



引文格式: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),2025,50(9):1802-1811.DOI:10.13203/j.whugis20240203

Citation: 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, 2025, 50(9): 1802-1811. 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^{1,2} 刘庆生^{1,3,4} 张新⁵ 黄翀¹ 李贺¹
车纯广⁶ 陈一⁷

1 中国科学院地理科学与资源研究所地理信息科学与技术全国重点实验室,北京,100101

2 中国科学院大学,北京,100049

3 自然资源部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 中国科学院空天信息创新研究院遥感与数字地球全国重点实验室,北京,100094

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: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。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4种特征变量并构建了不同的分类方案。利用随机森林方法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,选取最佳的优选特征改善植被分类的效果。结果表明,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;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。所提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。该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

中图分类号:P208

文献标识码:A

收稿日期:2024-09-25

DOI:10.13203/j.whugis20240203

文章编号:1671-8860(2025)09-1802-10

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

LIU Yixian^{1,2} LIU Qingsheng^{1,3,4} ZHANG Xin⁵ HUANG Chong¹ LI He¹
CHE Chunguang⁶ CHEN Yi⁷

1 State Key Laboratory of Resources and Environmental 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, Ministry of Natural Resources, Beijing 100055, China

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

5 State Key Laboratory of Remote Sensing and Digital Earth, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, 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 is the basis for protection and restoration of ecological functions in the Yellow River Delta (YRD) wetland. Due to the complexity of vegetation growth

基金项目:国家重点研发计划(2021YFB3901300);自然资源部自然资源要素耦合过程与效应重点实验室开放基金(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

通信作者:张新,博士,研究员。zhangxin@radi.ac.cn

conditions and the spread of the invasive species *Spartina alterniflora*, vegetation detection and classification are particularly important. **Methods:** This paper takes part of the YRD wetland as the study area, and chooses high-resolution aerial images as data source. Four feature variables are generated, including spectral features, color features, index features and texture features. Six different classification schemes are constructed. Scheme 1 with only spectral features is regarded as the control group. Index features, texture features and color features are 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. Random forest method is used to classify vegetation for each extraction scheme and the corresponding accuracies are verified, aiming to explore the influences and reasons of different feature variables on the classification. The best preferred features are selected to improve the effect of vegetation classification. **Results:** Based on the visible and near-infrared spectra, just adding different features to the experiment has different effects on the accuracies of vegetation classification. 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* shows dark red on the image, which is quite different from other vegetation. Scheme 4 divides non-vegetated tidal flat areas and water body edges into *Phragmites australis*, which are similar in color, and adds color features results in some misclassifications. Scheme 5 and scheme 6 are well classified, but due to the mixed nature of vegetation, all schemes have varying degrees of misclassification of *Tamarix chinensis*, *Phragmites australis* and *Spartina alterniflora*. Index features have positive effects on *Suaeda salsa* extraction, texture features reduce the accuracy of vegetation classification, and the integration of color features is the key to improve overall accuracy of classification. Based on multi-feature optimization of random forest, the extraction effect is the best, with overall accuracy of 88% and 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 is verified and multi-feature optimization of random forest by integrating color features is a feasible method to classify vegetation information in the YRD. The proposed 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 YRD.

Key words: the Yellow River Delta; aerial image; 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^[1-2]。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^[3]。植被作为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^[4],各类植被通过竞争、共生等相互作用,形成了复杂的生态关系。它们在长期协同进化过程中不断演变,反映着湿地生态系统的发展方向。随着气候的变化、人口与经济的增长、工业和港口的建设以及水体的污染,湿地面临着巨大压力。在此背景下,对植被信息进行精细分类、准确获取黄河三角洲湿地典型盐沼植被的分布情况,是进行

滨海湿地生态系统恢复和保护的重要基础。

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,遥感成为重要的农林业监测手段。目前对于黄河三角洲湿地植被的分类与提取多涉及 Landsat 数据^[5]。文献[6]选取 3 个时相的 Landsat 数据,分别对芦苇、碱蓬与互花米草进行提取;文献[7]基于多年 Landsat 数据,以芦苇、互花米草、碱蓬与柽柳典型 4 种植被为对象,采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。文献[8]利用多时相 Sentinel-2 影像对黄河三角洲湿地信息进行提取;文献[9]基于 Sentinel-2 数据对滨海湿地植被进行分类研究;文献[10]获取 12 个月份的 Sentinel-2 数据提取 264 个特征,采用

