第37卷，第3期
2017年3月

光 谱 学 与 光 谱 分 析
Spectroscopy and Spectral Analysis

Vol. 37, No. 3, pp836-840
March, 2017

基于近红外高光谱成像技术的干制红枣品种鉴别

樊阳阳，裘正军*，陈 俭，吴 翔，何 勇

浙江大学生物系统工程与食品科学学院，浙江 杭州 310058

摘 要 为实现干制红枣的快速鉴别，提出了一种基于近红外高光谱成像技术的鉴别方法。采集四个品种共240个样本干制红枣的近红外高光谱图像(1 000～1 600 nm)。通过主成分分析法(principal component analysis，PCA)、载荷系数法(x-Loading Weights，x-LW)和连续投影算法(successive projections algorithm，SPA)分别提取7个、8个和10个特征波长；基于灰度共生矩阵(gray level co-occurrence matrix，GLCM)提取第一主成分图像的纹理特征。分别以光谱特征、纹理特征、光谱和纹理融合特征作为输入，建立偏最小二乘判别分析(partial least squares-discriminant analysis，PLS-DA)、反向传播神经网络(back-propagation neural network，BPNN)和最小二乘支持向量机(least squares support vector machines，LS-SVM)模型。结果显示，基于融合特征的模型鉴别率高于分别基于光谱特征或纹理特征的模型鉴别率；基于融合特征的BPNN模型的结果最优，对预测集样本鉴别正确率为100%。说明近红外高光谱成像技术可用于干制红枣品种的快速鉴别。

关键词 近红外高光谱成像；干制红枣；鉴别；纹理特征；特征融合

中图分类号：TP391.4 **文献标识码**：A **DOI**：10.3964/j.issn.1000-0593(2017)03-0836-05

引 言

干制红枣不仅保留了鲜枣的营养成分，而且食用方式多样，更有利于长期存储和长途运输。不同品种的红枣营养成分含量不同，价格也相差很多。鲜枣干制后，颜色、形状较为接近，通过肉眼很难区分，很多黑心商贩将低价品种混入优质品种中以欺骗消费者，因此迫切需要建立一种快速有效的干制红枣分类方法。

高光谱成像技术可以同时获取样本的光谱信息和图像信息，具有快速、准确性高等特点，近年来被广泛应用于农产品的品种鉴别和品质检测中。程术希等[1]利用近红外高光谱成像技术实现了八个品种的大白菜种子的有效鉴别，以载荷系数法获取的特征波长为输入，建立极限学习机(extreme learning machine，ELM)模型，识别正确率达到100%。Ashabahebwa Ambrose等[2]利用高光谱成像技术对玉米种子活力进行判别，建立的PLS-DA模型的预测集正确率达到95.6%。纹理特征是一种体现图像中样本表面结构变化的图像特征[3]，可用于品种鉴别[4-5]。将纹理特征与光谱特征融合，能够获得更完整的样本信息，提高识别率。章海亮等基于光谱信息和纹理信息的融合对六种品牌的绿茶进行鉴别，鉴别率从单独基于光谱信息的93.3%和基于纹理信息的90%提高到了100%。Wang[6]等基于特征波长的光谱信息和纹理信息的融合建立对玉米种子的分类模型，效果优于仅基于特征光谱或纹理特征所建立的模型。

本研究以四种干制红枣为对象，基于近红外高光谱成像技术，提取特征波长和纹理信息，结合不同的化学计量学方法建立模型，实现对干制红枣品种的快速鉴别。

1 实验部分

1.1 材料

试验分别采集山西的干制壶瓶枣和干制滩枣，新疆的干制若羌红枣、河北的干制金丝红枣四个品种各60个样本，共240个。基于Kennard-Stone(K-S)算法[7]按2∶1比例把样本划分为建模集和预测集，其中建模集样本个数为160，每个品种各40个，预测集样本个数为80，每个品种20个。

