

高光谱技术融合图像信息的牛肉品种识别方法研究

王彩霞, 王松磊*, 贺晓光, 董欢

宁夏大学农学院, 宁夏 银川 750021

摘要 高光谱图像包含了大量的光谱信息和图像信息, 采用高光谱成像技术对牛肉品种进行识别。获取可见-近红外(400~1 000 nm)光谱范围内的安格斯牛、利木赞牛、秦川牛、西门塔尔牛、荷斯坦奶牛五个品种共 252 个牛肉样本的高光谱图像。在 ENVI 软件中对高光谱图像进行阈值分割并构建掩膜图像, 获取样本的感兴趣区域(ROI), 并结合伪彩色图对牛肉样本的反射率指数进行可视化表达; 采用 Kennard-Stone(KS)法对样本集进行划分以提高模型的预测性能; 对原始光谱采用卷积平滑(SG)、区域归一化(Area normalize)、基线校正(Baseline)、一阶导数(FD)、标准正态变量变换(SNV)及多元散射校正(MSC)等 6 种方法进行预处理; 采用竞争性自适应重加权算法(CARS)提取特征波长。然后利用颜色矩对不同牛肉样本的颜色特征进行提取; 对原始光谱图像进行主成分分析, 结合灰度共生矩阵(GLCM)算法, 提取主要纹理特征。最后结合偏最小二乘判别(PLS-DA)算法建立牛肉样本基于特征波长、颜色特征以及纹理特征的识别模型。KS 法将牛肉样本划分为校正集 190 个, 预测集 62 个; 将未经预处理的光谱数据与经过 6 种不用预处理的光谱数据进行建模分析, 结果发现经 FD 法处理后的光谱数据所建模型的识别率最高; 结合 CARS 法对经 FD 法预处理后的光谱数据进行特征波长提取, 共提取出 22 个波长; 利用颜色矩和 GLCM 算法分别提取出每个牛肉样本的 9 个颜色特征、48 个纹理特征。将特征波长数据与颜色、纹理特征信息进行融合建模, 结果表明, 基于特征光谱+纹理特征的模型识别效果最佳, 其校正集与预测集识别率分别为 98.42% 和 93.55%, 均高于特征光谱数据模型识别率, 说明融合纹理特征后使样本分类信息的表达更加全面; 融合颜色特征后模型的校正集识别率均有所增加, 但预测集识别率稍逊, 颜色特征虽携带了部分有效信息, 但这些信息与牛肉样本的相关性不大。因此, 寻找与牛肉样本相关性更大的颜色特征是提高模型识别率的重要途径之一。该研究结果为牛肉品种的快速无损识别提供了一定的参考。

关键词 高光谱成像技术; 图像特征; 牛肉品种; 识别; 偏最小二乘判别

中图分类号: TP391.41 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)03-0911-06

引言

牛肉味道鲜美、营养丰富, 是我国消费最为普遍的肉制品之一。不同品种的牛肉在口感和品质上存在很大差异, 但肉品性状和颜色又极为相似, 肉眼难以区分。为保护一些优良的牛肉品种及消费者利益, 需对不同品种的牛肉进行分类识别。传统的肉类识别方法操作繁琐、耗时耗力^[1-3], 难以满足现实需求。因此, 建立牛肉品种的快速无损识别方法, 具有重要的现实意义。

高光谱成像技术具有连续多波段、高分辨率和图谱合一等特点。融合高光谱成像技术的光谱信息和图像信息, 可

以提高模型的准确性和可靠性^[4-6]。赵娟等^[7]基于高光谱图像纹理特征建立的牛肉嫩度判别模型, 其预测集判别精度达 94.44%; Liu^[8]等结合偏最小二乘回归算法(PLSR)建立了基于光谱、纹理及融合信息的腌肉 pH 预测模型, 结果表明, 基于数据融合的预测样本的决定系数(R_p^2)为 0.794, 优于仅基于光谱(R_p^2)或纹理(R_p^2)的模型结果。但当过多的特征直接叠加时, 会产生大量冗余信息, 影响结果的稳定性。因此, 本研究在利用较少的特征波长表征光谱信息的基础上, 结合图像处理技术, 分别对牛肉样本的纹理特征及颜色特征进行提取分析, 探讨基于高光谱图谱融合技术对牛肉品种分类识别的可行性。

