

# 基于邻域粗糙集和高光谱散射图像的苹果粉质化检测<sup>\*</sup>

朱启兵 黄敏 赵桂林

(江南大学轻工过程先进控制教育部重点实验室, 无锡 214122)

**【摘要】** 研究了基于邻域粗糙集理论的高光谱散射图像苹果粉质化无损检测方法。以 576 幅波长范围为 600 ~ 1 000 nm 的苹果高光谱数据为研究对象, 利用邻域粗糙集模型对 81 个原始波段进行选择, 从中选择出最优波长子集; 利用支持向量机建立分类模型, 随机选择 526 个样本作为训练集, 其余 50 个样本作为测试集, 重复仿真 10 次验证分类能力。仿真结果表明邻域粗糙集能够得到充分表述粉质化程度的 14 个最优波长, 测试模型的平均精度为 75%, 高于全波长模型的 71% 和采用主成分分析法的 74%。

**关键词:** 苹果 粉质化 无损检测 邻域粗糙集 高光谱散射图像

**中图分类号:** O657.3 **文献标识码:** A **文章编号:** 1000-1298(2011)10-0154-04

## Apple Mealiness Detection Based on Neighborhood Rough Set and Hyperspectral Scattering Image

Zhu Qibing Huang Min Zhao Guilin

(Key Laboratory of Advanced Process Control for Light Industry, Ministry of Education, Jiangnan University, Wuxi 214122, China)

### Abstract

Nondestructive detection of apple mealiness was investigated by using neighborhood rough set theory and hyperspectral scattering image technology. Spectral scattering profiles between 600 nm and 1 000 nm were acquired by hyperspectral scattering image system for 576 apple samples. The optimal wavelength sets were chosen from 81 raw characteristic attributes by neighborhood rough set. 526 samples were selected randomly for calibration set and 50 samples were selected for validation set to develop classification model using optimal wavelengths coupled with support vector machine (SVM). Simulation was repeated 10 times to investigate the ability of classification model. Results showed that neighborhood rough set could select 14 optimal wavelengths effectively. The validation model using 14 optimal wavelengths yielded better result (classification accuracy 75%) than the full spectrum model (classification accuracy 71%) and the principle component analysis algorithm (classification accuracy 74%).

**Key words** Apple, Mealiness, Nondestructive detection, Neighborhood rough set, Hyperspectral scattering image

### 引言

苹果粉质化是影响苹果等级的重要口感参数, 是苹果内部综合指标的反映<sup>[1]</sup>。传统的基于人工品尝和破坏性仪器测量方法难以满足无损测量要求。目前, 利用高光谱图像技术进行水果的内外部

品质检测已成为国内外的一个研究热点, 已有大量的文献报道<sup>[2-3]</sup>。

由于高光谱图像数据存在很强的谱间相关性, 在线获取如此多的信息所需要的时间长, 很难满足快速检测的目的, 解决这一问题的途径是选取最能表述水果内部品质的高光谱图像的最优波长, 然后

收稿日期: 2011-05-11 修回日期: 2011-05-17

<sup>\*</sup> 国家自然科学基金资助项目(60805014)和中央高校基本科研业务费专项资金资助项目(JUSRP20913、JUSRP21132)

作者简介: 朱启兵, 副教授, 博士, 主要从事农产品无损检测研究, E-mail: zhuqib@163.com

通过设计具有若干个波长的光谱图像系统,应用到生产实际中。这样设计出的多光谱图像检测系统能大大提高检测效率<sup>[7]</sup>。但是如何进行高光谱图像的波长选择目前还没有一个很好的理论指导。

本文利用高光谱散射图像技术检测苹果的粉质化。首先利用邻域粗糙集 (neighborhood rough sets, 简称 NRS) 对苹果样本的高光谱波段进行选择, 选择出决定分类准确率的最优波长, 然后利用支持向量机 (support vector machine, 简称 SVM) 建立分类模型。

