中国常电信息

第31卷第5期2019年10月

China Agricultural Informatics

2019, 31 (5): 25-37

Vol.31, No.5 Oct., 2019

农业遥感

基于Rapideye数据的棉花特征光谱指数构建及类型识别*

王利民,刘 佳*,姚保民,高建孟,季富华(中国农业科学院农业资源与农业区划研究所,北京100081)

摘要:【目的】特征光谱指数构建是提高农作物类型识别效率的前提,也是构建通用型识别方法的基础。【方法】基于 2017 年 8 月 8 日河北省冀州市 5 m 空间分辨率的 Rapideye 影像,采用红边、近红外波段反射率之和构建了棉花提取指数(Cotton Extraction Index,CEI),结合同期水体、裸地(含城镇建筑)掩模处理,实现了棉花类型识别。CEI 指数方法具体过程是: 样本点选择、水体及裸地(含城镇建筑)掩模处理、CEI 指数构建、棉花识别及精度验证。其中,样本点采用随机方法获取,共获取了 5 144 个样本点,作为最佳 CEI 阈值以及精度验证的依据。【结果】基于 CEI 方法提取结果的总体精度达到 88.80%,Kappa 系数达到 0.751 7。为了评价 CEI 指数方法的相对精度,分别采用最大似然分类方法和随机森林分类方法对影像进行分类和精度验证。结果表明基于最大似然分类方法的棉花提取总体精度和 Kappa 系数分别为 86.53% 和 0.698 3,采用随机森林方法的棉花提取总体精度和 Kappa 系数分别为 90.12% 和 0.766 7。【结论】对比 3 类方法,基于 CEI 指数方法获得的棉花提取结果,与其他 2 种常用方法的提取精度相当。该方法具有简单、高效,对样本依赖性较小的优势,在时效性要求较高的农作物面积识别中具有较高的应用价值。

关键词:棉花;遥感识别; Rapideye;特征光谱指数;随机森林;最大似然分类 DOI: 10.12105/j.issn.1672-0423.20190503

0 引言

棉花是中国重要的经济作物之一,黄淮海流域、长江流域、新疆是我国三大产棉区^[1-2],近年来棉花种植面积有下滑趋势^[3],稳定棉花种植面积是我国农业生产的重要目标。快速、准确地监测中国棉花主产区的棉花种植情况,可为农业政策的制定提供可靠的依据,有利于维护农民的经济利益。遥感技术具有监测面积大、客观性强的特点,特别是近年来国内外中高分辨率卫星数据的不断增多,如 Landsat-8^[4]、Sentinel-2^[5]、Worldview 系列卫星^[6]、Rapideye 系列卫星^[7-8],以及国内的资源系列卫星^[9]、环境系列卫星^[10]和高分系列卫星^[11]等,利用遥感卫星技术正逐步成为农作物种植面积获取的

收稿日期: 2019-09-16

第一作者简介:王利民(1968—),博士、研究员。研究方向:农业遥感监测业务运行。Email: wanglimin01@caas.cn ※ 通信作者简介:刘佳(1968—),硕士、研究员。研究方向:农业遥感监测业务运行。Email: liujia06@caas.cn

^{*}基金项目:中国农业科学院创新工程领军人才 C 类(农业遥感)

重要途径[12-16]。

农作物种植面积遥感监测技术主要包括监督分类、非监督分类、决策树分类^[17]、面向对象分类^[18]等,其中监督分类方法又包括最大似然分类^[19]、支持向量机^[20]、随机森林^[21]等,非监督分类又包括 ISOData^[22]、Kmeans^[23]方法等。这些技术方法在棉花遥感识别中都有不同程度的应用^[24-28],并获得了较好的精度。如 Jia 等^[24]利用 HJ-1A/B CCD 影像采用支持向量机的方法比较了单时相和多时相不同组合情况下提取小麦、棉花及其他几种地物的面积精度,在多时相组合情况下总体精度优于 91%; Li 等^[29]利用 CBERS01 和 HJ1B 影像通过决策树分类模型提取山东省广饶县丁庄镇棉花信息,棉花面积的提取精度在 90% 以上。

为提高棉花识别精度,众多学者也开展了综合棉花遥感波段特征分类方法的研究。如曹卫彬等^[30]对不同生育时期地物光谱曲线特征进行地面测量,并在 TM 影像光谱特征分析基础上,提出了棉花识别方法,棉花识别精度达到 96.0%。为进一步简化识别步骤,便于监测业务的开展,加权指数的方法在农作物识别中也有应用。如王利民等^[11]提出了一种冬小麦 NDVI 加权指数的分类方法,利用不同时期冬小麦与其他地物 NDVI 的差异,通过加权指数方式增强这一差异,从而实现河北安平冬小麦的快速识别,识别总体精度达 94.4%; Song 等^[31]基于 AMSR-E 数据,提出了一种水稻种植指数,基于该指数对全国水稻的种植面积进行了提取,将提取结果与统计数据进行对比,两者相关系数达到 0.94。

