

基于近红外高光谱图像的黄瓜叶片色素含量快速检测*

邹小波 陈正伟 石吉勇 黄晓玮 张德涛

(江苏大学食品与生物工程学院, 镇江 212013)

【摘要】 利用高光谱图像技术和高效液相色谱法(HPLC)快速检测了新鲜黄瓜叶中叶绿素 a、叶绿素 b、β-胡萝卜素和叶黄素 4 种色素含量。采集了 120 片黄瓜叶的近红外高光谱图像数据以及用 HPLC 精确测定黄瓜叶中色素含量;提取高光谱图像中 50×50 像素感兴趣区域(ROI)的平均光谱与 4 种色素含量分别建立偏最小二乘(PLS)预测模型;为了提高模型的稳定性和预测精度,分别采用区间偏最小二乘(iPLS)、向后区间偏最小二乘(BiPLS)和联合区间偏最小二乘(SiPLS)对各种色素对应的特征波段进行优选,同时对光谱划分数进行了优化。结果表明 BiPLS 和 SiPLS 对应模型的预测效果较好,对叶绿素 a、叶绿素 b、β-胡萝卜素和叶黄素 4 种色素的预测集相关系数 R_p 分别为 0.825 7、0.813 4、0.811 6、0.826 2。

关键词: 黄瓜叶片 色素 高光谱图像 高效液相色谱 偏最小二乘法

中图分类号: TP391.41; Q94-3 文献标识码: A 文章编号: 1000-1298(2012)05-0152-05

Rapid Detection of Cucumber Leaves Pigments Based on Near Infrared Hyper-spectral Image Technology

Zou Xiaobo Chen Zhengwei Shi Jiyong Huang Xiaowei Zhang Detao

(School of Food and Biological Engineering, Jiangsu University, Zhenjiang 212013, China)

Abstract

Rapid detection the content of chlorophyll a, chlorophyll b, β-carotene and lutein on fresh cucumber leaves by using hyper-spectral image and high performance liquid chromatography (HPLC) was done. One hundred and twenty hyper-spectral images of cucumber leaves were collected by near infrared hyper-spectral camera, then using HPLC to accurate detect the content of chlorophyll a, chlorophyll b, β-carotene and lutein. After hyper-spectral images have been corrected, a 50×50 pixels region as region of interest (ROI) was defined. Average spectrum from ROI was extracted and prediction model of partial least squares (PLS) with the content of four pigments was built. In order to advance the stability and prediction accuracy, interval partial least squares (iPLS), backwards interval partial least squares (BiPLS) and synergy interval partial least squares(SiPLS) algorithm was used to select diagnostic bands of each pigments after optimized the number of spectrum interval. The result showed that using BiPLS and SiPLS would get the best PLS model, the best prediction coefficient of chlorophyll a, chlorophyll b, β-carotene and lutein is 0.825 7, 0.813 4, 0.811 6 and 0.826 2, respectively.

Key words Cucumber leaves, Pigment, Hyper-spectral image, High performance liquid chromatography, Partial least squares

引言

作物叶片色素含量与作物的光合能力、营养状

况和生长发育之间存在较好的相关性,是评价作物生长情况和营养状况的一个可靠指标^[1]。作物色素检测一般采用有机溶剂将组织中的色素溶出,然

收稿日期: 2011-07-26 修回日期: 2011-08-04

* 国家自然科学基金资助项目(60901079)和国家高技术研究发展计划(863 计划)资助项目(2008AA10Z208)

作者简介: 邹小波,教授,主要从事农产品、食品无损检测研究,E-mail: zou_xiaobo@ujs.edu.cn

后用分光光度计或高效液相色谱 (HPLC) 进行测量^[2-3]。分光光度法操作简便、快速;HPLC 具有分辨率高、重复性好、结果精确等优点,但化学检测法会破坏检测对象,难以实现现场快速检测。为了克服化学检测方法的不足,许多学者提出了色素含量的无损检测方法^[4-5],其中高光谱图像技术是图像处理技术和光谱分析技术在低技术层面的融合技术,兼有这两种技术的优势,既能对研究对象的外部特征进行可视化分析,也能对内部有效成分进行定量预测。这些研究所采用的高光谱波段多集中在可见光区域,基于近红外波段的研究较少。