收稿日期: 2019-03-08, 修订日期: 2019-06-12

基金项目: 国家自然科学基金项目(31660484)资助

作者简介: 王彩霞, 1994 年生, 宁夏大学农学院硕士研究生 e-mail: 18295671639@163.com

* 通讯联系人 e-mail: wangsonglei163@126.com

1 实验部分

1.1 样本与处理

采集 3 岁左右的安格斯牛、利木赞牛、秦川牛、西门塔尔牛、荷斯坦奶牛各 3~5 头,肉样采自宁夏固原市与宁夏吴忠市。牛经屠宰后在 0 °C 下冷藏 48 h 完成排酸处理。排酸结束后,对牛肉样品进行分割。取 5 个部位肉(脖肉、眼肉、里脊肉、瓜条肉、臀肉),放入保温箱运至实验室,贮藏在 4 °C 冷柜备用。光谱扫描前将肉样整形切块(40 mm×30 mm×10 mm),室温下放置 2 h,待肉样中心温度达到室温水平后,用滤纸吸干样品表面的水分,依次进行光谱扫描。本实验共获得牛肉样本 252 个(其中,安格斯牛、利木赞牛、秦川牛、西门塔尔牛、荷斯坦奶牛样本数分别为 59, 31, 62, 34 和 66 个)。

1.2 光谱采集与校正

实验采用美国 Headwall Photonics 公司生产的 Hyper-Spec Vis-NIR 高光谱成像系统。主要包括光谱相机、光源、VT-80 精密电控位移平台、计算机和数据采集软件等。其中,光谱相机由 Inspector N 系列成像光谱仪和 G4-232 增强型 EMCCD 相机组成,成像光谱仪在 400~1 000 nm 波段范围内共有 125 个波段,光谱分辨率为 2.8 nm,入射狭缝宽度为 25 μm;光源系统由 2 个 150 W 的光纤式卤素灯光源组成;电控位移平台背景设置为黑色以消除反射杂散光干扰。实验采用 Hyperspec-N 软件控制高光谱成像系统采集牛肉样本的成像信息。

光谱扫描前需进行黑白校正,并设置合理的采集参数。经实验确定采集参数为:相机曝光时间 30 ms,物距 380 mm,扫描速度 15 mm·s⁻¹,线扫描实际长度 60 mm。

1.3 数据处理

1.3.1 感兴趣区域选取及光谱反射率指数可视化

光谱反射率指数可视化指利用不同颜色将每个像素点的化学成分差异显现出来,并产生每个样本的平均光谱反射率值。在进行光谱反射率指数可视化之前,需提取合理的感兴趣区域(ROI)。ROI 区域的选取采用 ENVI 软件中的波段阈值进行设置,经多次尝试,最终设定二值化阈值为 0.25。对

采集到的图像依次进行阈值分割,得到二值化掩膜图像用于确定 ROI。然后将 ROI 图像与高光谱图像进行耦合,生成光谱反射率指数的可视化图^[9],具体方法见参考文献[9]。

1.3.2 光谱预处理及特征波长提取

采用 KS 算法对样本集进行划分。由于原始光谱含有大量的噪声及无关信息,需对划分后的样本数据进行预处理。所尝试的预处理方法有卷积平滑(SG)、一阶导数(FD)、区域归一化(Area normalize)、基线校正(Baseline)、标准正态变量变换(SNV)及多元散射校正(MSC)等。由于牛肉样本的全波段光谱数据量大、信息混杂,需选用适当的特征波长提取方法剔除不相关或者非线性变量,降低模型运算量、提高模型稳健性。本工作采用 CARS 法^[10]进行特征波长提取。

