

基于高光谱数据的主要农作物类型信息提取

于成龙

(中国气象局东北地区生态气象创新开放实验室/黑龙江省气象院士工作站/黑龙江省气象科学研究所, 哈尔滨 150030)

摘要: 黑龙江省是我国最大的商品粮生产基地, 其农作物种植信息对于粮食估产, 防灾减灾, 乃至保障国家粮食安全均具有重要意义。为提高黑龙江省主要农作物种植类型信息的提取精度提供有参考意义的方法, 以黑龙江省齐齐哈尔市为研究区, 针对多光谱数据在区分玉米和大豆种植信息方面的局限性, 基于环境减灾星高光谱数据(HSI), 利用决策树法提取主要农作物类型信息。结果表明, 齐齐哈尔市的土地覆盖类型以农耕地(包括大豆、玉米、水稻和其它作物)为主, 占该市土地面积的70%以上, 其中玉米种植区是该市最主要的土地利用类型, 各作物分类精度均超过80%, 总体精度达到88.71%。

关键词: 高光谱; 农作物; 光谱特征

中图分类号: TP79

文献标识码: A

文章编号: 2096-5877(2019)03-0045-07

Extracting Crop Types from Hyperspectral Data

YU Chenglong

(*Innovation and Opening Laboratory of Regional Eco-Meteorology in Northeast, China Meteorological Administration/Meteorological Academician Workstation of Heilongjiang Province/Heilongjiang Province Institute of Meteorological Sciences, Harbin 150030, China*)

Abstract: Heilongjiang Province is the largest commodity grain production base in China. Its crop information is of great significance for grain estimating, disaster prevention and reduction, and food security. This study explored the method of the planting information extraction to base on the hyperspectral data to help get plant distribution faster and better. Multi-spectral data have some limitation in distinguishing the planting information of corn and soybean, so in this paper we chose hyperspectral data to study the method of extracting main dry crop types in Qiqihaer City of Heilongjiang Province. After data preprocessing and optimum band selection, the crop types of corn, soybean, rice and other crops were extracted using the decision tree method. The results showed that the classification accuracy of each crop was more than 80% and its overall accuracy reached up to 88.71%. The classification results showed that the main land cover type was cultivated land in Qiqihaer City, which accounted for 70% of the city. And the plant area of corn was the main land cover type.

Key words: Hyperspectral; Crop; Spectral features

黑龙江省是我国粮食主产区, 粮食总产居全国首位, 玉米和大豆是黑龙江省的主要旱作物, 种植面积占该省全部粮食作物的近70%。依赖于地面调查的耕地面积统计方法成本高、周期长, 遥感技术为提取农作物类型信息提供了一条新的途径, 然而由于黑龙江省玉米和大豆的生长特

点, 传统的多光谱数据对其分辨不够敏感。

高光谱数据以其光谱响应范围广, 光谱分辨率高的特点, 为在较窄的光谱波段范围内寻找不同植被的光谱差异, 实现不同类别的划分提供了新的数据源。Peng等^[1]基于高光谱数据, 利用最佳高光谱指数评价了退化温带植被叶片叶绿素含量; Tian等^[2]利用IKONOS (Multispectral satellite) 影像评估了遥感技术在监测水库中水生植物(大型植物和丝状藻类)的种类组成、面积范围和密度的有效性; Sankey等^[3]耦合高光谱数据和无人机激光雷达影像, 对干旱和半干旱植被稀疏的生态系统中植被种类进行监测; Abdullah等^[4]基于高光谱

收稿日期: 2018-12-18

基金项目: 2018年中国气象局东北地区生态气象创新开放实验室开放研究基金项目(stqx2018zd03); 黑龙江省科学基金面上项目(D2018006)