本文拟采用近红外高光谱图像和 HPLC 同时对新鲜黄瓜叶中叶绿素 a、叶绿素 b、β-胡萝卜素和叶黄素进行快速检测,并通过测定两模型的相关系数验证近红外高光谱图像无损检测黄瓜叶片色素含量的可行性。为提高检测的精度和稳定性,采用 iPLS、BiPLS 和 SiPLS 对各个色素对应的特征波段进行优选,并用 PLS 建立特征波段对应的预测模型。

1 材料与方法

1.1 实验材料

实验所用黄瓜叶片于 2010 年 11 月采自江苏大学温室大棚,品种为碧玉三号水果黄瓜。随机采集黄瓜老叶、中叶、嫩叶共 120 片叶片作为实验样本,随机挑取 90 片作为校正集样本,剩余 30 片组成预测集样本。

1.2 仪器与试剂

高光谱图像采集系统如图 1 所示,由近红外光谱仪 (ImSpector VI7E 型, Spectra Imaging, 芬兰)、高光谱摄像机 (XEVA-FPA-1.7-320 型, XenICs, 比利时)、光纤卤素灯 (2900 型, Illumination Technologies, 美国)、精密电控平移台 (Zolix TS200AB 型, Zolix, 北京) 和计算机等部件组成。

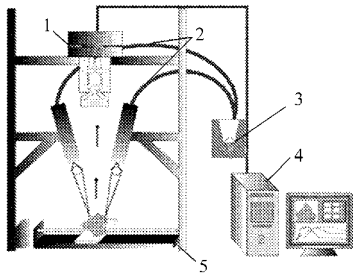


图 1 高光谱图像系统

Fig. 1 Hyper-spectral imaging system

1. 摄像机 2. 光纤 3. 光源 4. 计算机 5. 输送装置

高效液相色谱检测系统包括:LC-20A 型岛津高效液相色谱仪 (SHIMAD-ZU 型, 日本);C18 反相色谱柱 (Spherisorb ODS-2 型, 250 × 4.6 mm,

5 μm); 紫外可见检测器 (Prominence SPD-20A 型);梯度系统 (LC-20AT 型)。

主要化学试剂:叶绿素 a 标准品 (美国 Sigma 公司);甲醇 (色谱纯);乙酸乙酯、丙酮和三乙胺 (分析纯, 国药集团化学试剂有限公司);二丁基羟基甲苯 (国药集团化学试剂有限公司)。

1.3 高光谱数据采集

为了防止基线漂移,数据采集前将高光谱图像采集系统打开,预热 30 min。高光谱数据采集前需进行黑白标定^[6]。在与样品采集相同的条件下,采集标准白色校正板得到全白的标定图像 W , 关闭相机快门采集得到全黑的标定图像 B 。此后采集得到的绝对图像 I 变成相对图像 R 完成标定,即

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

将黄瓜叶片平铺在白色底板的输送台上进行高光谱图像采集。图像扫描的行数根据叶片实际大小确定,每行扫描的像素数为 320,光谱范围是 871 ~ 1 766 nm,采样间隔为 4.4 nm,采集得到 256 个波长下的高光谱图像数据块。获取叶片高光谱图片后立即用 HPLC 测定色素含量。

1.4 HPLC 检测

准确剪取黄瓜叶片上的感兴趣区域叶片 30 ~ 40 mg 置于研钵中,加入适量液氮进行冷冻干燥并研磨成粉。加入 4 mL 体积分数为 80% 的丙酮水溶液 (含质量分数 0.01% 二丁基羟基甲苯) 进行色素提取,随后将提取液倒入离心管中。色素提取步骤重复一次至叶片粉末成无色。离心管搅拌摇匀后静置 1 min,将样品放入离心分离机中以 4 000 r/min 的速度离心 10 min,收集上清液倒入 10 mL 棕色容量瓶中并用 80% 丙酮溶液定容。取约 1 mL 萃取液,经 0.45 μm 微膜过滤器过滤,滤液装入棕色色谱瓶中进行 HPLC 分析。以上所有程序都在昏暗的橘色光下进行以防止色素见光分解。

HPLC 实验参考文献[3]进行并适当优化,根据文献对色谱图中色素成分定性。流动相为 80% 甲醇水溶液 (含体积分数 0.05% 三乙胺) 和 100% 乙酸乙酯溶液 (含体积分数 0.05% 三乙胺) 体系,研究采用梯度洗脱,可在 30 min 内完成样品测定,梯度洗脱条件如表 1 所示。进样量: 25 μL,柱温: (30 ± 0.5) °C,检测波长: 440 nm。

