

苗荣慧, 黄锋华, 杨 华, 等. 基于空谱一体化的农田高光谱图像分类[J]. 江苏农业学报, 2018, 34(4): 818-824.

doi: 10.3969/j.issn.1000-4440.2018.04.015

基于空谱一体化的农田高光谱图像分类

苗荣慧¹, 黄锋华¹, 杨 华¹, 邓雪峰¹, 陈晓倩²

(1. 山西农业大学信息科学与工程学院, 山西 太谷 030801; 2. 西北农林科技大学信息工程学院, 陕西 杨凌 712100)

摘要: 为强化高光谱成像技术在近地农业方面的应用, 以农田近红外高光谱图像为研究对象, 利用高光谱成像技术, 结合光谱分析方法和监督分类方法, 对农田图像进行分类。针对高光谱图像数据量大、非线性等特点, 采用主成分分析(PCA)和支持向量机(SVM)法建立农田图像分类器。在利用光谱信息分类的基础上, 采用空谱一体化方法对光谱分类结果进行修正, 去除孤立点和噪声的影响。基于支持向量机的总体分类精度为 88.4%, 采用空谱一体化方法的总体分类精度最高达 89.7%, 说明利用空间信息修正光谱信息可以提高近地农田对象的分类精度, 为基于高光谱图像的农田识别提供理论依据。

关键词: 高光谱图像; 空谱一体化; 农田图像分类; 主成分分析; 支持向量机

中图分类号: TP751; TP391 **文献标识码:** A **文章编号:** 1000-4440(2018)04-0818-07

Farmland classification of hyperspectral image based on spatial-spectral integration method

MIAO Rong-hui¹, HUANG Feng-hua¹, YANG hua¹, DENG Xue-feng¹, CHEN Xiao-qian²

(1. College of Information Science and Engineering, Shanxi Agricultural University, Taigu 030801, China; 2. College of Information Engineering, Northwest A&F University, Yangling 712100, China)

Abstract: In order to intensify the application of hyperspectral imaging technology in near field agriculture, near-infrared hyperspectral images were selected as research objects, and the hyperspectral imaging technology combining with spectral analysis method and supervised classification method was used to classify the farmland images. Since the hyperspectral data had the characteristic of huge and nonlinear, principal component analysis (PCA) and support vector machine (SVM) were adopted for classifier establishing. On the basis of spectral classification, spatial-spectral integration method was used to amend the spectral classification results, the isolated points and noise were removed. The results showed that the overall classification accuracy by SVM could reach 88.4%, and the highest overall classification accuracy by spatial-spectral integration method was up to 89.7%, indicating that using spatial information to modify spectral information could improve the classification accuracy of farmland objects, which would provide theoretical basis for hyperspectral image identification of near field farmland.

Key words: hyperspectral image; spatial-spectral integration; farmland image classification; principal component analysis (PCA); support vector machine (SVM)

收稿日期: 2017-10-31

基金项目: 国家自然科学基金项目(31671571); 山西农业大学青年科技创新基金项目(2017013)

作者简介: 苗荣慧(1990-), 女, 山西晋城人, 硕士, 助教, 主要从事图像分析与机器视觉、农产品无损检测研究。(Tel) 18306828214; (E-mail) ronghui092@163.com

高光谱成像技术将二维空间图像技术和一维光谱技术相结合, 形成三维光谱图像, 它不仅可以表征地物的图像特征, 还可以表征某一像素点的光谱特性^[1]。与传统图像相比, 高光谱图像具有波段多、波段窄、图谱合一等特性, 能以较高的可信度识别地

物目标^[2]。因此,它在农业环境监测、作物长势监测、品种识别、内部品质检测、土壤识别和监测等方面都得到了广泛应用^[3]。由于不同农作物在同一光谱波段上的反射率存在较大差异,所以利用不同物不同谱,同物一定同谱的原理不仅可以区分不同农作物,而且识别率也较高。农作物遥感是农业遥感的基础与前提,没有高精度的农作物信息提取就无法进行农业资源调查,作物评估和灾害监测^[4]。

