

文章编号: 0253-2239(2008)08-1618-05

# 作物和杂草叶片的可见-近红外反射光谱特性

吴迪<sup>1</sup> 黄凌霞<sup>2</sup> 何勇<sup>1</sup> 潘家志<sup>1</sup> 张贇<sup>1</sup>  
(<sup>1</sup> 浙江大学生物系统工程与食品科学学院, 浙江 杭州 310029)  
(<sup>2</sup> 浙江大学动物科学学院, 浙江 杭州 310029)

**摘要** 为了进行快速实时的杂草识别,研究了作物和杂草叶片的可见-近红外反射光谱特性。选择了两种常见的田间作物大豆(*Glycine max*)和玉米(*Zea mays*),以及铁苋菜(*Acalypha australis* L.)和田字草(*Marsilea quadrifolia* L.)两种杂草作为研究对象,每种各 30 个样本,共 120 个样本。采用 ASD Fieldspec 便携式光谱仪进行光谱采集。在对 400~1000 nm 的光谱数据进行平滑和一阶求导预处理,通过主成份分析,去除了一个奇异样本。最后用 79 个样本组成的建模集进行偏最小二乘法建模,对剩余的 40 个样本进行预测。预测模型结果的相关性达到 0.986,识别率达到 100%。说明研究中选用的作物和杂草叶片的可见-近红外反射光谱特性之间有明显的区别,可以用于进行杂草和作物的区分。

**关键词** 光谱学; 杂草识别; 可见-近红外反射光谱; 偏最小二乘法

**中图分类号** S123; TH744.1 **文献标识码** A **doi:** 10.3788/AOS20082808.1618

## Visible-Near Infrared reflection Spectroscopy for Crop-Weed Discrimination

Wu Di<sup>1</sup> Huang Lingxia<sup>2</sup> He Yong<sup>1</sup> Pan Jiazhi<sup>1</sup> Zhang Yun<sup>1</sup>

(<sup>1</sup> College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, Zhejiang 310029, China)  
(<sup>2</sup> College of Animal Sciences, Zhejiang University, Hangzhou, Zhejiang 310029, China)

**Abstract** Visible and near infrared reflection spectroscopy (Vis-NIRS) was studied as a fast and promising techniques for weed detection. Two kinds of weeds [threeseeded mercury (*Acalypha australis* L.) and quatrefoil duckweed (*Marsilea quadrifolia* L.)] and two kinds of crops [soybean (*Glycine mas*) and corn (*Zea mays*)] were investigated. A total 120 samples (30 for each species) of the plant leaves were prepared. The spectra were acquired by an ASD Fieldspec handheld spectrometer. After data pre-processing and principle component analysis on spectra from 400 to 1000 nm, one influential outlier was detected. After full-cross calibration of the 79 samples using partial least squares regression, the remaining 40 samples were predicted by the established model. Correlation coefficient of prediction set was 0.987. 100% recognition rate was obtained based on artificial thresholds. So, Vis-NIR spectroscopy is an available alternative for recognizing crops from weeds.

**Key words** spectroscopy; weed recognition; visible-near infrared reflection spectroscopy; partial least squares

## 1 引言

传统的单一的除草剂使用方式<sup>[1,2]</sup>导致了过多的化学药剂使用和农产品中的残留。在作物生长早期对杂草有选择性地使用除草剂,在经济上和环境中都是十分重要,关键在于如何快速识别杂草。

可见-近红外光谱分析技术,可以充分利用全波段或多波长下的光谱数据进行定性或定量分析,是一种快速、无损、低成本、无污染的分析技术,已广泛

应用于农业、食品、医学等领域<sup>[3~8]</sup>。Zwiggelaar 的研究指出,可以用杂草和作物在特定波段内的光谱反射率来区分它们,并有几项研究重点讨论了不同种类植物叶片的光学特性<sup>[9]</sup>。Koger 等介绍了用小波方法分析高光谱反射率信号,以检测豆苗地中杂草,并且达到了 83% 的正确率<sup>[10]</sup>。Smith 和 Blackshaw 试验了 7 种离体植物在 350~2500 nm 的反射率,用高光谱遥感识别这 5 种作物和 2 种杂