实验采用叶绿素 a 标准曲线作为外标测量叶绿素 a、叶绿素 b、β-胡萝卜素和叶黄素的含量。1 mg 叶绿素 a 标准品 (美国 Sigma 公司) 溶于 80% 丙酮水溶液 (含 0.01% 二丁基羟基甲苯) 并配置成不同质量浓度,以各组分的色谱面积 Y 对质量浓度 X 作线性回归得标准曲线。叶绿素 a 标准曲线方程为

$$Y = 28\,590X - 36\,446 \quad (R^2 = 0.995) \quad (2)$$

表 1 HPLC 流动相梯度洗脱条件

Tab.1 HPLC mobile phase gradient program for the analysis of pigment			%
时间/min	80% 甲醇水溶液 体积分数	100% 乙酸乙酯 体积分数	
0	70	30	
5	40	60	
25	0	100	
30	0	100	

1.5 不同区间偏最小二乘算法

iPLS^[7] 优选区间基本原理为:将全光谱均匀划分为 n 个子区间,对每个子区间分别建立 PLS 模型,最优模型对应的区间为入选区间;SiPLS^[8] 优选区间基本原理为:将全光谱均匀划分为 n 个子区间后,联合其中 m 个区间进行 PLS 建模,建模时按排列组合的思想对 C_n^m 种组合方式分别进行考察,挑选最优模型对应的 m 个子区间组合作为入选区间,由于此方法的计算量非常大,因此一般取 $m < 5$;BiPLS^[9] 优选区间基本原理为:将全光谱均匀划分为 n 个子区间,每次剔除一个区间使得剩余区间对应的 PLS 模型最优,如此重复直至剩余 1 个子区间,全局最优 PLS 模型对应的区间组合为入选区间。

从 iPLS、BiPLS、SiPLS 优选区间的基本原理可以看出,光谱子区间数 n 是一个重要参数,决定了每个子区间的窗口宽度。子区间划分数太少,光谱的特征区间可能会包含过多的干扰噪声;子区间数过多时,特征光谱区间建立的色素模型预测精度也不再提高。且色素分子式比较复杂,难以确定其特征区间的宽度,因此本文对光谱子区间数 $n \in [10, 24]$ 范围内进行了优化,综合考虑校正集和预测集样本实测值与预测值的相关系数 R_C 和 R_P 、交互验证均方根误差 (RMSECV) 及预测均方根误差 (RMSECP) 以确定最佳子区间划分数。

1.6 软件

高光谱数据采集和处理软件分别为 SpectraCube (Auto Vision, 美国) 和 ENVI 4.5 (ITT Visual Information Solutions, 美国)。光谱数据分析软件为 Matlab 7.4 (MathWorks, 美国)。

2 实验结果与分析

2.1 样品测定结果

梯度洗脱控制得当,黄瓜叶片中各主要色素可以得到较好的分离,液相色谱图如图 2 所示。根据色谱图得出叶绿素 a、叶绿素 b、 β -胡萝卜素和叶黄素的峰面积,再由标准曲线回归得出其质量分数作

为该样品的化学描述值。所有样品的 HPLC 测定统计结果如表 2 所示。

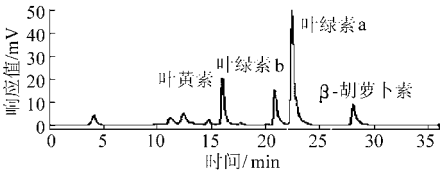


图 2 黄瓜叶片色素的 HPLC 图

Fig.2 HPLC of pigment at cucumber leaves

表 2 叶片色素含量统计

Tab.2 Descriptive statistics of pigment for leaves					μg/mg
色素	最大值	最小值	均值	方差	
叶绿素 a	1.465 8	0.026 9	0.656 1	0.332 2	
叶绿素 b	0.454 8	0.024 3	0.213 5	0.095 6	
β-胡萝卜素	0.311 8	0.024 3	0.147 7	0.069 8	
叶黄素	0.493 6	0.040 3	0.278 1	0.106 3	

2.2 叶片光谱信息提取与预处理

用 HPLC 测定的黄瓜叶片色素是叶片取样区域色素的均值,因此应提取高光谱图像中色素浓度测量区域的平均光谱进行分析。图 3 为叶片 HPLC 取样区域和该区域 120 个样品的原始光谱图。在操作中,尽量逼近取样区域,取 50×50 像素的矩形区域的平均光谱。光谱在 $1\,722.7\text{ nm}$ 以上存在明显噪声,若将此部分也加入到模型中将降低整个模型的精度,应舍弃。光谱线由于受到黄瓜叶片表面不均匀的影响而比较分散,采用标准正态变量法 (SNV) 来消除表面散射的影响。经 SNV 处理后的光谱曲线仍然会有部分噪声,为消除噪声对之后色素浓度模型带来的影响,采用 Savitzky - Golay 滤波法^[10] 使光谱曲线更加均匀,点数取 7。预处理后的光谱曲线如图 4 所示。

