

本,再使用卷积平滑(Savitzky-Golay Smoothing, SGS)、标准正态变量变换(Standardized Normal Variate, SNV)、基线校准(Baseline)和去趋势(Detrend)法对原始光谱进行预处理,根据建立的PLSR模型效果优选预处理方法。然后应用CARS、GAPLS和IRF提取特征波长,分别建立基于全光谱和特征波长的PLSR和LS-SVM的鸡蛋新鲜度预测模型;分析比较不同特征波长选取和建模方法对鸡蛋新鲜度的预测效果,选出最优模型,为高光谱成像技术对鸡蛋新鲜度检测提供参考和技术支持。

1 材料与方法

1.1 材料与仪器

鸡蛋:为同批海兰褐鸡所产新鲜蛋,购于宁夏某鸡场,样本共150个,常温储藏;

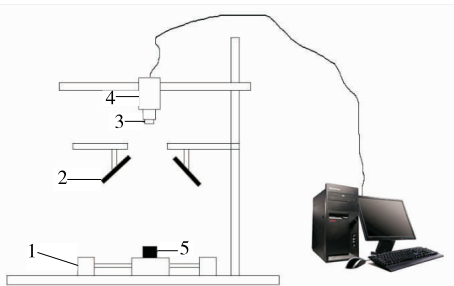
高光谱成像仪:V10E-QE型,芬兰Specim公司;

CCD相机:C8484-05G型,日本Hamamatsu公司;

光纤卤素灯:DCRⅢ型,150 W,美国Schott公司;

电控位移平台:SC300-1A型,北京卓立汉光仪器有限公司;

可见-近红外高光谱成像系统:400~1000 nm,光谱分辨率2.5 nm,125个波段。该系统由高光谱成像仪、CCD相机、光纤卤素灯、电控位移平台组成,见图1。



1. 转动发动机 2. 光源 3. 镜头 4. 相机 5. 样品

图1 可视近红外高光谱成像系统

Figure 1 Vis/NIR hyperspectral imaging system

1.2 试验方法

1.2.1 高光谱图像采集 选取大小均匀,表面无杂物的鸡蛋作为最终试验样本,共选出126个。每天随机取出10个鸡蛋,编号,扫描其高光谱图像。为保证图像清晰,需试验确定高光谱参数,最终确定参数:相机曝光时间设为10 ms,物镜高度为385 mm,电控位移平台速度为15 mm/s。由于暗电流和噪声的影响,需要对采集的高光谱图像进行黑白校正^[13],首先采集聚四氟乙烯板的全反射图像 R_w ,然后盖上镜盖,采集全黑图像 R_d 。黑白校正公式为:

$$I = \frac{R - R_d}{R_w - R_d} \times 100\% \quad (1)$$

式中:

I ——校正后的高光谱反射图像,%;

R ——样本原始的高光谱反射图像;

R_d ——全黑图像;

R_w ——白板的高光谱反射图像。

获得样本高光谱图像后,利用ENVI 4.8软件选取鸡蛋椭圆表面作为感兴趣区域(Region of Interest, ROI),计算出每张ROI的平均反射光谱作为样本的反射光谱。

1.2.2 鸡蛋新鲜度测定 光谱采集后的鸡蛋,按编号逐个放入精度为0.001 g的电子天平称重,然后破壳,用0.02 mm的游标卡尺测量距离蛋黄1 cm处蛋白高度3次,测定时要选准位置,取平均值为最终蛋白高度,代入哈夫值(H_a)公式计算新鲜度^[14-15]:

$$H_a = 100 \times \lg(h - 1.7w^{0.37} + 7.57) \quad (2)$$

式中:

H_a ——哈夫值;

h ——平均蛋白高度,mm;

w ——鸡蛋重量,g。

1.2.3 样本划分方法 采用Galvao等^[16]提出的SPXY(Sample Set Partitioning Based on Joint X-Y distance)法。该方法的优点是将变量 X 和 Y 均考虑在内,能够有效地覆盖多维向量空间,从而改善所建模型的预测能力。

