

基于高光谱图像分割的固态发酵中 不同杆菌的识别方法

刘定操

(河南经贸职业学院计算机工程学院, 河南郑州 450000)

摘要: 本文研究了基于高光谱图像分割的固态发酵中不同杆菌快速识别问题, 实验采用马铃薯葡萄糖琼脂培养基对芽孢杆菌、乳杆菌、红螺菌样本进行制备, 通过黑色镜头盖在镜头上获取全黑的反射图像, 对图像进行去噪、分割处理, 获取高光谱图像梯度及分布峰值, 利用掩模图像中菌落所处位置对光谱数据信息进行提取, 把获取的三种细菌的光谱数据分割成校正集与测试集, 依次用于模型构建与验证。通过标准正态变量变化方法完成对原始光谱数据的预处理, 利用主成分分析法进行降维处理并区分不同菌落, 并采用偏最小二乘判别分析法建立识别模型, 分析细菌高光谱图像和高光谱响应值主成分, 建立并验证模型。结果表明, 采用标准正态变量变化法进行预处理后, 高光谱分析总误差与分析误差的比值是 1.12, 低于规定稳健性的参数; 不同细菌菌落的高光谱波峰存在差异, 对光的反射率也有较大不同; 当培养时长为 36 h 时, 三种细菌可被有效区分; 本文模型比其它两种模型更优, 对测试集的认识率为 98.25%。综上所述, 采用本文方法具有分析误差低、识别率高的优点。

关键词: 高光谱图像技术; 固态发酵; 芽孢杆菌; 识别

文章编号: 1673-9078(2019)09-317-321

DOI: 10.13982/j.mfst.1673-9078.2019.9.042

Recognition of Different Bacteria in Solid State Fermentation Based on Hyperspectral Image Segmentation

LIU Ding-cao

(Henan Vocational College of Economics and Trade, College of Computer Engineering, Zhengzhou 450000, China)

Abstract: In this work, the fast identification of different bacteria in solid-state fermentation based on hyperspectral image segmentation was investigated. The samples of *Bacillus*, *Lactobacillus* and *Rhodospirillum* were prepared by using potato glucose agar medium. The full-black reflective images were obtained on the lens through the black lens cover. The image was denoised and segmented to obtain the gradient and distribution peak value of hyperspectral image. The information of spectral data was extracted from the position of colonies in the mask image, and the spectral data of three bacteria were divided into correction set and test set, which were used to build and validate the model in turn. The pre-processing of original spectral data was completed by the standard normal variable change method. The principal component analysis method was used to reduce the dimension and distinguish different colonies. The identification model was established by partial least squares discriminant analysis. The principal components of hyperspectral image and hyperspectral response value of bacteria were analyzed, and the model was established and validated. The results showed that the ratio of total error to analysis error was 1.12 after pretreatment with standard normal variable change method, which was lower than the prescribed robustness parameter. The hyperspectral peaks of different bacterial colonies were different, and the reflectivity of light was quite different. When the culture time was 36 h, the three bacteria could be distinguished effectively. The model in this work was better than the other two models. The recognition rate of test set was 98.25%. In conclusion, this method had the advantages of low analysis error and high recognition rate.

Key words: hyperspectral image technology; solid state fermentation; bacillus; recognition

固态发酵是以惰性底物或自然底物做碳源及能源, 其体系无水或接近于无水的任何发酵过程, 本质上来讲, 即微生物是在某种湿度固体基质上生长发酵

收稿日期: 2019-05-08

基金项目: 河南科技厅科技攻关项目 (182102210572)

作者简介: 刘定操 (1976-), 男, 讲师, 研究方向: 计算机图像处理

的过程。固态发酵被广泛应用在食品加工的过程中, 传统通过固态发酵制作食品口感好, 风味独特, 受很多人的欢迎。镇江香醋为其中的典型代表, 其主要原料是糯米与麸皮等, 通过特有的固态分层发酵技术, 酿成品醋。在香醋进行固态发酵的过程中, 微生物菌群的代谢产物对香醋的营养与功能产生直接影响,

产酸菌在发酵微生物中起到非常关键的作用,芽孢杆菌为固态发酵过程中微生物优势产酸菌之一,有助于提升香醋的口味,同时来源广泛容易培养。为了获取固态发酵过程中产酸菌活动状态,研究不同杆菌的菌群发展,并对其进行快速识别^[1,2]。