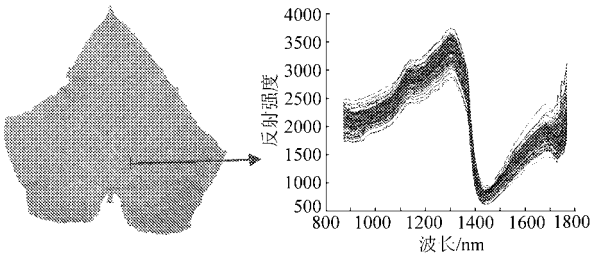


图 3 黄瓜叶片高光谱图像感兴趣区域的光谱曲线

Fig.3 Spectral profile from hyperspectral image of ROI region in cucumber leaves

2.3 特征波段选择及建模

用交互验证法确定主成分数和筛选子区间,以校正集和预测集样本实测值与预测值的相关系数 R_C 和 R_P 、交互验证均方根误差 RMSECV 及预测均方根误差 RMSECP 作为评价各种建模方法的有效

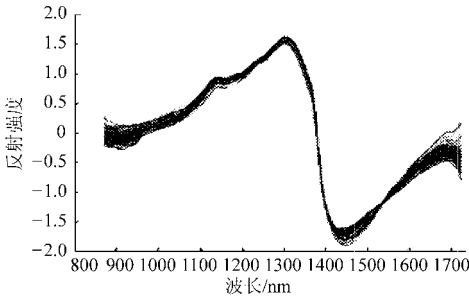


图 4 预处理后光谱曲线
Fig. 4 Spectral after pretreatment

指标。为了更准确寻找 4 种色素的特征波段,将整个光谱区域划分为 10 ~ 24 个子区间并分别应用 iPLS、BiPLS 和 SiPLS 建立 PLS 模型。以叶绿素 a 为例,图 5 显示了不同子区间数目下通过 BiPLS 挑选特征波段后建立的 PLS 模型的 R_p ,当将整个光谱划分为 14 个子区间时值最大。

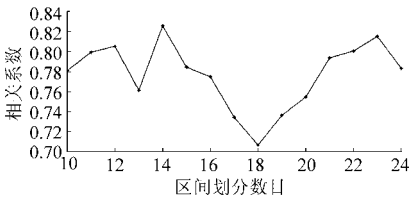


图 5 不同区间划分数目下的相关系数曲线
Fig. 5 Effect of interval number

表 3 列出了叶绿素 a、叶绿素 b、 β -胡萝卜素和叶黄素 4 种色素全光谱 PLS 模型及通过 iPLS、BiPLS 和 SiPLS 挑选特征波段后建立的 PLS 模型结果。由表 3 可见,叶绿素 a 和叶绿素 b 通过 BiPLS 或 SiPLS 挑选特征波段建立的模型效果最好,对应的光谱划分区间分别为 14 和 21。叶绿素 a 挑选第 1、8、13、14 个区间(波长范围为 871 ~ 931.8 nm、1 296.9 ~ 1 357.8 nm和 1 601 ~ 1 722.7 nm)建立 PLS 模型。叶绿素 b 建模入选区间为第 2、15、19、20、21 个区间(波长范围为 911.6 ~ 952.1 nm、1 438.8 ~ 1 479.4 nm和 1 601 ~ 1 722.7 nm)。 β -胡萝卜素在光谱划分为 23 个区间时通过 BiPLS 挑选第 1、7、18、21、22 区间(波长范围为 871 ~ 908.1 nm、1 093.2 ~ 1 130.2 nm、1 500.5 ~ 1 537.5 nm和 1 611.61 ~ 1 685.7 nm)建立的模型效果最好。叶黄素应用 SiPLS 在光谱划分为 24 个区间并联合第 1、10、23 和 24 区间(波长范围为 871 ~ 906.5 nm、1 190.4 ~ 1 225.9 nm、1 651.72 ~ 1 722.7 nm)建立的 PLS 模型预测性能最佳。

