

基于太赫兹时域透射成像技术的葵花籽内部品质无损检测研究

刘翠玲^{1,2}, 王少敏^{1,2}, 吴静珠^{1,2}, 孙晓荣^{1,2}

1. 北京工商大学计算机与信息工程学院, 北京 100048

2. 北京工商大学食品安全大数据技术北京市重点实验室, 北京 100048

摘要 葵花籽壳内籽仁的品质直接影响产出食用油的质量, 利用太赫兹时域透射成像技术, 结合形态学滤波和 K-均值图像分割的方法, 对葵花籽破损粒、虫蚀粒、空壳粒三种异常样本进行带壳无损检测, 实现对葵花籽壳内品质的探索研究。参考国家标准和前人经验制备破损粒、虫蚀粒和空壳粒三种带壳葵花籽样本, 利用太赫兹时域光谱仪 TeraPulse 4000 及透射成像附件, 以 0.2 mm 的分辨率分别采集上述三种异常样本的光谱图像, 并采集一颗正常葵花籽的光谱图像作为参照, 以峰峰值成像的方式对四种葵花籽太赫兹图像进行重构。通过太赫兹图像可初步判断壳内籽仁的形态, 但仍存在对比度低、边缘信息模糊等问题, 需要进一步优化处理。采用形态学滤波算法对葵花籽的太赫兹图像进行滤波, 选择边长为 3 的平坦型菱形结构元素作为核对图像进行一次膨胀, 再计算图像的外部梯度, 完成对图像的滤波处理; 将形态学滤波结果分别与中值滤波、均值滤波和非局部均值滤波结果进行比较, 对比发现, 形态学滤波不仅保证了图像的清晰度, 保留了边缘信息, 同时能使葵花籽样本与背景之间存在明显界限, 有利于后续进行图像分割处理。为了更加准确的检测葵花籽籽仁的形态, 对滤波后图像进行图像分割。采用 K-均值聚类算法对滤波后的太赫兹图像进行图像分割; 为提高分割结果的准确性, 对不同样本的图像分别确定了不同的初始聚类中心个数 K , 其中破损粒 $K=4$ 、虫蚀粒 $K=5$ 、空壳粒 $K=3$ 、正常粒 $K=4$, 分割后的图像能准确呈现葵花籽壳内籽仁的形态。研究表明, 太赫兹时域透射成像技术结合形态学滤波及 K-均值图像分割方法对实现葵花籽内部品质的带壳无损检测具有可行性, 为后期建立带壳葵花籽品质无损检测模型奠定基础, 为油料作物内部品质的带壳无损检测提供参考。

关键词 太赫兹; 时域成像; 形态学滤波; K-均值图像分割; 葵花籽; 无损检测

中图分类号: O433.1 **文献标识码**: A **DOI**: 10.3964/j.issn.1000-0593(2020)11-3384-06

引言

葵花籽是仅次于棕榈、大豆、菜籽的世界第四大油料作物, 其品质的好坏对后期产油及相关产品的加工至关重要^[1]。由于种植或储存不当, 葵花籽壳内籽仁常会出现破损、虫蚀、空壳等异常情况, 若在榨油时选用的葵花籽存在以上异常颗粒, 将影响后期出油率及生产油脂的品质。因此, 从源头把控, 对葵花籽内部品质进行检测是很有必要的。

葵花籽异常颗粒的出现将会影响出油率和油脂的品质, 传统的检测方法主要有手工检验和机器筛选, 手工检验耗费人力, 主观性较大, 可能存在漏检等问题; 机器筛选会破坏葵花籽外壳, 易造成原材料污染, 不利于后期贮藏加工。Ma 等^[2]利用可见光和近红外区域的多光谱成像技术对葵花

