

林 萍,蒋善超,陈永明,等. 多源异构大米品质等级非线性高光谱信息耦合识别模型研究[J]. 江苏农业科学,2018,46(20):261-263.
doi:10.15889/j.issn.1002-1302.2018.20.066

多源异构大米品质等级非线性高光谱 信息耦合识别模型研究

林 萍, 蒋善超, 陈永明, 辅小荣, 顾春雷

(盐城工学院电气工程学院,江苏盐城 224051)

摘要:通过构建有效耦合多源异构高光谱波谱信息和影像信息模型来定性和定量划分大米品质等级。首先提取大米高光谱感兴趣区域中全部像素点上的光谱反射率特征值和最优分辨率配置下多尺度滤波器输出的大米高光谱影像特征,通过异构特征无纲量化函数消除异质波谱特征和影像特征量纲影响,在高维 Hilbert 空间中构建有效耦合机制,消除大米非线性高光谱多源品质等级信息异构特性的影响,在高维耦合空间中根据样本之间的度量对多源异构大米高光谱非线性特征品质等级信息进行分类。结果表明,多源异构大米非线性高光谱品质等级信息耦合识别模型获得 84.5% 的训练精度和 82.2% 的测试精度,提出多源异构大米品质等级非线性高光谱信息耦合识别模型性能优于传统的直接利用单源特征信息分类模型。表明提出模型可被用于更精确地确定大米品质等级。

关键词:大米;多源异构;高光谱;信息耦合;品质等级

中图分类号:S127;TP391.4 **文献标志码:**A **文章编号:**1002-1302(2018)20-0261-03

大米是世界上最重要的粮食作物之一,我国大米总产量为世界大米产量之首,约占全球大米总产量的 30%。大米的品质直接决定其市场销售价格,因此对大米品质进行研究具有重要的意义^[1]。大米品质受多种因素影响,靠某单一手段或某一指标很难从整体上实现对大米综合品质的准确评价。常规大米品质检测方法主要是基于可见光与近红外光谱技术的大米内部理化品质检测技术和基于感官视觉或机器视觉的大米外观品质检测技术。基于可见光与近红外光谱技术的大米品质检测技术,缺乏同步获取大米外部感官品质信息的能力。为了适应国际市场上对大米品质快速无损检测的需求,国内外一些学者尝试用可见光与近红外光谱技术对大米品质进行快速检测^[2-3],例如利用近红外光谱技术建立不同类型和粒度的大米样品蛋白质含量预测模型;结合近红外光谱技术和化学计量学方法建立大米脂肪酸值的近红外分析模型并对模型进行预测准确性评价;使用偏最小二乘法建立近红外光谱和水稻糙米直链淀粉含量的数学模型,用于糙米直链淀粉含量的预测。以上利用可见光谱和近红外光谱技术建立的大米理化品质参数的预测模型表明,光谱反射率与蛋白质含量、脂肪酸值、直链淀粉含量、氨基酸含量指标间有较好的相关性,但建立的可见近红外光谱预测模型无法定量分析大米外部感官品质参数,即模型无法关联大米的物理参数,品质评价手段单一,因此品质评价结果存在片面性。另外传统的感官检测方法主要是由人工对大米的碎米率、完整率、黄粒米、

垩白率、互混率等进行评价,对检验人员能力要求较高,同时存在主观性和片面性,并且检测结果难以量化^[4-5]。传统的感官检测或实验室理化分析方法无法满足快速、定量检测要求。利用新技术研究大米内部理化和外部感官综合品质指标信息的快速、无损、同步获取方法并对获取的异构品质指标进行有效融合,可以有效避免单一指标检测的片面性,必然有助于解决大米综合品质的快速准确检测问题,是实现贮藏、流通过程中大米综合品质动态评估、等级标注和实时监控的基础和关键,对于大米综合品质快速、无损检测具有较强的理论和现实意义。