当前常用的细菌识别方法主要有酶联免疫吸附、生物传感器等,对细菌的识别均有一定的弊端,较为耗时耗力,需专业人员进行复杂的实验操作,无法达到当前杆菌快速识别的要求^[3,4]。高光谱图像分割技术是一种将高光谱技术与图像技术结合在一起的技术,将一幅图像分割成不同的区域,这些区域的某些图像特征如边缘、纹理、颜色、亮度等是一致的或相似的,能够完成对物料的快速识别,在细菌快速识别中有广阔的应用前景^[5]。有学者通过高光谱图像分割技术对弯曲杆菌与非弯曲杆菌进行识别通过波段对比法得到的识别准确率高达95%以上,他们还通过高光谱图像技术对六种大肠杆菌进行识别,在像素水平上得到的识别准确率为95%,在菌落水平上的识别准确率是97%。

因此本研究通过高光谱图像分割方法对固态发酵中不同杆菌进行快速识别,利用标准变量变换方法对原始高光谱图像进行预处理,通过主成分分析法对其进行数据降维处理,过滤重叠信息,区分不同细菌菌落,采用偏最小二乘判别分析法对固态发酵中不同杆菌进行快速识别^[6,7]。对模型性能进行比较,建立培养基上不同杆菌菌落识别的快速方法。

1 材料与方法

1.1 材料及仪器

实验采用的枯草芽孢杆菌、谷氨酸棒杆菌、苏云金杆菌都在某大学微生物实验室采集。选择的培养基是马铃薯葡萄糖琼脂培养基,其主要成分是马铃薯浸粉8 g、葡萄糖25 g、NaCl 8 g、氯霉素0.2 g、琼脂20 g、水1 L。把三种菌依次接种于马铃薯葡萄糖琼脂培养基中,在30℃,70%相对湿度环境下活化1周,重新接种并二次培养。

对二次培养的菌种通过无菌水进行清洗,制成菌悬浮液,把菌液滴在血球计数板上,在显微镜下进行计数。依据计数结果,获取菌液浓度,同时进行稀释,实现样本制备。每种菌群制备100个培养基,培养36 h,在进行培养时,针对三组样本每12 h抽取20个培养基进行高光谱图像分割。

实验采用的高光谱成像系统主要由背后照明式CCD相机及其控制模块、图像光谱仪、通过石英卤钨

灯与稳压电源构成的光源系统、载物台以及聚光镜等。

1.2 高光谱图像采集及分割

在使用高光谱成像系统前,对其进行空间与光谱维校正处理,将样本放在载物台上,在光源照射至样本的情况下,漫反射光经聚光镜与光谱仪,利用CCD相机产生光谱图像。为了防止外界光线对高光谱图像采集产生干扰,将仪器放在暗箱中,背景是黑色不反光^[8]。

把培养基放在载物台上,对高光谱图像信息进行采集,采集到的是400~1000 nm范围内共420个波长处的图像。对样本进行检测后即将其丢弃,共获取450个样本的高光谱图像数据。因为受相机暗电流与外界环境的干扰,采集得到的高光谱图像存在一定噪声,需对其进行去噪。通过黑色镜头盖在镜头上获取全黑的反射图像,采集聚四氟乙烯白板获取白色反射图像。通过下式求出经去噪的相对图像值:

$$R_c = \frac{R_0 - D}{W - D} \quad (1)$$

其中: R_0 用于描述原始高光谱透射图像值; D 用于描述盖上镜头后获取的全黑标定图像值; W 用于描述通过聚四氟乙烯白板获取的全白标定图像值; R_c 用于描述去噪后高光谱透射图像值。

对获取到的去噪后的去噪后高光谱透射图像进行分割,获取高光谱图像梯度及分布峰值,并以此为依据得到培养基上细菌光谱信息,为模型建立提供依据,在MATLAB中利用波段减法获取细菌菌落与背景培养基差异显著的灰度图像,通过大津阈值分割法对图像进行二值化处理,获取原始光谱图像的掩模图像。利用掩模图像中菌落所处位置对光谱数据信息进行提取,同时把获取的三种细菌的光谱数据分割成校正集与预测集,通过校正集对模型进行构建,通过预测集对模型性能进行验证。样本集组成用表1进行描述。