籽昆虫侵染、发霉、异色、酸败等异常情况进行了非破坏性区分, 利用 Fisher's 线性判别函数进行特征波长提取, 结合 Wilks lambda 逐步法, 对完好无损的葵花籽进行分类, 准确率最低可达到 97%; 同时利用多光谱成像技术结合主成分分析聚类分析(PCA-CA), 对不同酸败程度的完整向日葵种子进行精准区分。李艳茹等^[3]利用低场核磁共振技术结合主成分分析(PCA)方法对 3 种不同品种的葵花籽进行区分, 并建立偏最小二乘回归(PLSR)模型对 3 种葵花籽的含油量进行预测, 实现了葵花籽品质的检测。Jayabrintha 等^[4]通过机器视觉技术对 10 个不同品种的葵花籽进行品种鉴定, 利用 SVM 分类算法结合最佳顺序级联方法, 分类准确率达到 98.82%。以上研究内容均实现对葵花籽品质的检测, 但由于实验仪器或方法的限制, 均通过葵花籽的外部特征或剥除外壳后对品质进行检测研究, 无法实现对葵花籽壳内籽仁品质的带壳无损

收稿日期: 2019-11-22, 修订日期: 2020-03-09

基金项目: 北京市自然科学基金项目(4182017), 国家自然科学基金青年科学基金项目(61807001)资助

作者简介: 刘翠玲, 女, 1963 年生, 北京工商大学计算机与信息工程学院教授 e-mail: liucl@btbu.edu.cn

检测。

太赫兹波介于红外和微波之间,具有较低能量和较高穿透性,不会引起有害的电离反应,且对绝大部分非极性物质有较强穿透能力,近年来在食品^[5]、农业^[6]、工业^[7]、医学^[8]、安检^[9]及考古^[10]等检测领域得到了广泛应用。相比现有检测手段,太赫兹时域光谱成像技术能在不破坏葵花籽外壳的情况下将内部籽仁的形态以图像的形式呈现出来。因此,本文利用太赫兹时域透射成像技术分别获得带壳葵花籽正常粒、破损粒、虫蚀粒和空壳粒的透射图像,再通过形态学滤波和 K-均值图像分割算法,实现对葵花籽壳内籽仁品质的带壳无损检测,为带壳油料作物内部品质的无损检测提供新思路。

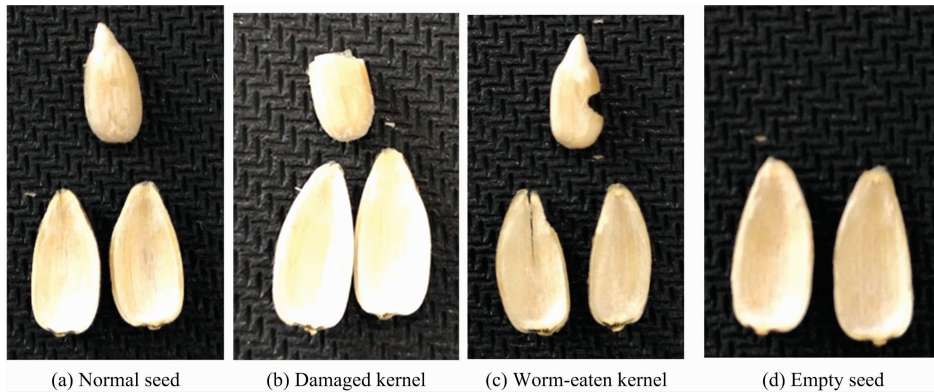


图 1 葵花籽样品图

(a): 正常粒; (b): 破损粒; (c): 虫蚀粒; (d): 空壳粒

Fig. 1 Sample of sunflower seeds

(a): Normal seed; (b): Damaged kernel; (c): Worm-eaten kernel; (d): Empty seed

1.2 仪器

实验仪器采用剑桥 Tera View 公司生产的太赫兹时域脉冲光谱仪,型号为 TeraPulse 4000,如图 2(a)所示,成像方式为脉冲成像,可发射频率从 60 GHz 到 4 THz($2 \sim 133$

cm^{-1})的太赫兹波,信噪比最高可达到 70 dB。图像数据采集部分利用太赫兹透射成像附件进行,如图 2(b)所示,其扫描范围为 $20 \text{ mm} \times 20 \text{ mm}$ 。