本研究利用高光谱图谱合一技术结合多源异构信息融合技术,探索由大米内部理化指标和外部感官参数构成的综合品质快速无损准确检测新方法。本研究立足于当前大米质量检测技术发展的总体趋势,将信息科学与食品科学相结合,采用高光谱成像技术开展大米综合品质的快速、无损、准确检测研究,这对大米产品质量的快速无损检测具有较强的理论和现实意义。通过提出的理论模型,可以进一步开发出基于高光谱成像技术的大米综合品质专用检测仪器,直接应用于大米产品质量检测实践,具有广阔的应用前景。

1 理论和算法

1.1 高光谱影像特征提取算法

在最优分辨率配置下利用多尺度滤波器提取大米高光谱影像多分辨率特征;其中,滤波函数采用余弦函数调制的二维的带通 Gabor 函数^[6],见公式(1)。

$$g(x, y) = \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right] \cdot \cos(2\pi u_0 x) \quad (1)$$

式中:变量 x, y 表示像素的横、纵坐标值; σ_x, σ_y 为高斯函数核宽; u_0 为调制频率。将上述函数变换到 Fourier 空间,为公式(2)。

收稿日期:2018-01-10

基金项目:国家自然科学基金(编号:31601227、31501221);江苏省自然科学基金(编号:BK20161310)。

作者简介:林 萍(1982—),女,江苏盐城人,博士,副教授,主要从事农业大数据分析研究。E-mail:binglvcha007@126.com。

通信作者:陈永明,博士,讲师,主要从事农业智能系统研究。E-mail:billrange@126.com。

$$G(u, v) = \lambda \cdot \exp \left[-\frac{1}{2} \left(\frac{(u - u_0)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} \right) \right] \quad (2)$$

式中: u, v 为空间频率变量; $\sigma_u = 1/2\pi\sigma_x$, $\lambda = 2\pi\sigma_x\sigma_y$; u_0 为调制频率。二维的带通 Gabor 滤波函数无法实现正交分解, 因此, 基于二维的带通 Gabor 滤波函数变换后的图像中会有频带交叠的重复信息。

本研究在空频域中构建“菊花状”的最优多分辨率二维的带通 Gabor 滤波函数, 该带通 Gabor 滤波函数呈现近似正交的形状, 即滤波器在空间频率域的上半峰幅值接触但不出现重叠, 这样使得多分辨率下的二维的带通 Gabor 滤波函数的输出重复信息量最小^[7]。把函数 $g(x, y)$ 作为母函数, 通过适当的尺度和旋转变换建立自相仿滤波函数:

$$g'_{\text{mm}}(x, y) = \tau^m g(x', y') \quad (3)$$

式中: $x' = \tau^m (x \cos \theta + y \sin \theta)$, $y' = \tau^m (-x \sin \theta + y \cos \theta)$, $\theta = n\pi/K$, $n \in [0, 1, \dots, N]$, N 为旋转的方向总数; $m \in [0, 1, \dots, M]$, M 为在同一旋转方向上的变换尺度总数。尺度因子 τ 可通过公式(4)获得。

$$\tau = \left(\frac{U_{\max}}{U_{\min}} \right)^{\frac{1}{M-1}} \quad (4)$$

式中: U_{\max}, U_{\min} 分别为最大、最小径向中心频率。变换后的频率函数为 $(u', v') = (u \cos \theta + v \sin \theta, -u \sin \theta + v \cos \theta)$ 。方程中 Fourier 域中的尺度参数 $\sigma_{u|m}$ 和 $\sigma_{v|m}$ 可以按公式(5)计算。

$$\begin{cases} \sigma_{u|m} = \tau^{m-1} \cdot \frac{(a-1)}{(a+1)} \cdot \frac{U_{\min}}{\sqrt{2 \ln 2}} \\ \sigma_{v|m} = \tan(\pi/2N) \sqrt{\frac{(U_{\min} \cdot \tau^{m-1})^2}{2 \ln 2} - \sigma_{u|m}^2} \end{cases} \quad (5)$$

