

引用格式: Yan Runzhou, Li Liwei, Wang Tao, *et al.* Spectral Clustering based Feature Extraction for Parcel Classification Using High Spatial Resolution Remote Sensing Images [J]. Remote Sensing Technology and Application, 2021, 36(3): 705-712. [闫润州, 李利伟, 王涛, 等. 基于光谱聚类的高分影像复杂地块特征提取方法研究[J]. 遥感技术与应用, 2021, 36(3): 705-712.]  
doi: 10.11873/j.issn.1004-0323.2021.3.0705

## 基于光谱聚类的高分影像复杂地块特征提取方法研究

闫润州<sup>1,2</sup>, 李利伟<sup>2</sup>, 王涛<sup>3</sup>, 陈俊奇<sup>3</sup>, 赖健<sup>3</sup>, 张兵<sup>1,2</sup>

(1. 中国科学院大学, 北京 100049;

2. 中国科学院空天信息创新研究院 中国科学院数字地球重点实验室 北京 100094;

3. 上海卫星工程研究所 高分上海数据与应用中心 上海 201109)

**摘要:** 高空间分辨率遥感技术为大范围判识农用地利用类型提供了丰富的数据源。农用地类型多样性和复杂性给高效应用高分影像识别农用地类型带来很大挑战。地块矢量的引入可以帮助更好综合利用多元影像特征, 进而提高农用地类型判识精度。但是, 传统地块特征提取方法将地块视为一个整体, 通过对地块内部像元特征平均实现地块特征表达, 不能很好适用于地块内部像元光谱具有较强异质性的情况。针对内部光谱异质但具有较强规律的地块, 提出一种基于光谱聚类的特征提取方法, 将地块内部的光谱聚类特征作为地块的特征之一, 对地块进行分类。利用上海崇明区内2个典型样区的BJ-2卫星影像和地面调查数据进行实验验证分析, 结果表明: ①该方法相对利用地块内部所有像元光谱平均的方法, 能够有效提升地块分类精度; ②通过引入地块内部光谱聚类特征到传统地块特征组合, 可以进一步提升地块分类精度, 对菜地和廊道等内部像元光谱混合比例变化较大的类别提升效果最为明显。该方法为复杂地块分类提供了新思路。

**关键词:** 北京2号; 地块; 特征; 聚类

**中图分类号:** TP753 **文献标志码:** A **文章编号:** 1004-0323(2021)03-0705-08

### 1 引言

准确及时地获取农用地利用类型, 对于预测农作物产量, 优化农作物种植空间格局, 辅助政府决策实现农用地的科学管理具有重要意义<sup>[1]</sup>。遥感技术因其覆盖面积大、高效、省时省力的优点, 相比传统逐级汇报和抽样调查方法更能及时、准确、直观地掌握农用地空间分布信息。受到作物种植类型、地块大小形状、成像条件等因素影响, 农用地遥感特征十分复杂, 如何充分挖掘遥感影像特征实现农

用地类型快速准确识别是当前研究的一个热点<sup>[2-3]</sup>。

依据观察空间粒度不同, 农用地影像特征可以在像元、对象、地块3个层次提取。基于像元的特征提取方法依靠像元光谱特征描述地物类型, 计算简单, 适用于中低空间分辨率光学影像<sup>[4-6]</sup>, 但是, 像元特征不能有效刻画地物空间结构, 随着空间分辨率提高, 容易造成误分类; 面向对象的特征提取方法以对象为单元构建高层语义特征, 能够有效弥补像元特征的不足<sup>[7-8]</sup>。但是, 该方法依赖于图像对象分

收稿日期: 2020-01-14; 修订日期: 2021-04-14

基金项目: 国家自然科学基金项目“中高分辨率多时相多源光学影像分类模型与方法研究”(41971327), 国家重点研发计划项目“地球资源环境动态监测技术”(2016YFB0501501)。

作者简介: 闫润州(1995—), 男, 浙江宁波人, 硕士研究生, 主要从事高分遥感图像与机器学习、深度学习结合方面的研究。

E-mail: azhouazhou5@163.com

通讯作者: 李利伟(1992—), 男, 河南洛阳人, 副研究员, 主要从事遥感信息提取理论与方法研究。E-mail: lilw@radi.ac.cn

割质量,给面向对象特征提取方法应用带来困难。面向地块的特征提取引入人工矢量化地块信息作为先验知识,相比于面向对象的方法,人工地块对象信息更加稳定且符合实际应用需求<sup>[3]</sup>。