1.2 高光谱图像采集与校正

利用如图1所示的近红外高光谱成像系统对样本进行光谱采集，系统包括成像光谱仪、线光源、镜头、电控位移平

收稿日期：2016-05-05，**修订日期**：2016-10-23

基金项目：国家科技支撑计划项目(2014BAD04B04)资助

作者简介：樊阳阳，1994年生，浙江大学生物系统工程与食品科学学院博士研究生 e-mail：fanyangy9836@163.com

*通讯联系人 e-mail：zjqiu@zju.edu.cn

台等组成部件，其中成像光谱仪是由芬兰奥卢的 Spectral imaging Ltd 公司制造，光谱波长范围为 874～1 734 nm，光谱分辨率为 5 nm。

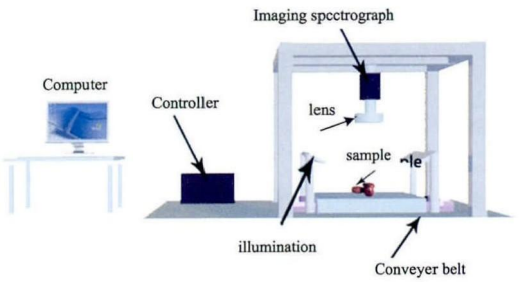

图 1　高光谱成像系统

Fig. 1　Hyperspectral imaging system

采集过程中，将样本表面距镜头高度设为 31.9 cm，移动平台速度为 22.5 mm·s^{-1}，曝光时间为 5 ms。共获得 256 个波段的高光谱信息。相机产生的暗电流会对高光谱图像造成影响，需要按下式进行校正

$$I = \frac{I_{raw} - B}{W - B}$$

式中，I_{raw} 为校正前高光谱图像，B 为黑色参照图像，W 为白色参照图像。

在校正后的高光谱图像中，为每个红枣样本选取一个感兴趣区域(region of interest，ROI)，以感兴趣区域内的平均光谱反射率作为该样本的光谱数据。为减小噪声，选取 1 000～1 600 nm 波段范围内光谱并采用 SG 卷积平滑法进行预处理。

1.3　数据处理

1.3.1　特征波长提取

光谱数据之间存在大量的冗余和共线性信息，且大量数据会增加计算量和模型的复杂度。本研究采用主成分分析法(PCA)、载荷系数法(x-LW)和连续投影算法(SPA)三种方法来提取特征波长，以简化后续建模分析。

PCA 是对光谱数据进行主成分分析[8]，在累计贡献率大的前几个主成分相应的权值系数曲线中，权值系数的绝对值与其相对应波长的贡献程度成正比，故选择曲线中波峰和波谷处对应的波长为特征波长。

x-LW 通过建立偏最小二乘回归模型[9](partial least squares regression，PLSR)选取特征波长。波长对模型性能影响可通过载荷系数绝对值的大小体现，在 PLSR 模型的隐含变量(latent variable，LV)的载荷系数曲线中，绝对值最大处对应的波长为特征波长。

SPA 是一种特征变量前向选择算法，近几年被广泛应用于光谱特征波长提取[10]。设定波长个数选择范围为 5～30，当均方根误差达到最小值时，得到特征波长个数。

1.3.2　纹理特征提取

灰度共生矩阵(GLCM)基于像素灰度的空间相关性表示纹理特征。为降低特征维度，首先对样本高光谱图像中的 ROI 进行主成分变换，选取第一主成分图像，然后基于灰度共生矩阵提取图像中的平均值(Mean)、方差(Variance)、同质性(Homogeneity)、对比度(Contrast)、非相似度(Dissimilarity)、熵(Entropy)、二阶矩(Second Moment)、相关性(Correlation)作为纹理特征值。

1.4　鉴别模型

分别基于特征波长、纹理特征、特征波长和纹理的融合特征，对比了偏最小二乘判别分析(PLS-DA)、反向传播神经网络(BPNN)、最小二乘支持向量机(LS-SVM)三种建模方法的效果，根据鉴别率筛选出最优模型。