从已有研究报道来看,针对棉花遥感识别的光谱特征指数的研究尚不多见。而根据以往的研究,棉花波谱和同期播种的作物之间存在较大差异^[30,32],因此可以考虑通过构建特征指数增强这种差异,进而实现棉花快速、准确识别和提取。已有研究的棉花波段主要集中在常见的红、绿、蓝、近红外波段,对红边棉花识别能力的研究相对较少,而红边波段已证明在部分作物(玉米、大豆、水稻等)分类中具有重要的作用^[7,33]。针对以上问题,文章以包含红边波段的 Rapideye 卫星影像为数据源,以棉花为识别对象,通过地面样本数据的收集分析,结合红边波段构建相应的特征光谱植被指数,建立相应的处理技术流程,实现县级尺度棉花的高精度提取,为区域作物的监测提供了一个较为可靠的方法。

1 研究区概况

研究区包括河北省冀州市全部,以及周边的深州、辛集、宁晋等市辖区等部分区域。 冀州市属河北省衡水市,地处华北平原腹地,位于115°11′~115°41′E,37°19′~37°44′W, 研究区总面积约为918 km²。

研究区位于华北平原北部,气候类型属于典型的温带大陆性季风气候,四季分明、雨热同期。全年平均气温 12℃左右,最冷月份通常为 1 月,平均气温 -4℃左右,最热月份通常为 7 月,平均气温 27℃左右。全年降水量约为 500 mm,且降水主要集中在夏季。研究区光热资源充足,年光照时数约为 2 600 h,无霜期 190 d,研究区地势平坦,土壤以壤土为主^[34]。

2019年10月-

研究区气候和土壤条件尤其适宜农作物生长,研究区内广泛种植冬小麦、玉米等粮食作物,以及棉花等经济作物,棉花的种植面积约占研究区耕地面积的一半,逐渐成为国家重点棉花生产区域。研究区区位如图 1 所示。

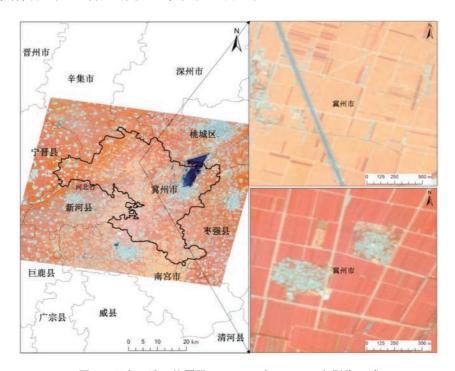


图 1 研究区地理位置及 Rapideye(B5/B4/B3)影像区域 Fig.1 Location and Rapideye data of study area

2 试验数据获取与处理

2.1 遥感数据获取及预处理

研究数据采用 2017 年 8 月 8 日的 Rapideye 影像,该数据空间分辨率为 5 m,包括蓝(440~510 nm)、绿(520~590 nm)、红(630~685 nm)、红边(690~730 nm)、近红外(760~850 nm) 5 个波段,影像数据如图 1 所示。对获取的 Rapideye 数据进行辐射定标、大气校正和几何精校正等预处理工作。

辐射定标采用的线性转换公式、定标斜率、定标截距都按照卫星方提供的公式与参数计算。大气校正采用 ENVI/FLAASH 大气校正模块进行,通过 RESA 网站^[35] 获取 Rapideye 卫星传感器的光谱响应函数,制作成波谱库文件,与卫星观测几何及气溶胶模式、光学厚度等参数,以及辐射定标后的影像一并输入 FLAASH 模型,模型输出结果就是消除大气影响的反射率影像。几何校正则在无控制点条件下采用影像自带的有理多项式系数(Rational polynomial coefficients,RPC)参数进行,同时与研究区的本底遥感影像数据进行几何精配准,使其定位精度达到亚像素级,满足遥感影像分类定位精度要求。经大气校正和几何校正后的研究区域影像如图 2 所示。