现有基于地块的影像特征提取方法通常将地块视为一个整体对象,通过对地块内部像元的光谱均值、纹理均值、植被指数均值等作为地块影像特征<sup>[9-10]</sup>。该方法可以抑制地块内少量背景噪声的影响,适用于地块内部比较一致的情况。然而,实际应用中,由于地块边界时效性或划分方法的限制,具有单一功能标识的地块内部像元可能存在规律性的光谱异质特点,例如,稀疏林地地块通常是裸土和树木的组合;作物地块内部可能存在田埂;菜地块内只有部分蔬菜覆盖等。该情况下将地块作为一个整体进行平均特征提取,可能会造成有效影像信息丢失,不利于复杂地块判别<sup>[11]</sup>。

现有针对混合地块分类的研究较少。顾晓鹤等<sup>[10]</sup>通过计算地块内部变异系数筛选出光谱均一地块,进而仅针对纯地块进行分类;Kussul等<sup>[12]</sup>通过对地块内部类别连通斑块的形状和面积建立规则,融合基于像元的分类结果和基于地块平均的分类结果,应用于 Sentinel-1A 和 Landsat 8 数据地块分类。由于遥感影像地物特征的不同空间尺度差异较大,现有方法尚不能满足高分影像中复杂地块分类的需求。

针对亚米级光学影像复杂地块分类,结合地面调查和高分影像判读经验,借鉴现有基于聚类分析的复杂地物遥感影像特征提取策略<sup>[13]</sup>,提出一种基于光谱聚类的地块特征提取方法,将地块内部不同光谱簇地物分离后作为一种增强特征,进行地块分类,并利用上海崇明区2个典型样区的 BJ-2 影像和地块调查矢量对提出的聚类特征进行实验验证。

## 2 研究方法

### 2.1 地块内部光谱异质特点

地块在高分遥感影像中常以混合地物形式出现,即一个地块内部不止存在一种光谱相似的地物,并且存在规律性的地物组合,如图1所示。其中,水稻、水产地、大棚等地块内包含功能地物与田埂或田间小道;林地地块包含裸土与林木;廊道包含水体与林木;菜地地块内部包含蔬菜与裸土。对于混合地块,不适宜用简单平均的途径表达地块影像特征,一种直观的方法是采用光谱聚类将地块内

部不同光谱特征的像元簇分开表达,实现复杂地块的精细特征表达。

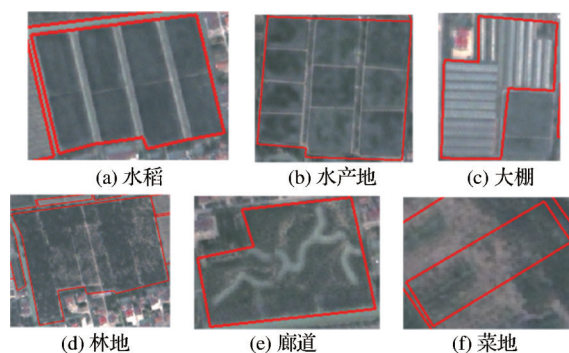


图1 光谱异质性地块

Fig.1 Parcels with spectral heterogeneity

### 2.2 地块像元光谱聚类方法

研究采用一种改进的 k-means 算法进行地块内部像元聚类。传统 k-means 算法容易受到初始聚类中心随机选择的影响,不能收敛到最优解。为了获取更加稳定聚类结果,引入一种聚类中心优选机制<sup>[14]</sup>,使得初始聚类中心之间距离尽可能远。假设输入向量集合为  $X = \{x_1, x_2, x_3, \dots, x_{n-1}, x_n\}$ , 聚类数目为  $k$ , 改进后 k-means 算法步骤如下:

(1) 从输入的向量集合中随机选取一个向量  $x_i \in X$  作为第一个聚类中心。

(2) 对于满足  $\{x_j | x_j \in X\}$  的所有向量,计算其与最近的聚类中心的距离  $D(x_j)$ , 选择  $D(x_j)$  最大的向量作为新聚类中心,迭代直到选择出  $k$  个聚类中心。