作者简介: 于成龙(1973-), 男, 高级工程师, 博士, 主要从事生态气象方向研究。

数据评估了科威特沙巴阿艾哈迈德自然保护区的植被变化。而对在全发育期内的光谱特征均很相似的黑龙江省的玉米和大豆的区分研究较少,本研究正是针对这一问题,利用高光谱遥感数据特点,以齐齐哈尔市为例,研究基于环境减灾星 HSI 高光谱数据提取包括玉米和大豆在内的主要农作物类型信息的方法,从而提供更为精细化的遥感分类产品,进而为防灾减灾、灾情评估以及作物估产等工作提供准确、可靠的基础数据。

1 研究区及数据

1.1 研究区概况

齐齐哈尔市位于黑龙江省中西部松嫩平原,东经 $122.39^{\circ} \sim 126.67^{\circ}$, 北纬 $46.21^{\circ} \sim 48.92^{\circ}$ 之间。全市土地面积约 4.2 万 km^2 , 该市是黑龙江省主要农区, 地势东北、西南高, 中间低, 主要由台地和平原组成, 平均海拔高度 146 m , 土壤类型主要有暗棕壤、黑土、黑钙土、草甸土等。该市属温带大陆性季风气候, 年降水量 $400 \sim 550 \text{ mm}$, 年平均气温 3.2°C , 年日照时数 $2\ 600 \sim 2\ 900 \text{ h}$ 。

1.2 数据来源

1.2.1 遥感数据

遥感资料来源于中国资源卫星应用中心 (<http://www.cresda.com/n16/index.html>), 环境减灾卫星 A 星 (HJ-1A 星) 的高光谱成像仪的光谱波段达到 115 个, 光谱分辨率达到 4 nm , 本研究应用 2013 年 6 月 29 日、7 月 13 日和 7 月 14 日无云污染的 HSI 数据和 CCD 相机数据进行分类。

与 E0-1Hyperion (光谱范围 $356 \sim 2\ 577 \text{ nm}$, 重返周期 16 d) 数据相比, HSI 数据 (光谱范围 $450 \sim 950 \text{ nm}$, 重返周期 4 d) 虽然光谱范围较窄, 但重返周期短, 既可以满足短时间的监测需求, 也增加了获取可用数据的概率, 而对于植被分类研究来说, 波长范围在 $400 \sim 1\ 500 \text{ nm}$ 之间即可以满足需要, 因此 HSI 数据应用于本研究中具有独特优势。

1.2.2 地面调查数据

地面控制点数据用于对卫星数据的地理校正, 获取方法是结合 Google earth 软件, 确定研究区主要标志物大致位置, 然后用手持 GPS 进行地面实际调查, 获取实际经纬度, 标志物选取以主要高速公路拐角处为主, 完成有效定位点 21 个。

了解地物光谱特征是分类的基础, 根据齐齐哈尔市土地利用特点和分类需要, 本研究借助 GPS, 实地调查自然植被、建筑用地和裸地、水田、玉米、大豆、其它作物 6 种土地覆盖类型, 得到有

效地面调查样点 635 个; 其中水体的样点是根据 2015 年 HJ-1A 星的 CCD 相机数据, 通过目视解译选取的, 共 52 个样点, 各样点分布如图 1。