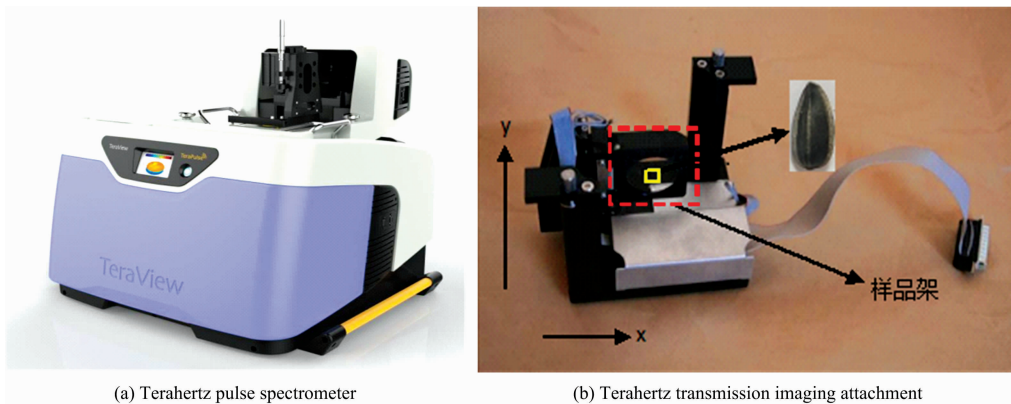


图 2 实验仪器

(a): 太赫兹脉冲光谱仪; (b) 太赫兹透射成像附件

Fig. 2 Experimental apparatus

(a): Terahertz pulse spectrometer; (b): Terahertz transmission imaging attachment

1.3 光谱图像采集

葵花籽光谱图像采集具体过程如下:将透射附件安装到

太赫兹时域脉冲光谱仪上,采集无样品时的太赫兹脉冲波作为参考信号,确认信号源无误后开始正式实验。将样品固定

在样品架中心位置,设置相关参数后开始采集,采集过程由 TeraPulse 软件控制执行。根据样本大小设置合适的采样范围,以 0.2 mm 分辨率进行光谱图像采集,采集一个完整的样本图像所需时间约 13~14 min。图像重构方式选择峰峰值成像。

1.4 数据处理方法

太赫兹图像含有噪声,分辨率较低,边缘模糊不光滑,需要对原始图像进行相关优化处理,保证葵花籽品质检测的准确性。

1.4.1 图像滤波处理

为保留葵花籽透射图像中的边缘信息,选用形态学滤波器(Morphological filter)中的外部梯度^[11]进行图像滤波处理。外部梯度先对原始图像进行一定次数的膨胀操作,再将膨胀结果与原始图像做差,实现对原始图像的滤波。

1.4.2 图像分割算法

图像分割方法包括基于阈值、基于区域、基于聚类及基于特定理论四大类,本研究选用基于聚类的 K-均值(K-means)图像分割算法对滤波后的葵花籽图像进行分割。K-means 算法可以实现类间相似度最低和类内相似度最高^[12],其基本原理^[13]为:

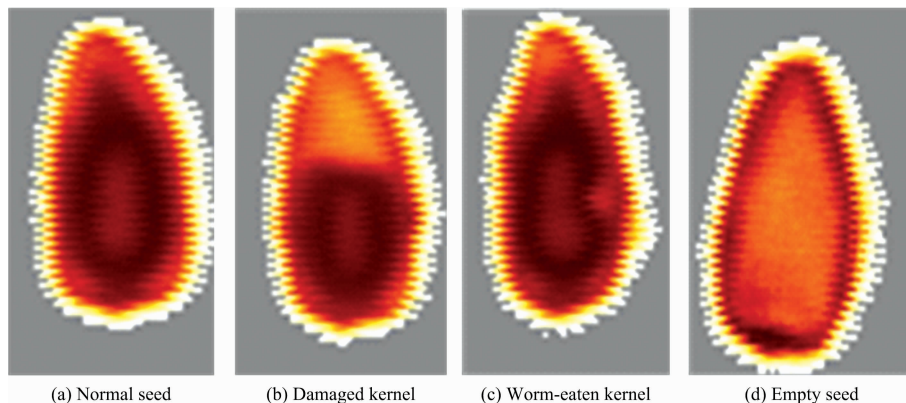


图 3 四类葵花籽的太赫兹透射图像

(a): 正常粒; (b): 破损粒; (c): 虫蚀粒; (d): 空壳粒

Fig. 3 Terahertz transmission images of four categories of sunflower seeds

(a): Normal seed; (b): Damaged kernel; (c): Worm-eaten kernel; (d): Empty seed

此外,葵花籽的不同位置成像后颜色不同,以虫蚀粒为例,对比籽仁部分和外壳部分太赫兹光谱可以发现有明显区别,如图 4 所示。籽仁主要成分为脂肪和蛋白质,所得太赫兹光谱曲线图有明显的两个峰,并且籽仁部分相比空壳部分较厚,对太赫兹波的吸收较强,因此太赫兹信号的强度最大只达到 2 左右;空壳部分为干燥的木质化外壳,对太赫兹信号吸收较少,太赫兹信号的强度最大可达到 5 左右。由于不同位置物质不同,对太赫兹光的吸收强度不同,太赫兹光透射率不同,最终探测到返回的信号强度不同,使得重构后图像的不同位置有明显的颜色差异。因此,通过对葵花籽的太赫兹图像进行分析处理,实现对葵花籽品质的带壳无损检测是可行的。

2.1 图像滤波结果

为实现对葵花籽壳内籽仁异常情况的检测,需要提高图

像的对比度,同时尽可能保证边缘信息被保留,因此首先对葵花籽太赫兹图像进行滤波处理。图 5 为采用中值滤波、均值滤波及非局部均值(non-local means, NLM)滤波对虫蚀粒的太赫兹图像进行滤波所得的结果,其中图 5(a)为利用 5×5 模板进行均值滤波所得结果,图 5(b)为 5×5 模板进行中值滤波的结果,图 5(c)为搜索窗口为 5×5 、邻域窗口为 2×2 、滤波程度为 10 时 NLM 滤波结果。

以上数据处理过程均通过 Matlab2017 软件完成。

2 结果与讨论

太赫兹图像不仅包含样品的空间信息,也包括样品时间轴上的信息,图像中的每个像素点代表一条光谱。通过太赫兹透射成像附件扫描所得样本的太赫兹图像如图 3 所示,观察葵花籽的太赫兹图像可发现,葵花籽内部籽仁的形态可初步分辨,但仍存在分辨率较低,边缘不明确的问题。

图 6 为采用外部梯度对葵花籽虫蚀粒的太赫兹图像进行滤波的结果。在利用外部梯度进行滤波时,应先考虑对原始图像的膨胀。由膨胀的原理可知,与原始图像进行卷积的核的选择会直接影响滤波后图像的质量。观察采集的太赫兹图像可发现由于仪器精度限制和噪声影响,图像边缘为锯齿状。为保证样本形状不发生巨大变化,同时使边缘光滑,本文选择的核为平坦型菱形结构元素,其尺寸大小为 3。同时,

将待分割图像中所有像素点看作数据集合中的所有样本点,随机选取 K 个样本点作为初始聚类中心,计算每个样本点到各聚类中心的欧氏距离,找出其中的最小值,将该样本点划入该聚类中心所属类,遍历完所有样本点以后,重新确定聚类中心,一般以所在区域内样本点的均值作为新的聚类中心;然后再次计算每个样本点到新的聚类中心的距离,对样本重新进行归类,再次调整聚类中心;重复以上步骤,直到聚类中心不再改变,即当前所属类区域内的样本点到该聚类中心的欧式距离平方和最小,样本点划分完毕,图像分割完成。

为确保图像不被过度膨胀，本文只对各葵花籽的透射图像进行一次膨胀，之后将膨胀结果与原始图像做差，求得外部梯度。对比虫蚀粒的外部梯度结果[如图 6(c)]和基础形态学梯度结果[如图 6(d)]可发现，外部梯度的滤波结果相比基本形态学梯度的滤波结果有明显提升。