1.2.4 异常样本剔除 由于高光谱成像仪的精度限制和噪声等因素的影响,获得的高光谱图像不可避免地存在一小部分异常样本。异常样本的存在影响了模型稳定性和预测能力,所以剔除异常样本对提高模型效果非常重要。本试验采用蒙特卡洛采样法^[17-18]对全部样本进行异常样本检测。

1.2.5 光谱数据处理方法 鸡蛋新鲜度快速无损检测的实现需要一个稳定性高、预测能力强的模型,本试验选用PLSR和LS-SVM 2种建模方法对鸡蛋新鲜度进行预测。

由于在不同波长下光源的强度不均匀、仪器噪音等因素的影响,需要对原始光谱数据进行预处理。本试验选取SGS、SNV、Baseline和Detrend预处理方法对光谱进行处理。

由于高光谱采集过程受到多种外界因素干扰,获得的光谱信息中会存在一些像基线漂移、噪声等无用信息,并且全光谱共125个波长,信息量大,处理速度慢,选用适当的方法剔除不相关或者非线性变量,实现用少数关键变量代替全光谱,达到降低模型运算量和复杂度、提高模型稳定性和预测准确性的目的。本试验选用CARS、GAPLS和IRF法提取特征波长。其中光谱预处理和建立PLSR模型在The Unscrambler X 10.4软件上实现,其余算法在Matlab R2014a软件上完成。

2 结果与分析

2.1 异常样本检测

采用蒙特卡洛方法检测异常样本,运行之前首先建立基于全部126个样本的PLSR模型,确定RMSECV最小时对应的主成分数最优为11,预处理方法设置为Mean center;抽样次数设置为2500次,蒙特卡洛抽样所得校正集与测试集比例为3:1,预测误差均值和标准差阈值分别取自各自平均值的2.5倍。鸡蛋哈夫值蒙特卡洛异常样本检测结果见图2。

由图2可知,10号、22号、31号、79号和123号的预测

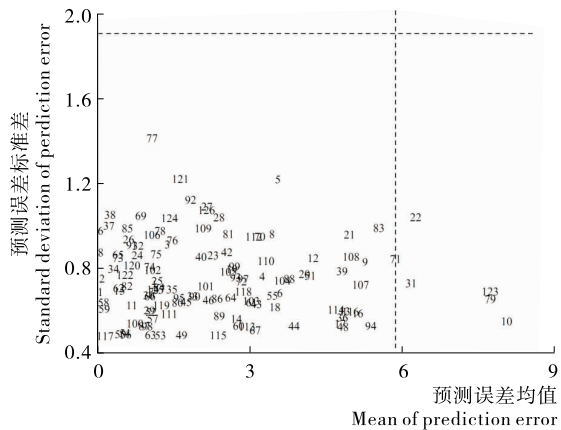


图 2 基于蒙特卡洛方法的鸡蛋哈夫值异常样本检测结果
Figure 2 Outlier detection result of Haugh value of eggs content by MC

表 1 鸡蛋哈夫值及重量的统计数据

Table 1 Statistics data of Haugh value and weight of eggs

样本集	样本数	哈夫值				蛋重/g			
		最小值	最大值	平均值	标准偏差	最小值	最大值	平均值	标准偏差
校正集	90	74.7	94.6	83.1	4.7	45.4	72.1	56.2	6.0
预测集	30	77.4	92.7	83.8	4.3	47.8	69.5	55.5	5.4

2.3 光谱预处理

基于 4 种预处理方法的鸡蛋哈夫值的 PLSR 模型结果统计见表 2。由表 2 可知,所有预处理后的光谱的 PLSR 模型的 R_c 均小于原始光谱的 PLSR 模型,经 SNV 预处理的 PLSR 模型的 RMSECV 最低,且最优主成分数为 7 低于其他模型,说明经 SNV 预处理的 PLSR 模型较为稳定,确定 SNV 为最优预处理方法。图 3 为经 SNV 预处理后的反射光谱。