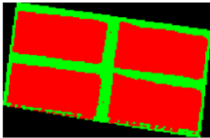
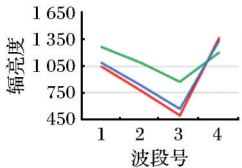
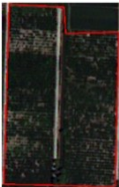
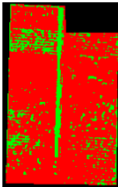
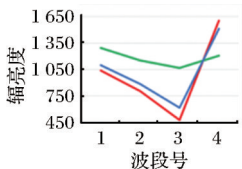

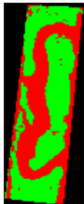
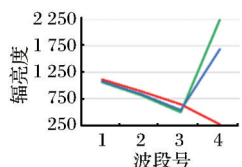


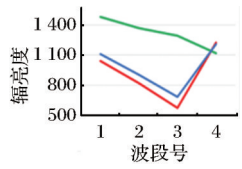

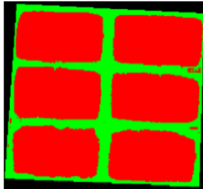
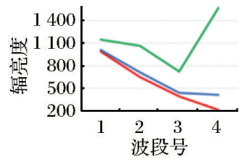

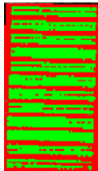
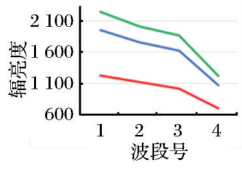
(3) 对以上 2 个步骤重复 10 次,选择聚类中心距离最大的一组聚类中心作为初始聚类中心。

(4) 依据优选出的  $k$  个初始聚类中心执行 k-means 算法。实验中, k-means 迭代终止条件为聚类中心变化幅度小于 0.01% 或者迭代次数大于 300 次。

为验证以上聚类方法描述复杂地块特征的有效性,选择 6 个不同类别典型地块进行 2 类聚类实验,如表 1 所示。通过光谱聚类结果可以看出,水稻地块内水稻和田埂或小道,林地地块内林地与裸土,廊道地块内水体与林木,菜地地块内蔬菜与裸土,水产地块内水面与小路,以及大棚地块内棚顶与小路都很好分离。同时,观察聚类中心光谱可知,当地块内单一地物像素占据主体时,平均光谱会接近此类地物光谱,类似于地块平均光谱的结果,然而,当地块中存在一种以上像元光谱不同的地物时,地块平均光谱的表达并不稳定,此时聚类特征呈现出更好的表达效果。

表 1 典型地块地物聚类结果示意

Table 1 Demonstration of typical parcel clustering results

类别	地块真彩色图像	地块聚类结果	光谱曲线示意
水稻			
林地			
廊道			
菜地			
水产地			
大棚			

注:—为绿簇平均光谱,—为红簇平均光谱,—为地块平均光谱。

2.3 基于光谱聚类的特征提取方法

给定卫星影像与矢量地块,基于光谱聚类的特征提取方法主要包含光谱聚类,聚类簇平均光谱计算,特征排序 3 个步骤,如图 2 所示。

(1)改进 k-means 光谱聚类。对每一个地块内部所有像元通过聚类方法分为两类,每个地块形成两个聚类簇。两类的设定是依据实际经验观察,复杂地块内部像元通常可以通过两个光谱簇进行表达,同时,混合地物主要为裸土地或水体。

(2)聚类簇平均光谱计算。依据地块内部像元光谱聚类结果,逐波段提取每一个聚类簇的平均光谱,用于精细描述不同簇的地物特征。

(3)聚类簇特征约束排序。将两个聚类簇特征按照近红外波段值升序排列,然后将排序后特征进行向量堆积,形成用于地块分类的聚类特征。这里特征排序是为了规范聚类过程中随机聚类中心选择带来的聚类簇次序的不确定性。近红外波段的选择是考虑到地块混合地物多为水体和稀疏植被