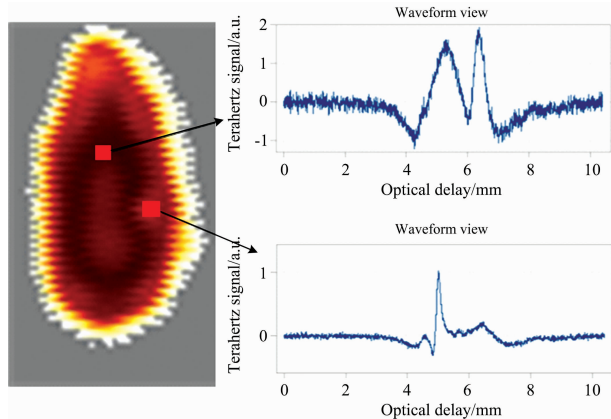


图 4 葵花籽不同位置对应的时域光谱

Fig. 4 Time-domain spectra corresponding to different positions of sunflower seed

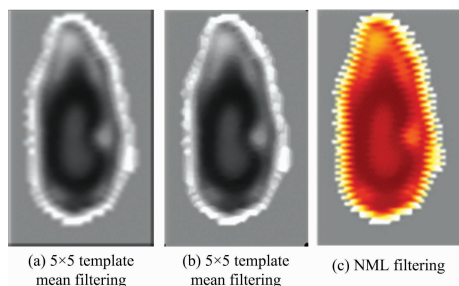


图 5 虫蚀葵花籽太赫兹图像滤波结果图

(a): 5×5 模板均值滤波; (b): 5×5 模板中值滤波; (c): NML 滤波

Fig. 5 Terahertz image filtering results of insect-eroded sunflower seeds

(a): 5×5 template mean filtering;
(b): 5×5 template median filtering; (c): NML filtering

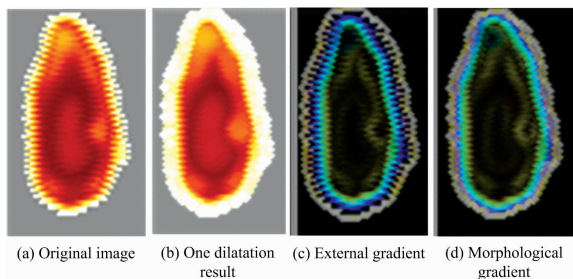


图 6 虫蚀葵花籽太赫兹图像形态学滤波结果图

(a): 原始图像; (b): 一次膨胀结果;
(c): 外部梯度; (d): 形态学梯度

Fig. 6 Results of morphological filtering of insect-eroded sunflower seeds terahertz image

(a): Original image; (b): One dilatation result;
(c): External gradient; (d): Morphological gradient

对比图 5 和图 6(c) 的滤波结果可以发现，中值滤波、均值滤波及 NLM 滤波能很好地去除噪声，但图像清晰度低，边缘部分并没有得到很好地保留，不利于葵花籽籽仁形态的检测。而形态学滤波明显提升了图像质量，在有效去除图像噪声的同时，提高图像对比度，还能准确保留图像边缘信息，使葵花籽与背景之间有明显界限，其滤波结果明显优于中值滤波、均值滤波及 NLM 滤波结果。四种不同状态的葵花籽对应的形态学滤波结果如图 7 所示。

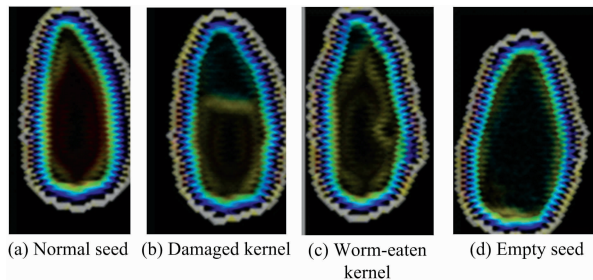


图 7 不同状态葵花籽图像形态学滤波结果

(a): 正常粒; (b): 破损粒; (c): 虫蚀粒; (d): 空壳粒

Fig. 7 Morphological filtering results of sunflower seed images with different states

