09-25  
网络首发日期：2024-10-12  
引用格式：刘逸娴，刘庆生，张新，黄翀，李贺，车纯广，陈一. 融合颜色特征的随机森林特征优选的黄河三角洲植被信息分类[J/OL]. 武汉大学学报(信息科学版). <https://doi.org/10.13203/j.whugis20240203>



**网络首发：**在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

**出版确认：**纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

DOI:10.13203/j.whugis20240203

引用格式：

刘逸娴, 刘庆生, 张新, 等. 融合颜色特征的随机森林特征优选的黄河三角洲植被信息分类[J].武汉大学学报(信息科学版),2024,DOI: 10.13203/j.whugis20240203 (LIU Yixian, LIU Qingsheng, ZHANG Xin, et al. Classification of Vegetation Information by Integrating Color Features with Multi-feature Optimization of Random Forest in the Yellow River Delta[J].Geomatics and Information Science of Wuhan University,2024,DOI: 10.13203/j.whugis20240203)

## 融合颜色特征的随机森林特征优选的黄河三角洲植被信息 分类

刘逸娴<sup>1,2</sup> 刘庆生<sup>1,3,4</sup> 张新<sup>5\*</sup> 黄 翀<sup>1</sup> 李 贺<sup>1</sup> 车纯广<sup>6</sup> 陈 一<sup>7</sup>

1 中国科学院地理科学与资源研究所 资源与环境信息系统国家重点实验室, 北京, 100101

2 中国科学院大学, 北京, 100049

3 自然资源要素耦合过程与效应重点实验室, 北京, 100055

4 江苏省地理信息资源开发与利用协同创新中心, 江苏 南京, 210023

5 空天信息创新研究 遥感科学国家重点实验室, 北京, 100101

6 黄河三角洲国家级自然保护区黄河口管理站, 山东 东营, 257500

7 北京市测绘设计研究院, 北京, 100038

**摘 要：**黄河三角洲湿地盐沼植被的监测是黄河三角洲湿地生态功能保护与恢复的基础。本文以黄河三角洲湿地部分区域为研究区, 以高分辨率航空影像为数据源, 生成了光谱特征、颜色特征、指数特征和纹理特征四种特征变量并构建不同的分类方案, 利用随机森林对每种提取方案进行植被分类并验证其精度, 旨在探求不同特征变量对分类的影响及原因, 以选取最佳的优选特征改善植被分类的效果。结果表明:(1) 指数特征对碱蓬提取有积极作用, 纹理特征会降低植被分类的精度, 融合颜色特征进行分类是提高总体精度的关键。(2) 基于随机森林特征优选提取效果最佳, 总体精度为 88%, Kappa 系数为 0.85。该方法能有效区分植被与非植被, 同时将各植被类型提取出来。本研究为黄河三角洲植被信息提取在特征选取与方法上提供了一种有效的技术路线。

**关键词：**黄河三角洲; 航空影像; 随机森林; 植被信息提取; 颜色特征

## Classification of Vegetation Information by Integrating Color Features with Multi-feature Optimization of Random Forest in the Yellow River Delta

LIU Yixian<sup>1,2</sup> LIU Qingsheng<sup>1,3,4</sup> ZHANG Xin<sup>5\*</sup> HUANG Chong<sup>1</sup> LI He<sup>1</sup> CHE Chunguang<sup>6</sup>  
CHEN Yi<sup>7</sup>

1 State Key Laboratory of Resources and Environment Information System, institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, 100101, China

2 University of Chinese Academy of Sciences, Beijing, 100049, China

3 Key Laboratory of Natural Resource Coupling Process and Effects, Beijing 100055, China

**收稿日期：**2024-09-25

**项目资助：**国家重点研发计划项目(2021YFB3901300); 自然资源要素耦合过程与效应重点实验室开放课题(2023KFKTB003)。

**第一作者：**刘逸娴, 博士生, 研究方向为植被生态遥感。liuyixian1420@igsrr.ac.cn

**通讯作者：**张新, 博士, 研究员。Zhangxin@radi.ac.cn

4 Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

5 State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Beijing 100101, China

6 Yellow River Mouth Management Station of Yellow River Delta National Nature Reserve, Dongying 257500, China

7 Beijing Institute of Surveying and Mapping, Beijing 100038, China

**Abstract: Objectives:** The monitoring of salt marsh vegetation in the Yellow River Delta wetlands is the basis for the protection and restoration of the ecological functions of the Yellow River Delta wetlands. The vegetation situation in the Yellow River Delta is complex, and the detection and classification of vegetation is particularly important due to the spread of the invasive species *Spartina alterniflora*, which has a significant impact on wetlands. **Methods:** This study takes part of the Yellow River Delta wetland as the study area, takes high-resolution aerial images as the data source. Four feature variables are generated: spectral features, color features, index features and texture features, and six different classification schemes are constructed. Scheme 1 with only spectral features was regarded as the control group, and index features, texture features and color features were integrated into scheme 2, scheme 3 and scheme 4, respectively. Scheme 5 contains all features, and scheme 6 constructs a multi-feature optimization feature set for all features. The random forest was used to classify vegetation for each extraction scheme and their accuracy were verified, aiming to explore the influence and reasons of different feature variables on the classification. Then the best preferred features were selected to improve the effect of vegetation classification. **Results:** Based on the visible and near-infrared spectra, just adding different features to the experiment had different effects on the accuracy of vegetation classification: (1) Scheme 1, scheme 2 and scheme 3 have unsatisfactory extraction effect on *Phragmites australis* and *Suaeda salsa*. Scheme 4, scheme 5 and scheme 6 with the addition of color features can better distinguish between the two, probably because *Suaeda salsa* is dark red on the image, which is quite different from other vegetation, so the color features can be better distinguished. (2) Scheme 4 divides the tidal flat area with non-vegetation information into *Phragmites australis* at the edge of the water body, which are similar in color, but partially misdivided when color features are added. Schemes 5 and 6 were well classified, but due to the mixed nature of vegetation, all schemes had varying degrees of misclassification of *Tamarix chinensis*, *Phragmites australis* and *Spartina alterniflora*. (3) The index features have a positive effect on *Suaeda salsa* extraction, texture features will reduce the accuracy of vegetation classification, and the integration of color features for classification is the key to improve the overall accuracy. (4) Based on multi-feature optimization of random forest, the extraction effect was the best, with an overall accuracy of 88% and a Kappa coefficient of 0.85. **Conclusions:** The main advantages of this study are the acquisition of new data sources, the introduction of multiple feature variables, and the experimental evaluation and classification accuracy analysis of different feature variables. The importance of color features are verified and multi-feature optimization of random forest by integrating color features is a feasible method to classify vegetation information in the Yellow River Delta. The method can effectively distinguish vegetation from non-vegetation and extract each vegetation type at the same time. This study provides an effective technical route in feature selection and methodology for vegetation information extraction in the Yellow River Delta.