1.3.3 图像主成分分析

主成分分析(PCA)法是根据方差最大原则,沿协方差最大的方向将高维变量投射到低维空间,从而得到能够表征高维数据空间和信息的低维分量的过程^[11]。采用基于 SVD 的 PCA 降维后协方差贡献率的大小来确定主成分数目。

1.3.4 颜色及纹理特征提取

图像中的颜色矩能有效表征图像中颜色的分布,颜色矩包括一阶矩、二阶矩和三阶矩。利用 R, G 和 B 三个颜色分量的一到三阶矩来表示不同牛肉样本图像的颜色特征。采用灰度共生矩阵法(GLCM)^[12]提取图像纹理信息,利用 Matlab 中的 graycomatrix 函数,设置像素间距离参数值为 1,对主成分图像依次取 0, 45°, 90° 和 135° 方向的能量、熵、惯性矩和相关性^[13]进行纹理特征提取。

1.4 建模方法

采用偏最小二乘判别(PLS-DA)^[14]法建立牛肉品种识别模型。PLS-DA 算法是在偏最小二乘回归算法基础上建立样本分类变量与光谱特征间的回归模型的分方法,适用于变量较多且干扰噪声大的情况,能有效解决多重共线性问题。

2 结果与讨论

2.1 光谱反射率曲线及反射率指数可视化

利用 ENVI 软件对高光谱图像进行阈值分割提取 ROI,并对光谱反射率指数进行可视化表达,结果如图 1 所示。

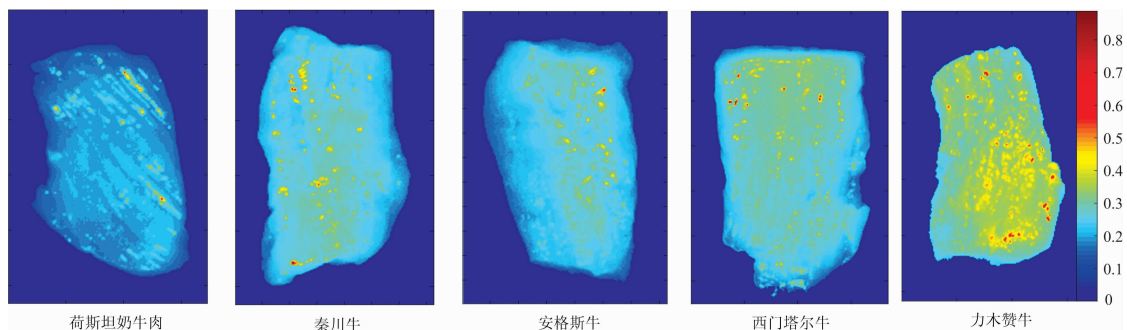


图 1 牛肉样本光谱反射率指数可视化图

Fig. 1 The visualization of spectral reflectance index in color map of beef samples

