

基于高光谱的 GA 和 SPA 算法对赣南脐橙叶绿素定量分析

刘燕德, 张光伟, 蔡丽君

华东交通大学机电学院光机电技术及应用研究所, 江西 南昌 330013

摘要 用遗传算法(GA)和连续投影算法(SPA)分别提取了赣南脐橙叶片高光谱图像的有效信息,对叶绿素的含量用偏小二乘法(PLS)进行建模定量分析。高光谱图像标定后,提取感兴趣区域(ROI)的平均光谱,用GA和SPA算法分别选出了27和8条特征波长,然后用PLS对叶绿素含量建模。GA-PLS与SPA-PLS模型得到的预测集相关系数分别为0.80和0.83,均方根误差分别为2.45和2.30。结果表明:SPA-PLS模型具有较高的优势,可以结合高光谱技术对赣南脐橙叶绿素含量快速、无损的定量分析。

关键词 叶绿素含量; 高光谱成像; 赣南脐橙; 遗传算法; 连续投影算法

中图分类号: TP391 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2012)12-3377-04

引言

叶绿素是一类与光合作用(photosynthesis)有关的最重要的色素,叶绿素的含量能间接反映植物生长情况,光合作用的能力与植物的营养状况密切相关^[1],也是赣南脐橙的重要指标之一。传统的分光光度法测量叶绿素含量,会损伤植物叶片且耗费大量时间^[2]。SPAD叶绿素仪测量叶片叶绿素的含量时需要反复接触叶片、多次测量,工作量大,且仅能检测某个点或每个小区域叶绿素的含量^[3]。高光谱具有波段多、图谱合一和分辨率高的特点,把二维图像和光谱技术融为三维的图谱信息。用高光谱成像技术检测的叶绿素具有安全、快速、无损、定量等特点,并能对叶片大区域叶绿素含量的检测,有较大的应用潜力。

石吉勇等^[4]用高光谱技术分析黄瓜叶片中叶绿素含量及分布,彭颜昆等^[5]研究了基于LS-SVM和高光谱技术的玉米叶片叶绿素含量检测。但大多使用宽的光谱波段进行叶绿素含量研究,模型较大。选取叶绿素光谱的特长波长进行分析研究的较少,鉴于此,用遗传算法(GA)和连续投影算法(SPA)对脐橙叶片的光谱信息进行特长波长的选择分析。

1 实验部分

1.1 材料

用的脐橙叶片样本于2012年4月23日采集于赣州市脐

橙果园,分别采集了14棵树,每棵树采集了10片,共采集了140片,叶片大小为长55 mm×宽35 mm×长130 mm×宽75 mm,尽量覆盖这个生长时期的所有脐橙叶片。

1.2 数据采集

叶绿素的高光谱图像数据是基于光谱仪的高光谱成像系统采集到的。高光谱成像系统如图1所示,是由基于图像光谱仪的高光谱摄像机(ImSpector, V10E, 芬兰)、两盏20 W卤素灯(OSRAM, DECOSTAR 51, MR16)、精密电动平台、运动控制器和计算机等部件组成。高光谱摄像机与位移平台的高度为50 cm,光源距位移平台的高度为65 cm,调整光源

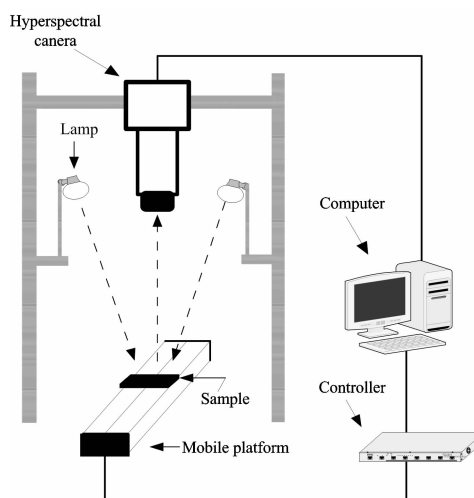


Fig. 1 Hyperspectral imaging system

收稿日期: 2012-05-26, 修订日期: 2012-08-08

基金项目: 科技部农业科技成果转化资金项目(2011GB2C500008), 国家(863计划)课题项目(2012AA101906)和江西省研究生创新专项资金项目(YC2011-S077)资助