在利用光谱技术获取农田信息研究方面,Williams等^[5]采用高光谱图像对玉米内核的3种硬度进行分类识别,结果显示基于积分直方图和光谱均值的分类结果较好。Mo等^[6]采用偏最小二乘回归映射苹果可溶性固形物含量模型,指出高光谱吸光度成像技术能对苹果可溶性固形物含量进行有效预测。Pan等^[7]利用白萝卜高光谱图像,对其凹陷部分成功构建了偏最小二乘判别分析和BP(Back propagation)反馈神经网络2种预测模型。孙俊等^[8]、黄双萍等^[9]分别采用高光谱成像技术检测掺假大米和预测柑橘叶片氮素含量。邹卓阳等^[10]、郑纪伟等^[11]分别将高光谱遥感技术应用在植被信息提取和农作物分类中。在这些研究中仅利用图像的光谱信息进行分类识别,没有考虑空间域信息。同时,室内研究大多只针对单类农产品的检测,室外研究大多为航拍遥感图像,而且室内光谱数据的获取对环境和设备要求不高,图像的室外采集受到大气辐射、环境等因素干扰。针对这些存在的问题,本研究拟以农田近红外高光谱图像为研究对象,通过光谱数据获取及预处理,光谱特征分类,空间信息分类等处理,提出一种针对农田对象的分类判别模型。在光谱特征分类的基础上,采用空谱一体化方法,基于不同物不同谱,同物一定同谱的原理将高光谱成像技术应用于室外近地农田中,以期提出一种利用农田高光谱图像进行地物识别与分类的有效方法。

1 材料与方法

1.1 高光谱成像系统

研究采用 HyperSIS 高光谱成像分析系统(SPECIM 公司产品),主要由推扫型近红外成像光谱仪(N17E)、面阵探测器、数据采集软件(SpectralSENS)、计算机以及室外地物扫描配件等组成。N17E为体相全息透射光栅分光型,波长900~1700 nm,光谱分辨率为5 nm,光谱采样点为4 nm。面阵

探测器,满帧像素数为320×256。SpectralSENS软件主要用于实现光谱及图像的实时采集,白光漫射型光源,光谱为350~2500 nm,功率100 W。高光谱图像采集装置如图1显示。

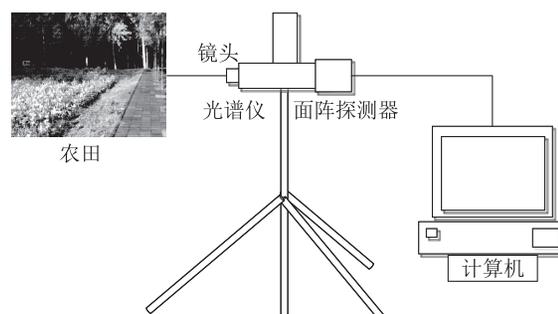


图1 高光谱图像采集装置

Fig.1 Acquisition system of hyperspectral image

1.2 高光谱图像获取

通过高光谱成像仪获得的原始光谱图像存在各波段光源强度分布不均,易受摄像头中暗电流噪声影响,以及不同采集系统对检测光的敏感程度不同等问题。采用不同的高光谱成像系统在相同的外界条件下采集同样的样本,所获取的高光谱图像也不一定相同^[12]。因此,必须对原始图像进行黑白标定。具体方法为:首先采集白板图像,获得全白标定图像,然后在完全避光条件下,获得全黑标定图像。采用公式(1)对图像进行处理,得到灰度分布在0到1之间的相对图像。

$$R = (I - B) / (W - B) \quad (1)$$

式中, W :全白标定图像; B :全黑标定图像; I :原始高光谱图像; R :标定后高光谱图像。

高光谱成像仪的各参数如下:垂直方向扫描视觉为40°,水平方向扫描视觉为70°,扫描农田的速度为20 mm/s,相机曝光时间为10 ms,采集到的高光谱图像大小为320×400×256。采用ENVI 4.7软件对高光谱图像进行处理和分析。采集到的图像既包含二维图像信息,也包含一维单个像素点的光谱信息。

1.3 样品来源

研究样本来源于西北农林科技大学北校区试验田,主要农田对象有农作物、道路、树木和土壤等,部分图像包含2~3个农田对象,需对其进行图像筛选。共获取了不同光照条件下的60幅高光谱图像,