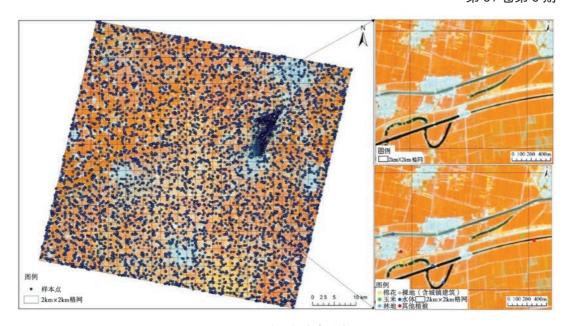


图 2 研究区样本点分布 Fig.2 The sample distribution

2.2 样本点数据获取

样本点数据的作用有 2 个,即作为训练样本及精度验证样本使用。为保证样本分布的代表性,采用随机方式获取样本点。同时将研究区划分为 313 个 2 km×2 km 格网,并将研究区地物类型划分为棉花、玉米、林地(含果园)、裸地(含城镇建筑)、水体和其他植被等 6 种,其中其他植被地物类型主要以植被为主,包括撂荒地以及其他作物等。采用目视解译方式逐个确定样本点地物类型。共选择了 5 144 个样本点,其中,棉花、玉米、林地(含果园)、裸地(含城镇建筑)、水体和其他植被 6 种地物的样本点数量分别为 1 343、617、776、1 208、561、639 个,占总样本数比例分别为 26.11%、11.99%、15.09%、23.48%、10.91%、12.42%。样本点分布如图 2 所示。

2.3 训练及精度验证样本

将 5 144 个点划分为棉花及非棉花 2 类, 分别有 1 343 和 3 801 个点, 各抽取 1/3 作为训练样本, 其他 2/3 作为验证样本,则共有 1 454 个训练样本, 3 690 个验证样本。精度验证数据空间分布如图 3 所示。

3 研究方法

3.1 技术思路

地物特征都会占有特定的光谱空间,构建主要光谱特征的综合特征,利用形式较为简单的指数形式,可以实现地物特征识别的目的。基于这一思想,针对研究区的6种地物类型,分析其光谱可分性,构建棉花识别指数,实现棉花类型的识别。

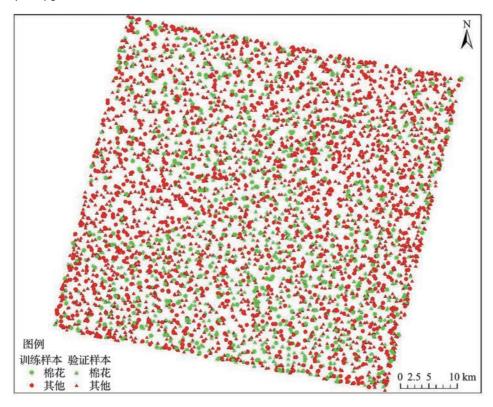


图 3 研究区训练样本集验证样本分布 Fig.3 Training sample and validation sample distribution

主要技术方案是,根据 6 种地物的光谱特征,对 6 种地物类型的分离程度定性排序,去除水体、裸地(含城镇建筑)2 种与棉花不易混淆的地物类型,对剩余的林地、玉米、其他植被和棉花 4 种地物类型构建棉花提取指数(Cotton Extraction Index, CEI),要求棉花识别指数的阈值范围要占据 4 种地物范围的高值区域。

将棉花识别指数由低到高排序,生成均匀分布于棉花指数范围内的系列阈值,并获得每个阈值下分类结果的混淆矩阵,取总体精度最高值作为识别阈值,利用该阈值对全部影像分类,并采用目视解译获得的本底数据进行精度验证。

为验证该算法的可靠性,将该算法的结果与最大似然、随机森林等 2 种方法的识别结果进行比较评价。

3.2 地物光谱特征分析

基于 5 144 个样本点, 计算 6 种地物不同波段反射率(为了便于说明和计算, 反射率值统一扩大 1 万倍, 下同)均值, 不同地物影像反射率值变化规律如图 4 所示。可以看出, 研究区主要地物光谱反射特征可以归纳为水体、植被、裸地(含城镇建筑)等 3 类低、中、高反射率地物。植被(棉花、玉米、林地及其他植被)的光谱特征显著, 其可见光波段的蓝光(B1)、绿光(B2)、红光(B3)波段的反射率均较低, 而红边(B4)和近红外(B5)则急速升高,表明 B4+B5 可以作为反映植被特征的指数; 水体的各波段反射率则普遍较低; 裸地(含城镇建筑)的各波段反射率普遍较高。