收稿日期: 2007-10-08; 收到修改稿日期: 2008-02-05

基金项目: 国家 863 计划(2007AA10Z210)和国家自然科学基金(30671213)资助课题。

作者简介: 吴迪(1984-),男,博士研究生,主要从事光谱与多光谱检测技术、数字农业等方面的研究。

E-mail: ilkb8mvp@yahoo.com.cn

导师简介: 何勇(1963-),男,教授,博士生导师,主要从事光谱与多光谱检测技术、数字农业等方面的研究。

E-mail: yhe@zju.edu.cn

草<sup>[11]</sup>。Slaughter 等用 400~2500 nm 的可见-红外波段识别西红柿和茄属杂草,建立了一个可见光区的窄波段的光谱识别模型,正确率达到 95%<sup>[12]</sup>。但是,除了图像处理技术外,国内使用可见-近红外反射光谱检测杂草的研究不多见<sup>[13]</sup>。所以,使用可见-近红外反射光谱技术以区分作物和杂草,开发可以在农田中使用的检测设备是很有前途的、可行的方法。

本文研究了 4 种作物与杂草的可见-近红外反射光谱特性。并应用偏最小二乘回归法(Partial least squares, PLS)进行了种类识别。探究是否能够通过作物和杂草叶片的可见-近红外反射光谱特性的不同进行区分。

## 2 材料与方法

### 2.1 植物与样品准备

作物样品选用中国东南部常见的豆苗(*Glycine max*)和玉米(*Zea mays*)。铁苋菜(*Acalypha australis* L.)和田字草(*Marsilea quadrifolia* L.)则是两种中国东南部最常见的杂草。实验于 2005 年 5 月 20 日在浙江大学华家池校区的实验田中进行。由于田间的豆苗、玉米、铁苋菜和田字草的冠层上方,顶端展开叶占据了主要范围。作物或杂草的其它部分的光谱信息很难被采集到。通过采集这些植物顶端的叶片能够代表其特征。由于外界环境会对光谱采集产生影响,为了研究不同对象间的光谱特征差异,需准确地研究作物和杂草的光谱特性。将叶片采集后立即到室内完成光谱采集。采集好的叶片样本先保存在密封的塑料薄膜口袋中,放入装有冰块盒子,然后拿到实验室保存在(4±1)℃的低温环境中。在获取光谱之前,所有叶片先恢复到 25℃的室温。为保持叶片的水分,整个光谱采集在叶片采摘后 3 h 内完成,减少光谱变化。

### 2.2 叶片光谱反射率测定

实验使用美国 Analytical Spectral Device 公司的 Handheld FieldSpec 光谱仪,光谱测定范围 325~1075 nm,采样间隔为 1.5 nm。光谱的采集操作通过 FieldSpec HH RS2 操作软件在 Windows XP 操作系统进行,数据按间隔 1 nm 保存。由于外界复杂的环境会对光谱采集造成一定的影响,因此选用在室内避光的环境内进行。先把叶片样本平整地铺在一个直径为 120 mm 的玻璃皿中,要求尽量无重叠地铺满整个底面。这使得样本之间有相对一致的照明环境。由于叶片完全覆盖玻璃皿底面,玻璃皿底面对叶片反射光谱采集的影响可忽略。铺满

玻璃皿所需的叶片数量主要由叶片的面积大小决定。对每种植物采集 30 个样本,四种植物总共有 120 个样本。采集 325~1075 nm 波段范围的光谱反射率。光谱采集时,光谱仪探头和样本中心表面的距离为 14 cm。采集过程中,采用 Lowell prolam 公司(大阪、日本)的 14.5 V Bulb/128690 卤素灯作为光源。卤素灯与样本中心间的距离设定为 30 cm。光谱仪和卤素灯之间的夹角大约是 45°。整个光谱采集前通过一个面积约为 65 cm<sup>2</sup>的圆形 BaSO<sub>4</sub> 白盘进行光谱平衡校正。每个样本采集 3 个方向,方向间夹角 120°,以减少由于入射角差异带来的影响。每个方向采集 10 次,每个样本一共 30 次。旋转过程中不要移动玻璃皿中心点的位置。3 个方向光谱的均值作为该样本的反射光谱值。

### 2.3 近红外数据分析

由于仪器采集精度限制,在采集到的光谱数据中首端与末端的噪声较大。因此,将 325~399 nm 和 1001~1075 nm 的光谱数据去除,以提高测量数据的信噪比。所有的光谱分析都是采用 400~1000 nm 波段的数据。所有的反射率数据都通过取反求对数的方法[lg(1/R)]转化为吸光度数据。