色素的全光谱 PLS 和 iPLS 模型结果与联合区间(BiPLS 和 SiPLS)建立的 PLS 模型相比结果较差,且有拟合不足的情况出现。这是因为全光谱波段中可能包含影响色素 PLS 模型精度的其他化合物的光谱信息或光谱噪声,iPLS 模型只有一个区间的光

谱信息,反映的色素信息不足;而 BiPLS 和 SiPLS 采用的是挑选几个子区间联合建模的方法,在保证模型精度的情况下又剔除了干扰过大的变量区域,同时,可以克服单个区间包含的光谱信息不足的问题,使最终建立模型的预测能力和精度更高。通过挑选特征光谱区间提高 PLS 模型精度这一结果亦与现有的文献结论一致^[8,11]。

表 3 4 种色素 PLS、iPLS、BiPLS 和 SiPLS 模型校正集与预测集结果

Tab. 3 Results of PLS, iPLS, BiPLS and SiPLS models

色素	模型	R_c	RMSECV	R_p	RMSECP
叶绿素 a	全光谱 PLS	0.798 5	0.201 7	0.782 4	0.216 2
	iPLS	0.733 8	0.226 6	0.632 6	0.260 5
	BiPLS	0.833 4	0.185 5	0.825 7	0.222 7
	SiPLS	0.802 8	0.200 2	0.789 0	0.218 3
	(3 区间联合)	0.833 4	0.185 5	0.825 7	0.222 7
	(4 区间联合)	0.833 4	0.185 5	0.825 7	0.222 7
叶绿素 b	全光谱 PLS	0.678 4	0.070 6	0.833	0.059 5
	iPLS	0.692 0	0.068 8	0.698 3	0.071 6
	BiPLS	0.814 5	0.055 3	0.813 4	0.060 6
	SiPLS	0.814 5	0.055 4	0.813 4	0.060 6
	(3 区间联合)	0.817 0	0.055 3	0.816 8	0.062 2
	(4 区间联合)	0.817 0	0.055 3	0.816 8	0.062 2
β -胡萝卜素	全光谱 PLS	0.737 0	0.048 0	0.808 4	0.041 8
	iPLS	0.691 7	0.050 5	0.689 5	0.049 9
	BiPLS	0.820 3	0.040 6	0.811 6	0.041 7
	SiPLS	0.817 0	0.040 7	0.776 6	0.047 7
	(3 区间联合)	0.810 7	0.041 2	0.810 6	0.043 0
	(4 区间联合)	0.810 7	0.041 2	0.810 6	0.043 0
叶黄素	全光谱 PLS	0.780 1	0.068 6	0.786 4	0.075 0
	iPLS	0.662 2	0.080 1	0.662 6	0.080 2
	BiPLS	0.810 9	0.061 9	0.810 4	0.068 7
	SiPLS	0.788 2	0.065 1	0.774 1	0.070 4
	(3 区间联合)	0.830 3	0.058 7	0.826 2	0.064 9
	(4 区间联合)	0.830 3	0.058 7	0.826 2	0.064 9

3 结束语

应用高光谱技术获取黄瓜叶片感兴趣区域的光谱信息,再利用 HPLC 测得相应区域的叶绿素 a、叶绿素 b、 β -胡萝卜素和叶黄素 4 种色素的含量,然后通过联合区间(BiPLS 和 SiPLS)挑选特征波段建立叶片光谱与色素含量的 PLS 预测模型,模型对应的预测集相关系数 R_p 分别为 0.825 7、0.813 4、0.811 6 和 0.826 2,这说明高光谱图像技术可无损、快速分析黄瓜叶片的叶绿素 a、叶绿素 b、 β -胡萝卜素和叶黄素含量。

