第39卷，第10期
2019年10月

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

Vol. 39, No. 10, pp3273-3280
October, 2019

基于高光谱成像技术的稻谷品种鉴别研究

杨思成[1,2]，舒在习[2]，曹 阳[1*]

1. 国家粮食局科学研究院，北京 100037
2. 武汉轻工大学食品科学与工程学院，湖北 武汉 430023

摘 要 许多不同的稻谷品种看起来很相似，但它们的化学成分和最终产品质量却有很大差别，每年因品种混淆而造成巨大的经济损失，对稻谷品种的鉴别是发展优质粮食工程的现实需要，为此提出了一种采用高光谱成像技术实现稻谷品种无损快速鉴别的方法。主要研究内容和结果如下：(1) 在全波段388～1 000 nm范围内采集5个品种共150粒的稻谷高光谱反射率数据，筛选出差异明显的波段（600～800 nm），将此波段内每个品种的反射率进行Stacked计算和curve-smoothing平滑处理以增加其区分度。(2) 对5种稻谷经平滑处理后的反射率数据做主成分分析，找到权值系数最大的波长位于680 nm，将其作为特征波长。加载特征波长下的纹理图像，计算每粒稻谷样品的纹理特征参数：均值（Mean）、方差（Variance）、信息熵（Entropy）和偏差（Skewness）。利用阈值分割的方法将目标与背景区分开，计算每粒稻谷形态特征参数：面积像素数/pixels^2、边界的周长/pixels、长轴长度/pixels、短轴长度/pixels。结合稻谷的纹理特征参数和形态特征参数，比较Fisher判别分析模型、偏最小二乘回归模型（PLSR）和人工神经网络模型（ANN）对稻谷品种鉴别的效果。(3) 结果显示，Fisher判别分析中函数1和函数2的累计方差贡献率达到93%，能够较好地解释稻谷的品种信息。将样本的函数值与组质心的平方马氏距离（Mahalanobis）做比较，值相近的作为同一分组类别，对稻谷品种的整体识别正确率能达到95.3%；偏最小二成回归模型：$Y_{品种}=0.03X_{均值}-0.36X_{方差}-0.24X_{信息熵}+0.37X_{偏差}+0.31X_{面积}-0.32X_{周长}-0.39X_{长轴长度}+0.45X_{短轴长度}$，该回归模型相关系数$r=0.98$，校正均方根RMESS=0.29，交叉验证均方根PMESSCV=0.32，对稻谷的品种鉴别正确率能达到95%；构建的ANN模型为具有sigmoid隐含和softmax输出神经元的双层前馈网络，对150个样品按70%∶15%∶15%的比例随机划分训练集、测试集、验证集，选择共轭梯度法（scaled conjugate gradient）作为训练算法，以交叉熵（cross-entropy）作为模型的评价指标，对稻谷品种鉴别的正确率可达到98%。稻谷品种鉴别的ANN模型在分类精度上优于Fisher判别和PLSR，选择特征波长下的图像信息建立稻谷品种识别的ANN模型，对稻谷品种的无损快速鉴别具有重要指导意义。

关键词 高光谱；稻谷品种；鉴别；Fisher判别分析；偏最小二乘回归；人工神经网络

中图分类号：TP391.4 **文献标识码**：A **DOI**：10.3964/j.issn.1000-0593(2019)10-3273-08

引 言

水稻是世界上最重要的粮食作物之一，也是我国播种面积位居第二位的粮食作物，近年来播种面积保持在30 000 000公顷左右(中国国家统计局数据)，是我国的主要口粮之一，依据粒形、粒质分为籼稻(早籼稻、晚籼稻)、粳稻、糯稻(籼糯稻谷、粳糯稻谷)。水稻属于一年生禾本植物，共有50 000多种品种，不同品种的稻谷其储藏特性、抗虫霉性以及品质变化等不尽相同，应该做到分类分品种储藏，也更利于后期的收购和加工。对不同稻谷品种的鉴别，也是规范粮食流通秩序，优化粮食供给结构的现实需要。

稻谷品种鉴别的传统方法是由专业人员根据谷粒性状、长宽比、大小、稃壳和稃尖色、稃毛长短、柱头夹持率等特征，对比标准样品或样品图片进行鉴定，该方法效率低且受主观影响大。用幼苗鉴别法和田间小区种植鉴定法也可鉴定稻谷品种，但该方法耗时较长。蛋白质电泳图谱也可以对稻谷品种加以鉴别，但基因的表达有时受发育阶段和环境因素

收稿日期：2018-07-17，**修订日期**：2018-12-02

基金项目：国家重点研发计划项目(2016YFD0401001-2)资助

作者简介：杨思成，1993年生，国家粮食局科学研究院硕士研究生 e-mail：631724555@qq.com

*通讯联系人 e-mail：cy@chinagrain.org

的影响，某些基因组相近的品种无法找到特异性蛋白而影响判别结果[1]。采用 SSR 分子标记法检测稻谷品种可靠性高，但该方法操作复杂，对操作人员技术水平要求较高。近年来，机器视觉和近红外在无损检测中发展快速，但受到识别精度的限制并无法解决化学成分不均匀的问题，这些方法都不适合稻谷品种的无损快速鉴别。