2.4 特征波长提取

2.4.1 利用 CARS 提取特征波长 运行 CARS 之前,首先确定 PLSR 模型中最优主成分数为 7。设定 CARS 参数:蒙特卡洛采样次数为 200,交叉验证组数为 5。对样本光谱进行筛选,过程见图 4。由图 4(a)可知,随着变量筛选过程的进行,挑选的波长数逐渐下降,速度由快变慢,说明波长变量筛选先粗选、后精选。图 4(b)为变量筛选过程中交互验证均方根误差的变化趋势,可知交互验证均方根误差先降低后增

表 2 基于不同预处理方法的鸡蛋哈夫值 PLSR 模型

Table 2 PLSR model of Haugh value of eggs based on different pretreatment methods

预处理方法	主成分数	RMSECV	校正模型	
			R_c	RMSEC
原始光谱	10	2.670	0.899	2.070
SG-smoothing	10	2.648	0.892	2.142
SNV	7	2.586	0.890	2.156
Baseline	9	2.637	0.896	2.105
Detrend	8	2.846	0.879	2.259

误差均值均大于总体样本预测误差均值阈值,判定上述样本为异常样本,剔除这些样本后,建立 PLSR 模型,所得模型的交互验证系数 R_{cv} 为 0.827 大于原始样本模型(0.793);交互验证均方根误差 RMSECV 为 2.642 小于原始样本模型(2.927)。71 号样本在阈值线上,假定它为异常样本,剔除 71 号样本建立 PLSR 模型发现模型交互验证系数 R_{cv} 为 0.833,交互验证均方根误差 RMSECV 为 2.589。因此,10 号、22 号、31 号、71 号、79 号和 123 号均为异常样本,剔除后剩余 120 个作为后续处理样本。

2.2 划分样本

利用 SPXY 法将样本按 3 : 1 划分校正集和预测集,校正样本 90 个,预测样本 30 个,鸡蛋样本哈夫值及重量统计见表 1。由表 1 可知,校正集样本中哈夫值最大值大于预测集,最小值小于预测集,即校正集哈夫值范围较大,说明划分合理。