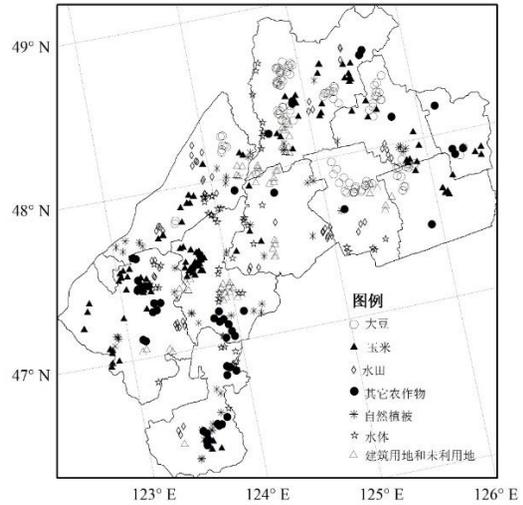


图 1 样点分布示意图

2 遥感数据预处理

2.1 图像旋转和去除条带噪声

由于卫星成像探元阵列的不一致和放大电路的频率响应等因素, 导致 HIS 数据产生条纹噪声, 较常用去除条带噪声的方法有空域滤波、频域滤波、根据图像自身特征的校正方法等^[5-7], 对于已经过系统几何纠正的 HSI 数据 2 级产品, 图像上的条带噪声不是呈垂直分布, 本研究首选采用钮立明等^[8]的方法进行图像旋转: 确定图像两个顶点坐标 $A(x_1, y_1)$ 和 $B(x_2, y_2)$ (如图 2), 求得两点所在直线与法线夹角的正切值 $\tan\alpha$, 进而得到图像旋转的角度 α , 将各波段数据旋转角度 α 后, 图像上的条纹基本上均沿垂直方向分布。然后再采用

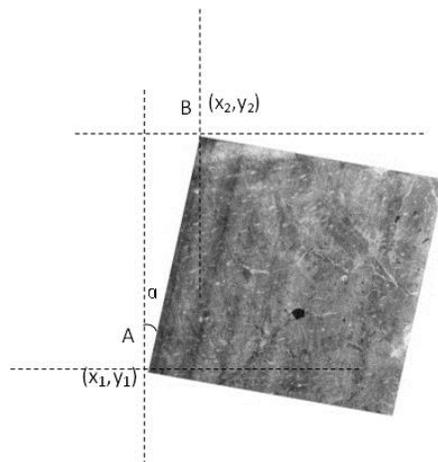


图 2 图像旋转角度计算示意图

“全局去条纹法”^[8-10]去除条带噪声。

图3为第30波段去除条带噪声处理结果,可

见去条带前后两幅图差别明显,原有的倾斜条带噪声在去条带处理后已不明显。

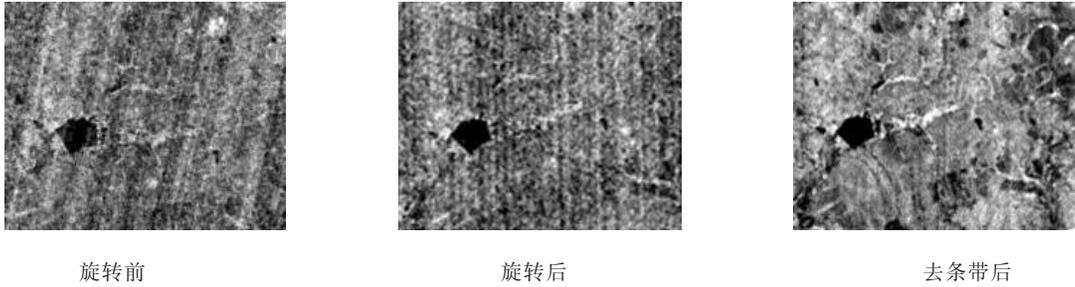


图3 条带噪声去除前后对比图

为了解上述方法对 HIS 数据条带噪声去除的效果,采用郭兴杰等^[11]的评价方法,通过最大值、最小值、灰度均值、标准差、信息熵、平均梯度指

标(公式见表1,表中: $\Delta x f(i, j), \Delta y f(i, j)$ 分别为像素 (i, j) 点在 x, y 方向上的一阶差分)来评价条带去除效果。

表1 评价指标公式

灰度均值	标准差	信息熵	平均梯度
$\bar{g} = \frac{1}{mn} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} F(x, y)$	$std = \sqrt{\frac{1}{mn} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} (F(x, y) - \bar{F}(x, y))^2}$	$H(x) = -\sum_{i=0}^{255} p(i) \log_2(p(i))$	$G = \frac{1}{mn} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} \sqrt{\Delta x f(i, j)^2 + \Delta y f(i, j)^2}$