## 1 试验材料与方法

### 1.1 试验样本及试验装置

试验用 580 个“Red Delicious”苹果, 来自于美国密歇根州立大学实验站和商业包装厂, 为得到非粉质化和不同程度粉质化的样本, 在 4℃ 的环境下储藏 240 个样本, 其余 340 个样本储藏在相对湿度为 95%、温度为 20℃ 的环境中, 以加速苹果的粉质化过程。由于不是所有样本都会在高温、高湿条件下完全粉质化, 试验时非粉质化样本仍会多于粉质化样本。试验前, 所有样本需要在室温下至少放置 15 h。

试验时, 在每个样本的中心赤道位置采集二维高光谱散射图像。高光谱试验系统包括高光谱成像单元、直流稳压光源和样品平台。高光谱成像单元由电荷耦合装置 (C4880-21-24A 型, Hamamatsu Photonics Systems, Bridgewater, NJ, USA)、光谱范围为 400 ~ 1 100 nm 的图像光谱仪 (ImSpector V10 型, Spectral Imaging Ltd., Oulu, Finland) 和可变焦距透镜组成。直流稳压光源由石英卤钨灯及反馈控制器组成, 用来提供点光源, 为了保证光纤聚焦, 苹果到光谱仪的距离是不变的。样品平台由水平机动平台、垂直机动平台和一对通过光束的光电传感器组成。苹果每移动 1 mm, 线扫描一次。苹果在水平方向移动的同时, 高光谱成像单元在垂直方向获得苹果的高光谱图像。每个样本共采集 10 幅高光谱散射图像<sup>[4]</sup>。

### 1.2 高光谱散射图像及粉质化仪器标准值测量

一个典型“Red Delicious”苹果的高光谱散射图像如图 1 所示, 水平坐标代表空间位置信息, 垂直坐标代表光谱波长信息。图像空间分辨率为 0.20 mm, 系统的光谱分辨率为 4.54 nm, 实际试验数据记录时, 取近似值 5 nm。

苹果的粉质化描述可以采用压缩硬度和汁液含量两个指标。采集完光谱图像之后, 使用一种特殊的内置夹具和一个质构仪 (TA. XT2i 型, Stable

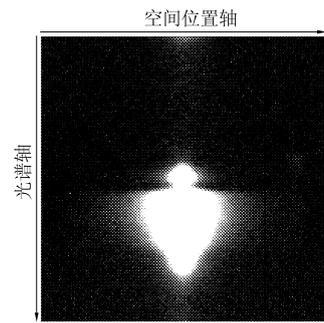


图 1 苹果的高光谱散射图像

Fig. 1 Hyperspectral scattering image of an apple

Micro Systems, Inc., Surrey, U. K.) 来测试苹果压缩硬度和汁液含量。苹果的切割圆柱长度为 16 mm, 直径为 18 mm, 放置在内置夹具中。压缩过程中, 压缩速度为 20 mm/min, 最大变形为 3 mm。硬度的单位为 N/mm, 取自压力变形曲线 1/3 ~ 2/3 的斜率。汁液含量的测定是在压缩过程中在样本底下放置滤纸, 最后测定滤纸上汁液的面积, 单位为 cm<sup>2</sup>。具体的试验过程参见文献[4]。

根据压缩硬度和汁液含量两个指标联合对样本进行分类, 分类标准<sup>[5]</sup>如表 1 所示。

表 1 苹果粉质化分类标准

Tab. 1 Standard of apple mealiness classification

参数	汁液含量	汁液含量
	(≥5 cm <sup>2</sup> )	(<5 cm <sup>2</sup> )
压缩硬度 (≥40 N/mm)	非粉质化	非粉质化
压缩硬度 (<40 N/mm)	非粉质化	粉质化

## 2 数据分析方法

### 2.1 高光谱散射图像特征生成

图 2 为高光谱散射图像在 3 个典型波长下的光谱强度曲线。从图中可以看出, 该空间散射图像是关于中心对称的, 且光强主要分布在 -10 ~ 10 mm, 因此在每一个波长下, 在 0 ~ 10 mm 空间范围内利用平均值法求取平均光强, 并计算其相对平均反射光强<sup>[6]</sup>为

$$\bar{R}_i = \frac{\bar{R}_{si}}{\bar{R}_{Ti}} \quad (1)$$

式中  $\bar{R}_{si}$ 、 $\bar{R}_{Ti}$ ——第  $i$  个波长下苹果、参考白板的平均反射光强

在波长 600 ~ 1 000 nm、光谱分辨率为 5 nm 的试验条件下, 每幅图像共有 81 个原始波长。在计算平均反射光强之前, 需要对图像进行预处理, 包括距离校正、果形校正以及光源波动影响校正<sup>[7]</sup>。