每幅样本图像的波段数为 256。选择农田对象类别多样的图像为研究对象,去除背景复杂、对象模糊的图像,最终选取 48 幅图像用于后续研究。

1.4 光谱数据获取及预处理

1.4.1 感兴趣区域光谱数据的获取 利用感兴趣区域获得特征光谱,然后用获得的特征光谱构建分类器。根据实地调查结果,决定采用交互式的感兴趣区域获取方式在高光谱图像上选取样本。利用 ENVI 4.7 软件获取多种类别农田对象的感兴趣区域,并将其作为样本集。考虑到农田图像的全局性,在图像中选取同一类别的多处不规则区域作为样本集。农田高光谱图像主要包括农作物、道路、树木和土壤,将这 4 类之外的对象统一归为其他。将获取的样本集按照 3 : 1 的比例随机分成训练集和测试集,训练集用来建立分类器,测试集用来测试分类器的性能。表 1 为不同农田对象获得的样本集统计结果,将获取的光谱数据保存为文档格式,便于后续分析。

表 1 不同农田对象样本集数据

Table 1 Sample sets of different farmland objects

数据集	样本集数目				
	农作物	土壤	树木	道路	其他
训练集	2 301	1 962	3 126	2 343	945
测试集	767	654	1 042	781	315

1.4.2 光谱数据预处理 采用多元散射校正(MSC)、Savitzky Golay (SG)平滑、标准正态变量变换(SNV)和导数处理等预处理方法对训练集和测试集光谱数据进行预处理。多元散射校正是一种多变量散射校正技术,该方法需要建立一个待测样品的理想光谱,然后以该光谱为标准对其他样品光谱进行修正^[13]。SG平滑方法基于最小二乘法原理,能够保留并分析信号中的有用信息,消除随机噪声^[14]。标准正态变量变换主要是用来消除固体颗粒大小、表面散射以及光程变化对近红外反射光谱的影响,是基于一条光谱来进行处理的^[15]。导数算法能够很好地消除基线漂移或减轻背景干扰造成的影响,提供比原始光谱分辨率更高、更清晰的轮廓变化^[16]。基于 Matlab7.5 平台来实现光谱数据的预处理。

1.5 分类方法

1.5.1 基于光谱信息的高光谱图像分类

1.5.1.1 基于主成分分析的光谱信息降维 采用主成分分析法实现光谱信息降维,对原始变量进行线性组合变换,得到新变量,利用新变量表征原始变量信息,去除光谱重叠信息。设 $X = \{x_1, x_2, \dots, x_p\}$ 为农田高光谱图像的光谱维矢量集,其中 $x_i = (x_{i1}, x_{i2}, \dots, x_{iN})^T$ 为任一样本像元的光谱维矢量, $i = 1, 2, \dots, p$, 其中 p 为矢量个数, N 为波段数。对 X 的协方差矩阵 $R_{N \times N}$ 进行特征值分解:

$$R_{N \times N} = U \Lambda U^T \quad (2)$$

其中, $U = (u_1, u_2, \dots, u_N)$ 为正交特征矢量矩阵, $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$ 为 $R_{N \times N}$ 的特征值矩阵,且 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N \geq 0$ 。对光谱矢量进行 PCA 变换,有

$$y_i = U^T(x_i - \mu) \quad (3)$$

其中 $y_i = (y_{i1}, y_{i2}, \dots, y_{iN})^T$, $i = 1, 2, \dots, p$ 。 $Y = \{y_1, y_2, \dots, y_p\}$ 为获得的光谱主成分矢量集。

1.5.1.2 基于支持向量机(SVM)的分类识别 SVM 是二分类器,研究选取“1-v-1”(One against one)方法实现多分类器的构建,即图像中存在 k 类农田对象,需建立 $\frac{k(k-1)}{2}$ 个二分类器,然后采用投票法解决

分类问题^[17]。由于图像中有农作物、道路、树木、土壤和其他 5 类对象,需建立 10 个二分类器,然后将测试样本输入分类器,得票数最多的类别则为该样本所属类别。在参数选择上采用网格搜索法实现 C 和 γ 的选择,其中 C 为惩罚参数, γ 为核函数参数。

1.5.2 基于空谱一体化的高光谱图像处理 经光谱分类后,图像存在明显的噪声和孤立点,因此需对经过光谱分类后的图像采用空间域处理方法进行处理。空谱一体化技术就是将空间信息引入到高光谱分类中,在光谱分类结果的基础上,对图像像素邻域进行决策,以期对分类结果进行修正,具体流程如图 2 显示。由于图像的灰度信息具有空间相关性,即图像具有明显的聚类效果,在光谱分类后,本研究分别采用主要和次要分析、类别集群法、类别筛选法对分类后的图像进行后处理。