表2为 HSI 数据第30波段去条带噪声前后各评价指标的对比结果,可见经过去条带处理后,最大值比去条带前有所减小,最小值未改变,灰度均值增加明显,标准差下降明显,信息熵有所

增加,平均梯度大幅度增加,说明经处理后图像的地物平均反射强度、影像信息的丰富程度有所增加,图像更加清晰。

表2 HSI 数据去条带噪声结果评价指标

	最大值	最小值	灰度均值	标准差	信息熵	平均梯度
去条带前	10 326	43	4 563.79	2 871.10	5.05	10.24
去条带后	10 318	43	4 799.63	2 288.80	5.11	158.46

2.2 大气校正

以水体和植被为例,各选取100个点,共400个点,提取各点的波谱曲线后分类平均,得到大气校正前后地物辐射亮度光谱曲线(图4),可见大气校正后水体和植被的辐射亮度光谱曲线较校正前发

生明显变化,大气校正前水体在波长460 nm(蓝光波段)~950 nm(近红外波段)范围内的辐射亮度从4 500 W/(m²·μm·sr)波动下降到350 W/(m²·μm·sr),而校正后水体基本在-420 W/(m²·μm·sr)~560 W/(m²·μm·sr)之间波动,这更符合水体在不同波

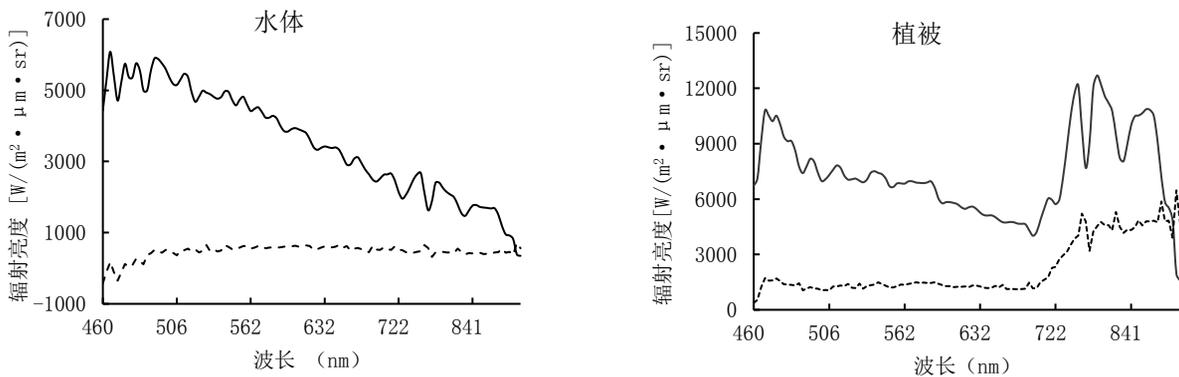


图4 HSI 数据大气校正前后地物辐射亮度光谱曲线

段范围内的辐射亮度变化规律;同样,大气校正前的植被辐射亮度在蓝光波段处于很高水平($>10\ 000\ \text{W}/(\text{m}^2 \cdot \mu\text{m} \cdot \text{sr})$),之后快速下降,到红光波段已下降到 $5\ 000\ \text{W}/(\text{m}^2 \cdot \mu\text{m} \cdot \text{sr})$ 以下,这段的光谱曲线不符

合植被光谱变化的正常规律,而大气校正后的蓝光到红光波段的辐射亮度值在 $1\ 100 \sim 1\ 300\ \text{W}/(\text{m}^2 \cdot \mu\text{m} \cdot \text{sr})$ 之间波动,到近红外波段才有明显上升,这种变化是符合正常植被光谱变化规律的。