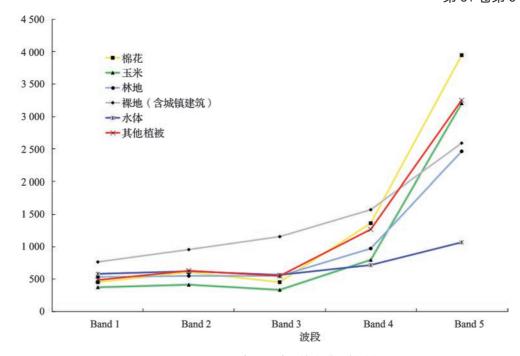


图 4 Rapideye 影像不同地物的光谱反射特征

Fig.4 Rapideye spectrum reflectance character of different ground objects

根据裸地(含城镇建筑)B1、B2 和B3等3个波段反射率值均较大,而植被和水体在B1、B2和B3波段反射率较小的特点,利用B1+B2+B3值大小可以对裸地(含城镇建筑)进行有效剔除。裸地(含城镇建筑)B1+B2+B3值大于3000,而植被和水体B1+B2+B3值不足2000,因此使用B1+B2+B3大于2500作为判别阈值剔除裸地(含城镇建筑)。

根据水体 5 个波段反射率值均较小的特点,利用 B1+B2+B3+B4+B5 值大小可以对水体像元进行有效剔除。由于仅有水体像元 B1+B2+B3+B4+B5 反射率值一般都在 3 000 以下,因此使用 B1+B2+B3+B4+B5 小于 3 000 作为判别阈值剔除水体。

3.3 棉花识别指数的构建

基于目视解译的研究区本底数据,绘制 6 种地物不同波段反射率组合下像元个数分布图,如图 5 所示。图 5a 为 B1+B2+B3+B4+B5 波段反射率组合下不同地物像元个数分布图,水体像元集中在 B1+B2+B3+B4+B5 波段组合的低反射率区,根据样本点确定的 3 000 阈值可以对水体进行有效剔除且不影响其他地物像元。图 5b 为 B1+B2+B3 波段反射率组合下不同地物像元个数分布图,仅有裸地(含城镇建筑)像元集中在 B1+B2+B3 波段组合的高反射率区,根据样本点确定的 2 500 阈值可以对多数裸地(含城镇建筑)像元进行剔除且不影响其他地物像元。由图 5c 可以看出 B4+B5 波段可以较好地反映植被特征,棉花、玉米、林地及其他植被均处于较大反射率值区域,预期可以用于棉花的识别。

剔除水体和裸地(含城镇建筑)像元后 B4+B5 波段组合下 4 种地物类型的像元个数分布如图 5d 所示。从上述分析结果可知,棉花的红边及近红外波段显著高于其他植被,在传统的考虑近红外波段基础上,叠加红边波段,可以进一步提高 B4+B5 数值,从而使得棉花与其他植被的区分更加显著,基于 B4+B5 值即可对棉花进行识别和提取。因此,将 B4+B5

2019年10月

定义即为棉花提取指数 CEI 值, 即:

$$CEI = Ref_{RE} + Ref_{NIR} \tag{1}$$

式(1)中, Ref_{RE} 表示红边波段反射率, Ref_{NIR} 表示近红外波段反射率。

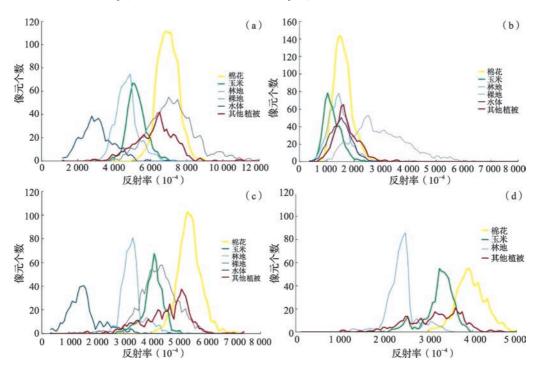


图 5 不同波段反射率组合下像元个数分布: (a)b1+b2+b3+b4+b5; (b)b1+b2+b3; (c)b4+b5; (d)b4+b5(去除水体及裸地/城镇建筑)

Fig.5 Pixel number distribution under different band reflectance combination

3.4 最大似然分类方法

最大似然法也称为最大概似估计或极大似然估计,是一种具有理论性的点估计法。最大似然分类假定每个波段每一类地物反射率统计呈正态分布,并计算给定像元属于特定类别的可能性。除非选择一个可能性阈值,所有像元都将参与分类。每一个像元被归到可能性最大的那一类里。该方法是作物遥感识别分类经常使用的分类算法,具有分类结果稳定可靠、精度较高的优点。具体算法可参考相关文献^[36]。

3.5 随机森林分类方法

随机森林分类(Random Forest Classification, RFC)是 Breiman 于 2001 年提出来的一种较新的多决策树分类方法^[21],该方法通过在数据上及特征变量上的随机重采样,构建多个 CART 类型决策树(不剪枝),通过多决策树投票的方式确定数据的类别归属。随机森林方法对于遥感影像分类具有很好的抗噪声性能,分类精度较高。该法利用样方数据自动构建分类决策树,属于监督分类的一种。相关定义及详细表述可参照相关文献^[37]。