图 1 中颜色越接近红色表示反射率指数越大,反之越小。荷斯坦奶牛样本反射率指数较低因而蓝色区域较多,而

利木赞牛样本中出现明显的黄色与少量红色,表明反射率指数较大。由于光谱反射率指数与样本的化学成分之间存在一

定的相关性,因此利木赞牛较荷斯坦奶牛,前者化学成分更丰富。对不同牛肉品种的平均光谱反射率曲线进行对比分析,结果如图 2 所示。

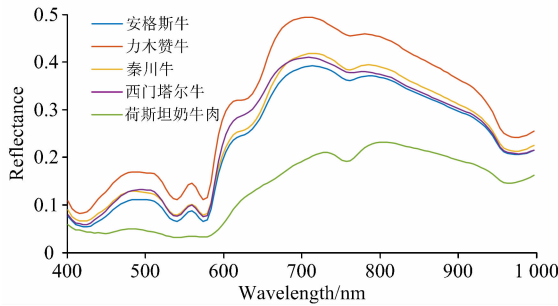


图 2 牛肉样本的平均光谱曲线

Fig. 2 The mean spectrum curves of beef samples

由图 2 可知,荷斯坦奶牛样本的平均光谱曲线较其他四种牛肉样本曲线,波峰波谷位置不明显且出现部分偏移现象,这与样本所含的化学成分有很大关系。牛肉含有大量的水分($\geq 70\%$)、脂肪和蛋白质,因此光谱吸收主要与其所含的—OH,—CH 和—NH₂ 等基团紧密相关。由图可知,在 400~580 nm 可见波段范围内光谱反射率值较低,在 610~780 nm 短波近红外区域内反射率较高。因在可见光区域,肉样中肌红蛋白与血红蛋白相互作用使样本表面呈红色,其互补色青绿色区域光谱吸收率较高,反射率较低;而在近红外区域,光谱吸收与物质分子基团能量吸收及能级跃迁有关。在部分波段,秦川牛与安格斯牛光谱出现重叠和谱峰交叉现象,但在 558~665 nm 波段,各曲线反射率差异较明显,这些波段为牛肉样本的识别提供了大量的有效信息。图 1 中反射率指数较低的样本在图 2 中其光谱反射率值也较低,因此,反射率指数也可以反应每个样本的平均光谱反射率。

2.2 光谱数据处理

2.2.1 光谱信息预处理

利用 KS 法按接近 3 : 1 的比例将牛肉样本划分为校正集 190 个,预测集 62 个。对划分后样本的光谱数据进行预处理,结果见表 1。

表 1 不同预处理方法的 PLS-DA 分类结果

Table 1 The results of PLS-DA models by different pretreatment methods

预处理方法	校正集		预测集	
	误判数 /个	识别率 /%	误判数 /个	识别率 /%
无	15	92.11	16	74.19
SG	15	92.11	10	83.87
Area normalize	9	94.37	9	82.25
Baseline	14	92.63	10	83.87
FD	6	96.84	5	91.94
SNV	10	94.74	10	83.87
MSC	9	94.37	7	88.71

由表 1 可知,与未经预处理的原始光谱识别结果相比,经预处理后模型校正集与预测集的识别率均有所增加。其中,经 FD 法预处理后的光谱数据其校正集与预测集识别率最高,分别为 96.84% 和 91.94%,说明 FD 预处理法可以有效消除基线漂移及背景干扰,提升了谱峰分辨率和灵敏度,提高了识别准确率。因此,FD 法为最佳预处理方法。

2.2.2 特征波长提取

CARS 法提取特征波长具体参数设置为:最大主成分数为 15,蒙特卡洛交互验证中采样次数为 50 次,采样率为 0.8,数据选用中心化处理方式,迭代次数为 2 000,阈值为 0.8,处理过程如图 3 所示。