高光谱作为无损快速检测技术，已被广泛应用到食品品质检测和定量分析中。相关研究从光谱的反射率信息实现了稻谷蛋白含量、肉制品新鲜度和咖啡豆品种的鉴别[2-4]。国内学者通过提取非转基因亲本及其转基因大豆判别分析的特征波长，结合化学计量学方法建立 PLS-DA 识别模型[5]，为转基因大豆的鉴别提供了一种新途径。因此，本工作拟利用高光谱技术，针对稻谷品种鉴别展开相关研究，探索一种适用于“中国好粮油”计划对优质稻谷的鉴别方法。

1 实验部分

1.1 材料

采用田间收集的 5 个品种稻谷作为研究对象，挑选色泽均匀、颗粒完好的稻谷，每个品种 30 粒共 150 个样本。将吉粳 108，深两优 5814，株两优 505，粤禾丝苗，D 两优 71 这 5 类依次标号为 1～5，第一类品种各样品标号为 1-1～1-30，依此类推，见表 1。

表 1 样品信息

Table 1 The information of samples

品种	产地	年份
吉粳 108	吉林	2017
深两优 5814	四川	2017
株两优 505	湖南	2017
粤禾丝苗	湖北	2017
D 两优 71	湖北	2017

1.2 仪器与设备

试验采用基于光谱仪式的可见/近红外高光谱成像系统(五铃光学，台湾)，系统主要包括成像光谱相机、卤素灯光源(Fiber-Lite DC950，Dolan-Jenner Industries Inc.，MA，美国)、镜头(镜头焦距 50 mm，厂商：NAVITAR，型号：1-19179)、步进电机(Zolix，SC300-1A，北京)，一个用于减少环境光干扰的暗箱箱体，一台用于控制移动平台并完成数据采集的计算机等配件(见图 1)。

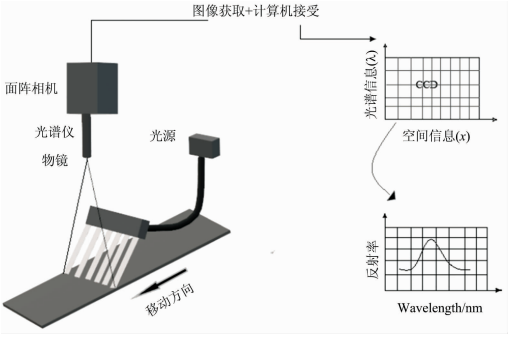

图 1 高光谱成像系统的主要组件

Fig. 1 Basic Spectral Imaging setup

1.3 高光谱图像采集

(1)参数设置

调节光源入射角度 40°和曝光时间 18 s；以画有黑线条的白纸片为参照，调节物距高度 15 cm，获得边缘有格与格之间黑白色差的参照物图像；调节光源强度，使样品 DN 值保持在最大值的 80%左右，为避免图像失真，设置位移平台运行速度 0.3 mm·s^{-1}；将稻谷颗粒均匀地固定在黑色硬卡纸上，在暗室环境中进行图像采集。

(2)图像校正

每一品种的稻谷都在同一起点位置沿着同一方向扫描。为避免光源强度分布不均匀造成的噪声，需要对采集到的图像进行校正，扫描标准白板采集的图像为 W，盖上镜头盖采集全黑图像 B，利用白板图像和全黑图像对原图进行校正，如式(1)

$$C = \frac{R - B}{W - B} \tag{1}$$

其中，C 为校正后图像，R 为原始图像，W 为白板图像，B 为全黑图像。

1.4 特征提取

(1)感兴趣区域选择

在高光谱的感兴趣区域(ROI)选取方面，ROI 形状大小和样品区域的选择直接影响着鉴别效果和模型的稳定性，本实验根据稻谷的形态，在稻谷颗粒的中心选定合适大小的范围 60 pixel×30 pixel 作为 ROI。

(2)光谱预处理

在 400～1 000 nm 范围内，ROI 内每一个像素点包含 600 个光谱反射率数据，每一个 ROI 区域即为一个 60×30×600 的空间反射率矩阵。分别计算 150 个样品在 ROI 内的反射率，再计算同品种稻谷的反射率平均值，得到 5 种稻谷的反射率曲线。提取此波段中差异相对明显的部分，用 Stacked 计算和 curve-smoothing 平滑对此波段数据进行预处理。

(3)特征波长选择

稻谷品种的不同，含有的水分、淀粉、脂肪、蛋白质等化学成分也不同，不同成分含量的分子结构具有特定的光学吸收特性[6]，在不同的波长照射下有不同的反射率。在主成分(PCA)分析中，权值系数绝对值越大所对应波长的贡献程度越大[7]，包含的品种特征信息越多，将权值系数峰值作为稻谷品种的特征波长。