作者简介: 刘燕德, 女, 1967年生, 华东交通大学教授 e-mail: jxliuyd@163.com

的角度,使两盏灯的光源正好射在位移平台中心上,并且使光照均匀。

实验数据采集前,为了避免基线漂移,将高光谱成像系统打开预热 30 min。将脐橙叶片平铺在平台的黑色底板上采集高光谱图像。设定的高光谱成像系统相机的曝光时间为 20 ms,摄像头分辨率为 $1\,344 \times 1\,024$,电动平台的移动速度为 $16\text{ mm} \cdot \text{s}^{-1}$,光谱范围是 $400 \sim 1\,000\text{ nm}$,光谱分辨率为 2.8 nm,光谱采样间隔为 2.44 nm,采集到 254 个波长下的图像,最终得到一个大小为 $1\,344 \times 1\,024 \times 254$ 的高光谱图像数据块。

1.3 高光谱图像黑白标定

采集到的高光谱图像在光源强度分布较弱的波段下含有大量的噪声,这是由暗电流噪声的存和样本受光强度不均匀造成的。需要对采集到的高光谱图像进行黑白标定^[6]。在与样品采集相同的条件下,关闭相机快门采集到全黑的标定图像 B ,打开快门扫描标准白色校正板得到全白的标定图像 W 。根据公式(1)使采集到的绝对图像 I 变成相对图像 R

$$R_{\lambda} = \frac{I_{\lambda} - B_{\lambda}}{W_{\lambda} - B_{\lambda}} \quad (1)$$

1.4 叶绿素含量的测定

脐橙叶绿素含量的测定使用 SPAD-520Plus 叶绿素仪测量,SPAD 值表示作物叶子中叶绿素的相对含量。SPAD-520Plus 叶绿素仪只能测量叶片一点的叶绿素含量,为了测量叶片一个区域的含量,本实验用 SPAD-520Plus 分别测量如图 2 所示叶片样本上圆形区域所示叶片 5 个位置的叶绿素含量,然后计算 5 个位置叶绿素含量的平均值。

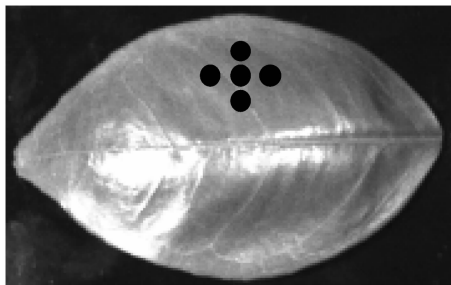


Fig. 2 Diagram of location selection for chlorophyll content measurement

2 结果与讨论

2.1 数据处理与分析

针对采集到的高光谱图像,用 ENVI 软件选取与 SPAD-520Plus 叶绿素仪测量位置对应的感兴趣区域 region of interesting(ROI),由于叶片的大小不一样,所以感兴趣区域面积也不相同,按叶片的大小比例选取矩形 ROI。感兴趣区域选取时,叶片的为彩色图像选择谱的通道 R 为 650 nm, G 为 550 nm, B 为 450 nm,如图 3(a)所示,由重塑的伪彩色图像可以看出,脐橙叶片具有本身叶片较厚,表面具有较厚油脂层的特点。得到感兴趣区域所有像素点的光谱曲线,然后计算该感兴趣区域 ROI 的平均光谱。

140 个脐橙叶片 ROI 的平均光谱图像如图 4 所示。全部

光谱信息包含较多的复杂信息,叶绿素主要吸收红光及蓝紫光(在 $640 \sim 660\text{ nm}$ 的红光部分和 $430 \sim 450\text{ nm}$ 的蓝紫光部分都有吸收峰),但由于高光谱摄像机的感应曲线的关系,采到的光谱与一般绿色植物的典型光谱曲线有区别。