为了降低由系统硬件带来的随机误差,所有的光谱吸光度数据都用 Savitzky-Golay 最小平方拟合进行平滑。然后用一阶 Savitzky-Golay 差分平滑光谱数据,消除基线漂移和背景噪声<sup>[14]</sup>带来的干扰。考虑到叶片样本上表面相对平坦,多重漫反射的影响相对较弱,可以忽略不计。

首先用主成份分析(PCA)方法来从光谱数据中产生 20 个主成份,以检验数据的相关性和剔除部分数据<sup>[15]</sup>。然后用一阶差分后的 400~1000 nm 光谱数据进行完全交叉验证的偏最小二乘回归法(PLS)以区分四种植物。随机选取 66%的样本进行模型校正,剩下的数据用来评估模型预测误差。所以,每个品种随机选取 20 个样本(共 80 个)用来作为建模数据集合,而剩下的共 40 个样本用来验证回归模型的有效性。所有的计算过程都在 Unscrambler (version 9.2, CAMO Process AS, Norway)统计软件上进行。

## 3 结论与讨论

### 3.1 植物的可见-近红外反射光谱

图 1 显示了每种植物在 400~1000 nm 的吸收光谱曲线。在“红边”<sup>[16]</sup>(680~740 nm)位置各种植物的光谱曲线发生了重叠。而在绿光区域(560~

600 nm, 玉米的光谱曲线与其他植物有一定的差异)能将玉米与其他三种植物区分开来。而蓝光区域(400~560 nm)、红光区域(600~680 nm)及近红外区域(740~1000 nm)的差异可用于进行作物和杂草的区分。然而当样本曲线增加时, 各类植物的光谱曲线会交叉重叠, 很难通过直接观察光谱曲区分作物和杂草。因此利用 PLS 模型进行光谱分析, 研究了作物和杂草的叶片反射光谱区特性。

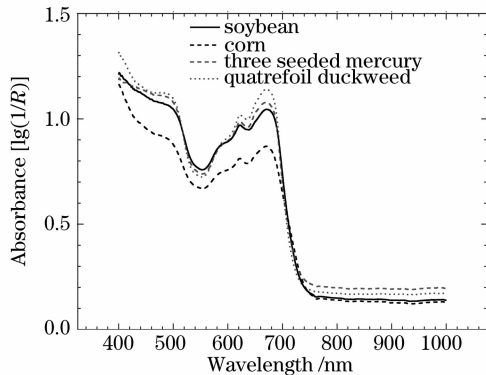


图1 大豆、玉米、铁苋菜、田字草样本的可见-近红外吸收光谱

Fig.1 Visible-near infrared absorption spectra for soybean, corn, three-seeded mercury and quatrifoil duckweed

## 3.2 统计分析

### 3.2.1 主成份分析

为了检测出奇异样本和在不同品种间作定性分析, 对训练集中的 80 个光谱样本在 400~1100 nm 之间的数据做了 PCA 分析。结果显示, 主成分 (Principal component), PC1 和 PC2 分别代表了 78% 和 13% 的分布差异性 (占总体的 91%)。而剩下的因子所占比例小得多。所以, 前两个主成份就近似于植物叶片的光谱特征描述。图 2 表示了 PCA 的聚类图 (PC1 和 PC2) 和四个品种的聚类。它可以解释植物品种的光谱反射率。

从图 2 中可以看到, 有一个豆苗样本被错误地分类到田字草当中, 这可能是由于实验中的随机误差导致的。所以, 它被作为对系统模型有较大影响的奇异样本剔除出训练集。80 个样本中的 79 个作为定标样本能够正确分类。此外, 20 个田字草样本的聚集区半径最小, 可能是由于它的叶片比较规则, 在摆放样本时叶片之间的空隙较小, 重叠因素的影响也小。而玉米的光谱样本在图中最为分散, 可能是由于玉米叶片中间有大的脊, 影响了入射光的角度, 造成了较大的反射率误差。