随机森林进行优选后进行黄河三角洲的3种植被分类。为了进行更细致的植被分类,部分作者使用中国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。文献[11]利用高分一号影像获取的3种指数特征共同提取黄河三角洲斑块状植被,有效区分了植被与非植被,分析了研究区斑块植被的变化;文献[12]利用资源一号02D卫星高光谱数据,进行黄河三角洲湿地景观分类,对166个波段进行筛选后分类,取得了良好的效果;文献[13]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,定量描述了整个黄河三角洲滨海湿地的空间分布。随着无人机(unmanned aerial vehicle, UAV)技术的快速发展,其数据采集系统凭借高度的灵活性和便携性,展现出卫星遥感无法比拟的优势。然而,现有研究主要集中于单一植被类型的提取与分析,且受限于UAV平台的载荷能力和续航时间,所获取影像的空间覆盖范围相对有限。文献[14-15]用UAV影像对互花米草进行提取和相关研究,但其影像局限于小区域。航空遥感比UAV获取影像空间范围广,同时具有高分辨率、受气象因素干扰小和时间灵活等特性,但目前基于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3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,山东省东营市大汶流内,地理位置为 $119^{\circ}14.325'E\sim 119^{\circ}15.051'E$, $37^{\circ}47.472'N\sim 37^{\circ}47.945'N$,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,是中国大型湿地之一。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。湿地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。由于互花米草入侵的迅速扩张破坏了生态系统^[16],为了保护黄河入海口的生态环境,中国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^[17],防止其挤占本地物种的生存空间。

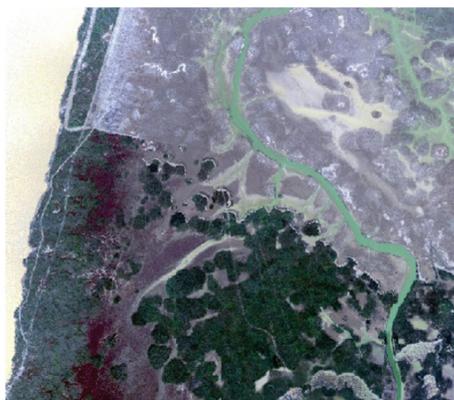


图1 研究区位置图

Fig. 1 Location of the 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塞斯纳公司制造的208B涡轮螺旋桨飞机,选用三线阵光学相机,其空间分辨率为0.2 m。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1 km,宽为0.9 km,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。为了统一影像数据的尺度,消除不同特征之间的量纲差异,对4个原始波段进行了数据归一化处理^[18]。

样本数据来源于实地采样数据和山东省东营市天地图数据(<https://shandong.tianditu.gov.cn/map.html>)。从历史高分辨率影像中目视解译出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。首先将地物分为植被区域与非植被区域,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 2 所示。研究区每种样本点数不低于

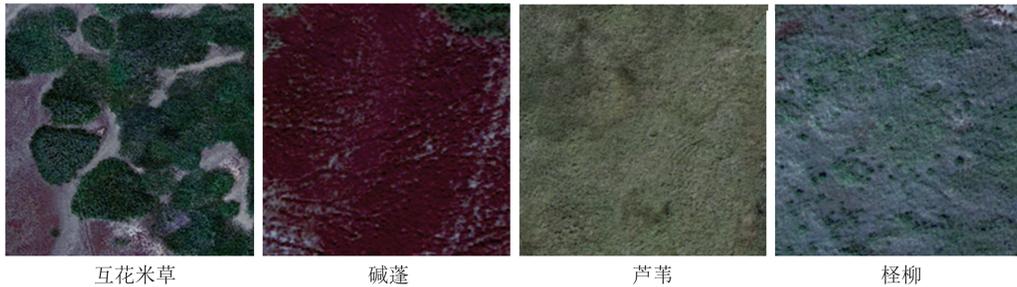


图 2 植被类型目视解译标志

Fig. 2 Visual Interpretation of Vegetation Types

100 个,样本数据质量可用 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,其中 70% 用于随机森林分类的训练,30% 用于精度验证。

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,具体内容如下:

1) 光谱特征选择影像的蓝(B)、绿(G)、红(R)、近红外(near infrared, NIR)4 个波段数据归一化后的像元值。

2) 指数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)等 5 种指数。

3) 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,可以很大程度提高遥感影像分类的精度。利用 ENVI 5.6 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^[19]。在进行纹理分析时,选取 5 个冗余度较小的二阶统计量对影像进行定量分析,包括角二阶矩、相关性、对比度、熵和方差,分别简称为 GLCM_A、GLCM_Cor、GLCM_Con、GLCM_E 和 GLCM_V。通常计算 GLCM 的参数时,为了避免得到冗余繁多的纹理特征,会将 0°、45°、90°、135° 4 个方向的特征值的平均值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 GLCM 提取了 5 个参数,共 10 个纹理特征。