**key words:** Yellow River Delta; aerial images; random forest; vegetation information extraction; color features

湿地是自然界中最富生物多样性的生态系统, 在保护生态环境、调节气候、控制土壤侵蚀和发展经济社会中发挥着不可替代的重要作用<sup>[1-2]</sup>。黄河三角洲拥有着中国最完整、最广泛、最年轻的河口湿地生态系统<sup>[3]</sup>。植被作为黄河三角洲生态系统功能的主体, 在生物多样性的恢复与保护中发挥着重要的作用。在黄河三角洲湿地中, 碱蓬、互花米草、芦苇和柽柳等群落是最典型的植被群落<sup>[4]</sup>。对湿地盐沼植被进行分类可以为黄河三角洲湿地生态功能保护与恢复提供科学依据与合理利用的基础。

原始的植被分类依靠人工实地勘察,不仅耗费人力与大量资源,且覆盖面低,由于湿地交通条件不佳,实地工作难以深入,人身安全也难以确保。随着对地观测系统的快速演进与分类算法的不断发展,因时间分辨率与空间分辨率的提高,遥感在农林业成为重要的监测手段。

目前对于湿地植被的分类与提取,多涉及 Landsat 数据<sup>[5]</sup>。例如学者张晨宇等<sup>[6]</sup>选取三个时相的 Landsat 数据分别对芦苇、碱蓬与互花米草进行提取;李燕妮等<sup>[7]</sup>基于多年 Landsat 数据以典型四种植被芦苇、互花米草、碱蓬与柽柳为对象用支持向量机进行分类。因 Sentinel-2 卫星的重访周期短且空间分辨率高,利用此卫星影像的学者逐渐增多。例如学者张磊等<sup>[8]</sup>利用多时相 Sentinel-2 影像对黄河三角洲湿地信息进行提取;宗影等<sup>[9]</sup>基于 Sentinel-2 数据对滨海湿地植被进行分类研究;Zhang 等<sup>[10]</sup>获取 12 个月份的 Sentinel-2 数据提取 264 个特征,用随机森林进行优选后进行黄河三角洲的三种植被分类。为了进行更细致的植被分类,部分作者使用国产影像对黄河三角洲植被类型进行分类,其中高分影像具有更高分辨率与细致的图谱信息,高光谱数据可以获取更多波段信息。例如学者刘庆生等<sup>[11]</sup>利用高分一号影像获取缨帽亮度和绿度与表土粒度指数共同提取黄河三角洲斑块状植被;韩月等<sup>[12]</sup>利用资源一号 02D 卫星高光谱数据黄河三角洲湿地景观分类;Tu 等<sup>[13]</sup>融合珠海 1 号高光谱卫星影像与高分 3 号全极化合成孔径雷达影像,提出了一种协同分类方法,定量描述整个黄河三角洲滨海湿地的空间分布。随着无人机技术的发展,数据采集获取时的灵活轻便使无人机遥感有着卫星遥感所不能比拟的优势,但目前研究集中在对某类植被的提取识别,且无人机获取影像空间范围有限。例如学者李岩舟等<sup>[14]</sup>、汪煜等<sup>[15]</sup>用 UAV 影像对互花米草进行识别提取研究。航空遥感比无人机获取影像空间范围广,但同时具有高分辨率,受气象因素干扰小、时间灵活以及等特性。目前航空遥感影像的湿地植被识别和监测的研究较少。

本文选取黄河三角洲的植被典型区域为研究区,利用航空遥感平台获取高分辨率影像数据,提取影像的光谱特征、指数特征、颜色特征与纹理特征,组合建立特征集,利用随机森林进行优选与分类,探求不同特征变量对分类的影响及原因,以期选取最佳的优选特征进行植被识别,为黄河三角洲典型植被的分类提供可行有效的技术支撑。

## 1. 研究区域与数据

### 1.1 研究区概况

黄河入海口处的黄河三角洲自然保护区,地处渤海之滨,是陆生野生动物重要栖息地。自然保护区可分为三个区域,分别是北部的黄河故道入海口一千二与南部的现行黄河入海口大汶流和黄河口。本文的研究区位于现行黄河入海口,中国山东省东营市大汶流内,地理位置为  $119^{\circ} 14.325' E-119^{\circ} 15.051' E$ ,  $37^{\circ} 47.472' N-37^{\circ} 47.945' N$  研究区位置如图 1 所示。

现行黄河入海口是黄河的终点,黄河携带泥沙流向大海在此处积留大面积新生陆地。随着时间的推移和自然环境的变化,加上人类活动和自然灾害的影响,黄河入海口处形成了一个宽阔、地势平坦的三角洲湿地。黄河入海口的湿地,是我国大型湿地之一。由于水源充足,植被丰富,水文条件独特,海淡水交汇,有机质含量丰富,生长了大片盐沼植物。其中最为典型的植被有碱蓬、芦苇、柽柳与互花米草。同时也提供了鸟类栖息繁衍的宜居之处,因此拥有了丰富的野生动植物生态资源和独特的生态系统。与此同时,该地拥有丰富的海洋资源,盐业与渔业蓬勃发展,承载着中华民族的历史和文化。由于互花米草的入侵的迅速扩张破坏了生态系统<sup>[16]</sup>,为了保护黄河入海口的生态环境,中国政府对黄河入海口的周边地区的互花米草进行了严格的清除与监测<sup>[17]</sup>,禁止任何破坏生态环境的行为。