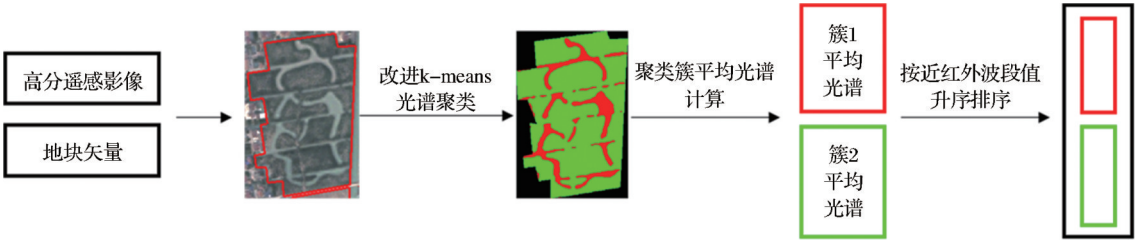


图 2 光谱聚类特征提取流程

Fig.2 Spectral clustering based feature extraction process

覆盖的裸土。

3 数据与实验

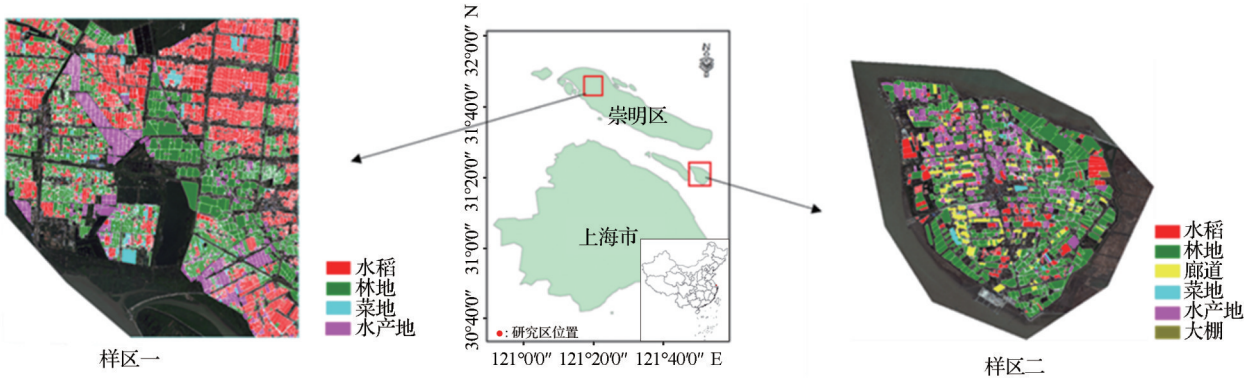
3.1 实验数据

选择上海市崇明区内两个典型实验样区,设计不同类别体系进行地块分类实验,如图 3 所示。

在实验样区 1 中,采用 2017 年 7 月 21 号获取的 BJ-2 影像和对应地块调查矢量进行单时相分类;样

区共计 6 729 个地块,包含水稻、林地、菜地和水产地 4 个类别,其中,林地包含果园、生态林等,不同类别的地块数量如表 2 所示。

在实验样区 2 中,采用 2017 年 7 月 25 号和 2018 年 2 月 23 号获取的 BJ-2 影像和对应地块调查矢量进行双时相分类。样区共计 1 795 个地块,包含水稻、林地、廊道、菜地、水产地和大棚 6 个类别,不同类别的地块数量如表 2 所示。



审图号:GS(2019)1822

图 3 研究区地理位置、遥感影像与地块真值图分布

Fig.3 Geographical location of the study area, remote sensing image and distribution of fields truth map

表 2 地块类别体系描述

Table 2 Description of parcel classification system

编号	类别名称	样区 1 地块数量	样区 2 地块数量
1	水稻	3222	257
2	林地	2487	852
3	廊道	-	129
4	菜地	287	133
5	水产地	733	391
6	大棚	-	33

两个实验样区采用了经过辐射定标,几何精校正以及全色多光谱融合的 BJ-2 影像,空间分辨率 0.8 m,包含蓝、绿、红、近红外 4 个波段。

在实验样本方面,两个实验样区各类别分别随机选取 50% 作为训练样本,剩余 50% 地块作为测试样本。为了平衡不同类别训练样本的数量,采用

了自适应合成抽样方法 (Adaptive synthetic sampling) 对训练样本中类别进行过采样处理<sup>[15-16]</sup>。

3.2 实验方案设计

结合 2 个实验样区影像和地块调查矢量,选择不同地块特征组合进行监督分类,利用分类结果精度客观评价聚类特征进行地块分类的效果。下面对实验设计及其关键环节进行具体描述。

实验选用的地块特征包括光谱特征、指数特征、纹理特征和聚类特征,具体描述如表 3 所示。每类特征都包含多个特征值,具体实验中,将一类特征内的所有特征值按照固定次序进行向量堆积叠加,形成用于分类的特征向量。