表1 校正集与预测集样本

数据集	谷氨酸棒杆菌	苏云金杆菌	枯草芽孢杆菌
校正集/个数	60	60	60
测试集/个数	90	90	90
总量/个数	150	150	150

1.3 研究方法

预处理方法主要有归一化方法、标准变量变换方法和附加散射校正方法,文章选用标准变量变换方法。

高光谱图像中含很多冗余信息,通过主成分分析

法对其进行数据降维处理, 过滤重叠信息。主成分分析法把方差当成评价信息量的指标, 方差越大, 则提供的信息量越大, 否则提供的信息量越小。通过主成分分析法对不同菌落进行区分, 其基本原理为利用线性变换将方差大, 信息量多的分量保留下来, 将信息量少的分量滤除, 以降低数据维数。在光谱判别过程中, 偏最小二乘判别分析法为常用方法, 文章通过该方法对固态发酵中不同杆菌进行快速识别^[9]。

S_{EP} 为高光谱分析总误差, 其涉及可靠性偏差、稳健性偏差以及处理时出现的误差, 可直接对预处理结果及识别结果的准确性进行衡量。 S_{EC} 为模型在建模样本范围内的分析误差, 不含稳健性偏差, 所以二者之差应超过 0。可将 S_{EP}/S_{EC} 当成衡量模型稳健性的参数, S_{EP}/S_{EC} 需低于 1.2, 超过该值则认为模型稳健性较差^[10]。在对模型性能进行衡量的过程中, 需同时符合上述两个条件, 且 S_{EP} 值越小越好。

1.4 数据处理

通过 ENVI 软件、MATLAB 统计工具箱与 SPSS 软件对得到的高光谱分割图像信息进行处理, 对不同杆菌进行快速识别^[11]。ENVI 软件是一种有效的遥感图像处理软件, 通过不同图像处理以及信息提取工具对数据进行处理, 是一种高光谱图像处理软件, 其中波谱分析工具可高效准确地提取不同重要信息。通过 ENVI 软件选择感兴趣区域进行研究, 求出高光谱图像感兴趣区域平均光谱值。

2 结果与讨论

2.1 光谱预处理

为了降低菌落样本内部化学信息不均匀性对识别结果产生的影响, 将 50×50 像素的矩形部分当成感兴趣区域, 将感兴趣区域平均反射光强当成不同杆菌分类的特征参数。在构建高光谱分割图像模型前需对原始光谱进行预处理, 从而降低谱线平移和噪声带来的

影响, 优化光谱信号, 增强信噪比, 从而有效实现对光谱信号的数据挖掘, 增强模型检测准确性与稳定性。实验通过标准正态变量变化完成对原始光谱数据的预处理, 以枯草芽孢杆菌、谷氨酸棒杆菌、苏云金杆菌为例, 结果用图 1 进行描述, 把采集的光谱数据添加至 MATLAB 软件构造模型。

采用标准正态变量变化法进行预处理前后, 性能评价参数用表 2 进行描述。

对光谱信息进行预处理, 并采用标准正态变量变化法进行预处理方法相比, 结果发现校正集相关系数无显著变化, 而测试集相关系数相对较低。采用标准正态变量变化法进行预处理后, S_{EP}/S_{EC} 值是 1.12, 低于 1.2, 也就是采用标准正态变量变化法进行预处理后有很高的稳定性与准确性。