2　结果与讨论

2.1　光谱特征分析

四个品种的干枣降噪后的平均光谱曲线如图 2 所示。不同品种干枣的光谱曲线趋势一致，在 1 200 和 1 450 nm 附近有明显的吸收峰。滩枣和若羌红枣光谱反射率曲线有部分重合，但与壶瓶枣、金丝红枣差异较大，表明这四种枣的有机组成存在一定的差异。对建模集光谱数据进行主成分定性分析，以观察样本的聚类趋势。得到前三个主成分 PC1，PC2，PC3 的累积贡献率为 99.93%，可以解释绝大部分变量。如图 3 所示，滩枣与其他三种枣区分明显；而壶瓶枣、若羌红枣和金丝红枣均有部分重合，需要进一步建模分析。

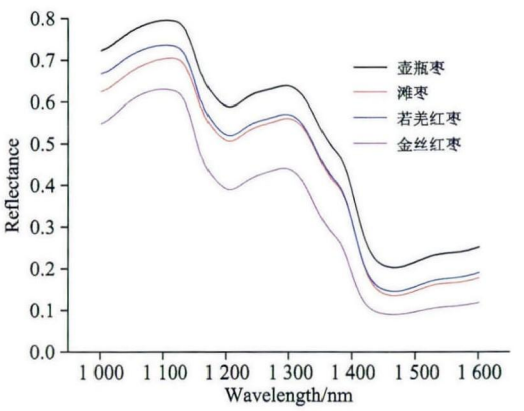

图 2　四种干制红枣的平均光谱曲线

Fig. 2　Average spectral curves of four varieties of jujubes

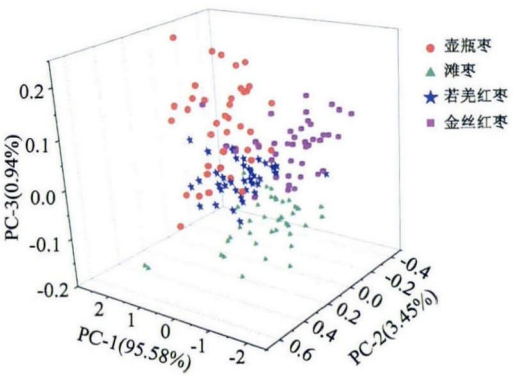

图 3　四种干制红枣在 PC1-PC2-PC3 的空间分布

Fig. 3　Distribution of four varieties of jujubes in PC1-PC2-PC3 space

2.2 基于特征波长的建模

以建模集光谱数据作为输入，分别采用 PCA，x-LW 和 SPA 选取特征波长。如图 4(a)所示，从 PCA 得到的前两个主成分的权值系数曲线中提取 7 个特征波长；x-LW 得到的 8 条 x-loading weights 曲线如图 4(b)所示，从每条曲线中提取特征波长；SPA 得到 10 个特征波长的空间分布如图 4(c)所示。

三种方法得到的特征波长如表 1 所示，不同提取方法原理不同，提取的特征波长的个数和种类会有所不同，但大多集中于红枣中有机物的吸收谱带。分布在 1 100～1 200 nm 范围的特征波长(1 146，1 207，1 210，1 123 和 1 224 nm)主要与烃类 C—H 键振动的二级倍频有关[11]；红枣中的氨基酸和蛋白质大都含有 N—H 键，如 PCA 得到的 1 470 nm 和 SPA 得到的 1 497 nm 分别属于胺的 N—H 键的振动组合频和振动吸收谱带[11]。

表 1 三种算法选择的特征波长

Table 1 Effective wavelengths selected by three algorithms

算法	特征波长数	特征波长/nm
PCA	7	1 086，1 173，1 207，1 234，1 291，1 409，1 470
x-LW	8	1 042，1 123，1 146，1 187，1 224，1 342，1 399，1 423
SPA	10	1 005，1 136，1 150，1 210，1 284，1 342，1 382，1 450，1 497，1 600