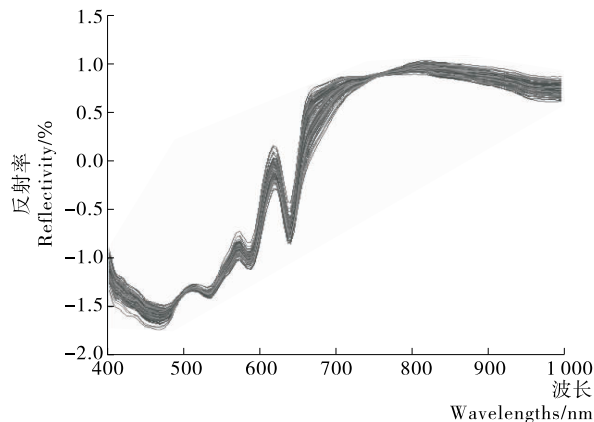


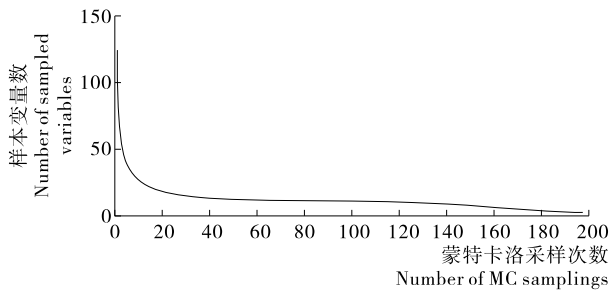
图 3 经过 SNV 处理的反射光谱

Figure 3 Reflection spectra after SNV

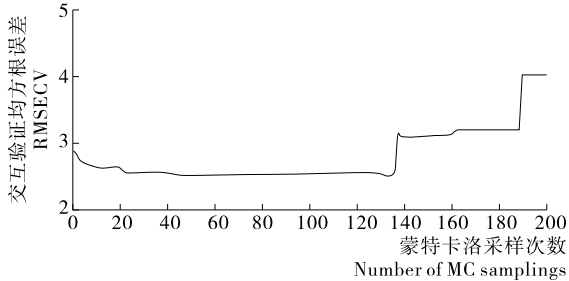
加,MC 采样次数为 131 时,交互验证均方根误差达到最小值 2.459,MC 采样次数继续增加后,交互验证均方误差逐渐增大,再结合筛选过程中波长变量回归系数变化趋势[见图 4(c),图中“*”是建模过程中最小 RMSECV 值对应的采样次数]。最终,在 MC 采样为 131 次时,挑选出 8 个特征波长,分别为:415.7, 449.4, 459.0, 487.8, 588.6, 771.1, 814.3, 996.7 nm。

2.4.2 利用 GAPLS 提取特征波长 设置 GAPLS 参数:种群数为 30,交叉概率为 50%,变异概率为 1%,最大遗传因子为 30,迭代次数 100 次。鸡蛋光谱通过 GAPLS 筛选的有效信息见图 5。运行 GAPLS 时,同时计算不同波长下的 RMSECV 值见图 6。最后结合 RMSECV 值和波长频次数选出最佳波长变量。

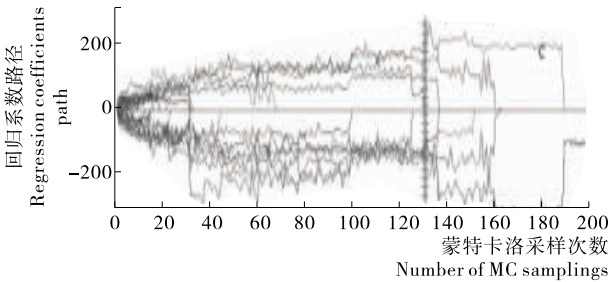
由图6可知,当选出35个波长变量时, RMSECV最低;结合图5中每个波长筛选频数,选出的35个波长的筛选频数 ≥ 7 。说明选出的35个波长与鸡蛋哈夫值相关性高,最终确定这35个波长为特征波长。主要分布在435.0~497.4, 545.4~593.4, 665.4~876.7 nm。



(a) 样本变量筛选过程



(b) 交叉验证均方根误差变化过程



(c) 回归系数路径

* 所对应的位置为131次MC采样

图4 鸡蛋哈夫值的竞争性自适应权重法特征波长筛选过程
Figure 4 Process of competitive adaptive reweighed sampling Characteristic wavelength selection for identification of Haugh value of eggs

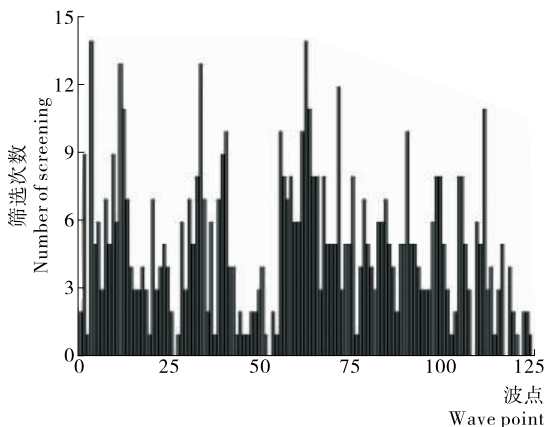
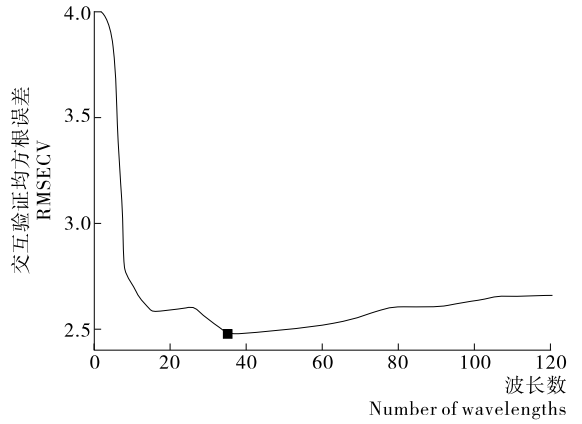