4) 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^[20]。将 RGB 通道数据分别除以图像像素的最大值以进行归一化处理。首先,对于 8 位 RGB 通道,选择其最大值为 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。但颜色指数也能通过其他方式生成^[21],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,部分学者直接将其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^[22]。根据解译标志可知,黄河三角洲植被的色彩具有明显差距,首先仿照处理 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 3 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式为:

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

本文选取的颜色特征有超绿指数(excess green index, ExG)^[23]、超红指数(excess red index, ExR)^[24]、超蓝指数(excess blue index, ExB)^[25]、绿-红差值指数(green-red difference index, GRDI)^[26]、

绿-蓝差值指数(green-blue difference index, GBDI)^[27]、超绿-超红差值指数(excess green-excess red difference index, ExGR)^[28]、三角绿度指数(triangular greenness index, TGI)^[29]、修正超绿指数(modified excess green index, MExG)^[30]、绿-蓝差值+修正超绿指数(green-blue difference+modified excess green index, GBDI+MExG)^[31]。计算式分别为:

$$I_{\text{ExG}} = 2g - r - b \quad (2)$$

$$I_{\text{ExR}} = 1.4r - g \quad (3)$$

$$I_{\text{ExB}} = 1.4b - g \quad (4)$$

$$I_{\text{GRDI}} = g - r \quad (5)$$

$$I_{\text{GBDI}} = g - b \quad (6)$$

$$I_{\text{ExGR}} = 3g - 2.4r - b \quad (7)$$

$$I_{\text{TGI}} = g - 0.39r - 0.61b \quad (8)$$

$$I_{\text{MExG}} = 1.262g - 0.884r - 0.311b \quad (9)$$

$$I_{\text{GBDI+MExG}} = 2.262g - 0.884r - 1.311b \quad (10)$$

表1列举了各种特征及其数量。根据表1可得到28种特征因子,即4种光谱特征、5种指数特征、10种纹理特征和9种颜色特征。

表1 航空影像特征集描述

特征	简称	特征数
光谱特征	B、G、R、NIR	4
指数特征	NDVI、RVI、SAVI、NDWI、DVI	5
纹理特征	GLCM_A、GLCM_Cor、GLCM_Con、GLCM_E、GLCM_V	10
颜色特征	ExG、ExR、ExB、GRDI、GBDI、ExGR、TGI、MExG、GBDI+MExG	9

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。

2.3 随机森林分类方法

随机森林是当今运用广泛、高度灵活的一种基于集成学习思想的经典机器学习算法,其本质是通过构建多棵决策树并集成其预测结果来实现更优的分类性能。该方法使用平均值来提高预测精度和控制过度拟合,以其出色的泛化能力

表2 实验方案信息

Table 2 The Information of Experimental Programs

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

和鲁棒性被广泛应用于各类数据挖掘任务中。随机森林通过以下步骤实现:

1) Bootstrap 抽样。在原始样本集中有放回地随机抽样 N 个样本组成训练子集。每轮随机采样中,训练子集中有 $(1-1/N)^N$ 比例的样本未被选中。当 N 足够大时,该比例约为 36.8%,这些样本形成袋外(out of bag, OOB)数据,可用于模型性能评估。

2) 随机特征选择。抽取 N 个训练子集生成 N 棵决策树组成随机森林。在构建每棵决策树时,如果原始特征维度为 M ,则从 M 个特征中随机选取常数 $m(m \ll M)$ 个特征子集,根据基尼指数,从 m 个特征子集中选择最佳特征在决策树中进行节点分裂。

3) 集成决策。对于分类任务,最终结果通过所有决策树的投票表决产生,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数据冗余,降低计算复杂度、提高模型效率。若直接采用表1中全部特征因子进行分类,将会导致高维特征空间计算复杂度增加、模型训练负担加重、冗余特征干扰模型学习、分类性能和效率下降等问题。为解决上述问题,本文采用随机森林特征重要性评估的方法筛选出对地物分类起关键作用的最优特征子集进行分类。该方法依据平均不纯度减少(mean decrease impurity, MDI)对所有特征的重要性进行排列,通过逐步剔除低重要性特征,观察模型分类精度的变化,在精度无明显下降的前提下,选择特征数量最少的组合作为最优特征子集。