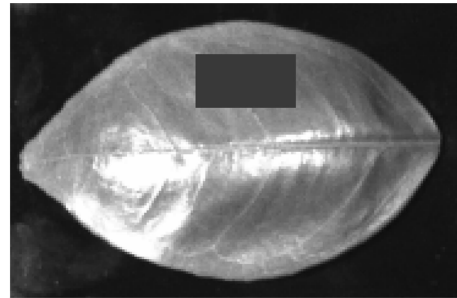


Fig. 3 ROI selection for navel orange leaf samples

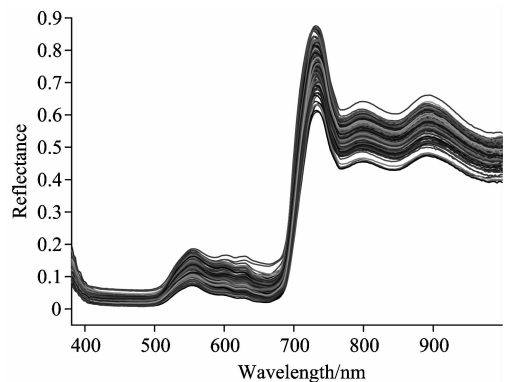


Fig. 4 ROI average spectral curve for 140 samples

将 140 个样本采用 K-S 分类,以 3 : 1 分为包含 93 个样本的校正集和包含 47 个样本的预测集,叶绿素 SPAD 值分类结果如表 1 所示。

Table 1 Descriptive statistics for sample measurement

	样本数	范围(SPAD 值)	均值	标准差
校正集	93	58.7~79.5	69.99	4.511 865
预测集	47	64.6~82.2	73.79	4.007 352

2.2 特征波长选择

遗传算法(genetic algorithm)^[7]是一类借鉴生物界的进化规律(适者生存,优胜劣汰遗传机制)演化而来的随机化搜索方法。它是由美国的 Holland 教授 1975 年首先提出,其主要特点是直接对结构对象进行操作,不存在求导和函数连续性的限定;具有内在的隐并行性和更好的全局寻优能力;采用概率化的寻优方法,能自动获取和指导优化的搜索空间,自适应地调整搜索方向,不需要确定的规则。将遗传算法(GA)应用在光谱分析中,可以实现全局搜索,选出特征波长,消除无关变量的干扰^[8]。

将 RIO 的平均光谱导入 matlab R2010a,参数设置为:群体数目为 30,交叉概率 0.50,编译概率 0.01,迭代次数为 100,独立运行 100 次。依照上述参数运行 GA100 次,每次输出 0-1 二进制编码字符串,计算波长点标识为“1”的概率。

挑选出频率较高的 27 个波长点 384, 396, 399, 406, 408, 411, 560, 562, 565, 567, 630, 648, 650, 652, 655, 667, 669, 691, 694, 696, 699, 701, 711, 713, 977, 979, 982 nm 为特征波长, 如图 5 所示。

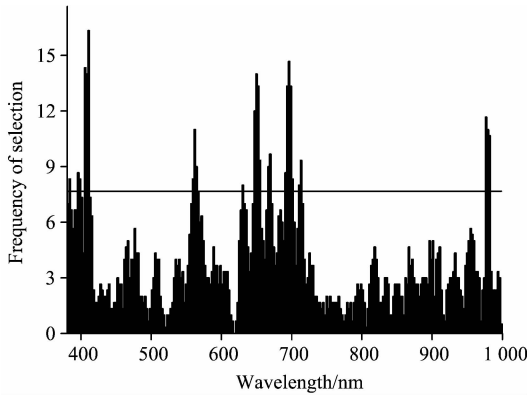


Fig. 5 Frequency of wavelength selection

连续投影算法 (SPA)^[9] 是一种新兴的敏感波长选取算法, 能够从光谱信息中充分寻找含有最低限度的冗余信息的变量组, 有效消除各波长变量之间的共线性影响, 使得变量之间的共线性达到最小, 降低模型的复杂度。其简便快速的特点得到越来越多的应用, 在多种样品波长的选取中取得了很好的效果^[10]。

用 SPA 算法对 ROI 区域的平均光谱进行筛选, 结果如图 6 所示, 得到 399, 528, 640, 672, 699, 738, 767, 979 nm 共 8 个波长作为叶绿素含量的优选波长组合。