参 考 文 献

- 1 Filella I, Penuelas J. Red edge position and shapes as indicators of plant chlorophyll content, biomass and hydric status[J]. International Journal of Remote Sensing, 1994, 15(7): 1 459 ~ 1 470.
- 2 王福民, 黄敬峰, 王秀珍. 水稻叶片叶绿素、类胡萝卜素含量估算的归一化色素指数研究[J]. 光谱学与光谱分析, 2009, 29(4): 1 064 ~ 1 068.
Wang Fumin, Huang Jingfeng, Wang Xiuzhen. Normalized difference ratio pigment index for estimating chlorophyll and carotenoid contents of in leaves of rice [J]. Spectroscopy and Spectral Analysis, 2009, 29(4): 1 064 ~ 1 068. (in Chinese)
- 3 Caldwell C R, Britz S J. Effect of supplemental ultraviolet radiation on the carotenoid and chlorophyll composition of green house-grown leaf lettuce (*Lactuca sativa* L.) cultivars [J]. Journal of Food Composition and Analysis, 2006, 19(6 ~ 7): 637 ~ 644.
- 4 汤旭光, 宋开山, 刘殿伟, 等. 基于可见/近红外反射光谱的大豆叶绿素含量估算方法比较[J]. 光谱学与光谱分析, 2011, 31(2): 371 ~ 374.
Tang Xuguang, Song Kaishan, Liu Dianwei, et al. Comparison of methods for estimating soybean chlorophyll content based on visual/near infrared reflection spectra [J]. Spectroscopy and Spectral Analysis, 2011, 31(2): 371 ~ 374. (in Chinese)
- 5 王伟, 彭彦昆, 马伟, 等. 冬小麦叶绿素含量高光谱检测技术[J]. 农业机械学报, 2010, 41(5): 172 ~ 177.
Wang Wei, Peng Yankun, Ma Wei, et al. Prediction of chlorophyll content of winter wheat using leaf-level hyperspectral data [J]. Transactions of the Chinese Society for Agricultural Machinery, 2010, 41(5): 172 ~ 177. (in Chinese)
- 6 陈全胜, 赵杰文, 蔡健荣, 等. 利用高光谱图像技术评判茶叶的质量等级[J]. 光学学报, 2008, 28(4): 669 ~ 674.
Chen Quansheng, Zhao Jiewen, Cai Jianrong, et al. Estimation of tea quality level using hyperspectral imaging technology [J]. Acta Optica Sinica, 2008, 28(4): 669 ~ 674. (in Chinese)
- 7 Nørgaard L, Saudland A, Wagner J, et al. Interval partial least squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy [J]. Applied Spectroscopy, 2000, 54(3): 413 ~ 419.
- 8 Chen Q S, Zhao J W, Liu M H, et al. Determination of total polyphenols content in green tea using FT-NIR spectroscopy and different PLS algorithms [J]. Journal of Pharmaceutical and Biomedical Analysis, 2008, 46(3): 568 ~ 573.
- 9 Zou X B, Zhao J W, Li Y X. Selection of the efficient wavelength regions in FT-NIR spectroscopy for determination of SSC of 'Fuji' apple based on BiPLS and FiPLS models [J]. Vibrational Spectroscopy, 2007, 44(2): 220 ~ 227.
- 10 王前, 束炯, 尹球. 高光谱图像光谱域噪声检测与去除的 DSGF 方法[J]. 红外与毫米波学报, 2006, 25(1): 29 ~ 32.
Wang Qian, Shu Jiong, Yin Qiu. DSGF method on detecting and removing spectral noise of hyperspectral image [J]. Journal of Infrared and Millimeter Waves, 2006, 25(1): 29 ~ 32. (in Chinese)
- 11 Zou Xiaobo, Zhao Jiewen, Malcolm J W P, et al. Variables selection methods in near-infrared spectroscopy [J]. Analytica Chimica Acta, 2010, 667(1 ~ 2): 14 ~ 32.

(上接第 167 页)

- 17 关海鸥, 黄燕. 大豆病斑智能识别无损预处理及其特征提取方法的研究[J]. 河北农业大学学报, 2011, 33(5): 123 ~ 128.
Guan Haiou, Huang Yan. Study on the method of non-loss preprocessing and feature extraction for intelligent recognition of soybean diseased spots[J]. Journal of Agricultural University of Hebei, 2011, 33(5): 123 ~ 128. (in Chinese)
- 18 关海鸥, 许少华, 谭峰. 基于遗传模糊神经网络的植物病斑区域图像分割模型[J]. 农业机械学报, 2010, 41(11): 163 ~ 168.
Guan Haiou, Xu Shaohua, Tan Feng. Image segmentation model of plant lesion based on genetic algorithm and fuzzy neural network [J]. Transactions of the Chinese Society for Agricultural Machinery, 2010, 41(11): 163 ~ 168. (in Chinese)
- 19 赵小勇, 付强, 邢贞相, 等. 投影寻踪模型的改进及其在生态农业建设综合评价中的应用[J]. 农业工程学报, 2006, 22(5): 222 ~ 225.
Zhao Xiaoyong, Fu Qiang, Xing Zhenxiang, et al. Improvement of projection pursuit model and the application in integrated evaluation of eco-agricultural construction [J]. Transactions of the CSAE, 2006, 22(5): 222 ~ 225. (in Chinese)