随机森林中决策树构建的核心环节在于节点分裂时的特征选择。理想情况下,节点分裂使得划分后的子节点样本属于同一类,即称分裂得到的每个子节点是纯的,基尼指数是衡量样本集合的不纯度的重要指标。集合 D 的基尼指数 G 定义如下:

$$G(D) = \sum_{b=1}^B p_b(1-p_b) = 1 - \sum_{b=1}^B p_b^2 \quad (11)$$

式中, B 为训练样本类别总数; 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 的某个取值 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 的基尼指数计算式为:

$$G(D, A) = \frac{|D_1|}{|D|} G(D_1) + \frac{|D_2|}{|D|} G(D_2) \quad (12)$$

式中, $|D|$ 、 $|D_1|$ 、 $|D_2|$ 分别表示集合 D 、 D_1 、 D_2 中的样本数。在随机森林中, 当使用某个特征进行节点分裂后平均基尼指数的减少程度越大, 表明该特征提升子节点纯度的程度越大, 对样本分类的贡献越大, 其重要性评分也就越高。平均不纯度减少表示为:

$$\Delta G = \frac{1}{K} \sum_{n=1}^K [G_n(D) - G_n(D, A)] \quad (13)$$

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

使用 OOB 误差进行模型评估以确定模型最优特征数量。首先对特征重要性进行降序排列, 依次构建包含前 1~28 个特征的模型, 计算各模型的 OOB 误差, 最优特征数量确定为 OOB 误差达到稳定最小值时所对应的特征子集规模。

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(overall accuracy, OA)、Kappa 系数、用户精度(user accuracy, UA)和生产者精度(producer accuracy, PA)来度量结果的准确度和可靠性。各指标的计算式分别为:

$$p_{OA} = \frac{\sum_{i=1}^N J_{ij}}{\sum_{i=1}^N \sum_{j=1}^N J_{ij}} \quad (14)$$

$$p_{Kappa} = \frac{\sum_{i=1}^N \sum_{j=1}^N J_{ij} \cdot \sum_{i=1}^N J_{ii} - \sum_{i=1}^N (\sum_{j=1}^N J_{ji} \cdot \sum_{j=1}^N J_{ij})}{\sum_{i=1}^N \sum_{j=1}^N J_{ij} \cdot \sum_{i=1}^N \sum_{j=1}^N J_{ij} - \sum_{i=1}^N (\sum_{j=1}^N J_{ji} \cdot \sum_{j=1}^N J_{ij})} \quad (15)$$

$$p_{UA} = J_{ii}/J_{i+} \quad (16)$$

$$p_{PA} = J_{ij}/J_{+i} \quad (17)$$

式中, N 为混淆矩阵的行列数; 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 1~5 作为对比实验, 不需要进行特征选择, 这 5 种方案用于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, 选择优选的特征组合构成方案 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, 其中, PCA1_GLCM_A 为第一主成分的角二阶矩, PCA2_GLCM_A 为第二主成分的角二阶矩, 依此类推。

由图 3 可知, 重要性位居前 10 的特征中, 航空影像的原始波段仅占 1 个, 指数特征和颜色特征分别占据 4 个和 5 个, 而纹理特征则未进入重要性前列。对植被提取分类重要性高的多为指数特征与颜色特征, 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, 能够被有效识别出来。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。

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 4 可知, 当特征数为 1~10 时, 精度逐渐增加, 在这个过程中, 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 开始增大, OOB 精度基本维持不变, 起伏很小。特征个数过小会导致精度不够, 使得分类结果不理想。特征数量过多则需要计算的工作量增大, 精度变化极小, 增长运算时间, 反而可能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%。

3.2 提取结果及精度评价分析

6 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 1~3 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 4~6 能将两者较好区分