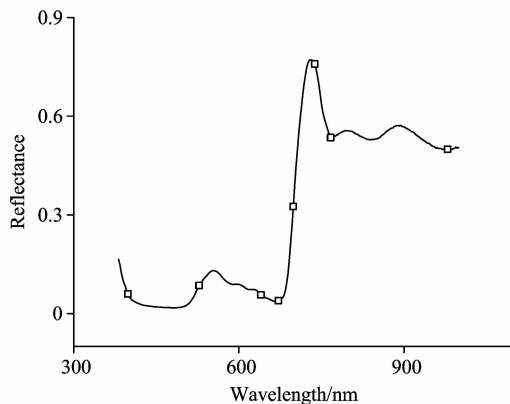


Fig. 6 Selected wavelength using SPA

比较 GA 和 SPA 算法, GA 算法挑选出的波长点主要集中在 550~720 nm 范围, SPA 选出的波长集中在 520~800 nm 波段。两种算法挑选的范围基本吻合, 都是叶绿素信息区域波段, 含有叶绿素的吸收峰。但是 GA 算法选择的波长点具有早熟的问题, 仅仅根据适应度函数决定所选波长点的优劣, 使得适应度较大的波长点在全谱内迅速扩散, 从而导致全谱失去多样性, 陷入局部最优。每次运行 GA 算法得到的特征波长数与波长点也不相同, 不利于模型的稳定。SPA 算法挑选的波长点主要是光谱的波峰或波谷, 且波长点较少, 每次运行 SPA 算法得到相同的结果。相对于 GA 算法,

SPA 有利于模型的简化、稳定, 便于进一步建模。

2.3 建模结果

偏最小二乘法是一种数学优化技术, 它通过最小化误差的平方和找到一组数据的最佳函数匹配, 并已被证明是从植被冠层光谱数据中提取叶面属性的最有效经验方法^[11]。全部光谱和 GA 与 SPA 算法选出的特征波长分别用 PLS 建模, 结果如表 2 所示。

Table 2 Prediction results for different algorithm-based models

Method	R	R _p	RMSEC	RMSEP	PCs
PLS	0.84	0.80	2.45	2.54	4
GA-PLS	0.85	0.80	2.39	2.45	4
SPA-PLS	0.85	0.83	2.34	2.30	4

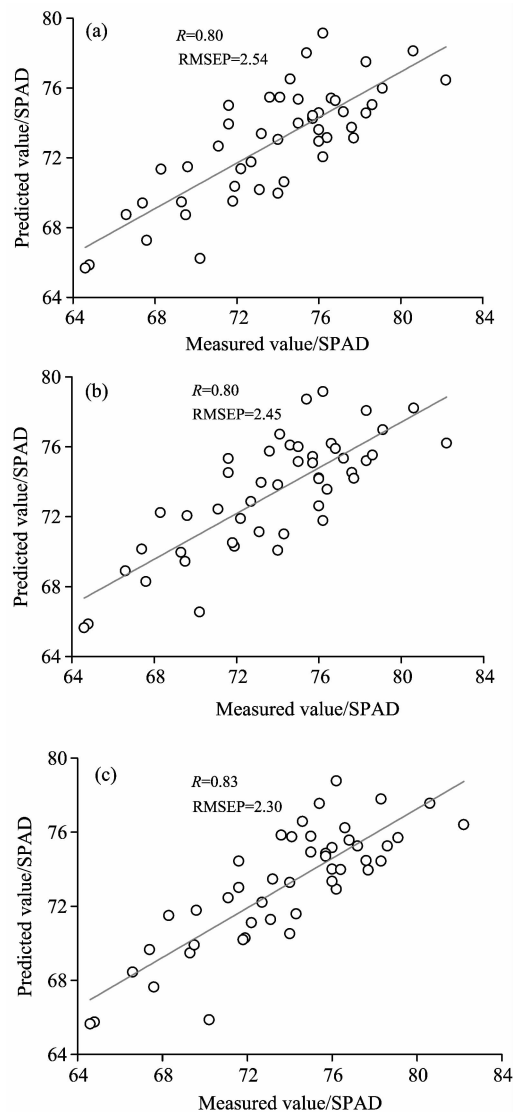


Fig. 7 Plots of measured SPAD value vs hyperspectral predicted values

(a): PLS; (b): GA-PLS; (c): SPA-PLS

基于高光谱成像技术可以有效的采集到脐橙叶的光谱信

息, 选取合适的 ROI 的平均光谱, 光谱含有大量与脐橙叶片叶绿素无关的光谱变量, 这些冗余信息的介入势必在一定程度上扩大估计方差, 导致模型稳定性下降、预测性能降低。因而用全光谱建立的偏最小二乘模型精度不高, 容易出现“过拟合”现象, 且模型显得过于复杂, 所以选择合适的特征波长简化模型是进行建模的关键。本研究讨论了 GA 和 SPA 两种算法对高光谱数据进行最优波段选择, 并用 PLS 进行建模。分析结果可知 SPA 算法选择的 8 个特征波长用 PLS 建模可以得到较好的结果 R_p 为 0.83, RMSEP 为 2.30, 均优于 GA 算法, 且 SPA 算法挑选的波长数较少有利于进一步简化模型。