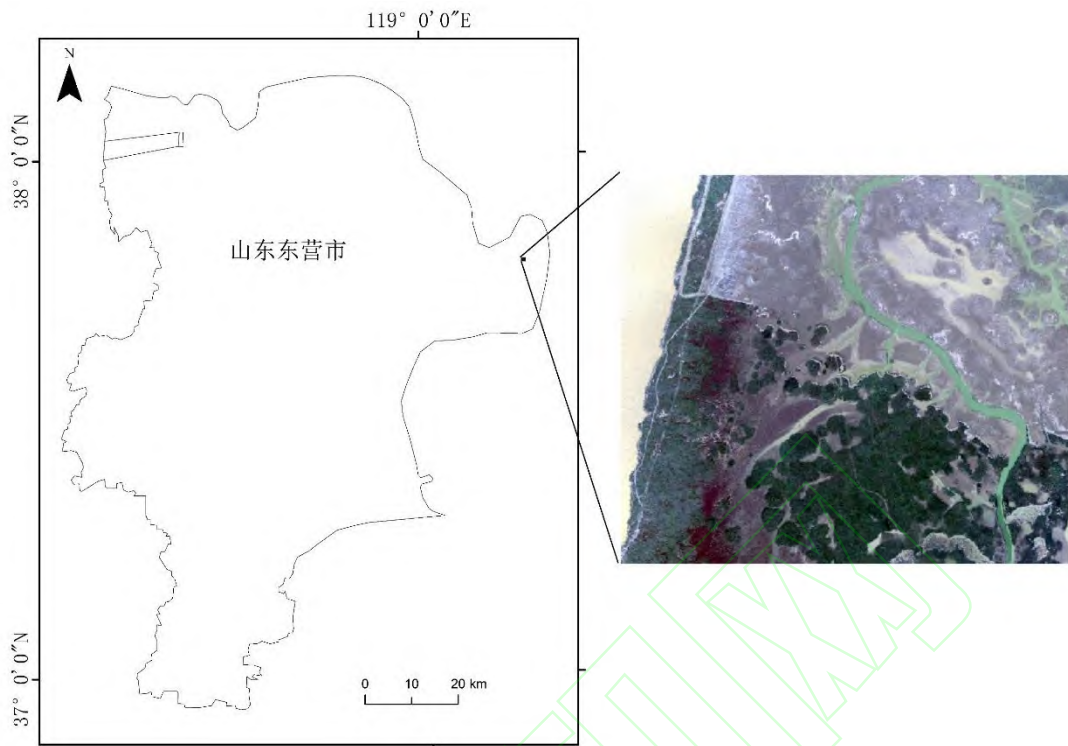


图 1 研究区位置图  
Fig. 1 Location of study area

## 1.2 数据获取及处理

本次航空数据获取任务使用美国 CESSNA 公司制造的塞斯纳 208B 涡轮螺旋桨飞机，选用三线阵光学相机，其空间分辨率为 0.2m。国家重点研发计划项目《北斗精准导航与高分辨率遥感集成技术及区域综合应用示范》课题一对获取的航空数据进行预处理。研究区航空影像长约 1km，宽为 0.9km，获取时间处于植被生长期 10 月初。本研究中对各波段中心波长的位置和波段范围并没有严格要求，所获取的影像没有进行严格的辐射定标。为了将统一影像数据的尺度，消除不同特征之间的量纲差异，对四个原始波段进行了数据归一化处理<sup>[18]</sup>。

样本数据来源实地采样数据和山东东营市天地图 (<https://shandong.tianditu.gov.cn/map.html>) 历史高分辨率影像中目视解译出具有代表性且典型的纯净像元作为样本点，尽量使其分布均匀，可用于分类训练和精度验证。本文将地物首先分为植被与非植被区域，根据研究区典型地物类别将植被分为互花米草、怪柳、碱蓬与芦苇，其解译标志如图所示。每个研究区每种样本点数不低于 100 个，样本数据质量可以用 J-M (Jerries-Matusita) 距离进行分离程度的判定，本文选取的样本点 J-M 距离在 1.9 以上，说明质量优秀，可用于训练样本和分类验证，其中 70% 用于随机森林分类的训练，30% 用于精度验证。



图 2 植被类型目视解译标志

Fig. 2 Visual Interpretation of Vegetation Types

## 2. 研究方法

### 2.1 特征提取

本文选取光谱特征、指数特征、颜色特征与纹理特征构建特征集，表 1 详细描述各种特征及其表达式。

表 1 航空影像特征集描述

Tab.1 Description of the Feature Sets of Aerial Images			
特征	名称	简称	特征说明
光谱特征	Blue	B	蓝波段
	Green	G	绿波段
	Red	R	红波段
	Near infrared	NIR	近红外波段
指数特征	Normalized Difference Vegetation Index	NDVI	$(NIR-R)/(R+NIR)$
	Ratio Vegetation Index	RVI	$NIR/R$
	Soil Adjusted Vegetation Index	SAVI	$\frac{(1+L)(NIR-R)}{NIR+R+L}, L=0.5$
	Normalized Difference Water Index	NDWI	$(G-NIR)/(G+NIR)$
	Difference Vegetation Index	DVI	$NIR-R$
纹理特征	Angular Second Moment	GLCM_A	角二阶矩
	Correlation	GLCM_Cor	相关度
	Contrast	GLCM_Con	对比度
	Entropy	GLCM_E	熵
	Variance	GLCM_V	方差
颜色特征	Excess green index	ExG	$2g-r-b$
	Excess red index	ExR	$1.4r-g$
	Excess blue index	ExB	$1.4b-g$
	Green-red difference index	GRDI	$g-r$
	Green-blue difference index	GBDI	$g-b$
	Excess green-excess red difference index	ExGR	$3g-2.4r-b$
	Triangular greenness index	TGI	$g-0.39r-0.61b$
	Modified excess green index	MExG	$1.262g-0.884r-0.311b$
Green-blue difference + modified excess green index	GBDI+MExG	$2.262g-0.884r-1.311b$	

实验选择影像的四个波段数据归一化后的像元值作为光谱特征，选择遥感领域常见的归一化植被指数（Normalized Difference Vegetation Index, NDVI），比值植被指数（Ratio Vegetation Index, RVI），土壤调节植被指数（Soil Adjusted Vegetation Index, SAVI），归一化水体指数（Normalized Difference Water Index, NDWI）、差值植被指数（Difference Vegetation Index, DVI）。

纹理特征是遥感影像的一种重要特征，包含着每种地物呈现自己特定纹理信息及其与环境的相互关系。合理利用纹理信息，可以很大提高遥感影像分类的精度。利用 ENVI5.6 影像纹理特征提取工具，基于灰度共生矩阵（Gray-level Co-occurrence Matrix, GLCM）提取影像的纹理特征<sup>[19]</sup>。纹理提取方法应用较为广泛。在进行纹理分析时，选取角二阶矩（Angular Second Moment, GLCM\_A）、相关性（Correlation, GLCM\_Cor）、对比度（Contrast, GLCM\_Con）、熵（Entropy, GLCM\_E）和方差（Variance, GLCM\_V）5 个冗余度较小的二阶统计量对影像进行定量分析。通常计算灰度共生矩阵的参数时，为了避免得到冗余繁多的纹理特征，会综合 0°，45°，90°，135°四个方向的特征值取其平均作为最终

的参数特征。为了避免对影像进行纹理分析时产生的统计分量过多，造成数据冗余，增加计算难度与提取难度，因此对原始影像进行主成分分析之后再获取纹理特征信息。其中第一主成分累计贡献率为 69.95%，第二主成分累计贡献率为 99.34%，因此取前两个主成分来提取纹理特征。通过多次对比实验分析，设置滑动窗口大小为 3，步长为 1，利用灰度共生矩阵提取了 5 个参数，共 10 个纹理特征。

颜色指数本主要是用于 24 位彩色图像，需要计算每个像素点的红色、绿色、蓝色的分量<sup>[20]</sup>。将 R、G、B 通道数据分别除以图像像素的最大值以归一化处理。对于 8 位 R、G、B 通道而言，这个值为 255。然后进行色彩空间归一化处理，得到光谱 r、g、b 色彩分量，最后生成颜色指数。但颜色指数同样能通过其他方式生成<sup>[21]</sup>，例如对于多光谱图像而言，图像采集相机硬件不同，其 RGB 颜色通道的值发生改变，也有学者直接用于颜色指数的计算，但有研究表明利用归一化色彩通道的指数可能更具有鲁棒性<sup>[22]</sup>。由于指根据解译标志可知，黄河三角洲植被的色彩具有明显差距，因此仿照处理 24 位彩色图像的方法对可见光谱进行了归一化处理，然后进行色彩空间归一化，得到三个色彩分量后生成颜色特征，用以提高分类精度。色彩空间归一化计算公式如下：

$$\begin{cases} r = \frac{R}{R+G+B} \\ g = \frac{G}{R+G+B} \\ b = \frac{B}{R+G+B} \end{cases} \quad (1)$$