图5 鸡蛋光谱的GA筛选图

Figure 5 GA screening of spectrum of egg

横坐标为125个波点;纵坐标为筛选的频数,频数越高表示适应性越强,与哈夫值相关性越高



最大主成分数为15

图6 不同波数下的交叉验证均方根误差值

Figure 6 Root mean square error of cross validation of different number of wavelengths

2.4.3 利用IRF提取特征波长 设置IRF参数:迭代次数N为1000,间隔宽度W为10,子间隔初始值Q为50,最大主成分数为15。运行IRF,得到116个间隔中排名前10的间隔见表3;同时计算每个间隔的RMSECV,见图7。

由表3可知,前10名区间选出的波点是从22号到49号,但图7显示,当选择前21个间隔时, RMSECV最低,所以选择排名前21个间隔的波长作为特征波长,这些波长是

表3 鸡蛋光谱数据排名前10的波长间隔

Table 3 The top ranked 10 wavelength intervals of egg spectral data

排名	间隔	排名	间隔
1	26~35	6	37~46
2	27~36	7	24~33
3	22~31	8	40~49
4	38~47	9	25~34
5	35~44	10	28~37

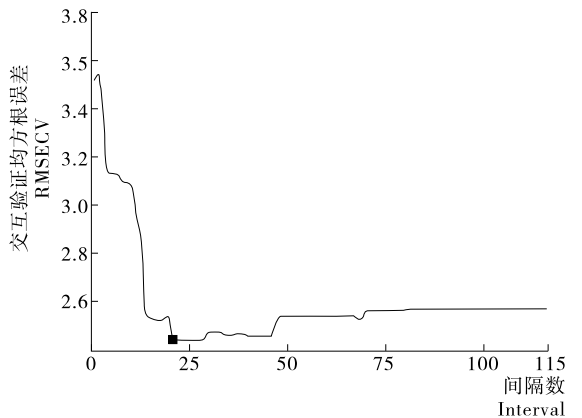


图7 排名从第一到最后—个波长间隔的交叉验证均方根误差值

Figure 7 The Root mean square error of cross validation of the union of the top ranked intervals from 1st to the last

17~50 号、77~93 号、96~107 号、113~123 号,具体是 478.2~636.6, 766.3~843.1, 857.5~910.3, 939.1~987.1 nm,共 74 个波长。

2.5 建模对比

原始光谱经 SNV 预处理后,分别建立基于全波段光谱(Full Spectrum, FS)和特征波长的 PLSR 和 LS-SVM 的鸡蛋新鲜度预测模型,结果见表 4。

建立 PLSR 模型时,首先根据 RMSECV 最低确定最优主成分数,确定 PLSR 建模结果。由表 4 可知,全波段和利

用 CARS、GAPLS、UVE 法提取的特征波长建立的预测模型的最佳主成分数分别是 7,6,6,7。在 PLSR 预测模型中,CARS-PLSR 和 IRF-PLSR 模型的 R_c 均小于 FS-PLSR 模型的,说明此 2 种模型不稳定;GAPLS-PLSR 与 FS-PLSR 模型的 R_c 均为 0.890,前者最优主成分数为 6 低于后者的,但前者的 RMSEC 大于后者的, R_p 为 0.800 小于后者的,说明 GAPLS-PLSR 与 FS-PLSR 校正性能一样,但预测能力弱于后者,说明 GAPLS-PLSR 较 FS-PLSR 模型不稳定。故认定 FS-PLSR 模型是所建立的 PLSR 模型中的最优模型。

表 4 基于全光谱和特征波长的 PLSR 和 LS-SVM 鸡蛋新鲜度预测模型

Table 4 Predicted value versus real value of egg freshness using different modeling methods