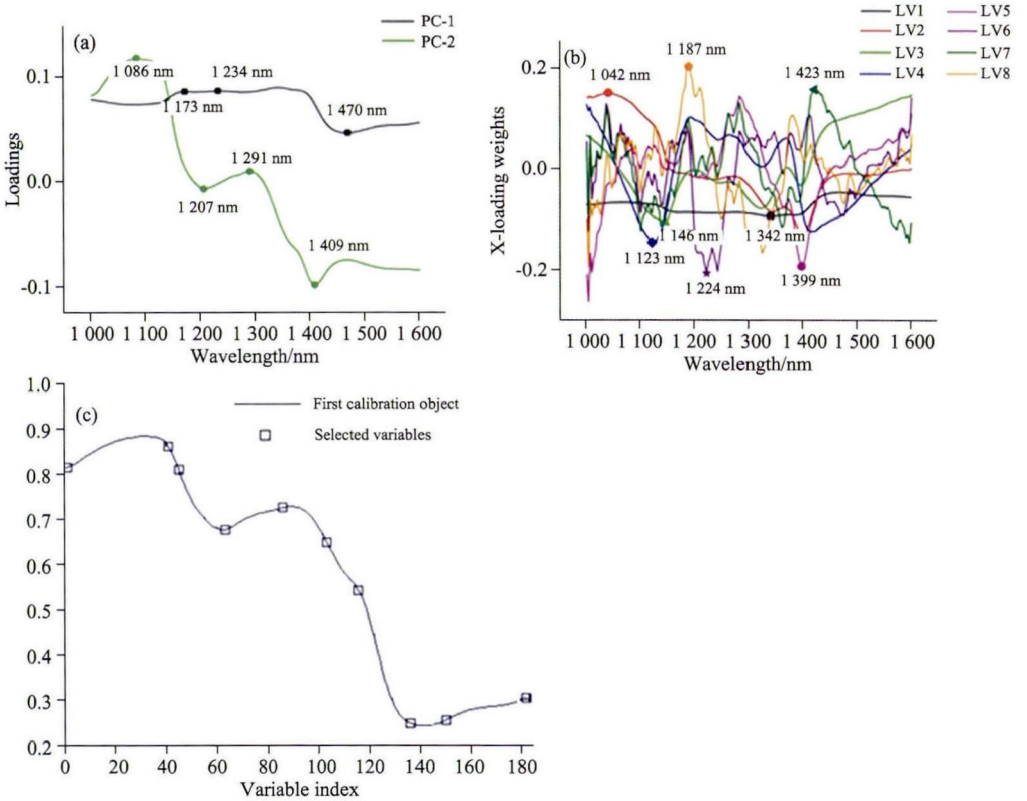

图 4 (a) PCA 法选择的特征波长；(b) x-LW 法通过 8 个隐含变量选择的特征波长；(c) SPA 法选择的特征波长分布

Fig. 4 (a) Effective wavelengths selected by PCA; (b) Effective wavelengths in 8 LVs selected by x-LW; (c) Distribution of effective wavelengths selected by SPA

表 2 基于特征波长的鉴别率(%)

Table 2 Identification rates based on effective wavelengths (%)

	PCA		x-LW		SPA	
	建模集	预测集	建模集	预测集	建模集	预测集
PLS-DA	73.13	61.25	78.75	61.25	77.50	63.75
BPNN	90.63	85	92.50	85	98.13	90
LS-SVM	98.13	78.75	96.88	80	97.50	85

以特征波长作为输入，分别构建 PLS-DA，BPNN，LS-SVM 鉴别模型，鉴别率结果见表 2。由表 2 可知，SPA-BPNN 模型的预测集分类效果最好，鉴别率为 90%；PLS-DA 模型正确率最低。基于 SPA 提取的特征波长建立的模型正确率最高，PCA 和 x-LW 的效果接近；用三种方法提取特征波长，均使输入变量维度减少了 90%以上，提高了运算效率。

2.3 基于纹理特征的建模

对感兴趣区域的高光谱图像进行主成分变换，实现高光谱图像降维。图 5 为壶瓶枣一个样本的前三个主成分图，可以看出 PC2 与 PC3 图像有效信息较少且包含有很多噪声点，会对纹理特征提取产生干扰，因此只对 PC1 图像提取纹理特

征，得到的 8 个纹理特征图像如图 6 所示。以纹理特征为输入，基于 PLS-DA，BPNN 和 LS-SVM 建立品种鉴别模型，结果如表 3 所示，可以看出，模型整体鉴别率低于基于光谱数据建立的模型，BPNN 模型取得了最优的预测集鉴别率 86.25%，PLS-DA 和 LS-SVM 模型的鉴别率均为 75%。