图 3、4 分别为 10 个粉质化和非粉质化的苹果在波长 600 ~ 1 000 nm 范围内的提取的特征光谱曲

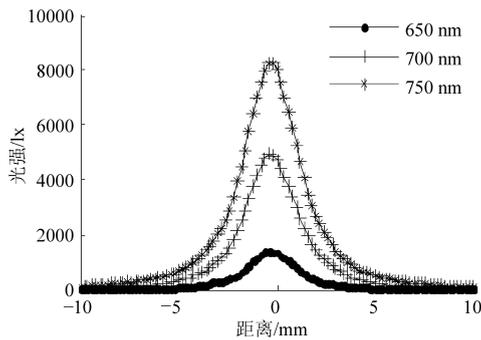


图2 3个波长下的空间散射图像

Fig. 2 Raw spatial scattering profiles at three different wavelengths

线。从图中可以看出,非粉质化的相对平均反射光强比粉质化的相对平均反射光强大。因此可以利用粉质化和非粉质化苹果的相对平均反射光强的差异性进行苹果的粉质化评价。

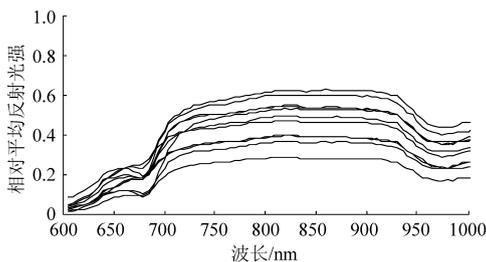


图3 波长为600~1000 nm 10个粉质化苹果的相对平均反射光强曲线

Fig. 3 Relative average reflectance for ten mealy apples

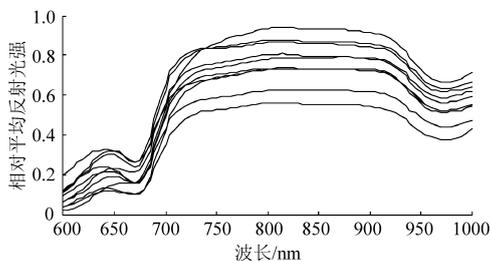


图4 波长为600~1000 nm 10个非粉质化苹果的相对平均反射光强曲线

Fig. 4 Relative average reflectance for ten nonmealy apples

## 2.2 邻域粗糙集模型

由于原始的81个光谱波段存在着冗余和矛盾信息,将所有的光谱特征信息作为分类特征,不仅降低了分类器的分类效率,而且也影响了分类的准确性。粗糙集(rough set,简称RS)是由波兰数学家Pawlak在1982年提出的一种数据分析理论,它无需提供问题所需处理的数据集合之外的任何先验信息,常用于处理模糊和不精确问题。经典RS用于处理数值型变量时必须先进行离散化,从而导致信息丢失,给波段选择的准确性带来影响。相比于经典的RS模型,邻域粗糙集模型省略了数值型变量的离散化过程,保证了特征选择的稳定性和准确性。

本文采用NRS模型对原始的81个波长进行最优选择,以达到减少波长、提高分类精度的目的。NRS的基本理论可参见文献[8]。

## 2.3 基于邻域粗糙集模型的苹果粉质化检测流程

检测流程如下:

(1) 将全部580个样本,利用式(1)计算每个苹果样本的相对平均反射光强,共得到81个原始波长。去除4个奇异样本后,共保留576个样本。

(2) 利用576个样本的81个原始波段构成决策系统的条件属性;利用破坏性仪器测量获得576个样本的具体分类结果作为其决策属性。

(3) 采用基于属性重要度的邻域粗糙集模型数值属性约简算法进行最优波长选择<sup>[9]</sup>。

(4) 采取随机采样的方法,将576个样本分为训练集和测试集,其中训练集共526个样本,测试集共50个样本。将选择后的最优波段作为支持向量机的输入特征进行分类,获得其分类预测结果。选择径向基核函数为SVM的核函数。其核参数 $\gamma$ 和惩罚系数 $C$ 对分类器的分类结果有很大的影响,本文采用交叉验证的网格搜索法确定<sup>[10]</sup>。搜索范围为 $\gamma \in [1, 1250]$ , $C \in [1, 1250]$ ,搜索步长选择50,交叉验证的子集个数为10,取最大平均校正精度对应的参数 $C$ 和 $\gamma$ 为SVM的最优参数。

## 3 结果与讨论

### 3.1 NRS-SVM模型和PCA-SVM模型分类结果

在邻域尺寸 $\delta = 0.05$ 、可变精度 $k = 0.85$ 条件下,共得到选择后的最优波长14个。所选择的最优波长为670、690、715、735、760、770、780、790、820、830、835、885、975、985 nm。

考虑到采样的随机性,共进行了10次随机采样,表2为利用原始81个波长、14个NRS选择的最优波长、以及利用PCA对原始波长进行线性组合获得的14个主分量,分别输入SVM后获得的分类结果比较。

表2 不同输入条件下的分类结果

模型	Table 2 Detection results with different inputs %					
	训练精度			测试精度		
	最大值	最小值	平均值	最大值	最小值	平均值
SVM	79.99	78.85	79.20	74.00	68.00	71.00
NRS-SVM	80.88	79.49	80.33	78.00	72.00	75.00
PCA-SVM	80.43	78.51	79.31	78.00	66.00	74.00

由表2可以看出,两种特征提取方法得到的平均测试精度均高于单纯使用SVM的测试精度,而训练精度变化不大。采用PCA-SVM方法得到的平

均测试精度为 74%, 说明相对于 NRS, 利用 PCA 提取出的主分量的属性集对于直接检测效果不是最好。除此之外, PCA 降维的结果是若干个波长的线性组合, 并不能明确显示出被选择的特征波长, 而 NRS 却能根据属性重要度挑选出最优波长, 可以利用这些最优波长指导多光谱图像的检测系统设计。

### 3.2 NRS 参数对检测结果影响

在 NRS 模型中, 邻域尺寸  $\delta$  和可变精度  $k$  决定了最优波长的数目及相关波长的选择, 从而影响了最终的分类模型的精度。借助已有研究文献[8]给出的  $k \in [0.8, 0.9]$  能获得较理想的结果, 本研究中  $k$  直接选取平均值 0.85, 针对本文的试验数据, 表 3 给出了不同邻域尺寸  $\delta$  下的粉质化检测结果。其中  $F$  代表经 NRS 约简后的波长数目,  $P_1$  代表 10 次重复仿真训练精度的平均值,  $P_2$  代表由  $P_1$  确定的最优参数组下的测试精度平均值, 平均精度为

$$P = \frac{526P_1 + 50P_2}{576} \quad (2)$$

从表 3 可以看出, 不同  $\delta$  值下, 分类精度均优于单纯使用 SVM 分类的结果, 说明利用 NRS 删除冗余波段后, 分类性能提高了。当  $\delta$  为 0.20 时, 由 NRS 选择出的波长数目为 4, 此时的  $P_2$  为 77%, 高于  $\delta$  取 0.05、波长数目为 14 时的  $P_2$ , 说明并不是波长数目多, 分类精度越高。当  $\delta$  为 0.05 时,  $P$  最大, 说明从训练精度和测试精度综合考虑, 此参数最优; 但是当  $\delta$  取 0.20 时,  $P$  次优且波长数目为 4, 说明综