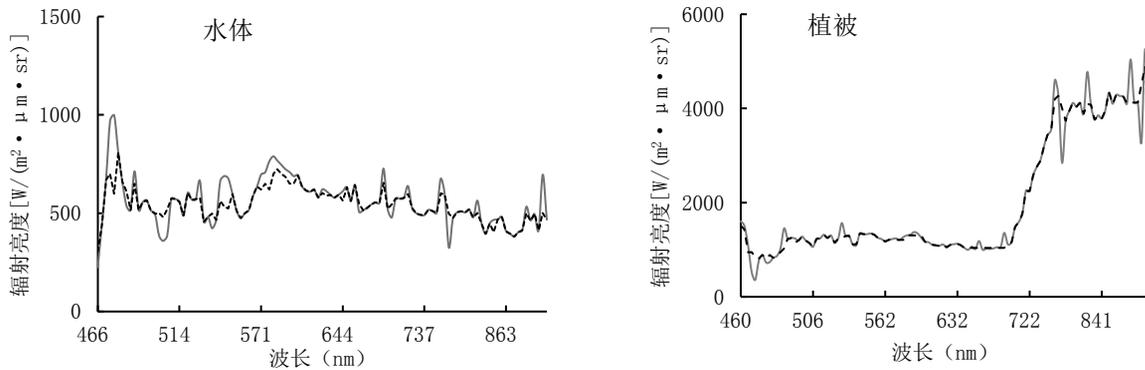


图5 HSI数据光谱优化前后地辐射亮度光谱曲线

由于数据在经过辐射定标、大气校正等处理过程中像元光谱产生连贯的锯齿噪音,因此需要对大气校正后的图像进行处理,本研究参考Datt等^[10]的处理方法,用MNF转换方法去除图像中像元波谱曲线上的锯齿,得到光谱优化图像,由图5可见,优化后曲线的抖动幅度明显降低。

另外,本研究选用在时间上与HSI数据最邻近的TM影像,选取40个左右的几何精校正控制点,采用图像对图像的方式进行几何精校正,校正误差控制在1个像元内。几何精校正后采用最邻近像元法对数据再进行空间分辨率为100 m的重采样处理。

2.3 数据融合

由于HSI数据的空间分辨率为100 m,分辨率较低,而与之同时搭载在HJ-1A星的还有CCD相机,该数据的空间分辨率为30 m,由于HJ-1A的CCD数据和HSI数据为同一时刻获取的数据,除

去成像信号形影方式不同等的影响外,两个光学系统定标后的数据应该基本一致,因此可以通过HSI数据与CCD数据融合,从而提高HSI数据的空间分辨率。

从图6可以看出,融合后的HSI图像无论从地物的纹理,地物边缘的清晰度、不同地物在亮度的对比上都有很大提升,融合后的HSI图像从空间分辨率上更接近于CCD数据。在融合前、后HSI图像和CCD图像中,分别从水体和植被各取100个点,共600个点,提取各点的波谱曲线后分类平均,得到分别不同图像和地类的平均波谱曲线(图7),从中可见且HSI具有更为精细的波谱特征;对比数据融合前后的HSI数据的波谱曲线可知,融合后的HSI数据比融合前的HSI数据的辐射亮度要高,二者在波峰和波谷处具有较高的一致性,且相关系数较高,水体的相关系数为0.68,植被的相关系数为0.87。