(4)纹理特征提取

稻谷颗粒表面的纹理具有缓慢变化或者周期性变化的排列，在一幅图像上规定一个方向(水平的、垂直的等)和一个距离(一个像素，两个像素等)，该物体共生矩阵 $\boldsymbol{P}$ 的第(i，j)个元素值等于灰度级 i 和 j 在物体内沿该方向和相距指定距离像素上同时出现的次数。基于概率统计的滤波(occurrence measures)将 ROI 中每一个灰度出现的次数用于纹理计算[8]，本实验中用于计算的类型有均值(Mean)[式(2)]、方差(Variance)[式(3)]、信息熵(Entropy)[式(4)]和偏差(Skewness)[式(5)]。

①均值

$$\text{Mean} = \sum_{i=0}^{\text{quant}_k} \binom{i}{j} \sum_{i=0}^{\text{quant}_k} p(i,j)i \tag{2}$$

均值反映纹理的规则程度，纹理杂乱无章，均值小；规律性强，均值大。

②方差

$$\text{Variance} = \sum_{i=0}^{\text{quant}_k} \binom{i}{j} \sum_{i=0}^{\text{quant}_k} p(i,j)(i-\text{Mean})^2 \tag{3}$$

方差反映图像的均匀性，当像元值与均值的偏差程度较大时，方差值大。

③信息熵

$$H(P) = -\sum_{i}^{j} X(i,j) \sum_{i}^{j} X(i,j) \ln P(i,j) \tag{4}$$

信息熵表征图像灰度分布的聚集特性，当体系的混乱程度大时，信息熵值大。

④偏差

$$\text{Skewness} = \sum_{i=1}^{n} (x_i - \bar{x})/SD^3 \frac{1}{n-1} \tag{5}$$

偏差描述的是图像的对称性，当纹理分布形态的偏斜程度越大时，偏差值大。

(5)形态特征提取

图像信息表征样品特性时，形态特征参数是农作物分类中常用的依据[9]。为了更好地找到稻谷的形态学边界，利用要提取的对象与背景在灰度特性上的差异，采用最大类间方差法(由 Otsu 于 1978 年提出)将目标从背景中区分开来，其简化公式为式(6)

$$(T) = W_A(\mu_a - \mu)^2 + W_B(\mu_b - \mu)^2 \tag{6}$$

式中：T 为两类间最大方差，W_A 为 A 类概率，μ_a 为 A 类平均灰度，W_B 为 B 类概率，μ_b 为 B 类平均灰度，μ 为图像总体平均灰度，阈值 T 将目标分成背景和样品两部分。

计算样品颗粒的投影面积像素数/pixels^2、边界的周长/pixels、长轴长度/pixels、短轴长度/pixels，将所求参数作为形态特征值。

1.5　多元分析方法

(1)主成分分析

主成分分析的核心思想是降维，通过数学变化将多变量归结为几个主要成分来表征原有信息，能够从一段连续的波长中提取出代表样品信息的特征波长。在数学变化中保持总量的方差不变，将变量中方差最大的作为第一主成分，主要公式为式(7)

$$\text{降维结果}_{(m\times k)} = \text{标准化矩阵}_{(m\times n)} \times \text{特征向量矩阵}_{(n\times k)} \tag{7}$$

对原始光谱数据 m 个样本的 n 个波长维随机变量进行标准化变化，得到标准化矩阵($m\times n$)；计算样本的协方差矩阵 nn^{T}，求解 nn^{T} 的特征值；将特征值从大到小排列，将前 k 个特征值对应的 k 个特征向量组成特征向量矩阵($n\times k$)；将样本点投影到特征向量上，这样实现了多光谱数据 n 到特征波长 k 的转变。

(2)Fisher 判别分析

Fisher 判别分析的分类方法是投影，把高维度的空间点投影到一条甚至多条直线上，直到投影构成的空间能够实现样本总体的区分。运用方差分析使不同组的组间离差最大，同时满足每个组的组内离差最小[10]。

给定 p 个 d 维特征的训练样例 x_n(n 从 1 到 p)，样例沿 w 方向的投影可计算为：$y_n = w^{\mathrm{T}}x_n$，y_n 是 x 投影在 w 直线上的点到原点的最小距离。w 方向的不同直接决定样品的分类效果，将 Fisher 判别准则函数取极大值时的向量记为 w^*，即 d 维 x 空间到一维 y 空间的最佳投影方向。Fisher 函数建立后，采用 mahanobis 距离法对检验样本归类

$$D_M(x) = \sqrt{(x-\mu)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(x-\mu)} \tag{8}$$

对于一个协方差矩阵 $\boldsymbol{\Sigma}$，μ 为均值，x 为多变量向量。将函数的 y_n 值与 $D_M(x)$ 值相近的作为同一组类别，从而达到判别分类效果。

(3)偏最小二乘回归

偏最小二乘回归研究的是 k 个因变量(y_1，y_2，y_3，…，y_k)与 j 个自变量(x_1，x_2，x_3，…，x_j)的回归模型，特别适用于内部高度线性相关的各变量。通过最小化误差的平方和寻找数据的最佳函数匹配，在光谱学的数据分类和建模分析中应用广泛。