合考虑最优波长子集的因素, 此参数最优。实际使用中可以根据不同的要求选择最优参数。

表 3 不同邻域尺寸下的检测结果  
Tab.3 Detection results varying with neighborhood size

$\delta$	$F$	$P_1/\%$	$P_2/\%$	$P/\%$
0.05	14	80.33	75.00	79.86
0.10	3	78.82	75.00	78.49
0.20	4	79.12	77.00	78.94
0.30	2	79.12	74.00	78.68

虽然基于 NRS - SVM 模型所获得的检测效果好于 SVM 和 PCA - SVM, 但分类精度还较低。造成此原因可能有样本的粉质化程度变化范围相对较小, 粉质化仪器测试标准中粉质化和非粉质化的汁液含量值存在重叠现象等。

## 4 结束语

采用邻域粗糙集法提取高光谱散射图像 600 ~ 1 000 nm 共 81 个波长的 14 个有效特征波长, 结合支持向量机模型实现苹果的粉质化检测。通过对 576 个“Red Delicious”苹果样本建模的结果分析表明, NRS - SVM 可以提高分类器的分类精度。其进一步的意义在于: 可以利用 NRS 模型选择出的特征波长, 设计可用于在线检测的多光谱系统, 从而提高检测的实时性。

## 参 考 文 献

- 赵桂林, 朱启兵, 黄敏. 基于高光谱图像技术的苹果粉质化 LLE - SVM 分类[J]. 光谱学与光谱分析, 2010, 30(10): 2 739 ~ 2 743.  
Zhao Guilin, Zhu Qibing, Huang Min. LLE - SVM classification of apple mealiness based on hyperspectral scattering image technique[J]. Spectroscopy and Spectral Analysis, 2010, 30(10): 2 739 ~ 2 743. (in Chinese)
- 单佳佳, 彭彦昆, 王伟, 等. 基于高光谱成像技术的苹果内外部品质同时检测[J]. 农业机械学报, 2011, 42(3): 140 ~ 144.  
Shan Jiajia, Peng Yankun, Wang Wei, et al. Simultaneous detection of external and internal quality parameters of apples using hyperspectral technology [J]. Transactions of the Chinese Society for Agricultural Machinery, 2011, 42(3): 140 ~ 144. (in Chinese)
- Huang M, Lu R. Optimal wavelength selection for hyperspectral scattering prediction of apple firmness and soluble solids content [J]. Transactions of the ASABE, 2010, 53(4): 1 175 ~ 1 182.
- Huang M, Lu R. Apple mealiness detection using hyperspectral scattering technique [J]. Postharvest Biology and Technology, 2010, 58(1): 168 ~ 175.
- Barreiro P, Ortiz C, Ruiz-Altisent M, et al. Comparison between sensory and instrumental measurements for mealiness assessment in apples; a collaborative experiment[J]. Journal of Texture Studies, 1998, 29(5): 509 ~ 525.
- Qin J, Lu R, Peng Y. Prediction of apple internal quality using spectral absorption and scattering properties[J]. Transactions of the ASABE, 2009, 52(2): 499 ~ 507.
- Peng Y, Lu R. An LCTF-based multispectral imaging system for estimation of apple fruit firmness: part II. selection of optimal wavelengths and development of prediction models [J]. Transactions of the ASABE, 2006, 49(1): 269 ~ 275.

主成分所建健康苹果和霉心病苹果的 Fisher 判别函数检验集正确判别率为 87.8%, 验证了近红外漫反射光谱技术应用于健康苹果和霉心病苹果检测的可行性。建模集正确判别率为 89.9%, 对