本研究选取的颜色特征有超绿指数 (Excess green index, ExG)<sup>[23]</sup>、超红指数 (Excess red index, ExR)<sup>[24]</sup>、超蓝指数 (Excess blue index, ExB)<sup>[25]</sup>、绿-红差值指数 (Green-red difference index, GRDI)<sup>[26]</sup>、绿-蓝差值指数 (Green-blue difference index, GBDI)<sup>[27]</sup>、超绿-超红差值指数 (Excess green-excess red difference index, ExGR)<sup>[28]</sup>、三角绿度指数 (Triangular greenness index, TGI)<sup>[29]</sup>、修正超绿指数 (Modified excess green index, MExG)<sup>[30]</sup>、绿-蓝差值+修正超绿指数 (Green-blue difference + modified excess green index, GBDI+MExG)<sup>[31]</sup>。

## 2.2 实验方案说明

为了确定不同特征对黄河三角洲植被精细分类的重要程度，研究不同的特征变量对分类结果的影响，探求提高植被分类精度的最佳特征组合方式，如表 2 所示，本文共设置了 6 种方案进行对比研究，以仅加入光谱特征的方案 1 为对照组，分别融入指数特征、纹理特征与颜色特征构成方案 2、方案 3 与方案 4。方案 5 包含所有特征，方案 6 对所有特征进行优选构建特征集。

表 2 实验方案信息

实验方案	特征组合
1	光谱特征
2	光谱特征+指数特征
3	光谱特征+纹理特征
4	光谱特征+颜色特征
5	光谱特征+指数特征+纹理特征+颜色特征
6	所有特征优选组合

## 2.3 随机森林分类方法

随机森林是当今运用广泛、高度灵活的一种机器学习算法。它的基本单元是决策树，

本质是将多棵决策树进行集成学习<sup>[32]</sup>。决策树参与分类的个数与分类结果个数一致，随机森林集成全部的分类结果，确定最佳的样本类别。这种方法具有较好的准确率，能够有效运用在大数据上。随机森林主要用于大量的样本基础上，使用平均值来提高预测精度和控制过度拟合。因此最后样本分类结果，不会脱离训练样本的范围。

随机森林的生成如下：

(1) 在原有的样本中选择  $N$  个样本组成训练样本集，选择方法为有放回地随机抽取。每轮随机采样中，训练集中有  $(1-1/N)^N$  的数据没有被选择。当  $N$  很大时，这个值趋近于三分之一，这些数据称为袋外数据。

(2) 抽取的  $N$  个样本集生成  $N$  棵决策树组成随机森林。决策树本质是分类回归树，节点引伸出两个分支过程中，如果每个样本的特征维度为  $M$ ，则选取常数  $m \ll M$ ，从  $M$  个特征中随机选取  $m$  个特征子集，根据 Gini 系数，从  $m$  个中选择最佳的特征在决策树中进行节点分裂。

(3) 对所有的决策树进行投票的过程就是对数据进行分类的过程，投票得分高的类别即为分类结果。

## 2.4 特征优选

特征优选的主要目的是减少计算过程中数据的冗余度，提高运算过程中的效率。在上述内容中，经过研究得到了 28 种特征因子，其中含有 4 种光谱特征、5 种指数特征、10 种纹理特征，9 种颜色特征。如果全部将这些特征全部应用于影像分类，所形成的高维特征空间将会增大计算难度，加重模型训练的负担，并且会造成“Hughes”现象使得分类效率大大降低。为解决该问题，基于随机森林筛选出对地物分类起关键作用的特征组合进行分类。本研究采用随机森林特征筛选算法选取最优特征子集，依据平均不纯度减少对所有特征的重要性进行排列，依据准确率降低确定最佳特征数量。

随机森林中决策树构建的关键是结点分裂时的特征选择。理想情况下，每个子节点的样本属于同一类，即称分裂得到的每个子节点是纯的，可以用基尼指数来度量样本集合的不纯度。集合  $D$  的基尼指数定义如下：

$$Gini(D) = \sum_{b=1}^B p_b (1-p_b) = 1 - \sum_{b=1}^B p_b^2 \quad (2)$$

式中： $B$  为训练样本中样本种类数， $p_b$  表示集合  $D$  中随机选中样本属于类别  $B$  的概率， $(1-p_b)$  表示样本被分错的概率。如果样本集合  $D$  根据特征  $A$  是否取某一可能值  $a$  被划分为  $D_1$  和  $D_2$  两个部分，则在特征  $A$  的条件下，集合  $D$  的基尼指数为：

$$Gini(D, A) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2) \quad (3)$$

式中， $|D|$  表示集合  $D$  中的样本数， $|D_1|$  表示集合  $D_1$  中的样本数， $|D_2|$  表示集合  $D_2$  中的样本数。在随机森林中，如果通过某个特征中划分后平均基尼指数减少的程度越大，表明通过此特征划分后的集合变纯的程度越大，则表明此特征分类能力越大，其特征重要性越大。平均不纯度减少表示为：

$$\Delta Gini = \frac{\sum_{n=1}^{n=K} [Gini_n(D) - Gini_n(D, A)]}{K} \quad (4)$$

式中： $K$  为随机森林中决策树的个数， $Gini_n(D)$  表明第  $n$  颗决策树  $Gini(D)$  划分前集合  $D$  的基尼指数， $Gini_n(D, A)$  表示第  $N$  棵决策树  $Gini(D, A)$  通过特征  $A$  划分后集合  $D$  的基尼指数。

使用袋外 (Out of Bag, OOB) 误差进行模型评估以确定模型最优特征数量。首先对特征重要性从高到低的顺序进行排序，从重要性最高的特征开始，每次增加一个特征，逐步