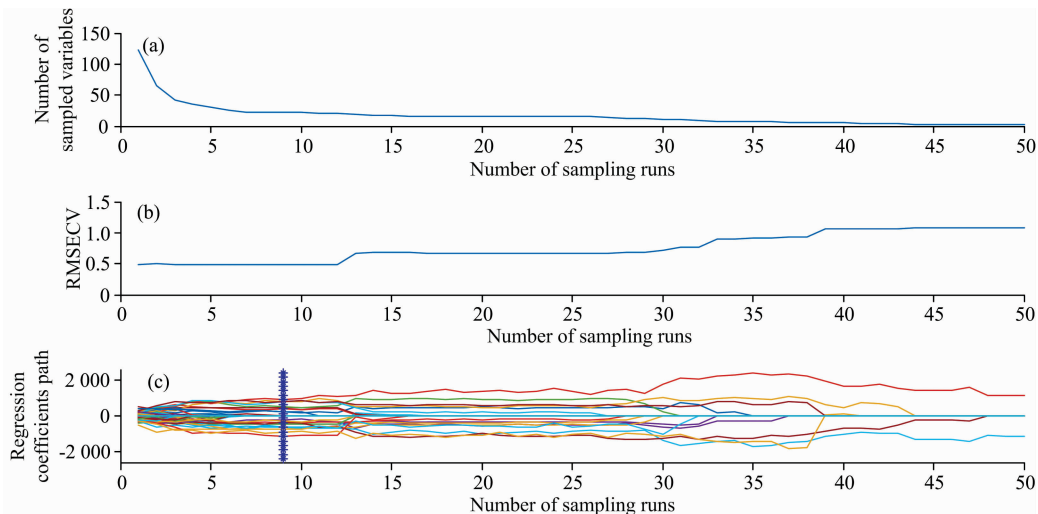


图 3 CARS 方法特征波长筛选过程

(a): 变量选择变化趋势; (b): 交互验证均方根误差变化过程; (c): 波长变量回归系数变化趋势

Fig. 3 Process of CARS characteristic wavelength selection

(a): Variation trend of variable selection; (b): Variation process of RMSECV;

(c): Variation trend of regression coefficient of wavelength variable

图 3(a)为变量选择变化趋势图,变量数随采样次数呈先快后慢的下降趋势;图 3(b)反映了筛选过程中交互验证均方根误差(RMSECV)的变化过程,RMSECV先降低后上升,最优化变量子集应位于欠拟合和过拟合交汇点处,即采样数为 9 次时,RMSECV 的值最小为 0.493 9;图 3(c)表示特征波长变量回归系数的变化趋势,“*”所对应的位置即为 RMSECV 最小处,依据 RMSECV 最小值原则共选出 22 个特征波长。分别为:449,469,473,483,526,574,589,598,613,622,694,709,733,747,761,781,862,910,915,934,949 和 973 nm,数据压缩率为 82.4%。

2.3 图像特征分析

2.3.1 图像主成分分析

利用 ENVI 软件对掩膜后的牛肉样本图像进行主成分分析。提取前 3 幅主成分(累计方差贡献率达 99%以上),主成分图像如图 4 所示。

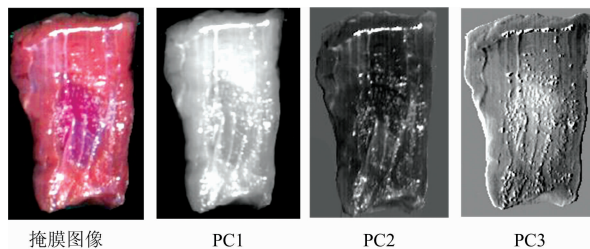


图 4 牛肉样本的前 3 个主成分图像

Fig. 4 Images of first three PCs of beef samples

2.3.2 颜色及纹理特征提取

对 252 个牛肉样本依次进行图像特征提取。提取掩膜图像的颜色特征,数据保存在 252×9 的矩阵中;提取前 3 幅主成分图像的纹理特征,数据保存在 252×48 的矩阵中。由于纹理特征数据维度较大,将提取的纹理特征数据做相关性分析以提取更加有效的信息。结果如表 2 所示。

表 2 主成分图像的纹理特征与牛肉品种的相关性分析

Table 2 Correlation between beef varieties and texture parameters from principal component images

PCs	能量				熵				惯性矩				相关性			
	0	45°	90°	135°	0	45°	90°	135°	0	45°	90°	135°	0	45°	90°	135°
PC1	-0.47	-0.46	-0.48	-0.45	0.96	0.95	0.96	0.96	0.92	0.90	0.91	0.95	0.68	0.67	0.63	0.68
PC2	-0.91	-0.91	-0.90	-0.93	0.68	0.66	0.66	0.68	0.43	0.36	0.35	0.40	-0.30	-0.27	-0.28	-0.31
PC3	-0.77	-0.76	-0.77	-0.76	0.90	0.88	0.87	0.90	0.62	0.52	0.50	0.61	-0.71	-0.68	-0.70	-0.69