1.5.2.1 主要和次要分析 主要和次要分析是用卷积模板对图像进行平移卷积运算,其中主要分析是将当前位置类别的像素值替换为 8 邻域空间中数目

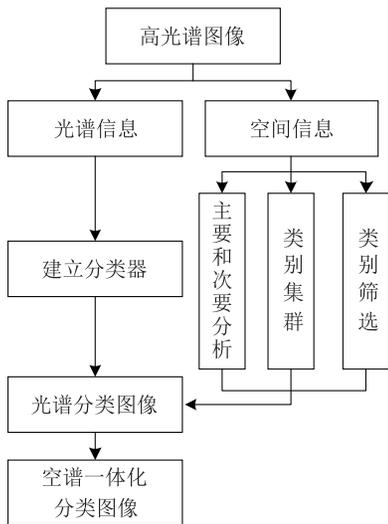
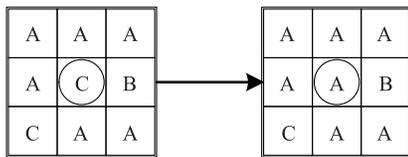


图2 空谱一体化原理图
Fig.2 Schematic diagram of spatial-spectral integration

较多类别的像素值,次要分析则相反。图3显示,假设A、B和C代表不同类别,则C将会被其邻域空间中类别数较多的A替换。



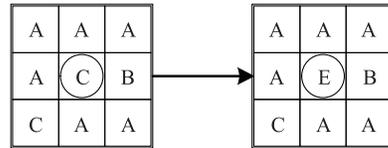
A、B、C 分别代表不同的类别。
图3 主要和次要分析原理图

Fig.3 Schematic diagram of majority and minority analysis

1.5.2.2 类别集群法 类别集群法是使用数学形态学方法将相邻相似的区域进行合并,以保证空间的连续性,然后将处理结果输出到内存中^[18]。主要利用形态学中的腐蚀操作实现,每次仅处理一个类,即对分类图中的一个类进行处理产生结果 m1,然后对 m1 中的另外一个类进行处理,以此类推实现多类的腐蚀操作。

1.5.2.3 类别筛选法 类别筛选法在统计分析的基础上进行,将孤立的像素从分类中独立出来。通过设置最小剔除像素个数剔除某些分类,如设置最小阈值为 2,即当分类的像素个数小于 2 时就剔除此分类,选择 8 邻域空间对像素进行统计。图 4 为类别筛选法原理图,当中心像元的