HSI融合前 第108波段

HSI融合后 第108波段

CCD数据 第4波段

图6 HSI与CCD数据融合前后图像对比

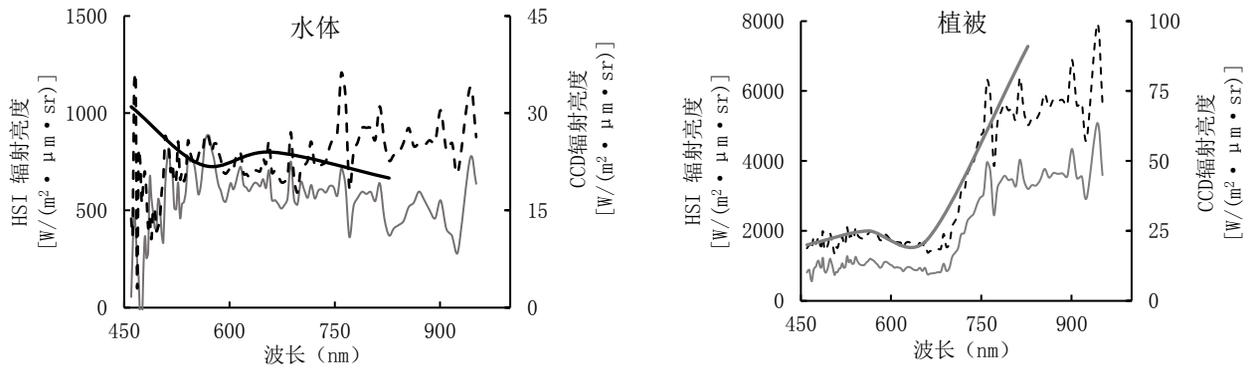


图7 HSI与CCD数据融合前后波普曲线特征对比

2.4 最佳波段选取

相对于多光谱数据来说,高光谱数据对地物光谱信息的提取能力增强的同时,也存在部分信噪比低和相关性高的波段,以及其它大量冗余的信息,增加了数据后续处理的计算量和计算的复杂性,因此需要针对信噪比低和相关性高的波段进行数据降维处理,从而提高数据的可利用性^[12-14],本研究利用ENVI和MATLAB软件,采用Chavez等^[15]在1982年提出的最佳指数因子法(Optimum Index Factor,简称OIF)的OIF值来判断最佳波段组合,OIF越大相应组合影像包含的信息量就越大。公式如下:

$$OIF = \frac{\sum_{i=1}^3 S_i}{\sum_{i=1}^3 |R_{ij}|}$$

式中, S_i 为第*i*个波段的标准差, R_{ij} 为*i*、*j*两波段的相关系数。

由于前20波段噪音过大,因此参加计算的HSI数据为21~115波段,计算每3波段组合的最佳指数,共有19656个结果,表3列出前5个最佳指数及对应波段组合,其中OIF最大的波段组合为40、78、108,这3个波段分别为绿光波段、红光波段和近红外波段,对植被类型的识别、植被覆盖状况和水体的识别具有很高的敏感性。

表3 HSI数据3波段组合的最佳指数

最佳指数	波段		
328.1285	40	78	108
327.9346	42	80	98
327.3574	50	98	105
326.5746	33	81	102
326.0254	54	76	106

3 分类实践与精度评价

3.1 分类结果

根据齐齐哈尔市农作物发育期和自然植被物候特点,本研究选取2015年5月中旬(大豆处于播种期、玉米处于出苗期、水田正在泡田、自然植被已经返青)、6月下旬(大豆处于三叶期、玉米处于七叶期、水田处于分蘖期、自然植被植被指数明显提高)和8月上旬(大豆、玉米、水田和自然植被植被指数明显高于建筑用地、水体和未利用地)的Landsat 8 OLI数据和HJ-1A/B HIS数据,把融合后的HIS数据第78波段和第108波段计算得到的NDVI、第40波段、78波段、108波段和ISODATA法分类的初始结果进行层堆栈,在地面调查样点中抽取包含各土地覆盖类型的200个点,确定不同土地覆盖类型的光谱特征,确定感兴趣区域,再