在自变量和因变量中分别提取第一成分 t_1，u_1，在 t_1 和 u_1 相关性达到最大的基础上，建立 y_1，y_2，…，y_k 与 t_1 的回归模型。若模型不满足分类要求，继续从自变量中提取 t_2，t_3，…，t_n 个成分，建立 y_1，y_2，…，y_p 与 t_1，t_2，…，t_n 的回归方程，将结果表示为 y 与原自变量 x 的回归式

$$y_b = a_{b1}x_1 + a_{b2}x_2 + \cdots + a_{bj}x_{bj}(b = 1,2,\cdots,k) \tag{9}$$

(4)人工神经网络

人工神经网络由多个节点组成，每个节点都是一个带有非线性激活函数的神经元，映射一层的输入向量到下一层的输出向量。神经元之间的连接赋予相关的权重，网络的输出则依赖于神经元的连接方式、激活函、权重值。训练学习算法在迭代过程中不断调整这些权重，从而使得分类误差最小化并给出预测精度[12]。

1.6　数据处理

本实验用到的数据分析软件有 ENVI 5.2(ITT Visual Information Solutions，Boulder，CO，USA)，IBM SPSS STAUISTICS 22.0，SIMCA 14.0，MATLAB 2016a(The Math Works，Natick，MA，USA)。ENVI 软件用于提取稻谷的光谱信息、纹理信息和形态信息，SPSS 用于 PCA 和 Fisher 判别分析，SIMCA 用于 PLSR，MATLAB 用于 ANN 模型的建立。

2　结果与讨论

从图 2 可以看出不同品种稻谷的反射率波形相似，在 400 nm 处出现波谷，随后呈现曲线上升的趋势，这客观地反映了不同的品种有着相同的化学组成。光谱数据的开始和结束包含相当大的随机噪声，且在 400～500 和 900～1 000 nm 波段范围存在相互的重合和交叉。为避免对稻谷品种进行区分造成的干扰，从全波段中选取 600～800 nm 的波段用于进一步的分析。

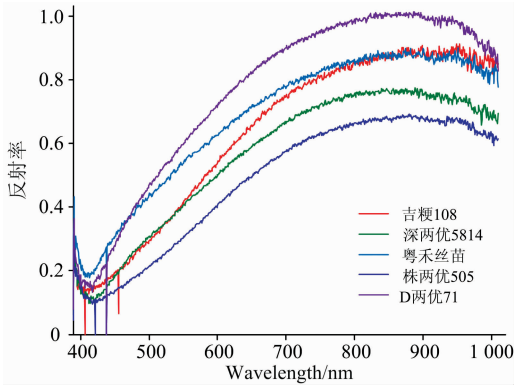

图 2 五种稻谷在全波段的平均光谱反射率

Fig. 2 The average spectral reflectance of five kinds of rice in the full band

2.1 前处理

为了增加在 600～800 nm 下各品种的光谱差异，将 ROI 内各品种的反射率进行 Stacked 堆叠计算，并对得到的光谱曲线进行 curve-smoothing 处理。从图 3 可以看出，各稻谷品种的曲线实现了较好的区分，用处理后的数据进行进一步特征提取分析。

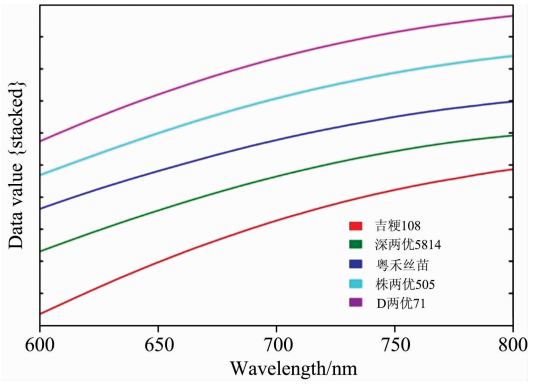

图 3 五种稻谷的 Stacked 曲线

Fig. 3 The stacked spectrum of five kinds of rice

2.2 特征提取结果

(1)特征波长提取结果

将稻谷的品种与 600～800 nm 下的反射率作主成分 PCA 分析，提取的前两个主成分累积贡献率已超过 98%，其中第一主成分的贡献率达到 95%，能较好地保持原有的光谱信息。从图 4 可得，在 680 nm 处第一主成分和第二主成分都具有明显的峰值，选取 680 nm 处作为特征波长。

(2)图像特征提取结果

加载稻谷在特征波长下的不同纹理类型图像，图 5 显示，吉粳 108 不同的纹理类型有特定的纹理结构，均值图像较好的保留了稻谷的整体形态并反映了稻谷的亮度差异，方差图像较好地提取了稻谷的轮廓，信息熵图像更好地反映了稻谷的纹理排列，偏差图像则呈现出稻谷颗粒整体的均匀性。各纹理类型中稻谷颗粒的像素点亮度存在差异，并且有着不同的排列规律，可以用来作为品种区分的重要信息。其他 4 种稻谷的均值图像、方差图像、信息熵图像和偏差图像与吉粳 108 相似，各品种稻谷都具有差异明显的 4 个纹理结构。