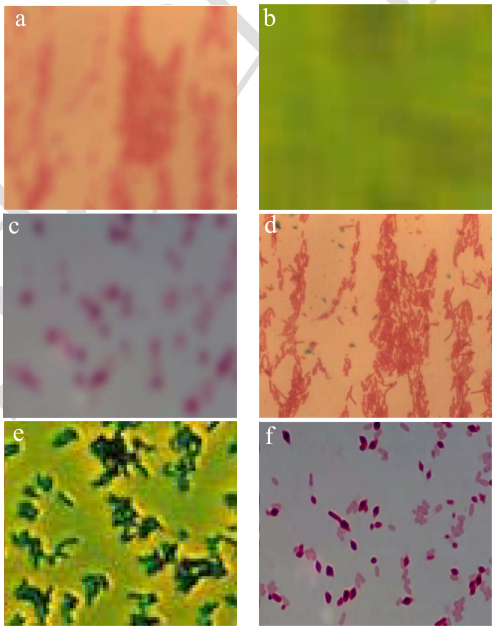


图 1 标准正态变量变化预处理法

Fig.1 Pretreatment of Normal Variables

注: a、b、c 分别为枯草芽孢杆菌、谷氨酸棒杆菌、苏云金杆菌的原始菌落样本图, d、e、f 分别为经预处理后枯草芽孢杆菌、谷氨酸棒杆菌、苏云金杆菌菌落样本图。

表 2 性能评价参数

Table 2 Performance evaluation parameters

项目	校正集相关系数	S_{EC}	测试集相关系数	S_{EP}	S_{EP}/S_{EC}	$S_{EP}-S_{EC}$
预处理前	0.92	0.43	0.79	0.66	1.53	0.23
预处理后	0.92	0.42	0.89	0.47	1.12	0.05

2.2 细菌高光谱图像分析

图 2 所示的是校正图像的掩膜图像, 黑色部分是被屏蔽的背景和噪声区域, 白色部分是分割获取的细菌菌落区域。通过掩膜图像中细菌菌落所处位置对原

始高光谱图像中细菌菌落所处位置进行定位, 对各细菌菌落高光谱信息进行提取。不同培养基样本细菌菌落平均光谱用图 3 进行描述。

由图 3 可知, 三种不同细菌菌落的高光谱在一定程度上有所不同, 波峰存在差异, 谷氨酸棒杆菌在 650

nm 到 700 nm 之间处有显著波峰, 当波长为 745 nm 时枯草芽孢杆菌的光谱出现峰值, 而苏云金杆菌光谱的波长达到 680 nm 时, 反射系数达到最大, 并在短期之内处于稳定状态。除此之外, 三种细菌菌落对光的反射率也有较大不同, 谷氨酸棒杆菌的反射率基本高于其它两种菌落。

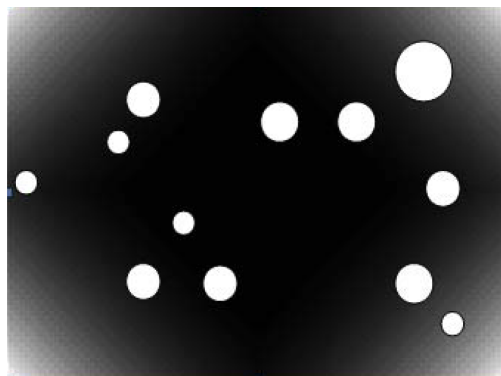


图2 校正图像的掩膜图像

Fig.2 Mask image of corrected image

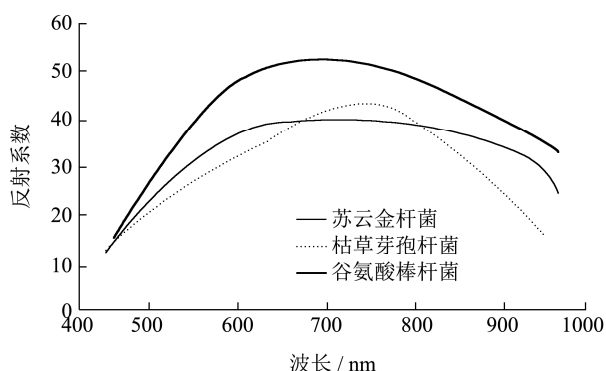


图3 不同培养基样本细菌菌落平均光谱

Fig.3 Average spectrum of bacterial colonies in different media samples

2.3 高光谱响应值主成分分析

在全波段区间中, 受光色散现象的影响, 使得细菌光谱值随菌落的生长逐渐增大, 光的色散逐渐形成, 不同菌落间色散程度会有很大不同, 所以利用高光谱图像分割技术能够对不同杆菌进行快速识别。对不同培养时间内全波段光谱值进行 PCA 分析, 对不同菌落进行区分, 见图 4。