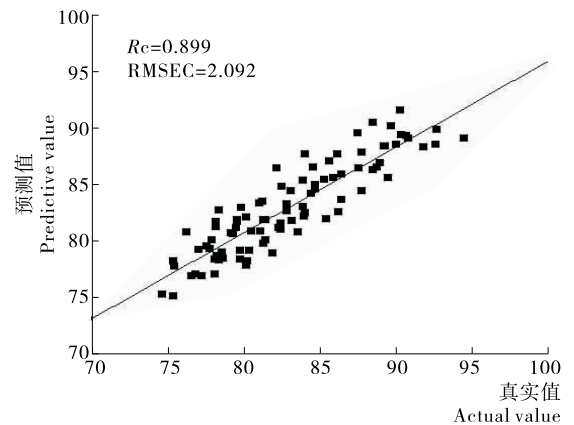
建模方法	特征波长提取方法	主成分数	波长数	校正集		预测集	
				R_c	RMSEC	R_p	RMSEP
PLSR	FS	7	125	0.890	2.156	0.813	2.546
	CARS	6	8	0.817	2.323	0.824	2.486
	GAPLS	6	35	0.890	2.162	0.800	2.638
	IRF	7	74	0.874	2.300	0.803	2.606
LS-SVM	FS	无	125	0.900	2.080	0.818	2.516
	CARS	无	8	0.880	2.259	0.832	2.430
	GAPLS	无	35	0.899	2.092	0.832	2.423
	IRF	无	74	0.907	2.012	0.814	2.537

由表 4 可知,IRF-LS-SVM 具有最大的 R_c ,但有最低的 R_p ,说明该模型的校正性能最优,预测能力最差;CARS-LS-SVM 和 GAPLS-LS-SVM 的 R_p 均为 0.832 且最大,但前者的 R_c 明显低于后者以及 FS-LS-SVM 模型的,说明 CARS-LS-SVM 模型预测能力较强,但校正性能较差;GAPLS-LS-SVM 的 R_c 为 0.899 略低于 FS-LS-SVM 模型的,前者的 RMSEC 略大于后者的,但 GAPLS-LS-SVM 模型预测集的 R_p 明显大于 FS-LS-SVM 模型的,且 RMSEP 最低,说明该模型校正性能较好,预测能力最强,模型稳定。最后确定 GAPLS-LS-SVM 为 LS-SVM 中的最优模型。

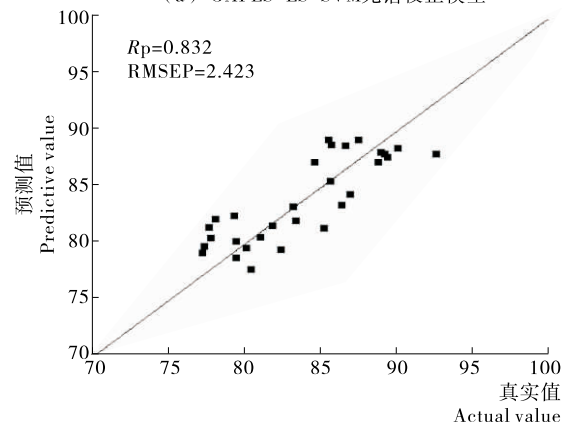
对比特征波长方法,CARS 法提取波长数最少,降维能力最强,GAPLS 法次之,IRF 法最差,但基于 CARS 特征波长建立的 PLSR 和 LS-SVM 模型效果均较差,GAPLS-LS-SVM 模型效果较优。对比 PLSR 和 LS-SVM 2 种模型中的最优模型,GAPLS-LS-SVM 的 R_c 和 R_p 均大于 FS-PLSR 模型的;且前者的 RMSEC 和 RMSEP 均小于后者的,故 GAPLS-LS-SVM 模型效果优于 FS-PLSR。本研究最终确定 GAPLS-LS-SVM 为最优模型,且 GAPLS 提取的光谱能代替全光谱建模,模型效果见图 8。