表 3　基于纹理特征的鉴别率(%)

Table 3　Identification rates based on texture features (%)

	建模集	预测集
PLS-DA	71.88	75.00
BPNN	91.25	86.25
LS-SVM	98.75	75.00

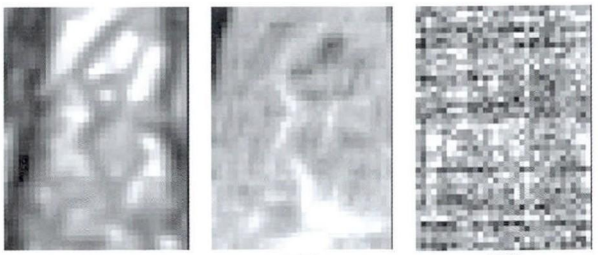

图 5　感兴趣区域的 PC1，PC2，PC3 图像

Fig. 5　PC1，PC2and PC3 images of ROI

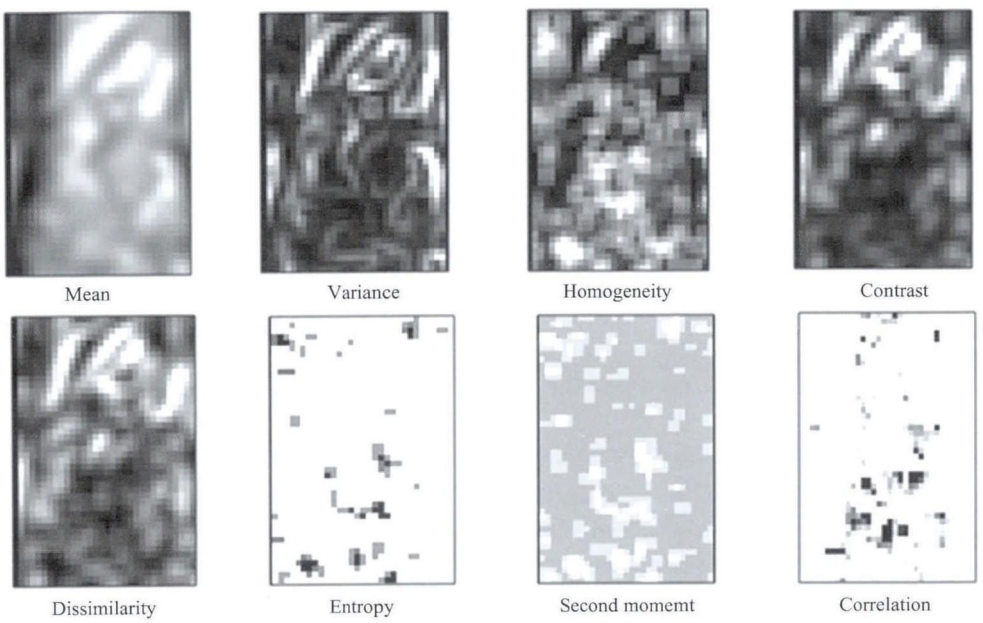

图 6　PC1 图像中得到的八个纹理特征

Fig. 6　Eight texture features extracted from PC1 image

2.4　基于光谱和纹理融合特征的建模

融合光谱特征和纹理特征作为输入建立模型，结果如表 4 所示，可以看出，融合特征的建模效果均优于单独使用光谱或纹理特征的建模效果。除了基于 PCA 提取的特征波长与纹理的融合特征建立的 PLS-DA 模型鉴别率为 86.25% 外，其他模型的鉴别率均大于 90%，其中，BPNN 对预测集样本的鉴别率达到 100%。结果表明将光谱和纹理特征融合可以弥补单一特征的局限性，从而显著提高鉴别准确率。

表 4　基于融合特征的鉴别率(%)

Table 4　Identification rates based on fusion features(%)

	PCA+纹理数据		x-LW+纹理数据		SPA+纹理数据	
	建模集	预测集	建模集	预测集	建模集	预测集
PLS DA	96.88	86.25	96.88	92.50	97.50	91.25
BPNN	100	100	100	100	100	100
LS-SVM	97.50	91.25	99.38	91.25	100	90.00