式中: a 为尺度因子。

将大米图像以 4×4 的网格划分为 16 块, 计算每一块图像在不同尺度和方向上滤波函数输出的均值并统计概率值, 其中滤波旋转方向数为 $N=8$ 个, 尺度数为 $M=4$ 个, 最终获得表述大米外观品质信息的自然度、粗糙度、膨胀度等 512 维语义特征向量。

1.2 异构高光谱特征无量纲化算法

将异构空间中的异质特征, 包括某一特征波段高光谱影像上选择区域中全部像素点上的光谱反射特征和高光谱影像空频域中最优分辨率配置特征 $\vec{X} = (\vec{x}_1, \vec{x}_2, \dots, \vec{x}_m)$ (其中 \vec{x}_i 是 d 维向量, $i=1, 2, \dots, g$, g 是大米总样本数) 进行无量纲化处理:

$$x_j = 2 \cdot \left(\frac{\vec{x}_j - \vec{x}_{\min}}{\vec{x}_{\max} - \vec{x}_{\min}} \right) - 1 \quad (6)$$

式中: \vec{x}_j 表示第 j 列待无量纲化特征向量; \vec{x}_{\min} 表示列向量中最小值; \vec{x}_{\max} 表示列向量中最大值; x_j 表示第 j 列被无量纲化后的列向量。如果列向量中最大值等于最小值, 即 $\vec{x}_{\max} = \vec{x}_{\min}$, 则列向量值不变 $\vec{x}_j = x_j$ 。使用中心化处理算法消除高光谱影像上选择区域中全部像素点上的光谱反射特征和高光谱影像空频域中最优分辨率配置特征均值漂移的影响。最终的输入样本数据是将一个变量的每个观测值(x_{ij})减去该变量的样本平均值, 具体算法见公式(7)。

$$\bar{x}_{ij} = x_{ij} - \frac{1}{J} \sum_{i=1}^J x_{ij} \quad (7)$$

式中: J 为输入样本总数。

1.3 非线性异构信息耦合算法

通过特征函数 δ 将异质数据 X 投影到高维特征空间 H 进行耦合, $\delta: x_i \in \mathfrak{R}^d, \delta(x_i) \in H$, 其中 \mathfrak{R}^d 表示 d 维实数空间。

在投影的 Hilbert 空间中估计耦合特征的协方差矩阵^[8]:

$$A = \frac{1}{m} \sum_{i=1}^m \delta(x_i) \delta(x_i)^T \quad (8)$$

建立其特征方程:

$$A v = \lambda v \quad (9)$$

式中: $v = \sum_{i=1}^m \gamma_i \delta(x_i)$ 和 λ 分别对应协方差矩阵的特征向量和特征值, 通过特征值分解即可求取其系数 γ 。最后, 对于任何一个测试样本 x 可求其特征变量:

$$[v, \delta(x)] = \sum_{i=1}^m \gamma_i [\delta(x_i)^T \delta(x)] \quad (10)$$

在特征空间根据样本之间的度量方法推断大米品质等级信息^[9], 首先要最小化模型样本 x_i 与对应的目标近邻样本平均度量:

$$\min_M \sum_{i,j \in N_i} 1[\delta(x_i), \delta(x_j)] = \sum_{i,j \in N_i} [\delta(x_i) - \delta(x_j)]^T \Omega [\delta(x_i) - \delta(x_j)] \quad (11)$$

式中: Ω 是半正定矩阵; N_i 为模型样本 x_i 目标邻近索引集合。

其次是让模型样本 x_i 到其目标近邻样本的度量与其到入侵近邻的距离保持单位间隔^[10]:

$$\forall_{i,j \in N_i, l, y_l \neq y_i}, 1[x_i, \delta(x_j^*)] + 1 \leq 1[x_i, \delta(x_j)] \quad (12)$$

式中: l 为入侵近邻样本索引; y_l 和 y_i 为类别标签集。

该求解问题可被重新表达为

$$\begin{cases} \min_M \sum_{i,j \in N_i} 1[\delta(x_i), \delta(x_j)] + \sum_{i,j \in N_i} \xi_{ij} \\ \forall_{i,j \in N_i, l, y_l \neq y_i}, 1[\delta(x_i), \delta(x_j^*)] + 1 \leq 1[\delta(x_i), \delta(x_j)] + \xi_{ij} \\ \xi_{ij} \geq 0 \\ \Omega \geq 0 \end{cases} \quad (13)$$