(a): Normal seed; (b): Damaged kernel;
(c): Worm-eaten kernel; (d): Empty seed

2.2 图像分割结果

为进一步准确识别葵花籽壳内籽仁状态，本文采用 K-means 聚类算法对滤波后的太赫兹图像进行图像分割。经形态学滤波后的太赫兹图像是 RGB 图像，构成的是一种非均匀颜色空间，用欧氏距离难以测量出颜色之间的特征差异，影响分割效果。而 CIE (Commission Internationale de L'Éclairage, 国际照明委员会) - Lab 颜色空间是一种均匀的颜色空间^[14]，在该颜色空间中，欧氏距离测量不同颜色之间差异性时具有不变性，即颜色相近的像素点其欧氏距离相应较小，颜色差异较大的其对应欧氏距离也较大^[15]。因此，为使图像分割结果准确，先将滤波后的太赫兹图像转换到 Lab 颜色空间再进行分割。

K-means 聚类算法是一种无导师监督的学习算法，初始聚类中心个数 K 的选择直接影响图像分割结果。本文针对籽仁的不同状态确定各葵花籽图像的初始聚类中心个数 K 如下：破损粒 $K=4$ ，虫蚀粒 $K=5$ ，空壳粒 $K=3$ 、正常粒 $K=4$ 。确定初始聚类中心的个数后，设置最大迭代次数为 500 次，即可对不同状态的葵花籽图像进行分割，结果如图 8 所示。

由图 8 可看出，空壳粒图像分割结果[图 8(d)]最好，能迅速准确的判断壳内无籽仁；正常粒[图 8(a)]、破损粒[图 8(b)]及虫蚀粒[图 8(c)]的分割结果能将葵花籽籽仁部分分割为一个类，但籽仁的最中间部分被分割到与背景同类，分析其原因是由于葵花籽籽仁的最中间位置没有边缘部分紧实，透过的太赫兹光强度相对较大，成像后的颜色经形态学滤波后与背景颜色接近，故而在图像分割时与背景归为一类，但最终的图像分割结果并未影响对葵花籽籽仁形态的判

定。因此, 尽管存在一些小的噪声类, 但 K-means 分割算法仍能准确呈现葵花籽壳内籽仁的形态, 实现对葵花籽品质的

无损检测。

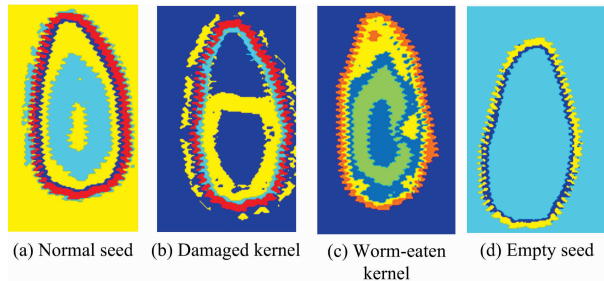


图 8 不同状态葵花籽图像 K-means 图像分割结果

(a): 正常粒; (b): 破损粒; (c): 虫蚀粒; (d): 空壳粒

Fig. 8 K-means image segmentation results of sunflower seed images with different states

(a): Normal seed; (b): Damaged kernel;
(c): Worm-eaten kernel; (d): Empty seed

3 结 论

利用太赫兹时域透射成像技术分别获得正常饱满的葵花籽和籽仁破损、虫蚀、缺失三种异常情况下葵花籽的太赫兹图像, 利用形态学滤波中的外部梯度对葵花籽太赫兹图像进行滤波处理, 在提高图像清晰度的同时, 保证了样品的形状及边缘信息不被改变或模糊。此外, 结合 K-means 图像分割算法对滤波后图像进行分割, 提高了对葵花籽壳内籽仁形态的检测准确性。实验结果表明, 相比传统检测方法费时费力, 易造成原材料污染等缺点, 太赫兹时域透射成像技术可以在不破坏葵花籽外壳的情况下将壳内籽仁的形态以图像的形式直观呈现, 结合形态学滤波算法及 K-means 图像分割算法, 能够对葵花籽壳内籽仁的品质实现快速无损绿色检测, 为建立基于太赫兹时域光谱图像的葵花籽品质检测模型奠定基础, 为油料作物内部品质的带壳无损检测提供新的方法参考。