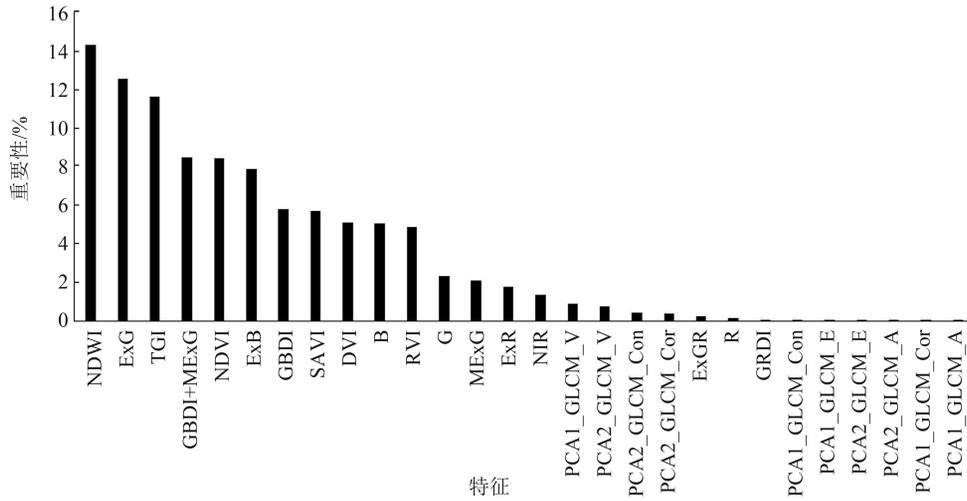


图3 特征重要性分布

Fig. 3 Distribution of Characteristic Importance

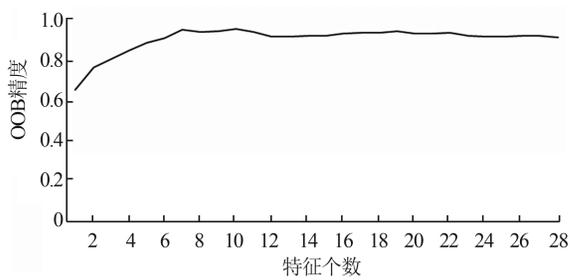


图4 特征个数与OOB精度的关系

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

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怪柳、芦苇与互花米草错分的情况。

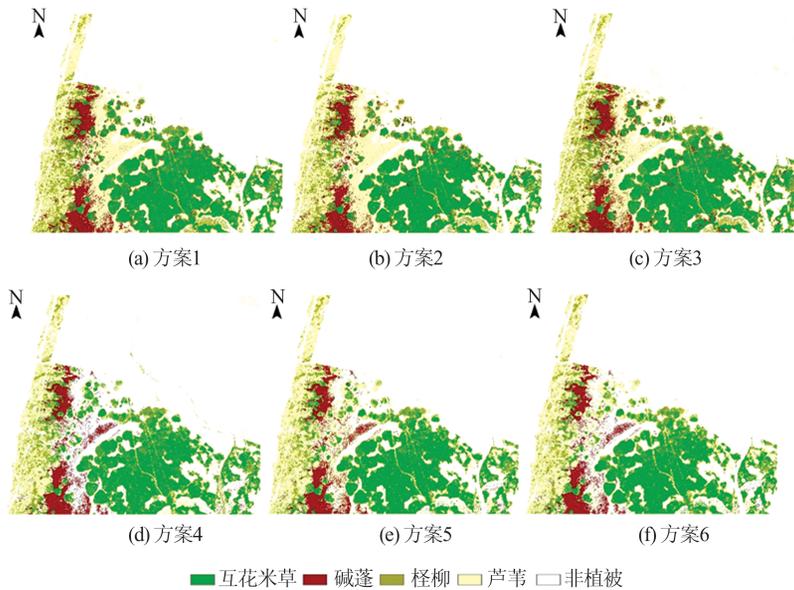


图5 不同方案分类结果

Fig. 5 Classification Results of Different Schemes

为了对所有方案进行定量分析,利用混淆矩阵的4个指标进行精度评价,结果见表3。由表3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系数分别为0.79、0.80与0.81。但是加入纹理特征的方

表 3 分类精度统计

Table 3 Statistics of Classification Accuracy

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

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 的总体精度与 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,尤其是怪柳与芦苇,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提高精度的主要原因。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,除怪柳以外类别的生产者精度均在 85% 以上。这是因为怪柳容易与芦苇混淆,被分类器分为芦苇。由于怪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。

4 结 语

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 4 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:

1) 以可见光谱与近红外光谱为基础,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。

2) 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 88.00%,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,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。但仍然存在一些不足:(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 4 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] 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(2): 230-237.
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(2): 230-237.
- [2] 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.
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.
- [3] 王仁卿, 张煜涵, 孙淑霞, 等. 黄河三角洲植被研究回顾与展望[J]. 山东大学学报(理学版), 2021, 56(10): 135-148.
WANG Renqing, ZHANG Yuhan, SUN Shuxia, et al. Review and Prospect of Vegetation Research in the Yellow River Delta [J]. *Journal of Shandong University (Natural Science)*, 2021, 56(10): 135-148.
- [4] 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]. 生态学杂志, 2024, 43(11): 3240-3245.
ZOU Yuhuan, 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, 43(11): 3240-3245.
- [5] 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.
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.
- [6] 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(1): 125-136.
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.
- [7] 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(1): 68-80.
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.
- [8] 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(2): 313-326.
ZHANG Lei, GONG Zhaoning, WANG Qiwei, 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.
- [9] 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(4): 47-55.
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(4): 47-55.
- [10] ZHANG C, GONG Z N, QIU H C, 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] LIU Q S, HUANG C, LI H. 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(2): 975-984.
- [12] 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(6): 1387-1399.
HAN Yue, KE Yinghai, WANG Zhanpeng, et al. Classification of the Yellow River Delta Wetland Landscape Based on ZY-1 02D Hyperspectral Imagery [J]. *National Remote Sensing Bulletin*, 2023, 27(6): 1387-1399.
- [13] TU C R, LI P, LI Z H, 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 Sensing*, 2021, 13(21): 4444.
- [14] 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(4): 159-166.
LI Yanzhou, HE Yanzhou, QIN Feng, et al. Identification of *Spartina Alterniflora* Based on Convolutional Neural Network [J]. *Journal of Chinese Agri-*

- cultural Mechanization*, 2023, 44(4): 159-166.
- [15] 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(1): 100-108.
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(1): 100-108.
- [16] 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(6): 1467-1479.
MIN Yukui, KE Yinghai, HAN Yue, et al. Dynamic Monitoring of Invasive *Spartina Alterniflora* Clearance via Fusion of Sentinel-2 and GF-1 Time Series Images [J]. *National Remote Sensing Bulletin*, 2023, 27(6): 1467-1479.
- [17] 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(4): 596-602.
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(4): 596-602.
- [18] 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(5): 152-157.
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(5): 152-157.
- [19] 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(7): 347-359.
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(7): 347-359.
- [20] HAMUDA E, GLAVIN M, JONES E. 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] MARDANISAMANI S, ERAMIAN M. Segmentation of Vegetation and Microplots in Aerial Agriculture Images: A Survey [J]. *The Plant Phenome Journal*, 2022, 5(1): e20042.
- [22] LEE M K, GOLZARIAN M R, KIM I. A New Color Index for Vegetation Segmentation and Classification [J]. *Precision Agriculture*, 2021, 22(1): 179-204.
- [23] WOEBBECKE D M, MEYER G E, VON BARGEN K, et al. Color Indices for Weed Identification Under Various Soil, Residue, and Lighting Conditions [J]. *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 [C] // Precision Agriculture and Biological Quality, Boston, USA, 1999.
- [25] GUIJARRO M, PAJARES G, RIOMOROS I, et al. Automatic Segmentation of Relevant Textures in Agricultural Images [J]. *Computers and Electronics in Agriculture*, 2011, 75(1): 75-83.
- [26] MOTOHKA T, NASAHARA K N, OGUMA H, et al. Applicability of Green-Red Vegetation Index for Remote Sensing of Vegetation Phenology [J]. *Remote Sensing*, 2010, 2(10): 2369-2387.
- [27] 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.
DAI Jianguo, XUE Jinli, ZHAO Qingzhan, et al. Extraction of Cotton Seedling Growth Information Using UAV Visible Light Remote Sensing Images [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2020, 36(4): 63-71.
- [28] MEYER G E, CAMARGO NETO J, JONES D D, et al. Intensified Fuzzy Clusters for Classifying Plant, Soil, and Residue Regions of Interest from Color Images [J]. *Computers and Electronics in Agriculture*, 2004, 42(3): 161-180.
- [29] HUNT E R, DAUGHTRY C S T, EITEL J U H, et al. Remote Sensing Leaf Chlorophyll Content Using a Visible Band Index [J]. *Agronomy Journal*, 2011, 103(4): 1090-1099.
- [30] BURGOS-ARTIZU X P, RIBEIRO A, GUIJARRO M, et al. Real-Time Image Processing for Crop/Weed Discrimination in Maize Fields [J]. *Computers and Electronics in Agriculture*, 2011, 75(2): 337-346.
- [31] 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(6): 508-522.
JIANG Jiale, LI Rongwang, MA Xuehua, et al. 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.