由表 2 可知,PC1 图像中当纹理特征为熵和惯性矩时,其值与牛肉种类呈正相关,且相关性均大于 0.90,因此,选取 PC1 图像中的熵和惯性矩作为最终纹理特征。以此类推,PC2 图像选取能量作为纹理特征值,PC3 图像选取熵作为最终纹理特征值,最终选取 16 个纹理特征,用于后续建模分析。

2.4 特征光谱与纹理数据融合建模分析

将特征光谱数据与颜色、纹理特征信息进行融合,利用 PLS-DA 算法对融合后的数据进行建模。PLS-DA 算法中设置最大主成分数为 20,进行数据归一化处理,并采用百叶窗交互验证,设置交互验证组数为 10。建模结果如表 3 所示。

表 3 4 种模型对牛肉样本的识别准确率

Table 3 Identification accuracy of four models for beef samples

模型	校正集		预测集	
	误判数/个	识别率/%	误判数/个	识别率/%
特征光谱数据	8	95.69	6	90.23
特征光谱数据+纹理特征	3	98.42	4	93.55
特征光谱数据+颜色特征	6	96.84	8	87.10
特征光谱数据+纹理特征+颜色特征	4	97.87	5	91.94

由表 3 可知,在 4 个模型中,基于特征光谱+纹理特征的模型识别效果最好,校正集与预测集识别率分别为 98.42%和 93.55%,说明纹理特征与牛肉品种之间较高的相关性在经过融合后使样本的信息量更加全面,因此模型的识别率提高;特征光谱+颜色特征模型校正集的识别率略高于单独的特征光谱模型,但预测集识别率较低,分析其原因,颜色特征虽携带了部分有效信息,但与牛肉品种的相关性较差,融合的同时也增加了部分无用信息,从而导致预测集识别率降低;这也是融合了颜色特征后,特征光谱+颜色特征

+纹理特征模型的识别率降低的原因。

3 结论

利用高光谱光谱和图像特征相融合的技术对牛肉品种进行识别研究,通过 CARS 法对经过 FD 法预处理后的光谱进行特征波长提取,共提取出了 22 个有效波长,降低了光谱数据量;通过主成分分析法取出前 3 幅主成分图像,然后利用颜色矩、GLCM 提取图像的颜色以及纹理信息;最后,分别

利用特征波长、颜色特征、纹理特征融合建立牛肉品种 PLS-DA 识别模型。主要结论:

(1) 基于特征光谱+纹理特征的模型识别效果最佳, 校正集与预测集识别率分别为 98.42% 和 93.55%, 均高于特征光谱数据模型识别率, 说明融合纹理特征使分类信息的表达更加全面。

(2) 融合颜色特征后的模型校正集识别率均有所增加, 但预测集识别率较差。可能原因是颜色特征虽携带了部分有效信息, 但这些信息与牛肉样本的相关性不大, 颜色特征的融合增加了部分冗余信息, 反而使识别率降低。因此, 寻找与牛肉样本相关性更大的颜色特征是提高模型识别率的重要途径。