## 参 考 文 献

- 1 Fan Q, Wang Y, Sun P, et al. Discrimination of Ephedra plants with diffuse reflectance FT-NIRS and multivariate analysis [J]. *Talanta*, 2010, 80(3): 1 245 ~ 1 250.
- 2 李亮, 丁武. 掺有植物性填充物牛奶的近红外光谱判别分析[J]. *光谱学与光谱分析*, 2010, 30(5): 1 238 ~ 1 242.  
Li Liang, Ding Wu. Discriminant analysis of raw milk adulterated with botanical filling material using near infrared spectroscopy [J]. *Spectroscopy and Spectral Analysis*, 2010, 30(5): 1 238 ~ 1 242. (in Chinese)
- 3 Clark C J, McGlone V A, Jordan R B. Detection of brownheart in "Braeburn" apple by transmission NIR spectroscopy [J]. *Postharvest Biology and Technology*, 2003, 28(1): 87 ~ 96.
- 4 Upchurch B L, Throop J A, Aneshansley D J. Detecting internal breakdown in apples using interreflectance measurements [J]. *Postharvest Biology and Technology*, 1997, 10(1): 15 ~ 19.
- 5 Han D, Tu R, Lu C, et al. Nondestructive detection of brown core in the Chinese pear 'Yali' by transmission visible-NIR spectroscopy [J]. *Food Control*, 2006, 17(8): 604 ~ 608.
- 6 周竹, 刘洁, 李小昱, 等. 霉变板栗的近红外光谱和神经网络方法判别[J]. *农业机械学报*, 2009, 40(增刊): 109 ~ 112.  
Zhou Zhu, Liu Jie, Li Xiaoyu, et al. Discrimination of moldy Chinese chestnut based on artificial neural network and near infrared spectra [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2009, 40(Supp.): 109 ~ 112. (in Chinese)
- 7 展慧, 李小昱, 周竹, 等. 基于近红外光谱和机器视觉融合技术的板栗缺陷检测[J]. *农业工程学报*, 2011, 27(2): 345 ~ 349.  
Zhan Hui, Li Xiaoyu, Zhou Zhu, et al. Detection of chestnut defect based on data fusion of near-infrared spectroscopy and machine vision [J]. *Transactions of the CSAE*, 2011, 27(2): 345 ~ 349. (in Chinese)
- 8 韩东海, 刘新鑫, 鲁超, 等. 苹果内部褐变的无损检测研究[J]. *农业机械学报*, 2006, 37(6): 86 ~ 88, 93.  
Han Donghai, Liu Xinxin, Lu Chao, et al. Study on optical-nondestructive detection of breakdown apples [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2006, 37(6): 86 ~ 88, 93. (in Chinese)
- 9 王加华, 孙旭东, 潘璐, 等. 基于可见/近红外能量光谱的苹果褐腐病和水心鉴别[J]. *光谱学与光谱分析*, 2008, 28(9): 2 098 ~ 2 102.  
Wang Jiahua, Sun Xudong, Pan Lu, et al. Discrimination of brownheart and watercore of apples based on energy spectrum of visible/near infrared transmittance [J]. *Spectroscopy and Spectral Analysis*, 2008, 28(9): 2 098 ~ 2 102. (in Chinese)
- 10 Shenderoy C, Shmulevich I, Alchanatis V, et al. NIRS detection of moldy core in apples [J]. *Food and Bioprocess Technology*, 2010, 3(1): 79 ~ 86.

(上接第 157 页)

- 8 胡清华, 于达仁, 谢宗霞. 基于邻域粒化和粗糙逼近的数值属性约简[J]. *软件学报*, 2008, 19(3): 640 ~ 649.  
Hu Qinghua, Yu Daren, Xie Zongxia. Numerical attribute reduction based on neighborhood granulation and rough approximation [J]. *Journal of Software*, 2008, 19(3): 640 ~ 649. (in Chinese)
- 9 赵勇, 方宗德, 王侃伟, 等. 邻域粗糙集在轮对踏面缺陷图像特征选择的应用[J]. *计算机测量与控制*, 2008, 16(11): 1 730 ~ 1 731.  
Zhao Yong, Fang Zongde, Wang Kanwei, et al. Application of neighborhood rough sets in features selection of wheelsets tread defect images [J]. *Computer Measurement & Control*, 2008, 16(11): 1 730 ~ 1 731. (in Chinese)
- 10 邓超, 吴军, 万紫娟. 基于支持向量机工具的性能劣化建模方法[J]. *计算机集成制造系统*, 2009, 15(4): 685 ~ 689.  
Deng Chao, Wu Jun, Wan Zijuan. Performance degradation model based on support vector machine [J]. *Computer Integrated Manufacturing Systems*, 2009, 15(4): 685 ~ 689. (in Chinese)