选择靠前的特征产生了不同的 28 个特征组合，利用每个组合的 OOB 误差确定最优特征个数。

## 2.5 精度评价

本文根据混淆矩阵对影像分类精度进行评价分析，用总体精度和 Kappa 系数、用户精度和生产者精度来度量结果的准确度和可靠性。各计算指标如下所示：

$$OA = \frac{\sum_{i=1}^n J_{ij}}{\sum_{i=1}^n \sum_{j=1}^n J_{ij}} \quad (5)$$

$$Kappa \text{ 系数} = \frac{\left( \sum_{i=1}^n \sum_{j=1}^n J_{ij} * \sum_{i=1}^n J_{ii} - \sum_{i=1}^n \left( \sum_{j=1}^n J_{ji} * \sum_{j=1}^n J_{ij} \right) \right)}{\left( \sum_{i=1}^n \sum_{j=1}^n J_{ij} * \sum_{i=1}^n \sum_{j=1}^n J_{ij} - \sum_{i=1}^n \left( \sum_{j=1}^n J_{ji} * \sum_{j=1}^n J_{ij} \right) \right)} \quad (6)$$

$$UA = \frac{J_{ii}}{J_{i+}} \quad (7)$$

$$PA = \frac{J_{ii}}{J_{+i}} \quad (8)$$

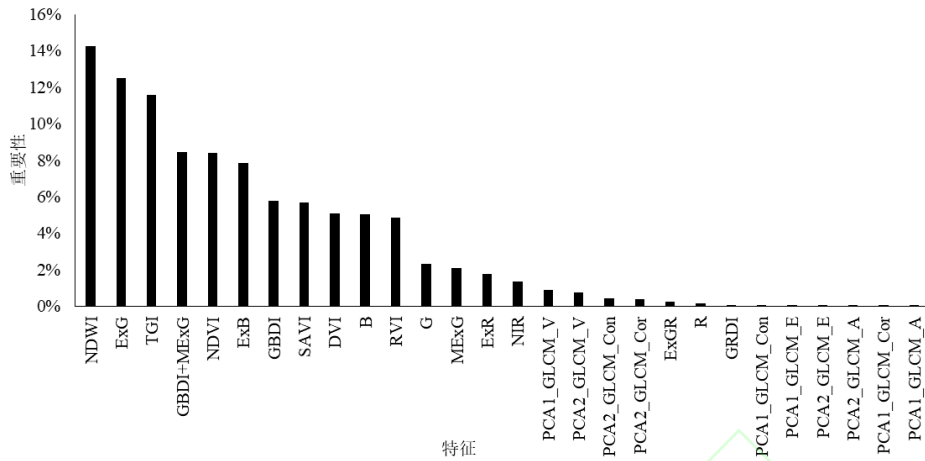
式中：n 为混淆矩阵的行列数；其中，J 表示样本总数情况下， $J_{ij}$  表示分类所得的第 i 类归属于实际类型的第 j 类的数量。 $J_{ii}$  为混淆矩阵第 i 行、第 i 列上的样本数； $J_{i+}$  和  $J_{+i}$  分别为第 i 行和第 i 列的样本总数。

## 3. 结果与分析

方案 1-方案 5 作为对比实验，不需要进行特征选择，前 5 种方案已在探求不同特征变量对黄河三角洲湿地信息提取精度的影响。在方案 5 的基础上，采用随机森林算法对 28 个特征变量进行筛选，选择优选的特征构成方案 6。

### 3.1 优选特征的确定

通过平均不纯度减少的方式可以计算出模型中每个特征的重要性，其重要性结果如图 3 所示。



注：PCA1\_GLCM\_V为第一主成分的角二阶矩，PCA2\_GLCM\_V为第二主成分的角二阶矩，依此类推。

图3 特征重要性分布

Fig. 3 The Distribution of Characteristic Importance

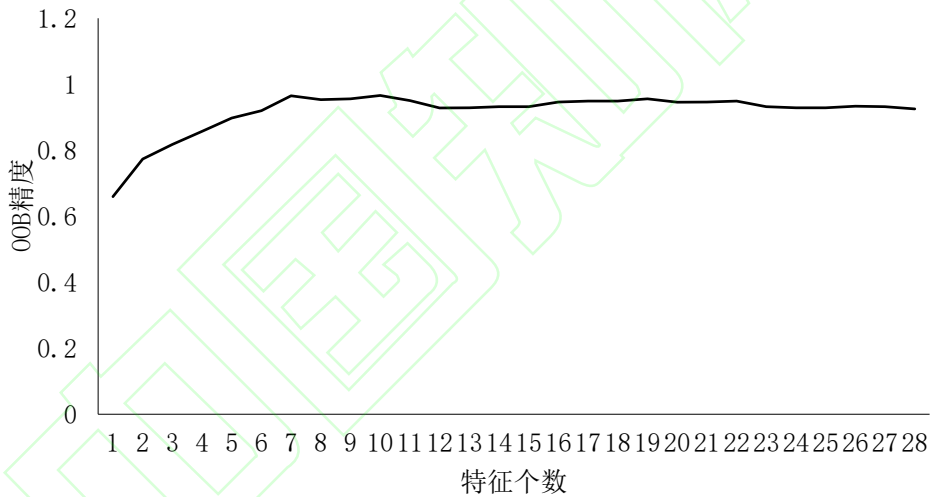


图4 特征个数与OOB精度的关系

Fig. 4 The Relation Between the Number of Characteristic Variables and the OOB Accuracy

以特征重要性评估结果为参考，逐步增加特征数量，以分析不同特征组合模型下的OOB精度，如图4所示。由图可知，当特征数从1至10时，精度逐渐增加，这个过程，特征变量相关性不高，冗余度小，从而能够提高分类模型的性能。但当特征个数从10开始，OOB精度基本维持不变，起伏很小。当特征个数过小时，会导致精度不够，使得分类结果不理想。特征数量过多时，需要计算的工作量增大，精度变化极小，增长运算时间，反而有时会出现过拟合的现象导致精度下降。本研究主要选取前10个特征顺序组合用以进行分类，此时OOB精度可达96.6%。

由图可知，重要性位居前10的特征中，航空影像的原始波段有1个，还有4个指数特征与5个颜色特征，而纹理特征不在前列。对植被提取分类重要性高的多为指数特征与颜色特征，这说明这两类特征对分类有较大的贡献。而唯一的原始波段位居第十，纹理特征不在前列，这说明影像的原始特征与纹理特征对植被精细提取分类重要性较低，具有较少的贡献度。NDWI特征的重要性得分最高，高达14.25%，这是因为NDWI对水体十分敏感，

能够有效识别出来。NDWI 在黄河三角洲地区可以很好区分植被与非植被区域，尤其是该非植被区域中包含大量的水体与潮滩。PCA1\_GLCM\_A 特征重要性得分最低，只有 0.01%，几乎不产生影响。

### 3.2 提取结果及精度评价分析

六种不同方案的提取结果如图 5 所示。与原始图像对比定性分析可以发现，方案 1、方案 2、方案 3 对芦苇与碱蓬的提取效果不理想，有较多混分情况。而加入了颜色特征的方案 4、方案 5、方案 6 能将两者较好区分出来，可能是因为碱蓬在影像上呈现深红色，与其他植被有较大区别，因此颜色特征能够更好区分出来。方案 4 会将非植被信息的滩涂区域与水体边缘部分分为芦苇，可能因为在滩涂和水体边缘处生长着部分芦苇，两者颜色相近，加入颜色特征后却产生部分错分。方案 5 与方案 6 的分类情况较好，但由于植被的混生特点，所有的方案都有不同程度的柽柳、芦苇与互花米草错分的情况。