实验分类器选用随机森林。随机森林作为一种集成学习算法模型,基本单元是决策树,每一

表 3 实验采用的地块特征说明

Table 3 Description of features in the experiments

特征名称	特征说明
光谱特征	地块内像元逐波段辐射亮度平均值,实验采用蓝、绿、红、近红外 4 个波段
指数特征	地块内像元逐个光谱指数平均值,实验采用指数包括 NDVI、NDWI
纹理特征	地块内像元逐波段纹理特征平均值,实验采用基于 GLCM 的熵和对比度(窗口为 5,步长为 1)
聚类特征	本文提出的聚类特征(聚类个数为 2)

棵决策树都是一个弱分类器,在训练过程中随机选取部分样本和部分特征进行决策树训练,训练完成后,所有的弱分类器结果组合投票决定分类结果。随机森林分类器能够适用于复杂高维空间特征表达,且计算效率高,在遥感影像分类中被广泛采用<sup>[17]</sup>。本实验中随机子树个数设定为 100,随机特征维度设定为特征数量平方根,特征评价采用 gini 系数。

实验采用定量为主定性为辅的综合方法对不同分类结果进行精度评价。定量指标主要基于分类混淆矩阵的总体精度、Kappa 系数、用户精度和制图精度。定性评价结合定量评价结果、实地调查和高分影像判读,分析分类结果误差。

为了有效评价聚类特征的地块分类性能,结合不同特征组合设计 5 组实验,如表 4 所示。实验 1 用于对比光谱特征与聚类特征的分类性能;实验 2~5 用于评价聚类特征作为一种新特征加入传统特征组合后的分类性能。

4 结果与分析

4.1 利用总体分类精度评价聚类特征效果

依据表 4 进行地块分类实验,结果验证总体精度和 Kappa 如表 5 所示。

根据实验 1 结果,单独对比光谱特征和聚类特

表 4 地块分类实验设计

Table 4 Experiment design for parcel classification

实验	特征组合 1	特征组合 2
1	光谱特征	聚类特征
2	光谱特征	光谱、聚类特征
3	光谱、纹理特征	光谱、纹理、聚类特征
4	光谱、指数特征	光谱、指数、聚类特征
5	光谱、纹理、指数特征	光谱、纹理、指数、聚类特征

征,聚类特征相比光谱特征在两个实验样区均取得更好总体分类效果,总体精度平均提高 4.12%,Kappa 平均提高 0.055 2,其中,实验样区 1 总体精度提高 4.46%,Kappa 提高 0.065 1;实验样区 2 总体精度提高 3.78%,Kappa 提高 0.045 3。根据实验 2~5 结果,聚类特征作为一种新特征加入传统特征组合,可以有效提高传统特征组合总体精度与 Kappa,总体精度平均提升 3.15%,Kappa 平均提升 0.043 6,其中,实验样区 1 总体精度平均提高 2.76%,Kappa 平均提高 0.040 8;实验样区 2 总体精度平均提高 3.54%,Kappa 平均提高 0.046 4。

4.2 利用各类别分类精度评价聚类特征效果

依据表 4 进行地块分类实验,实验结果中各类别的用户精度和制图精度如表 6、表 7 所示。

根据实验 1 结果,单独对比光谱特征,在用户精度与制图精度上,聚类特征整体优于光谱特征,其

表 5 实验总体精度

Table 5 Overall accuracy of experiments

样区	实验	组合 1	组合 2	组合 1	组合 2
		总体精度/%	总体精度/%	Kappa	Kappa
1	1	71.62	76.08	0.563 4	0.628 5
	2	71.62	76.73	0.563 4	0.636 8
	3	81.13	82.26	0.704 0	0.720 8
	4	73.05	76.64	0.582 0	0.636 0
	5	81.81	83.03	0.713 6	0.732 5
2	1	78.40	82.18	0.699 8	0.745 1
	2	78.40	82.52	0.699 8	0.752 1
	3	80.18	84.08	0.719 5	0.772 0
	4	79.06	82.96	0.704 9	0.754 6
	5	82.74	84.97	0.751 5	0.782 5

表 6 实验数据 1 中类别用户精度和制图精度(单位: %)