最后, 由输入大米样本附近的目标之间的度量来确定大米品质等级。

2 试验与分析

2.1 试验数据获取

本研究所用高光谱成像系统构件主要包括美国 Dolan Jenner Industries 有限公司研发的 Fiber-Lite DC950 线性光源, 芬兰光谱成像有限公司研发安装的 C-mount 成像高精密镜头 OLES22 的 N17E-QE 高光谱成像仪, 中国台湾 Isuzu 光学有限公司开发的型号为 IRCP0076 的电控移位载物台平台, 这些设备被安装在密闭暗室固定支架上, 设定物镜垂直高度参数为 278 mm, 设置高光谱相机曝光时间为 2 550 μs , 调整电控移位载物台平台运动速度为 30 mm/s, 设定高光谱图像分辨率大小为 310×260 像素点, 光谱分辨率为 5 nm。本高光谱仪成像系统试验平台搭建于农业部光谱学重点实验室, 大米高光谱试验数据采集时间为 2017 年年初。构建的高光谱仪成像系统用于获取陈米与新米 2 种大米的特征高光谱图, 其原理如图 1 所示。

2.2 光谱数据预处理

在获得的大米样本高光谱图像数据中心区域选取 $110 \times$

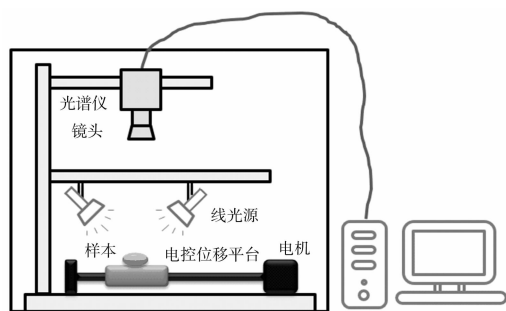


图1 高光谱仪成像结构系统原理

110 像素点面积区域作为感兴趣区域,计算每一波段图像上感兴趣区域中全部像素点上的光谱反射率均值,作为该样例的特征光谱曲线。高光谱分析软件 ENVI 被用于校正获得的大米高光谱图像:

$$\Sigma = (\eta_o - \eta_d) / (\eta_w - \eta_d). \quad (2)$$

式中: η_o 表示使用高光谱原始图像反射率强度值; η_d 表示高光谱图像黑色吸光板校正后反射率强度值; η_w 为高光谱图像白色反光板校正后反射率强度值; Σ 为使用黑色吸光板和白色反光板联合校正的高光谱图像反射率强度值。在使用高光谱仪扫描大米样本试验过程中得到的漫反射大米光谱曲线一般都会夹杂一定程度的噪声信号,本试验使用多元光谱散射校正法预处理大米高光谱反射数据,可消除随机噪声、光散射、杂散光、基线漂移等多重因素的影响^[11-12]。

2.3 算法性能

将2种不同品质等级大米(陈米与新米)的高光谱特征图作为输入特征来验证提出的模型。测试大米的样本总数为200个,其中不同等级的大米测试样本各100个,从同一等级的大米样本中随机抽取80个样本用于建模,其余20个大米样本用于验证。使用高光谱影像特征信息提取算法抽取给定最优分辨率配置下多尺度滤波器输出的大米高光谱影像特征和提取特征波段高光谱影像上选择区域中全部像素点上大米的光谱反射特征值;通过异构特征无量纲化函数将无量纲化异质最优分辨率配置下多尺度滤波器输出大米高光谱影像特征和特征波段高光谱影像上选择区域中全部像素点上的光谱反射特征作为非线性信息耦合模型的输入,将耦合特征映射到高维 Hilbert 空间中,消除由于非线性异质特征异构维度表达引起的影响,使得异质数据在高维线性同构空间中可以进行语义相似度度量,最后实现在高维特征空间中划分大米品质等级的语义归属。建模数据集的模型预测精度使用 N -fold 交叉验证法^[13]来估计,获得的单模态与多源异构模态大米品质信息等级建模集和预测集精度计算结果如表1所示。最终,可以根据返回的预测精度来了解大米品质等级的置信度。