由图 4 所示的结果可知, 三组细菌可有效区分。在培养时间是 12 h 的情况下, 还未有效区分, 这主要是因为三组细菌均处于生长阶段。在培养时间是 24 h 的情况下, 可将枯草芽孢杆菌杆菌区分出来, 因为枯草芽孢杆菌的生长速度更快。在培养时间是 36 h 的情况下, 三种细菌可被有效区分。所以可通过光谱响应值对培养 36 h 的菌落进行区分与识别。

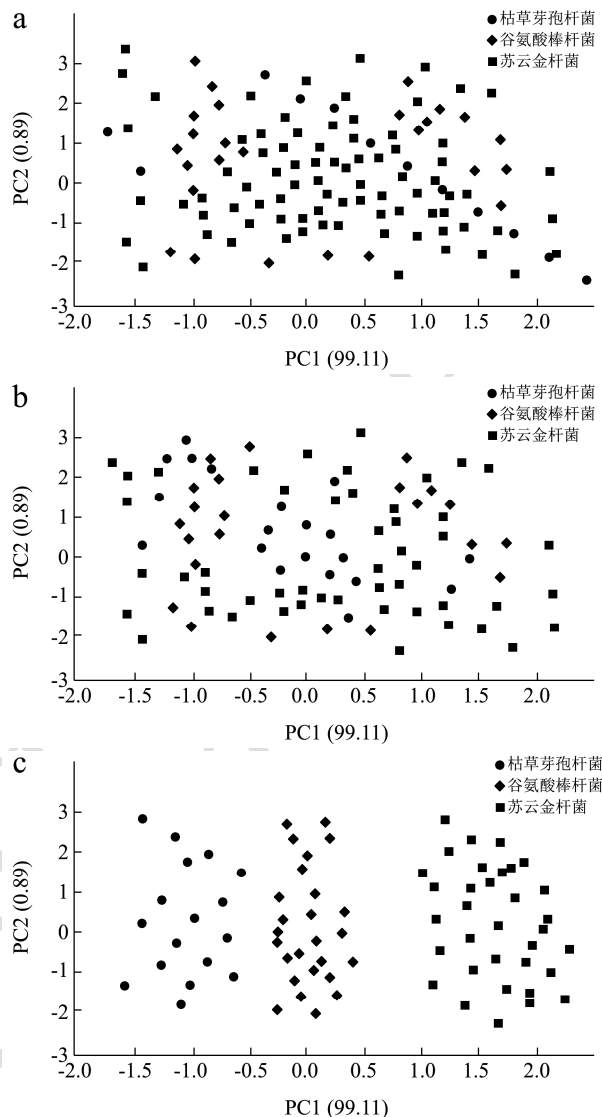


图4 PCA 分析结果

Fig.4 PCA analysis results

注: a: 培养 12 h; b: 培养 24 h; c: 培养 36 h。

2.4 识别模型构建和评价

完成 PCA 处理后, 对一定主成分因子进行提取, 将其当成识别模型的输入。针对校正集与测试集样本, 通过偏最小二乘判别分析法建立固态发酵中芽孢杆菌识别模型, 将基于最小角回归方法的识别模型和基于遗传算法的识别模型作为对比进行测试, 结果用表 3 进行描述。

分析表 3 可以看出, 本文采用模型比其它两种模型更优, 对测试集的识别率为 98.25%。基于最小角回归方法的识别模型和基于遗传算法的识别模型识别效果较差, 且以图像信息为依据对芽孢杆菌进行识别的识别精度明显低于高光谱信息, 无法识别纹理特征相似的细菌菌落, 产生这种情况的原因, 可能由于图像

信息主要反映菌落的外观形态,本研究选取的菌落为生长初期 12 h 的菌落,虽然菌落不同,但在菌落之间存在部分较为相似的纹理特征,难以凭借肉眼观察区分。高光谱信息主要反映菌落代谢物种类、组织结构以及内部化学组分,即使苏云金杆菌与枯草芽孢杆菌

都是芽孢杆菌属的一种,但是它们的内部结构和代谢产物有所不同,故依靠光谱信息能得到较好的区分。高光谱信息与图像信息的比较说明菌落的内在信息比外观信息在菌落的识别上更有决定性作用。