Table 6 User's accuracy and producer's accuracy in the No. 1 experiment( Unit: % )											
精度类别	地物类别	实验 1		实验 2		实验 3		实验 4		实验 5	
		组合 1	组合 2	组合 1	组合 2	组合 1	组合 2	组合 1	组合 2	组合 1	组合 2
制图精度	水稻	69.16	75.80	69.16	76.23	82.25	83.66	73.16	75.80	83.11	83.78
	林地	71.79	73.59	71.79	75.39	79.23	80.21	70.32	75.14	79.89	82.01
	菜地	53.74	59.86	53.74	56.46	56.46	57.14	51.70	59.86	55.78	56.46
	水产地	89.10	92.10	89.10	91.55	92.37	92.92	90.19	92.10	92.92	93.73
用户精度	水稻	83.04	85.16	83.04	85.76	88.50	89.08	83.29	85.40	88.32	88.57
	林地	68.97	74.26	68.97	73.76	81.09	82.02	69.75	73.93	81.83	82.05
	菜地	24.09	28.57	24.09	29.75	29.75	31.70	23.38	30.45	31.66	34.02
	水产地	80.15	85.35	80.15	86.38	89.68	90.93	87.80	87.11	89.74	89.82

中,在实验样区 1,聚类特征在所有类别明显优于光谱特征;在实验样区 2,聚类特征优于光谱特征,但在少数类别用户精度或制图精度上,光谱特征略有优势。

根据实验 2~实验 5 结果,聚类特征作为一种新特征加入传统特征组合时,几乎所有类别的用户精度和制图精度都得到提高。通过计算每个实验样区 4 个实验各类别的平均提高百分比,可知,在实验样区 1,制图精度与用户精度提升最大的类别均为菜地,提升幅度分别为 3.06% 与 4.26%;在实验样区 2,制图精度与用户精度提高最大的类别分别为廊道地与菜地,提升幅度分别为 8.33% 与 10.52%。结果表明聚类特征在地块描述方面具有传统地块特征不具备的能力,能够进一步帮助提高地块分类精度。

5 结 语

面向复杂地块分类,提出一种基于光谱聚类的

特征提取方法。利用 2 个代表性实验样区内 BJ-2 遥感影像和地块调研矢量,在随机森林分类器的支持下,结合光谱、纹理和植被指数等地块特征,进行了分类实验与结果分析,主要结论如下:

(1)该方法相对于地块内像元光谱平均的方法能够有效提升地块分类精度,具体到两个实验样区,总体精度平均提高 4.12%,Kappa 平均提高 0.055 2。

(2)引入聚类特征到传统特征组合均能进一步提升地块分类精度,具体到两个实验样区,总体精度平均提高 3.15%,Kappa 平均提高 0.043 6,其中,对于菜地、廊道和林地等内部地物混合比例多变的地块类别提升效果最为明显。

该方法为复杂地块分类提供了新思路。同时,为了能够更大范围推广应用该方法,仍然需要进一步分析方法中地块内部聚类个数的影响。实验中依据已有实验区地块特点,设定聚类个数为 2,验证了方法的有效性,但是,采用更多聚类个数的有效性需要进一步实验。理论上,采用更多聚类特征能

表 7 实验数据 2 中类别用户精度和制图精度(单位: %)

Table 7 User's accuracy and producer's accuracy in the No. 2 experiment( Unit: % )											
精度类别	地物类别	实验 1		实验 2		实验 3		实验 4		实验 5	
		组合 1	组合 2	组合 1	组合 2	组合 1	组合 2	组合 1	组合 2	组合 1	组合 2
制图精度	水稻	82.81	86.72	82.81	85.94	79.69	84.38	80.47	83.59	82.03	82.03
	林地	77.06	85.32	77.06	83.26	81.65	86.24	79.36	86.01	86.24	88.99
	廊道	75.00	71.67	75.00	78.33	61.67	76.67	71.67	76.67	66.67	76.67
	菜地	42.65	44.12	42.65	42.65	50.00	51.47	38.24	39.71	44.12	47.06
	水产地	93.75	91.15	93.75	95.83	94.27	94.79	96.35	94.27	95.83	95.83
	大棚	57.14	50.00	57.14	57.14	71.43	57.14	50.00	64.29	57.14	57.14
用户精度	水稻	74.13	80.43	74.13	80.29	77.27	82.44	78.03	82.31	83.33	85.37
	林地	92.56	91.18	92.56	91.90	90.82	91.71	90.34	88.86	89.95	90.87
	廊道	46.39	53.09	46.39	52.81	46.84	57.50	48.86	53.49	50.00	54.12
	菜地	34.94	48.39	34.94	43.94	44.16	54.69	36.11	48.21	53.57	64.00
	水产地	91.37	88.38	91.37	93.40	90.05	91.46	88.52	94.76	90.20	92.00
	大棚	53.33	63.64	53.33	57.14	58.82	57.14	50.00	69.23	57.14	61.54