3.6 精度验证

精度验证方法采用混淆矩阵的方法进行,包括总体精度、制图精度、用户精度、 Kappa 系数等 4 个精度衡量指标,相关定义及详细表述可参照相关文献^[38-40]。

4 结果与分析

4.1 基于样本点数据的分类阈值确定

按照 3.2 中所述方法去除水体和裸地(含城镇建筑)后,获得 CEI 指数影像。记录 5 144 个样本点对应的地物类型,分为"棉花"和"其他"两类,非棉花像元全部记为"其他"类别。经统计样本点的 CEI 值域范围为 0~6 800。按照下式得到一系列阈值。

$$T_i = 50 \times i \tag{2}$$

式 (2) 中, T_i 为第 i 个阈值,i 取值为 1~136 的自然数,为该范围内等间距分段的段数。

以系列阈值中的每一项作为分类阈值依次对 CEI 影像进行分类,高于阈值的判断为棉花,低于阈值的判断为其他,采用 5 144 个样本点中的 1 454 个样本点作为训练样本,3 690 个样本点作为验证样本,分别计算各阈值下棉花提取精度,从总体精度、Kappa 系数、棉花用户精度和棉花制图精度 4 个指标进行评价,精度评价指标随 CEI 指数的变化曲线如图 6 所示。

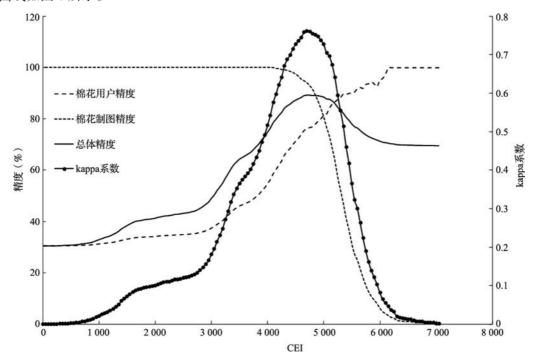


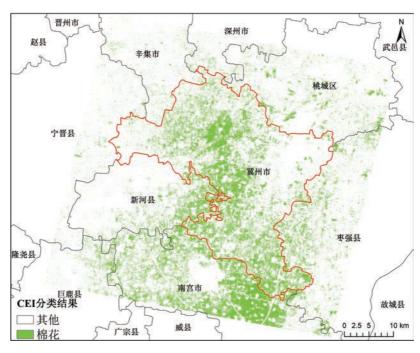
图 6 基于样本点数据的 CEI 指数法提取精度 Fig.6 Extraction accuracy of CEI method based on sample data

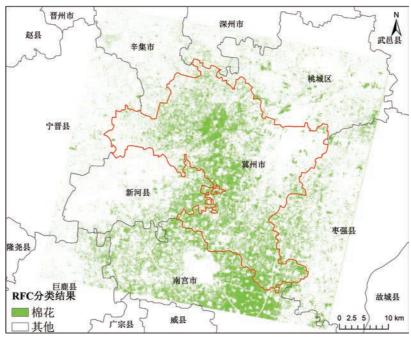
分类阈值较低时,棉花"错分"严重,无"漏分"现象,因此在 CEI 值较低时,棉花用户精度很低而制图精度很高。随着 CEI 值的增大,棉花"错分"现象减轻,逐渐出现"漏分"现象,因此,随着 CEI 值的增大,棉花用户精度逐渐增加而棉花制图精度逐渐降低。受棉花用户精度和制图精度变化规律差异的影响,总体精度和 Kappa 系数总体呈现先

增加后降低的趋势,且在 CEI 为 4 650 时,总体精度和 Kappa 系数同时达到了最大值,分别 91.3% 和 0.821 5。因此,基于样本点数据,确定 4 650 为棉花提取的最佳 CEI 阈值。

4.2 基于 CEI 指数的棉花类型识别

在去除水体、裸地(含城镇建筑)地物类型后,以 4 650 为分类阈值对整个研究区 CEI 的分布影像进行分类,高于阈值的判定为棉花,获得研究区的棉花分布结果,如图 7 所示,可以看出冀州市的棉花主要分布在中西部和南部等区域。