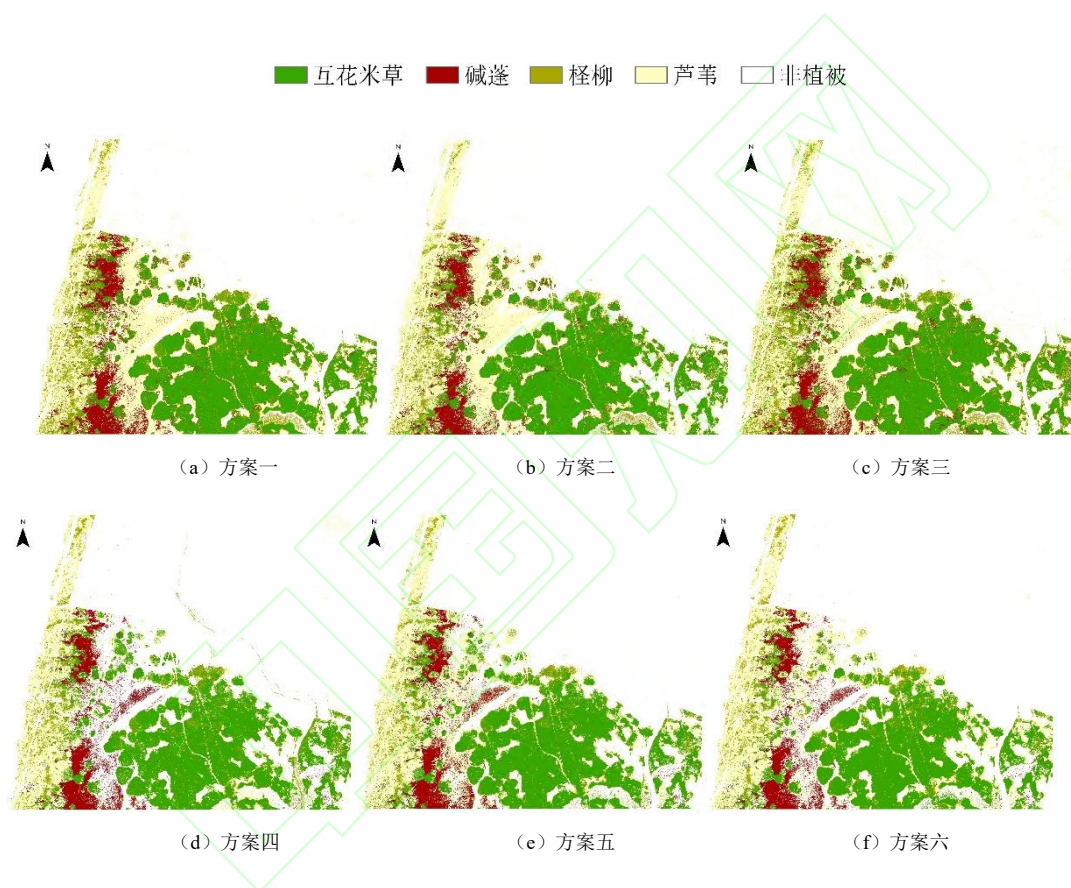


图 5 不同方案分类结果

Fig. 5 Classification Results of Different Plans

为了对所有方案进行定量分析，利用混淆矩阵的四个指标进行精度评价。由表可知，以光谱特征为基础的方案 1 总体精度为 82.00%，Kappa 系数为 0.76。在方案 1 的基础上加入指数特征、纹理特征与颜色特征均会对植被信息提取产生不同的程度的影响。分别加入指数特征、颜色特征的方案 2 和方案 4 以及所有特征都参与分类的之后的方案 5，分类精度略有提升，分别为 83.00%、84.00%与 85.00%，Kappa 系数为 0.79、0.80 与 0.81。但是加入纹理特征的方案 4 分类精度却在一定程度上有所降低，其总体精度为 78.00%，Kappa 系数为 0.73。这说明指数特征和颜色特征会提高植被信息的提取精度，但是纹理特征却会导致分类精度的降低，这也能在一定程度上解释在特征优选时指数特征与颜色特征重要性得分较高位于前列而纹理特征的重要性得分较低的原因。在随机森林优选特征后生成的方案 6

总体精度与 Kappa 系数有较大提升，总体精度为 88.00%，Kappa 系数为 0.85，相比较原光谱特征的基础方案总体精度提升了 6.00%，Kappa 系数增长了 0.09。

仅关注植被类型的用户精度与生产者精度可以发现，仅加入指数特征对碱蓬的提取精度较好，可能与碱蓬相比其他植被生长在高盐环境中且呈密集大片形状，而指数特征对植被生长环境与状况十分敏感有关。仅加入纹理特征反而降低了各类植被的提取精度，可能是因为各盐沼植被，尤其是柽柳与芦苇，并没有固定的纹理，分布不均。仅加入颜色特征的方案 4 与加入所有特征的方案 5 对各植被类型的分类提取均有一定程度的保持或提升作用，这说明融合了颜色特征的方案是主要提高精度的原因。基于特征优选随机森林分类结果中，所有类别的用户精度均在 80%以上，除柽柳以外类别的生产者精度均在 85%以上。这是因为柽柳容易与芦苇混淆，被分类器分为芦苇。由于柽柳与芦苇通常是混生在一起，交错分布，在影像中容易混淆，导致分类精度不高。总体而言，方案 6 的分类效果比较前 5 个方案均有提升，融合颜色特征的随机森林多特征优选对黄河三角洲植被精细分类具有较好的实用性。

表 3 分类精度统计

Tab.3 The Statistics of Classification Accuracy

类别		互花米草	碱蓬	柽柳	芦苇	非植被	OA	Kappa
方案 1	UA	82.61	92.86	93.33	60.71	95.00	82.00	0.76
	PA	95.00	65.00	70.00	85.00	95.00		
方案 2	UA	84.21	94.44	81.25	58.62	73.91	83.00	0.79
	PA	80.00	85.00	65.00	85.00	85.00		
方案 3	UA	82.61	100.00	90.00	52.94	95.00	78.00	0.73
	PA	95.00	65.00	45.00	90.00	95.00		
方案 4	UA	82.61	100.00	100.00	81.82	71.43	84.00	0.80
	PA	95.00	65.00	70.00	90.00	100.00		
方案 5	UA	86.36	100.00	100.00	67.86	86.96	85.00	0.81
	PA	95.00	65.00	70.00	95.00	100.00		
方案 6	UA	86.36	94.44	100.00	82.61	83.33	88.00	0.85
	PA	95.00	85.00	65.00	95.00	100.00		

#### 4. 结论与讨论

基于航空民用飞机的遥感信息获取技术逐渐应用于农业领域，航空遥感平台获取植被信息具有高分辨率、高时效、获取影像迅速、获取区域面积广的特点。本研究结合航空影像的四个原始波段，提取了指数特征与纹理特征，同时融合颜色特征，构建了 6 种不同的提取方案，对典型黄河三角洲植被信息进行分类提取并验证与对比不同方案的提取精度，探求不同特征变量对分类的影响及原因，选取最佳的优选特征改善植被精细分类的效果。主要得到以下结论：