3 结论

采集后的原始鸡蛋光谱经 4 种预处理方法,对比其 PLSR 模型效果,确定 SNV 法为最优预处理方法;对 SNV 预处理后的光谱经 CARS、GAPLS 和 IRF 法提取特征波长,分别获得 8,35,74 个特征波长,CARS 法降维效果最佳,GAPLS 法次之,IRF 法最差;分别建立基于特征波长和全光



(a) GAPLS-LS-SVM 光谱校正模型



(b) GAPLS-LS-SVM 光谱验证模型

图 8 GAPLS-LS-SVM 建模结果

Figure 8 GAPLS-LS-SVM modeling results

谱的 PLSR、LS-SVM 模型,结果表明,在 PLSR 模型中,FS-PLSR 模型较好,GAPLS-LS-SVM 模型效果最优;在 LS-SVM 模型中,GAPLS-LS-SVM 效果最优,利用 GAPLS 法提取的特征波长代替全波段建立 LS-SVM 模型是可行的。对比 2 种建模方法,确定 GAPLS-LS-SVM 鸡蛋新鲜度预测模型最优,其 R_c 为 0.899, RMSEC 为 2.092; R_p 为 0.832, RMSEP 为 2.423。

参考文献

[1] 毕夏坤, 赵杰文, 林颖, 等. 便携式近红外光谱仪判别鸡蛋的贮藏时间[J]. 食品科学, 2013(22): 281-285.

[2] 朱云鹏, 崔春利, 王兰娇, 等. 哈尔滨市售鸡蛋品质及营养成分分析[J]. 食品工业, 2017(5): 289-292.

[3] 王巧华, 任奕林, 文友先. 基于 BP 神经网络的鸡蛋新鲜度无损检测方法[J]. 农业机械学报, 2006(1): 104-106.

[4] 邢志中, 张海东, 王孟, 等. 基于计算机视觉和神经网络的鸡蛋新鲜度检测[J]. 江苏农业科学, 2017(11): 160-163.

[5] 赵杰文, 毕夏坤, 林颖, 等. 鸡蛋新鲜度的可见-近红外透射光谱快速识别[J]. 激光与光电子学进展, 2013(5): 213-220.

[6] LIN Hao, ZHAO Jie-wen, SUN Li, et al. Freshness measurement of eggs using near infrared (NIR) spectroscopy and multivariate data analysis[J]. Innovative Food Science & Emerging Technologies, 2011, 12(2): 182-186.

[7] GIUNCHI A L, BERARDINELLI A N, RAGNI L, et al. Non-destructive freshness assessment of shell eggs using FT-NIR spectroscopy[J]. Journal of Food Engineering, 2008, 89(2): 142-148.

[8] 李佳婷, 王俊, 李园, 等. 基于电子鼻的鸡蛋新鲜度检测[J]. 现代食品科技, 2017(4): 300-305, 188.

[9] SMITH D P, LAWRENCE K C, HEITSCHMIDT G W.

Fertility and embryo development of broiler hatching eggs evaluated with a hyperspectral imaging and predictive modeling system[J]. International Journal of Poultry Science, 2008, 7(10): 1 001-1 004.

[10] 王巧华, 周凯, 吴兰兰, 等. 基于高光谱的鸡蛋新鲜度检测[J]. 光谱学与光谱分析, 2016(8): 2 596-2 600.

[11] JOHNSON III Owen N, SLIDELL Mark, KREISHMAN Peter, et al. Hyperspectral imaging: an emerging technology as a potential novel adjunct in assessing peripheral perfusion deficits and success of lower extremity revascularizations[J]. Journal of the American College of Surgeons, 2008, 207(3): S114.

[12] 吴龙国, 何建国, 贺晓光, 等. 高光谱图像技术在水果无损检测中的研究进展[J]. 激光与红外, 2013, 43(9): 990-996.

[13] ELMASRY G, WANG Ning, ELSAYED A. Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry[J]. Journal of Food Engineering, 2007, 81: 98-107.

[14] 牡丹萌, 王风诺, 王世平. 鸡蛋新鲜度随储藏条件变化规律的研究[J]. 食品科技, 2014, 39(5): 26-29, 33.