为了研究四种植物光谱的差异, 进一步对主成份载荷图作分析 (图 3)。从图中可以看到, 主成份

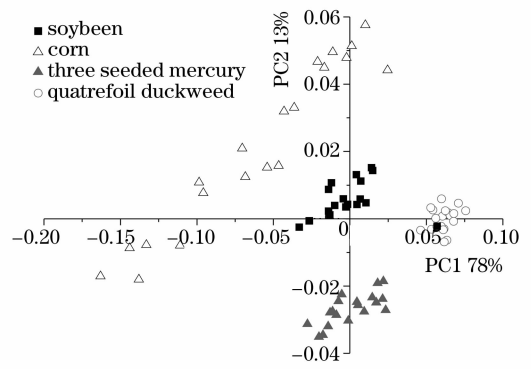


图2 大豆、玉米、铁苋菜、田字草的主成分 PC1 和主成分 PC2 的聚类图

Fig.2 Principal components cluster plot of crop and weed leaf samples

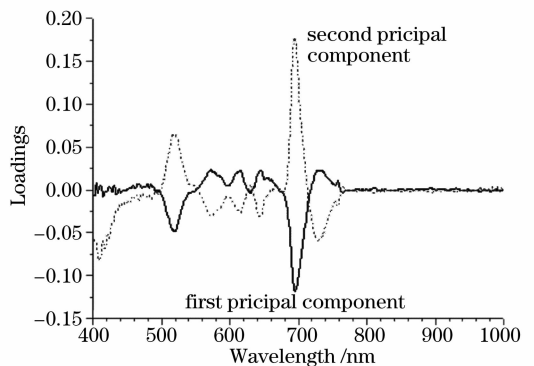


图3 作物和杂草叶片的可见-近红外反射光谱样本主成分 PC1 和主成分 PC2 的载荷图

Fig.3 Loadings weights of the first and second principal components at visible-near infrared spectra of crop and weed leaf sample

PC1 和 PC2 在可见光波段有较大差异, 并且波动明显, 而在近红外波段的载荷都很小。PC1 的最大的正向载荷出现在 638~665 nm 和 720~760 nm, 这些波段对应叶绿素吸收波段<sup>[10]</sup>。PC2 在大部分波段上与 PC1 反向变化, 它的最大正向载荷出现在 510~526 nm 和 687~706 nm, 分别对应蓝光吸收<sup>[9]</sup>和“红边”位置, 由于用一阶 Savitzky-Golay 差分进行光谱预处理而被放大。

### 3.2.2 PLS 校正和有效性分析

PLS 是一个双线性模型, 原始的独立变量信息 (X) 被投影到少数的 PCs 中。在作物与杂草的定性分析中, 由于没有相应的化学指标作为因变量 Y 值, 就人工赋以 1, 3, 5, 7, 分别代表大豆, 玉米, 铁苋菜和田字草。采用带交叉验证的 PLS 回归方法来建立在光谱数据与品种数值之间的定标模型。采用完全交叉验证的预测残差平方和 (Prediction residual error sum of squares, PRESS) 为评价标

准,确定最佳主成分个数,以防止模型过度拟合<sup>[15]</sup>。

根据 PCA 的结果,大豆的样本中有一个奇异样本被剔除。建模的光谱样本被随机分为训练集和验证集。用相关系数(correlation coefficient) $r$ ,建模标准差(standard error of calibration, SEC) $S_{EC}$ ,预测标准差(Standard error of prediction, SEP) $S_{EP}$ ,建模均方根误差(root mean square error of calibration, RMSEC) $R_{MSEC}$ ,预测均方根误差(Root mean square error of prediction, RMSEP) $R_{MSEP}$ ,偏差(bias) $b_{ias}$ ,斜率(slope) $s_{lope}$ 和偏移(offset) $o_{ffset}$ 等来衡量模型预测的精确性。好的模型有较小的 $b_{ias}$ , $S_{EC}$ , $S_{EP}$ , $R_{MSEC}$ 和 $R_{MSEP}$ ,以及较大的 $r$ 。 $b_{ias}$ , $S_{EC}$ <sup>[17]</sup>, $S_{EP}$ <sup>[15]</sup>分别由下式定义:

$$b_{ias} = \frac{1}{I} \sum_{i=1}^I (\hat{y}_i - y_i), \quad (1)$$

$$S_{EC} = \sqrt{\frac{1}{I-1} \sum_{i=1}^I (\hat{y}_i - y_i)^2}, \quad (2)$$

$$S_{EP} = \sqrt{\frac{1}{I-1} \sum_{i=1}^I (\hat{y}_i - y_i - b_{ias})^2}, \quad (3)$$

式中 $\hat{y}_i$ 和 $y_i$ 为每个样本的预测值和实际值, $I$ 是样本数。

经交互验证法判断,本研究 PLS 模型的最佳主成分数为 5。表 1 显示了四种植物共 79 个样本的 400~1000 nm 波段光谱的 PLS 模型建模和交叉验证结果。其中建模和交叉验证的 $r$ 分别为 0.987 和 0.984。建模和交叉验证的 $r$ 值较高,并且 $b_{ias}$ 较小。 $S_{EC}=0.356$ 和 $S_{EP}=0.396$ 值较小且差异不大,同时 $R_{MSEC}=0.354$ 和 $R_{MSEP}=0.394$ 值接近,说明模型没有出现“过度拟合”和“欠拟合”。因此,建立的 PLS 模型具有较好的区分作物好杂草光谱的能力。