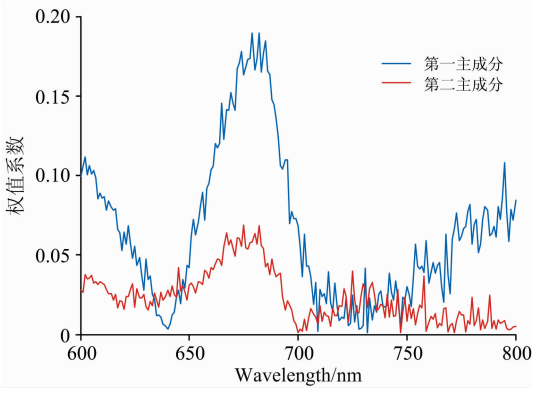

图 4 五种稻谷在 600～800 nm 处的主成分权值系数

Fig. 4 The weight of the principal component in the band of 600～800 nm

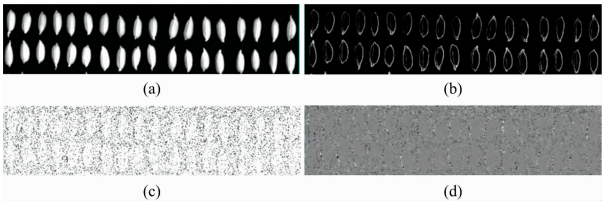

图 5 吉粳 108 在 680 nm 处的纹理图

(a)：纹理图-均值；(b)：纹理图-方差；

(c)：纹理图-信息熵；(d)：纹理图-偏差

Fig. 5 The texture of Jijing108 in the band of 680 nm

(a)：The texture of means；(b)：The texture of various；

(c)：The texture of entropy；(d)：The texture of skewness

图 6 以深两优 5814 和吉粳 108 为例，其他稻谷品种的阈值分割效果同样良好。阈值分割后的轮廓边缘并没有达到绝对的平滑效果，但对稻谷的形态特征提取影响不大，分割后的稻谷可以更好地提取稻谷的形态特征。深两优 5814 为南方种植的籼稻，吉粳 108 为北方种植的粳稻，二者在形态上差异明显。所选取的稻谷中还有其他 3 种类型的籼稻，尽管它们投影图像的结果在形态上与深两优 5814 相似，但形态上还存在着微小的差异，各形态特征参数可以为稻谷的品种分类作为参考。

图 6 两种稻谷的阈值分割图

(a)：深两优 5814；(b)：吉粳 108

Fig. 6 Threshold segmentation of two kinds of rice

(a)：Shenliangyou5814；(b)：Jijing108

2.3　识别结果

不同的稻谷品种光谱信息在空间的相似度高，在连续的波段中存在着有用信息冗杂的问题，不利于直接实现稻谷品种间的区分。提取特征光谱下的图像信息进行建模，弥补了基于单一光谱信息的识别模型在分类精度上的不足，有效利用了高光谱图像图谱合一的优势。

(1)基于图像信息的 Fisher 判别模型

在 SPSS 中导入 5 种稻谷的 150 个训练样本数据运行 Fisher 判别分析，设置最小分类值为 1，最大分类值为 5，建立了 4 个标准化的典型判别式函数。标准化的典型判别式函数所占的方差百分比反映了对品种信息的解释率，从表 2 可以看出前两个函数的累积贡献率已达到 93.3%，联合函数 1 和函数 2 可以较好地对稻谷品种进行分类。

表 2　标准化的典型判别式函数系数及特征值

Table 2　Standardized canonical discriminant functions and eigenvalues

函数	均值	方差	面积	周长	长轴长度	短轴长度	特征值	方差的/%	累计/%	正则相关性
1	−0.112	0.818	−0.08	0.072	0.295	−0.437	57.121[a]	74.4	74.4	0.991
2	0.604	0.405	0.358	−0.13	−0.193	0.463	14.456[a]	18.8	93.3	0.967
3	0.394	−0.245	1.060	0.134	0.097	−0.752	4.910[a]	6.4	99.7	0.911
4	0.304	−0.155	−1.293	1.083	0.321	0.848	0.248[a]	0.3	100	0.446