[15] 王巧华, 李小明, 段宇飞. 基于 CUVE-PLS-DA 的鸡蛋新鲜度在线检测分级[J]. 食品科学, 2016(22): 187-191.

[16] GALVÃO Roberto Kawakami Harrop, MÁRIO César Ugulino Araujo, GLEDSON Emidio Jose, et al. A method for calibration and validation subset partitioning[J]. Talanta, 2005, 67: 736-740.

[17] 张明锦, 杜一平. 蒙特卡罗-偏最小二乘回归系数法用于近红外光谱变量筛选[J]. 分析试验室, 2013(2): 12-16.

[18] 刘善梅. 基于高光谱成像技术的冷鲜猪肉品质无损检测方法研究[D]. 武汉: 华中农业大学, 2015; 26-41.

(上接第 130 页)

际的生产中需要对滚动的芒果进行实时连续的图像采集, 保证实现全方位的检测, 提高检测准确率。针对此不足之处, 下一步的研究需要针对芒果全部区域进行检测, 由综合试验结果判断芒果品质。

参考文献

[1] 刘静, 黄勇平, 章程辉. 视觉系统开发模块在芒果果面缺陷检测中的应用[J]. 食品与机械, 2009, 25(2): 82-85.

[2] 李甦, 谭永龙, 杨美英. 芒果分级与表面缺陷检测研究[J]. 计算机工程与设计, 2008(15): 3 954-3 957.

[3] 康志亮, 陈韵羽, 王思, 等. 便携式受损芒果检测装置的设计[J]. 农机化研究, 2010(12): 52-56.

[4] 赵杰文, 刘剑华, 陈全胜, 等. 利用高光谱图像技术检测水果轻微损伤[J]. 农业机械学报, 2008(1): 106-109.

[5] 党宏社, 宋晋国, 郭琴. 基于 ARM 的嵌入式芒果大小检测与分级系统的实现[J]. 四川农业大学学报, 2011(1): 89-93.

[6] 朱明, 陆小锋, 陆亨立, 等. AdaBoost 人脸检测算法在 DSP 上的移植与优化[J]. 计算机工程与应用, 2014(20): 197-201, 232.

[7] 吕颖. DSP 图像数据的可视化[J]. 福建电脑, 2011(3): 153-155.

[8] 肖剑雄峰. 局部对比度增强的彩色图像灰度化参数化算法研究[D]. 温州: 温州大学, 2016: 9-16.

[9] 姜小磊, 姚鸿勋, 赵思成. 一种极值约束的边缘保持图像平滑算法[J]. 计算机科学, 2014(10): 101-105, 127.

[10] 张琪. 结合边缘检测的图像二值化算法[D]. 长春: 吉林大学, 2011: 24-30.

[11] 张建光, 李永霞. 基于拉普拉斯边缘检测算子的图像分割[J]. 福建电脑, 2011(7): 99, 101.

[12] 周林妹. 数字图像边缘检测算法及其在农产品加工中的应用[J]. 食品与机械, 2009, 25(3): 139-142, 153.

[13] 赵源萌, 王岭雪, 金伟其, 等. 基于区域直方图统计的灰度图像色彩传递方法[J]. 北京理工大学学报, 2012(3): 322-326.

[14] 范晔. 基于 DM6437 的车牌识别系统的设计与实现[D]. 西安: 西安电子科技大学, 2012: 61-62.

[15] 杨方. 基于 TMS320C6678 的多核 DSP 并行处理应用技术研究[D]. 北京: 北京理工大学, 2014: 47-50.

[16] 刘德方, 邓明, 陈海燕, 等. 基于 DM6437 和 VLIB 的 Canny 边缘检测[J]. 安徽建筑工业学院学报: 自然科学版, 2013(2): 72-75.

[17] 黄震, 刘亚斌. TMS320C6000 系列 DSP 程序固化的研究[J]. 电子设计工程, 2016(12): 26-28, 32.