元素在邻域空间的统计数值小于 2 时,剔除该类将其分类为其他。



A、B、C 分别代表不同的类别;E:其他。

图4 类别筛选法原理图

Fig.4 Schematic diagram of classification screening

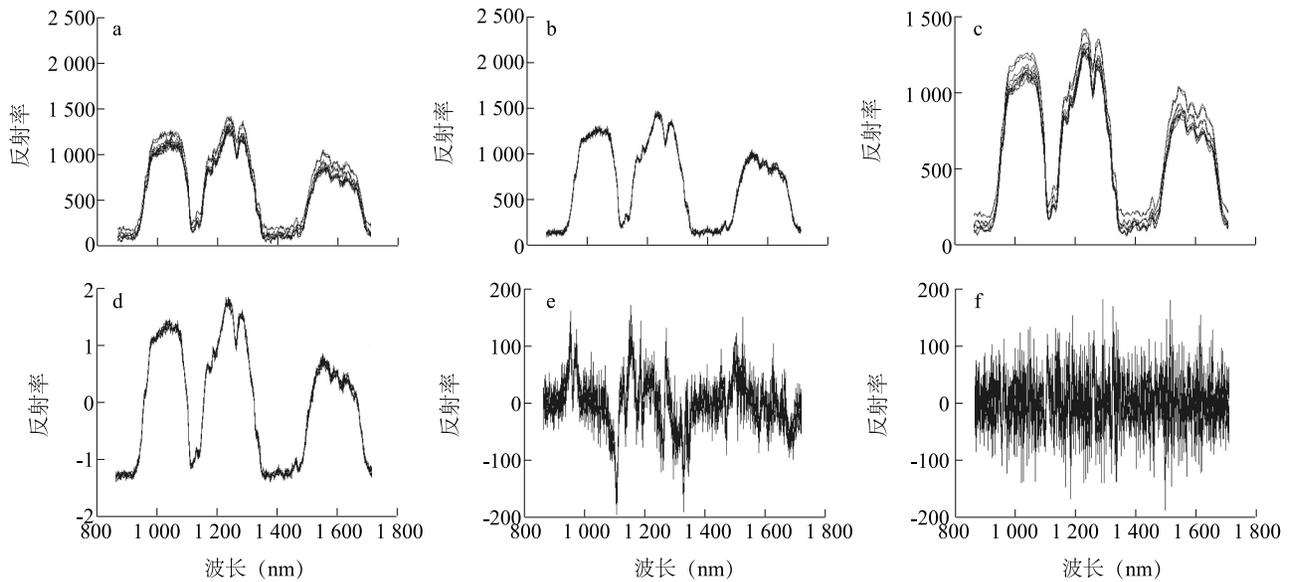
2 结果与分析

2.1 光谱数据预处理结果

图 5 显示,经过光谱预处理后光谱曲线基本保留了原始光谱数据的吸收特性,并有不同程度的强化。SG 平滑方法处理后的数据离散度较大,相比其他 2 种处理结果稍差,这是因为采用平滑方法进行预处理时,拟合次数和窗口移动参数的选择会影响预处理效果。多元散射校正和标准正态变量变换处理后的曲线较平滑,表明这 2 种方法均能有效去除光谱数据中由光谱散射引起的误差。导数处理能够有效解决光谱数据中基线漂移的现象。采用多元散射校正对获取的光谱数据进行预处理效果最好,因此本研究采用多元散射校正预处理之后的数据进行后续研究。

2.2 基于光谱特征的农田对象分类结果

主成分分析结果表明,主成分 PC1 和主成分 PC2 能够表征原始特征信息的 99%,因此选择前 2 个主成分作为分类器的输入向量,输入到 SVM 分类器中。采用总体分类精度和 Kappa 系数对分类结果进行验证。图 6 显示,各农田对象分类明显,并且比较符合人眼的视觉感知结果,但分类图像中也存在一些孔穴和噪声点。表 2 为 SVM 分类的验证结果,其中真实图像是通过单波段选取方式结合光谱图像反射率波峰处 1 128 波段,选择图像信息较清晰的单波段图像,然后通过人工分类标注进行手动分类,其分类结果接近理想化状态。SVM 的总体分类精度为 88.4%,这是通过人工手动分类获取真实像元的结果。



a:原始光谱;b:多元散射校正处理后的光谱;c:Savitzky Golay平滑处理后的光谱;d:标准正态变量变换处理后的光谱;e:一阶导数处理后的光谱;f:二阶导数处理后的光谱。

图5 预处理结果比较

Fig.5 Comparison of preprocessing results

2.3 基于空谱一体化的高光谱图像处理

主要和次要分析、类别集群法、类别筛选法作用于光谱分类后图像的处理结果(图7)显示,主要和次要分析、类别集群法效果较好,这2种方法能够利用图像的邻域信息将小类别归到大类别。类别筛选法使得图像存在更多的孤立点,一些孤立的点被分为其他,从而降低了分类图的空间连续性,效果较差。

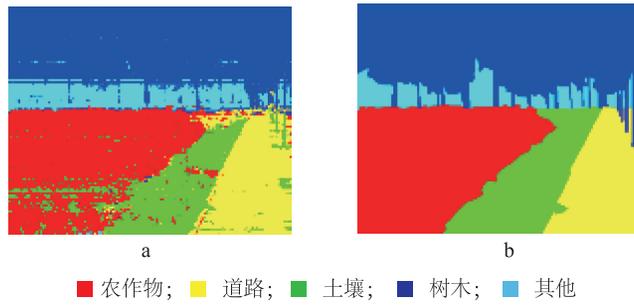
分类器在测试集上的精度验证结果(表2)显示,经主要和次要分析处理后图像的分类精度较高,最高可以达到89.7%。通常情况下,采用空间域图像处理方法能够去除孤立点,提高分类精度。但针对不同图像、不同方法,空间域图像处理方法并不是普遍适用的,分类精度也可能会降低,因此在分类过程中应结合图像特征合理使用图像处理方法。