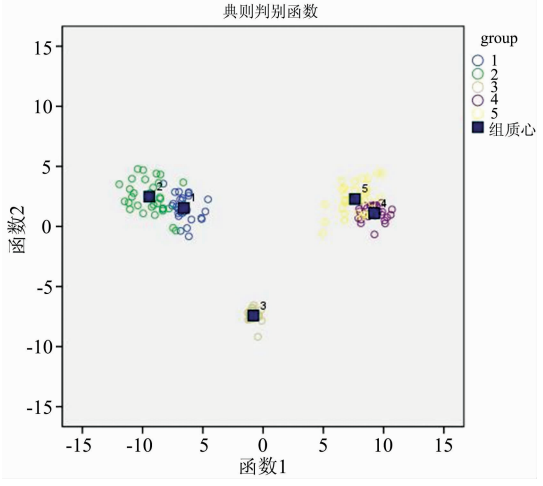

图 7　基于形态和纹理特征的 Fisher 识别结果

Fig. 7　Fisher recognition results based on morphological and textural features

通过比较样本函数值与组质心的平方 Mahalanobis 距离值可判定样本的类别，距某组质心值越近就判定为该类别。从图 7 可以看出，组 3(粤禾丝苗)与其他品种区分最为明显；组 1(吉粳 108)和组 2(深两优 5814)分别属于粳稻和籼稻类型，但两组间的距离很近，说明不同类型的稻谷存在图像特征上存在一定的共性；(组 4 株两优 505)和组 5(D 两优 71)属于同产地不同品种的籼稻，样本个体间存在相互交叉的现象，识别精度相对较低。Fisher 判别的分类结果显示，对初始案例分组中 95.3%和交叉验证分组案例中的 95.3%个样本进行了正确分类，说明该方法可以实现对稻谷品种的鉴别。

(2)基于图像信息的偏最小二乘回归模型

图像的单一信息无法全面地反映品种间的相似性与差异性，进行品种鉴别时需要综合考虑样本的多个特征。将因变量“品种”设为虚拟变量，数字 1～5 表示不同的品种类型，将稻谷的纹理特征和形态特征特征参数合并，共同作为输入变量与品种类型进行拟合，给出稻谷品种与 8 个自变量的标准化偏最小二乘回归模型

$$Y_{品种} = 0.03X_{均值} - 0.36X_{方差} - 0.24X_{信息熵} + 0.37X_{偏差} + 0.31X_{面积} - 0.32X_{周长} - 0.39X_{长轴长度} + 0.45X_{短轴长度}$$

该回归模型相关系数 $r=0.98$，校正均方根 RMESS=0.29，交叉验证均方根 PMESSCV=0.32，说明模型有较好的预测性。从回归系数看出，影响稻谷品种的纹理区别主要在方差和偏差，主要与稻谷颗粒各成分的分布不均匀性有关；而形态区别主要在短轴长度和长轴长度，可为稻谷品种的粗分类提供指导。

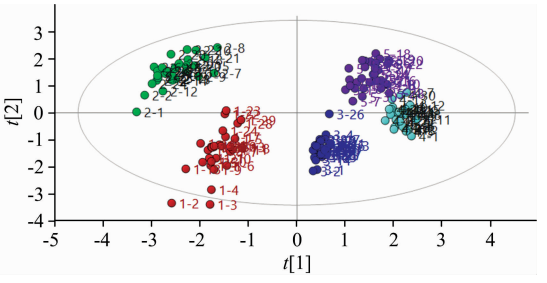

图 8　五种稻谷的偏最小二乘得分图

Fig. 8　Partial least squares score map of five kinds of rice

对该模型的鉴别效果进行评估，从图 8 可以看出，不同的稻谷品种形成簇状分布，排除位于椭圆外的异常点，同种稻谷相互间聚集紧密，甚至出现互相重叠的现象，只有少数样本处于相对离散的状态，表明同种稻谷间的相似度极大。其中 4 号和 5 号分别代表粤禾丝苗与 D 两优 71，均购于湖北省黄冈市马曹庙镇优质稻种植基地，二者在得分图上的距离相对较近，表现出较小的差异，这可能与种植条件与地理气候的影响相关。整体来看，该方法能实现稻谷品种的区分，且分类精度达到 95%以上。

(3)基于图像信息的神经网络模型

应用 MATLAB 中神经网络工具建立稻谷品种识别模型，该模型为一个具有 sigmoid 隐含和 softmax 输出神经元的双层前馈网络，选择共轭梯度反向传播(trainscg)对神经网络进行训练。Sigmoid 函数将输入值映射到(0，1)区间中，

输入值经 softmax 函数归一化处理后转入到概率测度空间，训练过程中引入交叉熵(cross-entropy)函数可以很好地衡量概率分布差异。交叉熵越小，模型的分类精度越高，在神经网络优化到一定程度时训练自动停止。

从图 9 可以看出，输入层形式为 150 个样品 8 个特征构成的[150×8]矩阵，输出层形式为 150 个样品 5 个类型的[150×5]矩阵。输入层的节点数对应特征变量的个数。隐含层的节点个数直接影响预测模型的识别精度，通过“试错法”确定隐藏层的节点为 10 个，与输出层的品种类型建立对应关系。

用于鉴别的 150 个样本，其中随机选取 70%用于训练，15%用于验证，15%用于测试。根据识别结果进行多次试验，最终建立以共轭梯度法作为训练算法，以交叉熵作为评价指标的神经网络模型，在很大程度上提高神了经网络的收敛速度和精度，如图 10 所示，5 种稻谷的整体识别率能达到 98%，说明该模型对稻谷品种的鉴别具有良好的预测能力。