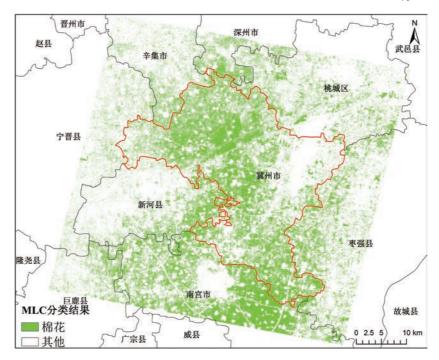


图 7 基于 CEI、RFC 和 MLC 3 种方法的棉花识别结果 Fig.7 Cotton extraction result based on CEI, RFC and MLC classification

利用目视解译判读 5 144 个样本点的地物类型对分类结果进行评价,结果表明,在 阈值为 4 650 时,总体精度达到 88.80%, kappa 系数达到 0.751 7。

4.3 与监督分类算法的比较

为了评价基于 CEI 指数方法的相对精度,采用同样样本点作为训练和精度验证样本,利用最大似然分类算法和随机森林分类算法对研究区域进行监督分类,并利用目视解译获得的本底数据对提取结果进行精度评价,从总体精度、Kappa 系数、棉花用户精度和棉花制图精度 4 个指标进行评价,基于 CEI 指数法、最大似然分类法和随机森林分类方法提取精度对比情况如表 1 所示。

表 1 基于 CEI 方法、最大似然分类方法和随机森林方法的棉花提取精度 Table 1 Cotton extraction accuracies of CEI, MLC and RFC methods

分类方法	总体精度(%)	Kappa 系数	制图精度(%)	用户精度(%)
CEI 法(阈值 4 650)	88.80	0.751 7	87.64	73.65
最大似然分类法	86.53	0.698 3	98.10	66.50
随机森林分类法	90.12	0.766 7	95.68	74.18

基于最大似然分类方法的棉花提取总体精度和 Kappa 分别为 86.53% 和 0.698 3。比较 CEI 方法和最大似然分类方法可以看出,基于 CEI 指数方法获得的棉花提取总体精度和 kappa 系数上均优于最大似然分类方法,总体精度提高了 2.27 个百分点,Kappa 系数提高了 7.65%。另外,CEI 指数方法提取的结果中,棉花的制图精度与用户精度差异相较

2019年10月-

最大似然分类方法更小,可以使得漏分和错分像元数量更为一致,减少遥感分类面积与实际面积值的差异。

基于随机森林分类方法的棉花提取总体精度和 Kappa 分别为 90.12% 和 0.766 7。比较 CEI 方法和随机森林分类方法可以看出,基于 CEI 指数方法获得的棉花提取总体精度与随机森林分类方法的提取精度相当,但 Kappa 系数较随机森林分类方法降低了 1.96%。随机森林方法提取的结果中,棉花的制图精度与用户精度均优于 CEI 指数方法,表明随机森林分类方法综合分类效果优于 CEI 指数分类方法和最大似然分类方法。

5 讨论与结论

该文利用棉花在红边波段与近红外波段反射率高的特点,构建棉花提取指数(CEI)。水体和裸地(含城镇建筑)掩膜后,棉花集中于CEI指数的高值区域。通过构建系列CEI阈值,并以样本点分类精度最大作为依据确定 4 650 为最佳阈值,实现了棉花像元的快速精确提取,对棉花提取结果进行精度验证,结果表明分类结果的总体精度达到88.80%,Kappa 系数达到 0.751 7。通过与最大似然分类方法和随机森林分类方法分类精度的对比发现,作为业务运行快速提取方案,该精度是可以接受的,表明基于CEI阈值方式进行棉花面积提取是可行的。

基于 CEI 指数的分类方法,分类结果优于最大似然分类方法,稍差于随机森林分类方法。随机森林分类方法分类精度虽然更高,但本质上是一个"黑箱"系统,用户输入训练样本并获得分类结果,分类过程中不同波段或指数信息对分类结果的权重信息不可知,不利于用户理解。在分类结果不理想时,用户很难有针对性地对样本进行筛选和优化。CEI 指数方法的最大优点在于物理意义明确、计算方便。CEI 指数即为提取棉花的特异性指数,用户易于理解。此外,用户可以根据实际情况适当调整阈值,获得棉花不同的"漏分"和"错分"效果。

CEI 指数原理简单,意义明确,计算方便,在区域棉花遥感识别中具有较好的应用潜力。该文仅仅通过单时相 CEI 指数,即可实现较高精度的棉花识别。在后期的研究中,可以进一步使用多时相的 CEI 指数,或结合其他类型的植被指数特征,从而实现更高精度的棉花识别甚至识别棉花的不同品种。同时,CEI 指数使用了红边波段,可以进一步推广应用到 Worldview、Sentinel-2A/B、GF-6 等具备红边波段的卫星影像中,有利于使用多源数据构建时序 CEI 指数,对于区域棉花的高精度识别具有较强的推广应用价值。