表1 单模态与多源异构模态大米品质等级信息建模和预测精度

方法	训练精度(%)	测试精度(%)
单模态(光谱)	82.3	79.4
单模态(图像)	78.5	75.7
异构模态(光谱+图像)	84.5	82.2

3 结论与讨论

本研究创新之处在于提出了通过构建有效耦合多源异构高光谱波谱信息和影像信息模型来精确定性和定量划分大米品质等级。试验提取大米高光谱感兴趣区域中全部像素点上的光谱反射率特征和最优分辨率配置下多尺度滤波器输出的大米高光谱影像特征,通过异构特征无量纲化函数消除异质波谱特征和影像特征量纲的影响,在高维 Hilbert 空间中构建有效耦合机制,消除大米非线性高光谱多源品质等级信息异构特性的影响,在高维耦合空间中根据样本之间的度量对多源异构大米高光谱非线性特征品质等级信息进行分类。本研究异构模态获得84.5%的训练精度和82.2%的测试精度,测试集和训练集精度比传统的基于单模态方法的平均识别精度分别提升了4.10、4.65个百分点,说明本研究提出的多源异构大米品质等级非线性高光谱信息耦合识别模型优于传统的仅基于单源特征信息分类的模型。根据构建的理论模型,可以进一步开发出基于高光谱成像技术的大米综合品质专用检测仪器,直接应用于大米产品品质质量检测,具有广阔的应用前景。

参考文献:

- [1]王永辉,张业辉,张名位,等.不同水稻品种大米直链淀粉含量对加工米粉品质的影响[J].中国农业科学,2013,46(1):109-120.
- [2]林萍,陈永明.利用可见近红外光谱技术快速鉴别大米品种[J].江苏农业科学,2015,43(12):320-323.
- [3]李路,黄汉英,赵思明,等.大米蛋白质、脂肪、总糖、水分近红外检测模型研究[J].中国粮油学报,2017,32(7):121-126.
- [4]梁明,孟大伟.基于图像的大米品质检测方法研究与仿真[J].计算机仿真,2012,29(6):244-247.
- [5]崔雯雯,周丽聪,黄碧竹,等.基于图像分析的大米透明度定量评价方法[J].食品工业,2015,36(7):126-129.
- [6]陈熙,张戈.基于非高斯二维 Gabor 滤波器的生物特征提取算法[J].计算机工程与应用,2017,53(18):170-175.
- [7]陈永明,林萍,何坚强,等.最优分辨率配置下典型湿地生境场景溯源方法[J].江苏农业学报,2015,31(6):1318-1324.
- [8]侯国林,阿拉坦仓,黄俊杰. Hilbert 空间线性二次最优控制问题中的一个算子的可逆性[J].数学学报,2007,50(2):473-480.
- [9]Assi K C, Labelle H, Cheriet F. Modified large margin nearest neighbor metric learning for regression[J]. IEEE Signal Processing Letters,2014,21(3):292-296.
- [10]Jayashankar S, Sridaran R. Superlative model using word cloud for short answers evaluation in eLearning[J]. Education and Information Technologies,2017,22(5):2383-2402.
- [11]杨天伟,张霁,史云东,等.红外光谱结合多元统计分析快速鉴别不同种类牛肝菌[J].食品科学,2015,36(24):116-121.
- [12]芦永军,曲艳玲,宋敏.近红外相关光谱的多元散射校正处理研究[J].光谱学与光谱分析,2007,27(5):877-880.
- [13]刘学艺,李平,邵传厚.极限学习机的快速留一交叉验证算法[J].上海交通大学学报,2011,45(8):1140-1145.