表 1 四种植物的 400~1000nm 波段光谱的 PLS 模型建模和交叉验证结果

Table 1 Calibration and cross-validation results of the four plants from PLS modeling at the spectral range of 400~1000 nm

Parameter	Calibration	Cross validation
Element	79	79
Slope	0.975	0.967
Offset	0.102	0.138
Correlation	0.987	0.984
Bias	$3.71 \times 10^{-7}$	0.00647
SEC	0.356	/
SEP	/	0.396
RMSEC	0.354	/
RMSEP	/	0.394

### 3.3 统计预测分析

根据建立好的 PLS 模型来预测每种植物中剩下的 10 个样本,图 4 显示了区分四种植物的预测结果。 $r=0.987$  表明预测值与真实值之间有极大的相关性。 $S_{EP}=0.366$  和建模时的 $S_{EC}=0.356$ 之间差异较小,表明模型有较高的预测精度。 $R_{MSEP}=0.364$ 和 $b_{ias}=-0.0436$ 表明了模型在作物与杂草分类上有较强的识别能力。从图 4 中可以看到,通过阈值 2.0,4.0 和 6.0 对预测的结果进行划分,可以 100%地区分四个植物品种的样本,说明了模型有较好的识别结果。

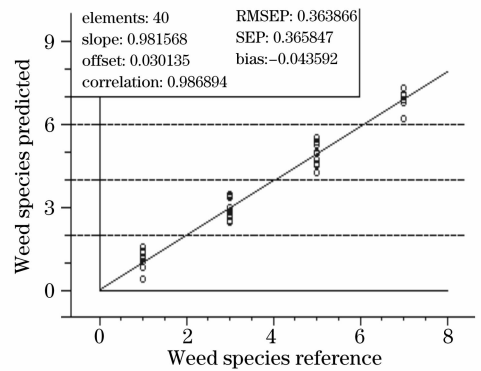


图 4 四种植物种类的可见-近红外光谱 PLS 预测结果  
Fig. 4 PLS prediction results of the four plants at visible and near-infrared spectra

## 4 结 论

对两种作物(玉米和豆苗)和两种杂草(铁苋菜和田字草)的叶片的可见-近红外反射光谱,建立了一个完全交叉验证的 PLS 模型。在对 80 个吸收度光谱做 Savitzky-Golay 平滑和一阶差分的预处理后,进行 400~1000 nm 光谱的 PCA 预分析,显示了前两个主成份可以代表超过 91%的总体变异,因此非常适合用来做种类鉴别。有一个豆苗样本被错误地归类为田字草,因此被作为奇异样本剔除了。PLS 模型的预测结果显示预测值与实际值高度相关( $r=98.7\%$ , $S_{EP}=0.366$ , $R_{MSEP}=0.364$ , $b_{ias}=-0.0436$ )。通过阈值的划分能够得到 100%的识别率。因此,只要用特定的数值来量化品种因子,作为因变量 Y 值,用可见-近红外反射光谱来做杂草与作物的品种鉴别,PLS 模型是一种有效的方法。

### 参 考 文 献

- Li Zhichen, Rao Honghui, Wang Yong *et al.*. Status quo and advance on research of variable-rate spraying technology[J]. *J. Northeast Agricultural University*, 2007, **38**(4): 563~567