参考文献

- [1] 张建华, 祁力钧, 冀荣华, 等. 基于粗糙集和 BP 神经网络的棉花病害识别. 农业工程学报, 2012, 28 (7): 161~167.
- [2] 焦险峰, 杨邦杰, 裴志远. 全国棉花种植面积遥感监测抽样方法设计. 农业工程学报, 2002, 18 (4): 159~162.
- [3] 国家统计局农村社会经济调查司.中国农村统计年鉴.北京:中国统计出版社,2017.
- [4] Khan Ahmad, Hansen Matthew C., Potapov Peter, et al. Landsat-based wheat mapping in the heterogeneous cropping system of Punjab, Pakistan. *International Journal of Remote Sensing*, 2016, 37 (6): 1391~1410.
- [5] Sharma Ram, Hara Keitarou, Tateishi Ryutaro. High-resolution vegetation mapping in Japan by combining Sentinel-2 and Landsat 8 based multi-temporal datasets through machine learning and cross-validation approach. Land, 2017, 6 (3): 50.

· 36 · 王利民等:基于 Rapideye 数据的棉花特征光谱指数构建及类型识别

第31卷第5期

- [6] Gaertner Julie, Genovese Vanessa Brooks, Potter Christopher, et al. Vegetation classification of coffea on Hawaii Island using WorldView-2 satellite imagery. *Journal of Applied Remote Sensing*, 2017, 11 (4): 046005.
- [7] 刘佳, 王利民, 滕飞, 等. RapidEye 卫星红边波段对农作物面积提取精度的影响. 农业工程学报, 2016, 32 (13): 140~148.
- [8] Tigges Jan, Lakes Tobia, Hostert Patrick. Urban vegetation classification: benefits of multitemporal RapidEye satellite data. Remote Sensing of Environment, 2013, 136 (5): 66~75.
- [9] 刘海启, 裴志远, 张松岭, 等. 资源一号卫星 CCD 图像在作物面积监测中的应用. 农业工程学报, 2001, 17 (4): 140~143.
- [10] Zhi Yang, Yun Shao, Kun Li, et al. An improved scheme for rice phenology estimation based on time-series multispectral HJ-1A/B and polarimetric RADARSAT-2 data. *Remote Sensing of Environment*, 2017 (195): 184~201.
- [11] 王利民, 刘佳, 杨玲波, 等. 基于 NDVI 加权指数的冬小麦种植面积遥感监测. 农业工程学报, 2016, 32 (17): 127~135.
- [12] Belgiu Mariana, Csillik Ovidiu. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. Remote Sensing of Environment, 2018 (204): 509~523.
- [13] Xiong D.. Crop growth remote sensing monitoring and its application. Sensors & Transducers, 2014, 169 (4): 174~178.
- [14] Graesser Jordan, Ramankutty Navin. Detection of cropland field parcels from Landsat imagery. Remote Sensing of Environment, 2017 (201): 165~180.
- [15] Macdonald R. B., Hall F. G., Global Crop Forecasting. Science, 1980, 208 (4445): 670~679.
- [16] Li Qingting, Wang Cuizhen, Zhang Bing, et al. Object-based crop classification with Landsat-MODIS enhanced timeseries data. Remote Sensing, 2015, 7 (12): 16091~16107.
- [17] Belward A. S., Hoyos A. De. A comparison of supervised maximum likelihood and decision tree classification for crop cover estimation from multitemporal LANDSAT MSS data. *International Journal of Remote Sensing*, 1987, 8 (2): 7.
- [18] Stumpf André, Kerle Norman. Object-oriented mapping of landslides using Random Forests. Remote Sensing of Environment, 2011, 115 (10): 2564-2577.
- [19] Raju P. V. Classification of wheat crop with multi-temporal images: performance of maximum likelihood and artificial neural networks. *International Journal of Remote Sensing*, 2003, 24 (23): 4871~4890.
- [20] Zhang Jiahua, Feng Lili, Yao Fengmei. Improved maize cultivated area estimation over a large scale combining MODIS– EVI time series data and crop phenological information. Isprs Journal of Photogrammetry & Remote Sensing, 2014(94): 102~113.
- [21] Breiman Leo. Random Forests. Machine Learning, 2001, 45 (1): 5~32.
- [22] Verma Amit Kumar, Garg Pradeep Kumar, Prasad K. S. Hari. Sugarcane crop identification from LISS IV data using ISODATA, MLC, and indices based decision tree approach. *Arabian Journal of Geosciences*, 2017, 10 (161).
- [23] Balakrishnan P. V., Cooper Martha C., Jacob Varghese S., et al. A study of the classification capabilities of neural networks using unsupervised learning: A comparison with K -means clustering, Psychometrika, 1994, 59 (4): 509~525.
- [24] Jia K., Wu B., Li Q.. Crop classification using HJ satellite multispectral data in the North China Plain. Journal of Applied Remote Sensing, 2013, 7(1): 73576.
- [25]张佳琪,武红旗.基于高分遥感数据的昌吉市棉花面积识别研究.天津农业科学,2017,23(10):55~60.
- [26] 曹卫彬, 刘姣娣, 马蓉. 新疆棉花遥感监测识别区域的划分. 农业工程学报, 2008, 24(4): 172~176.
- [27]李敏, 赵庚星, 蔡明庆, 等. 县域棉花信息遥感提取与棉田精确化管理分区研究. 自然资源学报, 2012, 27(11): 1971~1980.
- [28] Musande Vijaya. Cotton crop discrimination using fuzzy classification approach. Journal of the Indian Society of Remote Sensing, 2012, 40(4): 589~597.
- [29] Min L., Gengxing Z., Yuanwei Q., Extraction and monitoring of cotton area and growth information using remote sensing at small scale: a case study in Dingzhuang town of Guangrao county, China. *International Conference on Computer Distributed Control & Intelligent Environmental Monitoring, IEEE*, 2011.
- [30] 曹卫彬, 杨邦杰, 宋金鹏. TM 影像中基于光谱特征的棉花识别模型. 农业工程学报, 2004, 20(4): 112~116.
- [31] Song Peilin, Mansaray Lamin R., Huang Jingfeng, et al. Mapping paddy rice agriculture over China using AMSR-E time series data. *ISPRS journal of photogrammetry and remote sensing*, 2018(144): 469~482.
- [32] 李静, 柳钦火, 刘强, 等. 基于波谱知识的 CBERS-02 卫星遥感图像棉花像元识别方法研究. 中国科学: 技术科学, 2005, 35(s1): 141~155.
- [33] Kim Hyun Ok, Yeom Jong Min. Effect of red-edge and texture features for object-based paddy rice crop classification using RapidEye multi-spectral satellite image data. *International Journal of Remote Sensing*, 2014, 35(19): 7046~7068.