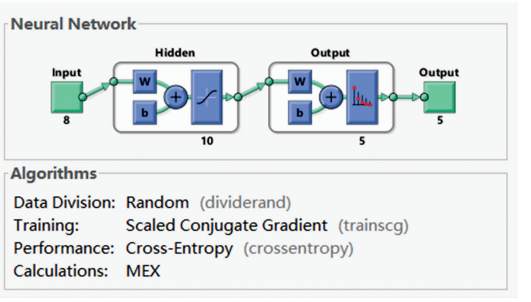

图 9 人工神经网络结构图

Fig. 9 The structure of ANN

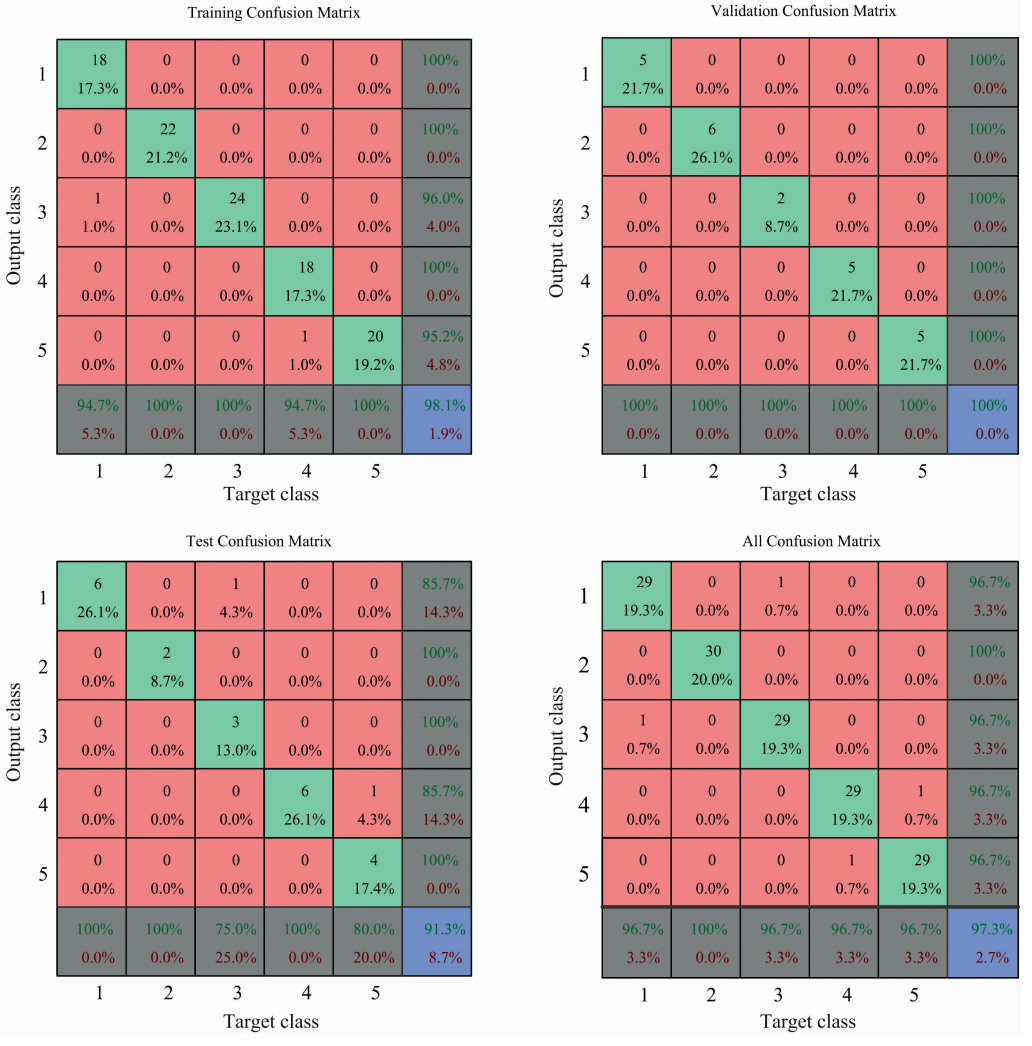

图 10 基于形态和纹理特征的神经网络模型识别率

Fig. 10 Recognition rate of multilayer perceptron model based on morphological and textural features

3 结 论

实验以 5 种不同品种的稻谷为研究对象，综合运用光谱分析技术和数据处理方法，针对样品不同的信息建立对应的模型，研究了高光谱成像技术在稻谷品种鉴别的检测方法，主要结论如下：

提出了基于图像信息的稻谷品种鉴别方法。通过 PCA 分析选取 680 nm 处作为特征波长，提取特征波长下的图像信息，将提取的纹理特征和形态特征参数合并作为输入量建立稻谷品种鉴别模型。结果显示，Fisher 判别分析、偏最小二乘回归模型和神经网络模型对稻谷品种的鉴别正确率分别达到 95.3%，95%和 98%，该方法将原始 150×200 的光谱数据矩阵降为 150×8 的特征分类矩阵，使用于分类建模过程中数据处理的工作量大大减少，并且满足对稻谷品种的鉴别精度要求。

在稻谷品种鉴别模型的优选上，基于图像特征参数的 ANN 模型比 Fisher 判别分析和 PLSR 的鉴别效果更好，识别正确率高达 98%。此方法可实现对稻谷品种的无损快速鉴别，有处理大量样本的潜力，可以用于大规模的尺度检测和智能分析。

References

[1] FU You-qiang, YU Xiao-li, YANG Xu-jian, et al(傅友强，于晓莉，杨旭健，等). Chinese Journal of Rice Science(中国水稻科学), 2017, 31(2): 133.