3 讨论

本研究以近红外农田高光谱图像为研究对象,对农田环境中不同农田对象的分类方法进行研究,综合应用高光谱成像技术、光谱分析技术、数字图像处理以及机器学习、模式识别等对农田高光谱图像

进行分析,最终结合光谱特征和空间信息实现了不同农田对象的分类,为室外农田近红外高光谱图像的研究奠定了基础。在分类器的建立中采用适合高光谱数据的支持向量机方法,完成了对农田对象的分类并对分类结果进行了分析,SVM的总体分类精度为88.4%。

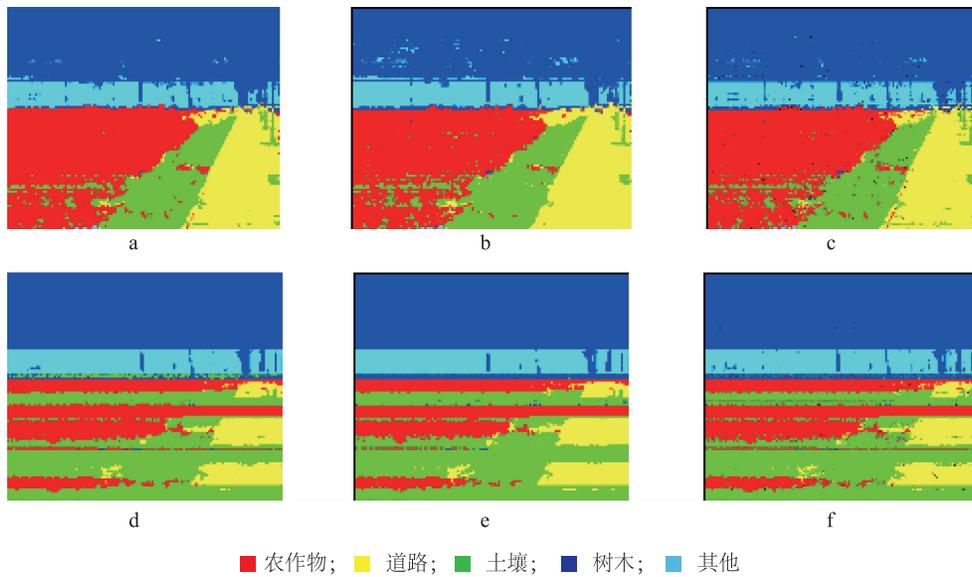
基于高光谱成像技术的地物分类识别研究多为航拍遥感图像,针对多类别的总体分类精度通常为80.0%~95.0%^[19-20]。本研究基本实现了不同农田对象的分类,表明运用高光谱成像技术和光谱分析技术可以实现农田对象的提取与分析,满足室外农田近红外高光谱图像研究的要求。与传统去噪技术相比,空谱一体化方法更适合分类图像,更能考虑到图像上下文信息。但是,由于高光谱图像的背景过于复杂,不同对象本身存在分类模糊现象,因此其分类精度会受到一定影响,本研究还可通过图像分割方法对高光谱图像进行分割,在图像分割的基础上提取分割区域的特征信息。本研究只选择了光谱特征来建立分类器,若选取合适的纹理特征和其他形状特征,将光谱与纹理特征等进行集成,可能会提高分类的精度。



a: SVM 分类结果; b: 人工手动分类真实地物图像。

图 6 SVM 和人工手动的分类结果图

Fig.6 Classification results of hyper-spectral image based on support vector machine (SVM) and manual classification method



a: 支持向量机+主要和次要分析; b: 支持向量机+类别集群法; c: 支持向量机+类别筛选法; d: 最大似然法+主要和次要分析; e: 最大似然法+类别集群法; f: 最大似然法+类别筛选法。

图 7 基于支持向量机和最大似然法的空间域处理结果

Fig.7 Spatial processing results based on SVM and maximum likelihood

表 2 测试集分类精度验证

Table 2 Evaluation of classification accuracy on testing set

分类结果	总体分类精度 (%)	Kappa 系数
SVM	88.4	0.85
SVM-M	89.7	0.86
SVM-C	87.6	0.84
SVM-S	88.6	0.85
ML	70.4	0.63
ML-M	70.5	0.63
ML-C	69.4	0.62
ML-S	70.3	0.63