3　结　论

研究了基于近红外高光谱成像技术的干制红枣快速鉴别方法。采集四个品种干制红枣的近红外高光谱数据，通过 PCA，x-LW 和 SPA 选取特征波长；采用主成分变换结合灰度共生矩阵的方法提取纹理特征。分别基于光谱特征、纹理特征以及光谱与纹理的融合特征建立了 PLS-DA，BPNN 和 LS-SVM 模型。在所有鉴别模型中，基于融合特征的 BPNN 模型取得最理想的效果，其建模集与预测集鉴别率均为 100%。表明基于近红外高光谱特征与纹理特征融合对干制红枣进行快速鉴别是可行的。在今后的研究中将考虑增加更多的红枣品种，建立适用范围更广的干制红枣品种鉴别模型。

References

[1] CHENG Shu-xi, KONG Wen-wen, ZHANG Chu, et al(程术希, 孔汶汶, 张 初, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2014, 34(9): 2519.

[2] Ashabahebwa Ambrose, Lalit Mohan Kandpal, Moon S Kim, et al. Infrared Physics & Technology, 2016, (75): 173.

[3] Wu D, Yang H Q, Chen X Y, et al. Journal of Food Engineering, 2008, 88(4): 474.

[4] Arvin R Yadav, R S Anand R S, Dewal M L, et al. Applied Soft Computing 2015, (32): 101.

[5] Alireza Pourreza, Hamidreza Pourreza, Mohammad-Hossein Abbaspour-Fard, et al. Computers and Electronics in Agriculture, 2012, (83): 102.

[6] Wang Lu, sun Dawen, Pu Hongbin, et al. Food Analytical Methods, 2016, (9): 225.

[7] Macho S, Iusa R, Callao M P, et al. Analytica Chimica Acta, 2001, 445(2): 213.

[8] Kamruzzaman M, Sun D W, ElMasry G, et al. Talanta, 2013, (103): 130.

[9] Liu F, He Y, Wang L. Analytica Chimica Acta, 2008, 615(1): 10.

[10] Wu D, Nie P C, He Y, et al. Analytica Chimica Acta, 2010, 659(1—2): 229.

[11] ZHU Xiao-li(褚小立). Molecular Spectroscopy Analytical Technology Combined with Chemometrics and Its Applications(化学计量学方法与分子光谱分析技术). Beijing: Chemical Industry Press(北京: 化学工业出版社), 2011.

Identification of Varieties of Dried Red Jujubes with Near-Infrared Hyperspectral Imaging

FAN Yang-yang, QIU Zheng-jun*, CHEN Jian, WU Xiang, HE Yong

College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China

Abstract In order to realize rapid identification of dried red jujubes, this paper proposes a method based on near-infrared hyperspectral imaging technology. The near-infrared hyperspectral images (1 000～1 600 nm) of 240 samples in total from 4 cultivars of dried red jujubes will be acquired. The samples are to be divided into the calibration set and the prediction set in the ratio of 2∶1. 7, 8, 10 effective wavelengths are to be selected by principal component analysis(PCA), x-loading weight(x-LW) and successive projection algorithm(SPA) respectively. The dimensionality of original hyperspectral images will be reduced with PCA, and texture features of the first principal component image are to be extracted with gray-level co-occurrence matrix(GLCM). The partial least squares-discriminant analysis(PLS-DA), back propagation neural network(BPNN) and least square support vector machine(LS-SVM) are to be applied to build identification models with the selected effective wavelengths, texture features and fusion of the former two features. The identification rates of the models based on fusion features will be higher than those of models based on the spectral features or texture features respectively. The BPNN models based on the fusion features will obtain the best results, whose identification rates of prediction set are to be 100%. The results in this paper indicate that the near-infrared hyperspectral imaging technology has great potential to identify the dried red jujubes rapidly.

Keywords Near-infrared hyperspectral imaging; Dried red jujube; Identification; Texture features; Features fusion

(Received May 5, 2016; accepted Oct. 23, 2016)

* Corresponding author