以可见光谱与近红外光谱为基础，仅仅加入不同特征进行实验对植被分类的精度影响有所区别。指数特征会提高碱蓬的分类精度，颜色特征会整体提高所有植被类型的分类精度，然而纹理特征会对黄河三角洲植被精细分类产生负面影响。这说明融合了颜色特征进行分类是提高精度的关键。

进行随机森林优选之后的分类结果有了很好的改善，不仅能降低分类维数，加快分类速度，同时也能提高分类精度和效率。融合了颜色特征的随机森林特征优选的分类方法在黄河三角洲植被分类中具有较高的分类精度，总体精度为 88%，Kappa 系数为 0.85。该方法能有效区分植被与非植被，同时将各植被类型提取出来。

本研究主要优势在于新数据源的获取、多种特征变量的引入，以及对不同特征变量进行实验评价与分类精度分析，本文验证了颜色特征的重要性与融合了颜色特征的随机森林特征优选分类方法的可行性。但是实仍然存在一些不足的地方：（1）本文只研究了单一时相的特征变量的影响，没有考虑到不同时相的特征变量分类产生的影响。后续将对不同月

份分类情况进行研究,利用时间变化特征区分光谱类似的植被类型,以提高分类精度。(2)本文只探讨了四种黄河三角洲典型植被区域,在不同的航空影像上植被类型或有减少与增加,为了实现植被类型分类提取的精细化,应该深入探讨特征变量对每类植被类型与几种植被类型组合的贡献与影响。

## 参考文献

- [1] Lang Qin, Niu Zhenguo, Hong Xiaoqi, et al. Remote Sensing Monitoring and Change Analysis of Wetlands in the Tibetan Plateau[J]. Geomatics and Information Science of Wuhan University, 2021, 46(02):230-237. (郎芹,牛振国,洪孝琪,等.青藏高原湿地遥感监测与变化分析[J].武汉大学学报(信息科学版),2021,46(02):230-237.)
- [2] Li Peng, Li Dahui, Li Zhenhong, et al. Wetland Classification Through Integration of GF-3 SAR and Sentinel-2B Multispectral Data over the Yellow River Delta[J]. Geomatics and Information Science of Wuhan University, 2019,44(11):1641-1649. DOI:10.13203/j.whugis20180258. (李鹏,黎达辉,李振洪,等.黄河三角洲地区 GF-3 雷达数据与 Sentinel-2 多光谱数据湿地协同分类研究[J].武汉大学学报(信息科学版),2019,44(11):1641-1649.)
- [3] WANG Renqing, ZHANG Yuhan, SUN Shuxia, et al. Journal of Shandong University(Natural Science), 2021,56(10):135-148.
- [4] ZOU Yuhan, LI Xue, ZHANG Xin, et al. Plant community composition and structure of the nascent wetlands of the Yellow River Estuary[J]. Chinese Journal of Ecology, 2024, 1-10.(邹雨函,李雪,张馨,等.黄河口新生湿地植物群落组成和结构[J/OL].生态学杂志,1-10[2024-05-22].<http://kns.cnki.net/kcms/detail/21.1148.Q.20230907.1250.012.html>.)
- [5] Xie Chunhua, Zhang Shuaiying, Cui Lizhen, et al. Progress and trend of remote sensing classification of land cover in the Yellow River Delta[J]. Science Technology and Engineering, 2022, 22( 33) : 14571-14583. (谢春华,张帅影,崔丽珍,等.黄河三角洲地物遥感分类研究进展与趋势[J].科学技术与工程,2022,22(33):14571-14583.)
- [6] Zhang Chenyu, Chen shenliang, Li Peng, et al. Spatiotemporal dynamic remote sensing monitoring of typical wetland vegetation in the Current Huanghe River Estuary Reserve[J]. Haiyang Xuebao, 2022, 44(1):125-136 (张晨宇,陈沈良,李鹏,等.现行黄河口保护区典型湿地植被时空动态遥感监测[J].海洋学报,2022,44(01):125-136.)
- [7] LI Yanni, LI Peng, WU Xiao, et al. Tempo-spatial variation of wetlands at the Yellow River Mouth and its control factors[J]. Marine Geology & Quaternary Geology, 2022, 42(1):68-80.(李燕妮,李鹏,吴晓,等.黄河口湿地时空变化过程及其主控因素[J].海洋地质与第四纪地质,2022,42(01):68-80.)
- [8] Zhang L, Gong Z N, Wang Q W, et al. Wetland mapping of Yellow River Delta wetlands based on multi-feature optimization of Sentinel-2 images[J]. Journal of Remote Sensing, 2019,23(2): 313-326.(张磊,宫兆宁,王启为,等.Sentinel-2 影像多特征优选的黄河三角洲湿地信息提取[J].遥感学报,2019,23(02):313-326)
- [9] Zong Ying, Li Yufeng, Liu Hongyu. A study of Coastal Wetland Vegetation Classification Based on Object-oriented Random Forest Method[J]. Journal of Nanjing Normal University(Engineering and Technology Edition), 2021,21(04):47-55.(宗影,李玉凤,刘红玉.基于面向对象随机森林方法的滨海湿地植被分类研究[J].南京师范大学学报(工程技术版),2021,21(04):47-55.)
- [10] Cheng Zhang, Zhaoning Gong, Huachang Qiu, et al. Mapping typical salt-marsh species in the Yellow River Delta wetland supported by temporal-spatial-spectral multidimensional features[J]. Science of the Total Environment, 2021,783:147061.
- [11]Qingsheng Liu, Chong Huang, He Li. Mapping plant communities within quasi-circular vegetation patches using tasseled cap brightness, greenness, and topsoil grain size index derived from GF-1 imagery[J]. Earth Science Informatics, 2021, 14:975-984
- [12] Han Y, Ke Y H, Wang Z P, at al. Classification of the Yellow River Delta wetland landscape based on ZY-1 02D hyperspectral imagery. National Remote Sensing Bulletin, 2023, 27 ( 6) : 1387-1399.(韩月,柯樱海,王展鹏,等.资源一号 02D 卫星高光谱数据黄河三角洲湿地景观分类[J].遥感学报,2023,27(06):1387-1399.)
- [13] Canran Tu, Peng Li, Zhenhong Li, et al. Synergetic Classification of Coastal Wetlands over the Yellow River Delta with GF-3 Full-Polarization SAR and Zhuhai-1 OHS Hyperspectral Remote Sensing[J]. Remote Sens, 2021,13(21):4444
- [14] Li Yanzhou, He Yanzhou, Qin Feng, et al. Identification of *Spartina alterniflora* based on convolutional neural network. Journal of Chinese Agricultural Mechanization, 2023,44(04):159-166. (李岩舟,何艳洲,覃锋,等.基于卷积神经网络的互花米草识别研究[J].中国

农机化学报,2023,44(04):159-166. )

[15] WANG Yu, TAN Junzhong, ZHANG Mei, et al. Extraction and correlation analysis of growth parameters of *Spartina alterniflora* based on UAV remote sensing[J]. *Marine Environmental Science*, 2024,43(01):100-108+118. (汪煜,谈俊忠,张梅,等.基于无人机遥感的互花米草长势参数提取与相关性分析[J].*海洋环境科学*,2024,43(01):100-108+118. )