2019年10月

- [34] 温雨鑫. 历史环境保护与再生中 BIM 技术应用研究. 天津: 河北工业大学, 2015.
- [35] Spectral Response Curves of the RapidEye Sensor. www. blackbridge. com, 2019.
- [36] 王增林, 朱大明. 基于遥感影像的最大似然分类算法的探讨. 河南科学, 2010(28): 1458~1461
- [37] Liaw A., Wiener M.. Classification and regression by Random Forest. R News, 2002: $18{\sim}22$.
- [38] Congalton R. G.. A Review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 1991(37): 35~46.
- [39] Hay A. M.. The derivation of global estimation from a confusion matrix. Int J Remote Sens, 1988(9): 1395~1398.
- [40] Congalton R. G.. A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensing data. *Photogramm Eng Remote Sens.*, 1988(54): 593~600.

Construction and application of cotton characteristic spectral index based on Rapideye image

Wang Limin, Liu Jia[™], Yao Baomin, Gao Jianmeng, Ji Fuhua (Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China)

Abstract: [Purpose] Construction of characteristic spectral index is the key to improve the extraction efficiency of agricultural crops, and it is also the basis of developing identification methods which can be commonly used. [Method] Cotton Extraction Index (CEI) was constructed by summing up the reflectance values in red-edge and near-infrared band of 5m resolution Rapideye image which covered Jizhou city in Hebei province and was acquired on 8 August 2017. Cotton area was identified with the CEI method after mask processing which aimed to eliminate the waterbody and bare land (including buildings). The whole process of CEI method consists of sample selection, mask treatment of water body and bare land, calculation of CEI, cotton identification and accuracy assessment. The threshold of CEI to separate cotton and other objects is considered to be the most important figure, which is obtained by 5 144 sample points. [Result] The overall accuracy and Kappa coefficient of the classification result based on CEI method were 88.80% and 0.751 7 respectively. In order to evaluate the effectiveness of CEI method, random forest (RF) and maximum likelihood classification (MLC) method were used. The result showed that the overall accuracy and Kappa coefficient of the result acquired by MLC were 86.53% and 0.698 3. The figures for RF were 90.12% and 0.766 7 respectively. [Conclusion] The cotton extraction accuracy based on CEI method was similar to those by MLC and RF methods and the CEI method was simpler, more efficient and less dependent on samples, which was applicable to the instant crop extraction.

Key words: cotton; remote sensing identification; Rapideye; characteristic spectrum index; random forest; maximum likelihoods