够更加精细描述地块内部聚类特点,有助于描述复杂光谱地块类型,但是,精细特征将增加特征空间的复杂性,可能降低特征分类结果的稳定性和可推广性。基于此,对于一个新的应用场景,建议首先结合实地调查和高分影像判读,获取典型地块内部光谱异质性特点,进而实验选择最佳的聚类个数。最后,由于地块多样性及其影像特征的复杂性,实验中最优总体分类精度仍不足90%,针对这一问题,下一步将继续围绕地块内部结构特点探索更有效的特征提取方法。

**致谢:**感谢中国科学院空天信息创新研究院申茜副研究员在论文撰写方面的帮助!

#### 参考文献(References):

- [1] Song Qian, Zhou Qingbo, Wu Wenbin, *et al.* Recent Progresses in Research of Integrating Multi-source Remote Sensing Data for Crop Mapping[J]. *Scientia Agricultura Sinica*, 2015, 48(6): 1122-1135.[宋茜,周清波,吴文斌,等.农作物遥感识别中的多源数据融合研究进展[J]. *中国农业科学*, 2015, 48(6): 1122-1135.]
- [2] Du Baojia, Zhang Jing, Wang Zongming, *et al.* Crop Mapping based on Sentinel-2A NDVI Time Series Using Object-oriented Classification and Decision Tree Model[J]. *Journal of Geo-information Science*, 2019, 21(5): 740-751.[杜保佳,张晶,王宗明,等.应用Sentinel-2A NDVI时间序列和面向对象决策树方法的农作物分类[J]. *地球信息科学学报*, 2019, 21(5): 740-751.]
- [3] Han Yanxin, Meng Jihua. A Review of Per-field Crop Classification Using Remote Sensing[J]. *Remote Sensing for Land and Resources*, 2019, 31(2): 1-9.[韩衍欣,蒙继华.面向地块的农作物遥感分类研究进展[J]. *国土资源遥感*, 2019, 31(2): 1-9.]
- [4] Smith G M, Fuller R M. An Integrated Approach to Land Cover Classification: An Example in the Island of Jersey[J]. *International Journal of Remote Sensing*, 2001, 22(16): 3123-3142. doi:10.1080/01431160152558288.
- [5] Lu D, Weng Q. A Survey of Image Classification Methods and Techniques for Improving Classification Performance[J]. *International Journal of Remote Sensing*, 2007, 28(5): 823-870. doi:10.1080/01431160600746456.
- [6] Wang Dejun, Jiang Qigang, Li Yuanhua, *et al.* Land Use Classification of Farming Areas based on Time Series Sentinel-2A/B Data and Random Forest Algorithm[J]. *Remote Sensing for Land and Resources*, 2020, 32(4): 236-243.[王德军,姜琦刚,李远华,等.基于Sentinel-2A/B时序数据与随机森林算法的农耕地土地利用分类[J]. *国土资源遥感*, 2020, 32(4): 239-246.]
- [7] Deng Yuanyuan, Wu Zhaocong, Yi Lina, *et al.* Research on Object-oriented Classification of Agricultural Land based on High Resolution Images[J]. *Remote Sensing for Land and Resources*, 2010, 22(4): 120-124.[邓媛媛,巫兆聪,易俐娜,等.面向对象的高分辨率影像农用地分类[J]. *国土资源遥感*, 2010, 22(4): 120-124.]
- [8] José M, Barragán P, Ngugi M K, *et al.* Object-based Crop Identification Using Multiple Vegetation Indices, Textural Features and Crop Phenology[J]. *Remote Sensing of Environment*, 2011, 115(6): 1301-1316. doi: 10.1016/j.rse. 2011. 01.009.
- [9] Nguyen-Thanh S, Chi-Farn C, Chen C, *et al.* Classification of Multitemporal Sentinel-2 Data for Field-level Monitoring of Rice Cropping Practices in Taiwan-Science Direct[J]. *Advances in Space Research*, 2020, 65(8): 1910-1921. doi: 10.1016/j.asr.2020.01.028.
- [10] Gu Xiaohe, Pan Yaozhong, He Xin, *et al.* Measurement of Sown Area of Winter Wheat based on Per-field Classification and Remote Sensing Imagery[J]. *Journal of Remote Sensing*, 2010, 14(4): 789-805.[顾晓鹤,潘耀忠,何馨,等.以地块分类为核心的冬小麦种植面积遥感估算[J]. *遥感学报*, 2010, 14(4): 789-805.]
- [11] Zhang B, Jia X, Chen Z, *et al.* A Patch-based Image Classification by Integrating Hyperspectral Data with GIS[J]. *International Journal of Remote Sensing*, 2006, 27(15): 3337-3346. doi:10.1080/01431160500409577.
- [12] Kussul N, Lemoine G, Gallego F J, *et al.* Parcel-based Crop Classification in Ukraine Using Landsat 8 Data and Sentinel-1A Data[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2016, 9(6): 2500-2508. doi: 10.1109/JSTARS.2016.2560141.
- [13] Lu Q, Huang X, Zhang L. A Novel Clustering-based Feature Representation for the Classification of Hyperspectral Imagery[J]. *Remote Sensing*, 2014, 6(6): 5732-5753. doi: 10.3390/rs6065732.
- [14] Arthur D, Vassilvitskii S. K-Means++: The Advantages of Careful Seeding[C]// *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2007, New Orleans, Louisiana, USA, January 7-9, 2007.* [ACM, 2007.]
- [15] He H, Bai Y, Garcia E A, *et al.* ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning[C]// *Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence).* IEEE International Joint Conference on. IEEE, 2008.
- [16] Fan Dongdong, Li Qiangzi, Wang Hongyan, *et al.* Improve-