表 3 识别模型性能评价结果

Table 3 Performance evaluation results of identification model

信息	偏最小二乘判别分析		最小角回归		遗传算法识别模型	
	校正集/%	测试集/%	校正集/%	测试集/%	校正集/%	测试集/%
高光谱信息	99.12	98.25	96.32	92.15	92.06	89.35
图像信息	68.31	68.22	52.15	56.39	49.35	42.15

3 结论

本文研究了固态发酵过程中常见的几种不同细菌菌群的图像,对其进行预处理,获取分割后的高光谱图像,并通过主成分分析法对其进行分析,构建基于偏最小二乘判别分析的不同杆菌快速识别模型。结果表明,所提模型可有效完成对不同杆菌的快速识别,校正准确率与识别准确率均高于 97%。除此之外,将基于最小角回归方法的识别模型和基于遗传算法的识别模型作为对比进行性能比较,发现所提模型是被准确性高于其它两种模型。上述结果表明:高光谱图像分割技术能够快速对不同杆菌进行识别,识别准确率高,为微生物识别提供了新的方向,有广阔的应用前景。

参考文献

[1] 柴玉华,毕文佳,谭克竹,等.基于高光谱图像技术的大豆品种无损鉴别[J].东北农业大学学报,2016,47(3):86-93
CHAI Yu-hua, BI Wen-jia, TAN Ke-zhu, et al. Nondestructive identification of soybean seed varieties based on hyperspectral image technology [J]. Journal of Northeast Agricultural University, 2016, 47(3): 86-93

[2] 谭克竹,柴玉华,宋伟先,等.基于高光谱图像处理的大豆品种识别[J].农业工程学报,2014,30(9):235-242
TAN Ke-zhu, CHAI Yu-hua, SONG Wei-xian, et al. Identification of soybean seed varieties based on hyperspectral image [J]. Transactions of the Chinese Society of Agricultural Engineering, 2014, 30(9): 235-242

[3] 肖慧,王振杰,孙晔,等.高光谱图像法对稻谷贮藏中五种常见真菌生长拟合及区分[J].食品工程,2016,37(13):276-281
XIAO Hui, WANG Zhen-jie, SUN Ye, et al. Growth simulation and discrimination of five fungi from rice storage using hyperspectral reflectance imaging technique [J].

Science and Technology of Food Industry, 2016, 37(13): 276-281

[4] 王庆国,黄敏,朱启兵,等.基于高光谱图像的玉米种子产地与年份鉴别[J].食品与生物技术学报,2014,33(2):163-170
WANG Qing-guo, HUANG Min, ZHU Qi-bing, et al. Geographical Origin and years identification of maize seeds based on the hyperspectral image [J]. Journal of Food Science and Biotechnology, 2014, 33(2): 163-170

[5] 楚秉泉,章海亮,罗微,等.柑橘真菌感染部位的高光谱成像快速检测[J].光谱学与光谱分析,2017,37(8):2551-2555
CHU Bing-quan, ZHANG Hai-liang, LUO Wei, et al. Nondestructive detecting rottenness defect of citrus by using hyper-spectra imaging technology [J]. Spectroscopy and Spectral Analysis, 2017, 37(8): 2551-2555

[6] Nair N, Raff H, Islam M T, et al. The *Bacillus subtilis* and *Bacillus halodurans* aspartyl-tRNA synthetases retain recognition of tRNA Asn [J]. Journal of Molecular Biology, 2016, 428(3): 618-630

[7] 孙晔,顾欣哲,王振杰,等.高光谱图像对灰葡萄孢霉、匍枝根霉、炭疽菌的生长拟合及区分[J].食品科学,2016,37(3):137-144
SUN Ye, GU Xin-zhe, WANG Zhen-jie, et al. Growth Simulation and discrimination of *Botrytis cinerea*, *Rhizopus stolonifer* and *Colletotrichum acutatum* by using hyperspectral reflectance imaging technique [J]. Food Science, 2016, 37(3): 137-144

[8] Bishop A H, Robinson C V. *Bacillus thuringiensis* HD-1 Cry(-): development of a safe, non-insecticidal simulant for *Bacillus anthracis* [J]. Journal of Applied Microbiology, 2015, 117(3): 654-662

[9] Kuo H, Huang W, Lin T, et al. The core-independent promoter-specific binding of *Bacillus subtilis* σ B [J]. Febs Journal, 2015, 282(7): 1307-1318