SVM: 支持向量机; ML: 最大似然法; SVM-M, SVM-C 和 SVM-S 表示利用支持向量机法对图像进行分类后, 再分别采用主要和次要分析、类别集群法以及类别筛选法进行处理; ML-M, ML-C 和 ML-S 表示利用最大似然法对图像进行分类后, 再分别采用主要和次要分析、类别集群法以及类别筛选法进行处理。

参考文献:

[1] 马本学, 应义斌, 饶秀勤, 等. 高光谱成像在水果内部品质无损检测中的研究进展[J]. 光谱学与光谱分析, 2009, 29(6): 1611-1615.

[2] 张文杰, 张钧萍, 张 晔. 基于区域活动轮廓模型的高光谱图像分割方法[J]. 遥感技术与应用, 2008, 23(3): 351-355.

[3] 冯 静, 舒 宁. 一种新的高光谱遥感图像纹理特征提取方法研究[J]. 武汉理工大学学报, 2009, 31(3): 10-14.

[4] 刘 亮, 姜小光, 李显彬, 等. 利用高光谱烟感数据进行农作物分类方法研究[J]. 中国科学院研究生院学报, 2006, 23(4): 484-488.

[5] WILLIAMS P J, KUCHERYAVSKIY S. Classification of maize kernels using NIR hyperspectral imaging[J]. Food Chemistry, 2016, 209: 131-138.

- [6] MO C, KIM M S, KIM G, et al. Spatial assessment of soluble solid contents on apple slices using hyperspectral imaging[J]. Biosystems Engineering, 2017, 159: 10-21.
- [7] PAN L, SUN Y, XIAO H, et al. Hyperspectral imaging with different illumination patterns for the hollowess classification of white radish[J]. Postharvest Biology and Technology, 2017, 126: 40-49.
- [8] 孙俊, 金夏明, 毛罕平, 等. 高光谱图像技术在掺假大米检测中的应用[J]. 农业工程学报, 2014, 30(21): 301-307.
- [9] 黄双萍, 洪添胜, 岳学军, 等. 基于高光谱的柑橘叶片氮素含量多元回归分析[J]. 农业工程学报, 2013, 29(5): 132-138.
- [10] 邹卓阳, 杨武年, 陈颖. 高光谱遥感技术在植被信息提取中的应用[J]. 测绘, 2010, 33(2): 55-57.
- [11] 郑纪伟, 潘泉, 赵永强, 等. 航拍高光谱图像中基于投影的自动目标检测算法[J]. 计测技术, 2005, 25(3): 4-7.
- [12] 张保华, 李江波, 樊书祥, 等. 高光谱成像技术在果蔬品质与安全无损检测中的原理及应用[J]. 光谱学与光谱分析, 2014, 34(10): 2743-2751.
- [13] ZHANG C, GUO C, LIU F, et al. Hyperspectral imaging analysis for ripeness evaluation of strawberry with support vector machine[J]. Journal of Food Engineering, 2016, 179: 11-18.
- [14] 陈华舟, 潘涛, 陈洁梅. 多元散射校正与 Savitzky-Golay 平滑模式的组合优选应用于土壤有机质的近红外光谱分析[J]. 计算机与应用化学, 2011, 28(5): 518-522.
- [15] 程术希. 基于光谱和成像技术的作物病害不同侵染期快速检测方法研究[D]. 杭州: 浙江大学, 2014.
- [16] 郭斗斗. 土壤有机质含量高光谱预测模型的构建[D]. 扬州: 扬州大学, 2014.
- [17] 郭腾霄, 丁学全, 董晓强, 等. 支持向量机在危险化学品被动红外遥测光谱鉴别中的应用[J]. 计算机与应用化学, 2013, 30(1): 18-20.
- [18] 王超. 基于 SPOT5 遥感图像的自动分类及处理方法研究[D]. 昆明: 云南大学, 2011.
- [19] 张倩. 基于决策树方法的航空高光谱遥感土地覆盖分类研究[D]. 青岛: 山东科技大学, 2005.
- [20] 赵春晖, 张焱, 王玉磊. 基于小波核主成分分析的相关向量机高光谱图像分类[J]. 电子与信息学报, 2012, 34(8): 1905-1910.

(责任编辑: 王妮)