ment in Recognition Accuracy of Minoritycrops by Resampling of Imbalanced Training Datasets of Remote Sensing[J]. Journal of Remote Sensing, 2019, 23(4): 730-742.[樊东东, 李强子, 王红岩, 等. 通过训练样本采样处理改善小宗作物遥感识别精度[J]. 遥感学报, 2019, 23(4): 730-742.]

[17] Tedros B, Charles L, Qiusheng W, *et al.* Decision-tree, Rule-based, and Random Forest Classification of High-resolution Multispectral Imagery for Wetland Mapping and Inventory [J]. Remote Sensing, 2018, 10 (4) : 580. doi: 10.3390/rs10040580.

## Spectral Clustering based Feature Extraction for Parcel Classification Using High Spatial Resolution Remote Sensing Images

Yan Runzhou<sup>1,2</sup>, Li Liwei<sup>2</sup>, Wang Tao<sup>3</sup>, Chen Junqi<sup>3</sup>, Lai Jian<sup>3</sup>, Zhang Bing<sup>1,2</sup>

(1. *University of Chinese Academy of Sciences, Beijing 100049, China;*

2. *Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China;*

3. *Shanghai Data & Application Center of High-resolution Earth Observation System, Shanghai Institute of Satellite Engineering, Shanghai 201149, China)*

**Abstract:** High spatial resolution satellite remote sensing provides redundant data for large-scale agricultural land management. Due to the variety and complexity of land parcels, it is still a great challenge to accurately and timely map land use types. Introducing parcel boundary has proven an effective strategy to integrate spatial and spectral features, to improve the classification accuracy. However, current feature extraction methods always treat each parcel as a whole and use only mean feature values of all pixels inside each parcel. This approach cannot well adapt to the scenarios that parcels include more than one types of spectral similar target. To this end, this paper proposes a spectral clustering based feature extraction method to better model the complexity of parcels. BJ-2 images and ground surveying data from 2 typical areas in the Chongming County in Shanghai were selected to experimentally evaluate the proposed method. The results show that: (1) Compared with the direct spectral averaging method, the proposed method can effectively improve the accuracy of land parcel classification; (2) Introducing the clustering features into the typical feature combination can further improve the accuracy of land parcel classification. And the improvement mainly lies on categories with unstable mixing ratio of internal pixel spectrum, such as vegetable field and corridor. The proposed method provides an effective alternative to classify parcel types especially for parcels including more than one spectral similar target.

**Key words:** BJ-2; Land parcels; Feature; Clustering