3 结 论

通过高光谱技术可以获得脐橙叶片的大量光谱数据, 对这些数据怎么样进行降维筛选及特征波长的提取一直是研究的难点。讨论了遗传算法(GA)和连续投影算法(SPA)对高光谱数据的特征波长选择的优劣, 然后用偏最小二乘法(PLS)对脐橙叶片叶绿素含量进行建模。结果表明: SPA-PLS 模型得到的预测集的相关系数和均方根差分别为 0.83 与 2.30, 相比较 GA-PLS 算法更简便稳定, 同时提高了模型的预测精度, 为以后高光谱技术的脐橙叶片叶绿素定量分析提供了良好的依据。

References

- [1] ZHANG Jin-heng, WANG Ke, WANG Ren-chao(张金恒, 王珂, 王人潮). Chinese Journal of Shanghai Jiaotong University(上海交通大学学报), 2003, 21(1): 74.
- [2] TONG Gui-feng, CHEN Zhi-fang, FAN Ying(童桂凤, 陈志芳, 范莹). The Administration and Technique of Environmental Monitoring(环境监测管理和技术), 2012, 24(1): 53.
- [3] Ling Q H, Huang W H, Jarvis P. Photosynthesis Research, 2011, 108(1): 89.
- [4] SHI Ji-yong, ZOU Xiao-bo, ZHAO Jie-wen(石吉勇, 邹小波, 赵杰文, 等). Chinese Journal of Analytical Chemistry(分析化学), 2011, 39(2): 243.
- [5] PENG Yan-kun, HUANG Hui, WANG Wei, et al(彭颜昆, 黄慧, 王伟, 等). Journal of Jiangsu University(江苏大学学报), 2011, 32(2): 125.
- [6] FANG Li-min, LIN Min(方利民, 林敏). Chinese J. Anal. Chem. (分析化学), 2008, 36(6): 815.
- [7] Leardi R, González A L. Chemometrics and Intelligent Laboratory Systems, 1998, 41: 195.
- [8] YAN Yan-lu, ZHAO Long-lian, HAN Dong-hai(严衍禄, 赵龙莲, 韩东海). The Foundation and Application of Near Infrared Spectroscopy Analysis(近红外光谱分析基础与应用). Beijing: China Light Industry Press(北京: 中国轻工业出版社), 2005.
- [9] Araujo M C U, Saldanha T C B, Galvao R K H, et al. Chemometrics and Intelligent Laboratory Systems, 2001, 57: 65.
- [10] QIAN Hai-bo, SUN Lai-jun, WANG Le-kai, et al(钱海波, 孙来军, 王乐凯, 等). Chinese Agricultural Science Bulletin(中国农学通报), 2011, 27(18): 51.
- [11] SONG Sha-lei, LI Ping-xiang, GONG Wei, et al(宋沙磊, 李平湘, 龚威, 等). Geomatics and Information Science of Wuhan University (武汉大学学报信息科学版), 2010, 35(2): 219.

Analysis of Chlorophyll in Gannan Navel Orange with Algorithm of GA and SPA Based on Hyperspectral

LIU Yan-de, ZHANG Guang-wei, CAI Li-jun

Institute of Mechanical and Electrical Machinery Technology and Application Research Institute, East China Jiaotong University, Nanchang 330013, China

Abstract The hyperspectral effective information of Gannan navel orange was extracted by genetic algorithm (GA) and successive projections algorithm (SPA) for partial least squares (PLS) model. Average spectral was extracted from region of interest (ROI) of hyperspectral images after preprocessing. GA and SPA were used to select 27 and 8 characteristic wavelengths for chlorophyll model with PLS. The correlation coefficients and rms error of GA-PLS were 0.80 and 2.45, and the correlation coefficients and rms error of SPA-PLS were 0.83 and 2.30, respectively. Overall results sufficiently demonstrate that SPA-PLS model has a greater advantage which can be combined with hyperspectral technique to be a nondestructive and rapid analytical method.

Keywords Chlorophyll content; Hyper-spectral imaging; Gannan navel orange; GA; SPA

(Received May 26, 2012; accepted Aug. 8, 2012)