[16] Min Y K, Ke Y H, Han Y, et al. Dynamic monitoring of invasive *Spartina alterniflora* clearance via fusion of Sentinel-2 and GF-1 time series images. *National Remote Sensing Bulletin*, 2023, 27(6):1467-1479.(闵钰魁,柯樱海,韩月,等.融合 Sentinel-2 和 GF-1 时序影像的入侵植物互花米草清除动态监测[J].*遥感学报*,2023,27(06):1467-1479. )

[17] YANG Junfang, MA Yi, REN Guangbo, et al. Monitoring method of invasive vegetation *Spartina alterniflora* in modern Yellow River delta based on gf remote sensing data[J]. *Marine Environmental Science*, 2017,36(04):596-602.(杨俊芳, 马毅, 任广波, 等. 基于国产高分卫星遥感数据的现代黄河三角洲入侵植物互花米草监测方法[J].*海洋环境科学*, 2017, 36(04):596-602.)

[18] Wang Xiaoqin, Wang Miaomiao, Wang Shaoqiang, et al. Extraction of vegetation information from visible unmanned aerial vehicle images[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2015,31(05):152-157+159+158. (汪小钦,王苗苗,王绍强,等.基于可见光波段无人机遥感的植被信息提取[J].*农业工程学报*,2015,31(05):152-157+159+158. )

[19] FAN Junliang, WANG Han, LIAO Zhenqi, et al. Winter Wheat Leaf Area Index Estimation Based on Texture-color Features and Vegetation Indices[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2023,54(07):347-359. (范军亮,王涵,廖振棋,等.基于纹理-颜色特征与植被指数融合的冬小麦 LAI 估测[J].*农业机械学报*,2023,54(07):347-359. )

[20] Esmael Hamuda, Martin Glavin, Edward Jones. A survey of image processing techniques for plant extraction and segmentation in the field[J]. *Computers and Electronics in Agriculture*,2016,125:184–199.

[21] Segmentation of vegetation and microplots in aerial agriculture images: A survey[J]. *The Plant Phenome Journal*, 2022,5: e20042.

[22] Lee, M. K., Golzarian, M. R., & Kim, I. (2020). A new color index for vegetation segmentation and classification. *Precision Agriculture*, 2020, 22(1), 179–204.

[23] Woebbecke, D. M., Meyer, G. E., Von Bargen, et al. Color indices for weed identification under various soil, residue, and lighting conditions. *Transactions of the ASAE*, 1995,38(1), 259–269.

[24] Meyer, G. E., Hindman, T. W., & Laksmi, K. Machine vision detection parameters for plant species identification[J]. *Precision Agriculture and Biological Quality*. 1999,3543:327-335.

[25] Guijarro M, Pajares G, Riomoros I, et al. Automatic segmentation of relevant textures in agricultural images[J/OL]. *Computers and Electronics in Agriculture*, 2011, 75(1): 75-83.

[26] Takeshi Motohka, Kenlo Nishida Nasahara, Hiroyuki Oguma, et al. Applicability of Green-Red Vegetation Index for Remote Sensing of Vegetation Phenology[J]. *Remote Sensing*,2010,2, 2369-2387.

[27] Dai Jianguo, Xue Jinli, Zhao Qingzhan, et al. Extraction of cotton seedling information using UAV visible light remote sensing images[J/OL]. *Transactions of the Chinese Society of Agricultural Engineering*, 2020, 36(4): 63-71.(戴建国, 薛金利, 赵庆展, 等. 利用无人机可见光遥感影像提取棉花苗情信息[J]. *农业工程学报*, 2020, 36(4): 63-71. )

[28] Meyer, G. E., Neto, J. C., Jones, D. D., et al. Intensified fuzzy clusters for classifying plant, soil, and residue regions of interest from color images. *Computers and Electronics in Agriculture*,2004,42(3),161–180. <https://doi.org/10.1016/j.compag.2003.08.002>.

[29] Hunt Jr, E. R., Daughtry, C., Eitel, J. U., et al. Remote sensing leaf chlorophyll content using a visible band index. *Agronomy Journal*, 2011,103(4), 1090–1099. <https://doi.org/10.2134/agronj2010.0395>.

[30] Burgos-Artizzu, X. P., Ribeiro, A., Guijarro, M., et al. Real-time image processing for crop/weed discrimination in maize fields. *Computers and Electronics in Agriculture*, 2011, 75(2), 337–346. <https://doi.org/10.1016/j.compag.2010.12.011>.

[31] Estimation of the quantity of drip-irrigated cotton seedling based on color and morphological features of UAV captured RGB images[J]. *Cotton Science*, 2022, 34(6): 508-522.(蒋家乐,李容望,马学花,等.基于无人机 RGB 图像颜色特征和形态特征的滴灌棉花苗期株数估算研究[J].*棉花学报*,2022,34(06):508-522. )

[32] BREIMAN L. Random forests[J]. *Machine learning*, 2001, 45(1):5-32.

**网络首发:**

标题：融合颜色特征的随机森林特征优选的黄河三角洲植被信息分类

作者：刘逸娴，刘庆生，张新，黄翀，李贺，车纯广，陈一

收稿日期：2024-09-25

DOI:10.13203/j.whugis20240203

#### 引用格式：

刘逸娴，刘庆生，张新，等. 融合颜色特征的随机森林特征优选的黄河三角洲植被信息分类[J].武汉大学学报(信息科学版),2024,DOI: 10.13203/j.whugis20240203 (LIU Yixian, LIU Qingsheng, ZHANG Xin, et al. Classification of Vegetation Information by Integrating Color Features with Multi-feature Optimization of Random Forest in the Yellow River Delta[J].Geomatics and Information Science of Wuhan University,2024,DOI: 10.13203/j.whugis20240203)

网络首发文章内容和格式与正式出版会有细微差别，请以正式出版文件为准！

---

#### 您感兴趣的其他相关论文：

##### 一种识别植被覆盖滑坡的多模态深度神经网络模型

唐小川, 涂子涵, 任绪清, 方成勇, 王宇, 刘鑫, 范宣梅

武汉大学学报(信息科学版), 2024, 49(9): 1566-1573.

<http://ch.whu.edu.cn/article/doi/10.13203/j.whugis20230099>

##### 高植被山区泥石流物源 LiDAR 遥感精细识别方法研究

刘小莎, 董秀军, 钱济人, 郭晨, 赵俊丞, 詹佳琪

武汉大学学报(信息科学版), 2024, 49(3): 400-410.

<http://ch.whu.edu.cn/article/doi/10.13203/j.whugis20210486>

##### 一种结合低级视觉特征和 PAPCNN 的 NSST 域遥感影像融合方法

侯昭阳, 吕开云, 龚循强, 支君豪, 王楠

武汉大学学报(信息科学版), 2023, 48(6): 960-969.

<http://ch.whu.edu.cn/article/doi/10.13203/j.whugis20220168>