[2] Onoyama H, Ryu C, Suguri M, et al. Precision Agriculture, 2017, (5): 1.

[3] Crichton S O, Kirchner S M, Porley V, et al. Meat Science, 2017, 129: 20.

[4] Bao Y D, Na C, Yong H E, et al. Optics & Precision Engineering, 2015, 23(2): 349.

[5] Wang L, Liu D, Pu H, et al. Food Analytical Methods, 2015, 8(2): 515.

[6] YU Hui-chun, WANG Run-bo, YIN Yong, et al(于慧春，王润博，殷 勇，等). Food Science(食品科学), 2017, 38(20): 292.

[7] Sun J, Jiang S, Mao H, et al. International Journal of Food Properties, 2016, 19(8): 1687.

[8] Roy A, Singha J, Manam L, et al. Let Image Processing, 2017, 11(6): 352.

[9] Isaza C, Anaya K, Paz J Z, et al. Multimed Tools & Applications, 2018, 77(2): 2593.

[10] Zhang C, Xie Y, Liu D, et al. IEEE Transactions on Image Processing, 2017, 26(3): 1355.

[11] ZHAN Bai-shao, ZHANG Hai-liang, YANG Jian-guo(詹白勺，章海亮，杨建国). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2017, 37(4): 1232.

Identification of Rice Varieties Based on Hyperspectral Image

YANG Si-cheng[1, 2], SHU Zai-xi[2], CAO Yang[1*]

1. Academy of State Administration of Grain, Beijing 100037, China

2. College of Food Science and Engineering, Wuhan Polytechnic University, Wuhan 430023, China

Abstract Many different varieties of rice look very similar, but their chemical composition and final product quality vary greatly, which causes huge economic losses each year as a result of variety confusion. Identification of rice varieties is the practical requirement for developing high quality grain engineering. In this paper, a fast and non-destructive method for rice variety identification using hyperspectral imaging technology was proposed. The main research contents and results were as follows: (1) Average spectrawere extracted from the region of total 150 samples with wavelength from 388～1 000 nm. In the full band, the reflectance was most obvious at 600～800 nm, which was calculated by Stacked stacking and curve-smoothing for increasing its differences. (2) Principal component analysis (PCA) was used to analyze the reflectance data smoothed. It was found that the wavelength with the largest weight coefficient was located at 680 nm and used as the characteristic wavelength. Loading the texture image of the characteristic wavelengths, the texture characteristic parameters of each rice sample were calculated as follows: Mean, Variance, Entropy and Skewness. Meanwhile, the thresholding method was used to separate the target from the background, and the morphological parameters of each grain werecalculated as follows: areas/pixels^2, perimeter/pixels, length of long axis/pixels, length of short axis/pixels. Based on the texture characteristics and morphological characteristics, the Fisher discriminant analysis model, partial least squares regression (PLSR) mode and Artificial neural network model (ANN) were established respectively for rice variety identification. (3) The results showed that the cumulative variance contribution rate of function 1 and function 2 established by Fisher discriminant analysis reached 93%, which could better explain the rice variety information. Comparing the function value of the sample with the square Mahalanobis distance of the group centroid, the individu-

als with similar values were taken as the same category. The overall recognition accuracy of the five rice varieties could reach 95.3%. The PLSR model: $Y_{\text{varieties}} = 0.03X_{\text{means}} - 0.36X_{\text{various}} - 0.24X_{\text{entropy}} + 0.37X_{\text{skewness}} + 0.31X_{\text{area}} - 0.32X_{\text{perimeter}} - 0.39X_{\text{length of long axis}} + 0.45X_{\text{length of short axis}}$, with correlation coefficient (r) = 0.98, corrected root mean square (RMESS) = 0.29, cross validation root mean square (RMESSCV) = 0.32, the accuracy of rice varieties identification could reach 95%. The neural network model is a two-layer feedforward network with sigmoid hidden and soft max output neurons, which randomly divides 150 samples into training samples, validation sets and test sets according to the ratio of 70% : 15% : 15%. With training algorithm of conjugate gradient method and evaluation index of Cross-Entropy method, the accuracy of rice variety identification can reach 98%. The overall results show that the neural network model of rice variety identification is superior to Fisher discriminant and PLSR in classification accuracy, which has an important guiding significance for rapid and non-destructive identification of rice varieties.

Keywords Hyperspectral; Rice variety; Identification; Fisher; Partial least squares regression; Artificial neural network

(Received Jul. 17, 2018; accepted Dec. 2, 2018)

* Corresponding author