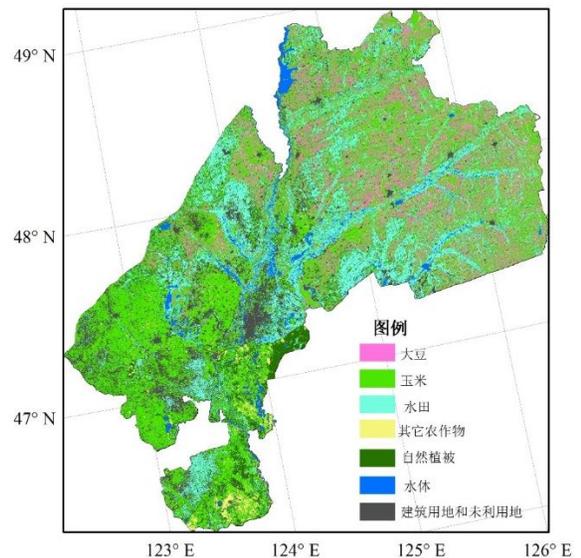


图8 基于决策树分类方法的分类结果

利用分类回归树 CART(Classification and Regression Tree)^[16]算法构建决策树进行分类,分类结果如图8所示。

表4 基于决策树分类方法的各土地覆盖类型面积统计

类别	面积(hm ²)	比例
其它作物	11.69	2.77%
大豆	42.69	10.12%
建筑用地及裸地	38.14	9.04%
水体	13.64	3.23%
水田	67.22	15.93%
玉米	186.88	44.29%
自然植被	61.66	14.61%

综合分析图8和各土地覆盖类型面积统计结果(表4)可见,齐齐哈尔市的土地覆盖类型以农耕地(包括大豆、玉米、水稻和其它作物)为主,总面积为308.48万hm²,占该市土地面积的73.11%,其中大豆主要分布在讷河、克山、甘南、依安、富裕、拜泉和齐齐哈尔市辖区北部,种植面积为42.69万hm²,占全市土地面积的10.12%;玉米分布于全市,是该市最主要的土地利用类型,面积

为186.88万hm²,占全市土地面积的44.29%;水田主要分布于河流、湖泊等水体附近,面积为67.22万hm²,占全市土地面积的15.93%;其它作物零星分布于全市,面积为11.69万hm²,占全市土地面积的2.77%。

3.2 精度验证

剩下的487个地面样点用于分类精度评价(具体结果如表5所示),总体分类精度为88.71%,Kappa系数为0.865。其中水体的分类精度最高,取的52个地面验证点只有1个点错分为水田。自然植被的分类精度最低,50个地面验证点中,有1个点错分为玉米,5个点错分为水体,3个错分为水田,原因是自然植被所选取的地面验证点多为耕地边的农田防护林和沼泽,由于所用数据的空间分辨率影响,出现了明显的错分现象。大豆的分类精度达84.26%,取了108个验证点,错分的17个点中,有15个点错分为玉米,而玉米的112个验证点中,有10个点错分为大豆,说明大豆和玉米之间还存在一定的同谱异物的现象。

表5 基于决策树分类混淆矩阵

类别	大豆	玉米	水田	其它作物	水体	自然植被	建筑用地及裸地	总计
大豆	91	10	1	0	0	0	0	102
玉米	16	99	0	4	0	1	0	120
水田	0	3	55	1	1	3	4	67
其它作物	0	0	0	45	0	0	0	45
水体	0	0	1	1	51	5	0	58
自然植被	0	0	0	1	0	40	0	41
建筑用地及裸地	1	0	0	1	0	1	51	54
总计	108	112	57	53	52	50	55	487
分类精度(%)	84.26	88.39	96.49	84.91	98.08	80.00	92.73	

4 结论与讨论

本研究利用高光谱遥感数据特点,结合同期多光谱数据、其他时期土地利用数据、地面调查数据等相关资料,以齐齐哈尔市为例,研究黑龙江省主要农作物类型信息提取方法,为防灾减灾、灾情评估以及作物估产等工作提供准确、可靠的基础数据,同时也可为更大区域农作物类型信息提取提供方法参考。

通过对HSI数据进行图像旋转、去除条带噪声、大气校正、光谱锯齿去除、地理精校正、与同期CCD数据融合等预处理,得到图像的地物平均

反射强度、影像信息的丰富程度均有所增加,图像更加清晰。应用决策树分类法,各作物分类精度均超过80%,总体精度达到88.71%。分类结果表明,齐齐哈尔市的土地覆盖类型以农耕地(包括大豆、玉米、水稻和其它作物)为主,占该市土地面积的70%以上,其中玉米分布于全市,是该市最主要的土地利用类型;大豆主要分布于讷河、克山、甘南、依安、富裕、拜泉和齐齐哈尔市辖区北部;水田主要分布于河流、湖泊等水体附近;其它作物零星分布于全市。