References

- [1] Montowska M, Fornal E. *Food Chemistry*, 2017, 237: 1092.
- [2] Vlachos A, Arvanitoyannis I S, Tserkezou P. *Critical Reviews in Food Science and Nutrition*, 2016, 56(7): 1061.
- [3] Ali M E, Razzak M A, Hamid S B A, et al. *Food Chemistry*, 2015, 177: 214.
- [4] LIU Huan, WANG Ya-qian, WANG Xiao-ming, et al(刘欢, 王雅倩, 王晓明, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2019, 39(1): 223.
- [5] LU Bing, SUN Jun, MAO Han-ping, et al(芦兵, 孙俊, 毛罕平, 等). *Jiangsu Journal of Agricultural Sciences(江苏农业学报)*, 2018, 34(6): 1254.
- [6] ZHANG Shuai-tang, WANG Zi-yan, ZOU Xiu-guo, et al(张帅堂, 王紫烟, 邹修国, 等). *Transactions of the Chinese Society of Agricultural Engineering(农业工程学报)*, 2017, 33(22): 200.
- [7] ZHAO Juan, PENG Yan-kun(赵娟, 彭彦昆). *Transactions of the Chinese Society of Agricultural Engineering(农业工程学报)*, 2015, 31(7): 279.
- [8] Liu D, Pu H, Sun D, et al. *Food Chemistry*, 2014, 160: 330.
- [9] Su W, Sun D. *Computers and Electronics in Agriculture*, 2016, 125: 113.
- [10] LIU Jin, XU Wen-li, SUN Tong, et al(刘津, 许文丽, 孙通, 等). *Chinese Journal of Analysis Laboratory(分析实验室)*, 2018, 37(1): 1.
- [11] HAN Man-li, HOU Wei-min, SUN Jing-guo, et al(韩嫚莉, 侯卫民, 孙靖国, 等). *Journal of University of Electronic Science and Technology of China(电子科技大学学报)*, 2019, 48(1): 117.
- [12] Haralick R M S K D I. *Studies in Media and Communication*, 1973, SMC-3(6): 610.
- [13] Huang X, Liu X, Zhang L. *Remote Sensing*, 2014, 6(9): 8424.
- [14] Su W, Sun D. *Talanta*, 2016, 155: 347.

The Identification of Beef Varieties by Fusing Image Information Based on Hypersepctral Image Technology

WANG Cai-xia, WANG Song-lei*, HE Xiao-guang, DONG Huan
School of Agriculture, Ningxia University, Yinchuan 750021, China

Abstract In this study, beef variety was identified by hyperspectral imaging technology which contains abundant spectral and spatial information in an object. Firstly, hyperspectral images of beef samples in the visible and near infrared (400-1000 nm) regions were acquired by the hyperspectral imaging system which contain 252 samples of five varieties of Angus, Limuzan, Qinchuan, Simmental, and Holstein cows. The binary mask image was successfully determined with a certain threshold from ENVI, and ROI (Region of Interest) of beef sample was determined by using the binary mask image. The visual distribution map of reflectance index in beef sample was plotted by pseudo-color map. Samples were divided by using KS method, which is to improve the prediction performance of the model; The spectral pretreatment method was utilized, such as SG, Area normalize, Bseline, FD, SNV, MSC and so on; Feature wavelengths were extracted by using competitive adaptive weighting algorithm (CARS). The color characteristics were represented by used color moment for different beef sample images; Principal component analysis was performed on the original hyperspectral image. The image textural information was described by extracting main texture features by the gray level co-occurrence matrix (GLCM) algorithm of the beef sample. Then spectral data from CARS, color feature and texture feature (from three principle component images) were utilized to develop different partial least squares discrimination (PLS-DA) models to identify beef samples respectively. The samples were divided into calibration set and prediction set by KS method, and calibration samples was 190, and prediction samples was 62; The spectral pretreatment was studied by the 7

methods. The results showed that the model effect of FD methods pretreatment was the best; A total of 22 characteristic wavelengths were extracted by the CARS method for spectral data using FD method; A total of 9 color features were extracted by color moments, and the GLCM algorithm was used to extract 48 texture features of each beef sample. Fusion models of spectral data, color feature, texture feature were established to identify beef samples. The results showed that, the model based on spectral data combined texture feature was the best with the correction set and prediction set recognition rate of 98.42% and 93.55%, respectively, which were higher than the recognition rate of feature spectral data. The texture feature made the expression of classification information more comprehensive. The recognition rate of the model correction set was increased by increasing color features, but the recognition rate of the prediction set was relatively poor. This meant the color features had some valid information, but the correlation between color features and the beef sample was not well, so the recognition rate of prediction set was reduced. Therefore, it is an important way to find color features that are more relevant to beef samples which could improve the recognition rate of models. This study provided valuable information for rapid destructive beef samples.

Keywords Hyperspectral imaging technology; Image feature; Beef samples; Identification; Partial least squares discrimination

(Received Mar. 8, 2019; accepted Jun. 12, 2019)

* Corresponding author