- 李志臣, 饶洪辉, 王 勇 等. 除草剂变量施药技术的研究现状与进展[J]. 东北农业大学学报, 2007, **38**(4): 563~567
- 2 Chen Derun, Wang Shumao, Wang Xiu. Status and prospect for recognition technology of farm weeds[J]. *Chinese Agriculture Mechanization*, 2005, **2**: 35~38  
陈德润, 王书茂, 王 秀. 农田杂草识别技术的现状与展望[J]. 中国农机化, 2005, **2**: 35~38
- 3 Wu Di, Feng Lei, Zhang Chuanqing *et al.*. Early detection of gray mold (cinerea) on eggplant leaves based on vis/near infrared spectra[J]. *J. Infrared and Millimeter Waves*, 2007, **26**(4): 269~273  
吴 迪, 冯 雷, 张传清 等. 基于可见/近红外光谱技术的茄子叶片灰霉病早期检测研究[J]. 红外与毫米波学报, 2007, **26**(4): 269~273
- 4 Chen Quansheng, Zhao Jiewen, Zhang Haidong *et al.*. Identification of authenticity of tea with near infrared spectroscopy based on support vector machine[J]. *Acta Optica Sinica*, 2006, **26**(6): 933~937  
陈全胜, 赵杰文, 张海东 等. 基于支持向量机的近红外光谱鉴别茶叶的真伪[J]. 光学学报, 2006, **26**(6): 933~937
- 5 Zhao Jiewen, Zhang Haidong, Liu Muhua. Preprocessing methods of near-infrared spectra for simplifying prediction model of sugar content of apples[J]. *Acta Optica Sinica*, 2006, **26**(1): 136~140  
赵文杰, 张海东, 刘木华. 简化苹果糖度预测模型的近红外光谱预处理方法[J]. 光学学报, 2006, **26**(1): 136~140
- 6 Shi Youming, Liu Gang, Liu Jianhong *et al.*. Identification of auricularia auricula from different regions by Fourier transform infrared spectroscopy[J]. *Acta Optica Sinica*, 2007, **27**(1): 129~132  
时有明, 刘 刚, 刘剑虹 等. 不同产地黑木耳的傅里叶变换红外光谱鉴别[J]. 光学学报, 2007, **27**(1): 129~132
- 7 Xu Xiangqun, Wu Liu. Dependence of optical clearing effect on tissue structure[J]. *Chin. J. Lasers*, 2006, **33**(7): 998~1002  
徐向群, 吴 柳. 不同结构生物组织光透明作用比较[J]. 中国激光, 2006, **33**(7): 998~1002
- 8 Wei Huajiang, Xing Da, Wu Guoyong *et al.*. Using spatially resolved reflectance to measure optical properties of stomach tissue[J]. *Chin. J. Lasers*, 2007, **34**(4): 582~588  
魏华江, 邢 达, 巫国勇 等. 采用空间分辨漫反射测定胃组织光学特性[J]. 中国激光, 2007, **34**(4): 582~587
- 9 Reyer Zwigelaar. A review of spectral properties of plants and their potential use for crop/weed discrimination in row-crops[J]. *Crop Protection*, 1998, **17**(3): 189~206
- 10 Cliff H. Koger, Lori M. Bruce, David R. Shaw *et al.*. Wavelet analysis of hyperspectral reflectance data for detecting pitted morningglory (Ipomoea lacunosa) in soybean (Glycine max)[J]. *Remote Sensing of Environment*, 2003, **86**(1): 108~119
- 11 Anne M. Smith, Robert E. Blackshaw. Weed-crop discrimination using remote sensing: A detached leaf experiment [J]. *Weed Technology*, 2003, **17**(4): 811~820
- 12 D. C. Slaughter, W. T. Lanini, D. K. Giles. Discriminating weeds from processing tomato using visible and near-infrared spectroscopy[J]. *Transactions of the ASAE*, 2004, **47**(6): 1907~1911
- 13 Ji Shouwen, Wang Rongben, Chen Jiajuan *et al.*. Research on recognizing weed from corn seedling by using computer image processing technology [J]. *Transactions of the CSAE*, 2001, **17**(2): 154~156  
纪寿文, 王荣本, 陈佳娟 等. 应用计算机图像处理技术识别玉米苗期田间杂草的研究 [J]. 农业工程学报, 2001, **17**(2): 154~156
- 14 M. Blanco, J. Coello, H. Iturriaga *et al.*. NIR calibration in non-linear systems; different PLS approaches and artificial neural networks[J]. *Chemometrics and Intelligent Laboratory Systems*, 2000, **50**(1): 75~82
- 15 T. Naes, T. Isaksson, T. Fearn *et al.*. *A user-Friendly Guide to Multivariate Calibration and Classification* [M]. UK: NIR Publications, 2002
- 16 Fang Hui, Song Haiyan, Cao Fang *et al.*. Study on the relationship between spectral properties of oilseed rape leaves and their chlorophyll content [J]. *Spectroscopy and Spectral Analysis*, 2007, **127**(19): 1731~1734  
方 慧, 宋海燕, 曹 芳 等. 油菜叶片的光谱特征与叶绿素含量之间的关系研究[J]. 光谱学与光谱分析, 2007, **127**(19): 1731~1734
- 17 Gamal ElMasry, Ning Wang, Adel ElSayed *et al.*. Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry [J]. *J. Food Engineering*, 2007, **81**(1): 98~107