由此可见,环境减灾星的HIS数据和CCD数据的融合,提高了HIS数据的空间分辨率,挖掘了

2种数据的应用潜力,证明了HIS数据在农作物种植信息提取中的业务支撑能力,但本研究所用的最佳波段选择方法都是针对整幅影像来考虑,没有考虑针对特定目标的最佳波段选择,因此分类精度会受到一定影响。另外遥感数据空间分辨率不足(HIS数据100 m, CCD数据30 m)也是限制分类精度的原因之一。

参考文献:

- [1] Peng Y, Fan M, Wang Q H, et al. Best hyperspectral indices for assessing leaf chlorophyll content in a degraded temperate vegetation [J]. *Ecology and Evolution*, 2018, 8(14): 7068-7078.
- [2] Tian Y Q, Yu Q, Zimmerman M J, et al. Differentiating aquatic plant communities in a eutrophic river using hyperspectral and multispectral remote sensing [J]. *Freshwater Biology*, 2010, 55(8): 1658-1673.
- [3] Sankey T T, Mcvay J, Swetnam T L, et al. UAV hyperspectral and lidar data and their fusion for arid and semi-arid land vegetation monitoring [J]. *Remote Sensing in Ecology and Conservation*, 2018, 4(1): 20-33.
- [4] Abdullah M M, Addae-Wireko L, Tena-Gonzalez G A. Assessing native desert vegetation recovery in a war-affected area using multispectral and hyperspectral imagery: a case study of the Sabah Al-Ahmad Nature Reserve, Kuwait [J]. *Restoration Ecology*, 2017, 25(6): 982-993.
- [5] 赵好好,肖鹏峰,冯学智.基于频谱分析的高分辨率遥感图像边缘特征检测研究[J]. *遥感信息*, 2013, 28(6): 96-115.
- [6] 王 显,张广友,李新涛,等.卫星遥感影像预处理中噪声去除方法的研究[J]. *遥感技术与应用*, 2007, 22(3): 455-459.
- [7] 谭炳香,李增元,陈尔学,等. EO-1Hyperion 高光谱数据的预处理[J]. *遥感信息*, 2005(6): 36-41.
- [8] 钮立明,蒙继华,吴炳方,等. HJ-1A星 HSI数据2级产品处理流程研究[J]. *国土资源遥感*, 2011(1): 77-82.
- [9] Buzzi J, Garcia-Melendez E, Vazquez I. Pyrite mine waste and water mapping using Hymap and Hyperion hyperspectral data [J]. *Environmental earth sciences*, 2012, 66(7): 1957-1971.
- [10] Datt B, McVicar T R, VanNiel T G, et al. Pre-processing EO-1 Hyperion Hyperspectral Data to Support the Application of Agricultural Indices [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2003, 41(6): 1246-1259.
- [11] 郭兴杰,王阳春,汪爱华,等. HJ-1A高光谱数据的条带噪声去除方法研究[J]. *遥感信息*, 2011, 26(1): 54-58.
- [12] 罗 琴,田 铮.高光谱图像无监督分类的非线性特征提取器[J]. *宇航学报*, 2007, 28(5): 1273-1277.
- [13] 尹继豪,王义松.高光谱遥感影像中最佳谱段的快速选择方法[J]. *遥感信息*, 2010, 25(3): 3-7.
- [14] Keshava N. Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2004, 42(7): 1552-1565.
- [15] Chavez P S J, Berlin G L, Sowers L B. Statistical method for selecting landsat MSS ratios[J]. *Journal of Applied Photographic Engineering*, 1982(8): 23-30.
- [16] 刘勇洪,牛 铮,王长耀.基于MODIS数据的决策树分类方法研究与应用[J]. *遥感学报*, 2005, 9(4): 405-412.