References

- [1] LUO Yin, DU Xuan-li, YANG Fan, et al(罗 寅, 杜宜利, 杨 帆, 等). Grain Processing(粮食加工), 2017, 42(1): 50.
- [2] Ma Fei, Wang Ju, Liu Changhong, et al. Food Analytical Methods, 2015, 8(7): 1629.
- [3] LI Yan-ru, GAO Xing-sheng, WU Jing(李艳茹, 高兴盛, 吴 晶). Chinese Journal of Analysis Laboratory(分析实验室), 2016, (7): 796.
- [4] Gunaseelan JayaBrindha, E. S. Gopi Subbu. IEEE Transactions on Emerging Topics in Computational Intelligence, 2018, 2(1): 78.
- [5] Seung Hyun Baek, Heung Bin Lim, Hyang Sook Chun. Journal of Agricultural and Food Chemistry, 2014, 62(24): 5403.
- [6] WANG Fang, GUO Shuai, ZHAO Jing-feng, et al(王 芳, 郭 帅, 赵景峰, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2018, 38(11): 3638.
- [7] Seiji Niijima, Masashi Shoyama, Kodo Kawase. IEEJ Transactions on Electronics Information and Systems, 2019, 139(2): 137.
- [8] ZHAO Feng, LI Jing-ling, YUAN Min-jie, et al(赵 峰, 李静玲, 袁敏杰, 等). Journal of Terahertz Science and Electronic Information Technology(太赫兹科学与电子信息学报), 2018, 16(2): 212.
- [9] DING Li, DING Xi, YE Yang-yang, et al(丁 丽, 丁 茜, 叶阳阳, 等). Chinese Optics(中国光学), 2017, 10(1): 114.
- [10] ZHOU Ping, QI Yang, LI Ze-yu, et al(周 萍, 齐 扬, 李赓宇, 等). Sciences of Conservation and Archaeology(文物保护与考古科学), 2016, 28(4): 133.
- [11] JIAO Jiao, WU Ling-da(焦 姣, 吴玲达). Journal of Image and Graphics(中国图象图形学报), 2019, 24(3): 435.
- [12] XU Li-ming, LÜ Ji-dong(徐黎明, 吕继东). Transaction of the Chinese Society of Agricultural Engineering(农业工程学报), 2015, 31(14): 202.
- [13] YANG Shu-ying, ZHANG Hua(杨淑莹, 张 桦). Pattern Recognition and Intelligent Calculation: MATLAB Technology Implementation(模式识别与智能计算: MATLAB 技术实现). Beijing: Publishing House of Electronics Industry (北京: 电子工业出版社), 2015. 254.
- [14] REN Zhi-bin, SUI Yong-xin, YANG Ying-hui, et al(任智斌, 隋永新, 杨英慧, 等). Optics and Precision Engineering(光学精密工程), 2002, 10(4): 340.
- [15] HUO Feng-cai, SUN Xue-ting, REN Wei-jian, et al(霍凤财, 孙雪婷, 任伟建, 等). Journal of Jilin University • Information Science Edition(吉林大学学报 • 信息科学版), 2019, 37(2): 148.

Study on Internal Quality Nondestructive Detection of Sunflower Seed Based on Terahertz Time-Domain Transmission Imaging Technology

LIU Cui-ling^{1, 2}, WANG Shao-min^{1, 2}, WU Jing-zhu^{1, 2}, SUN Xiao-rong^{1, 2}

1. School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China

2. Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing Technology and Business University, Beijing 100048, China

Abstract The quality of the kernels in the sunflower seed shell directly affects the quality of the edible oil. Using terahertz time-domain transmission imaging technology combined with morphological filtering and K-means image segmentation, the three abnormalities of sunflower seed like kernel damaged, kernel worm-eaten and empty seed were investigated to achieve to explore the quality of seed kernels in sunflower seed hull. According to the national standards and previous experience, three kinds of shelled sunflower seeds samples of kernel damaged, kernel worm-eaten and empty shells were prepared. The terahertz time domain spectrometer TeraPulse 4000 and transmission imaging accessory were used to acquire the spectral images of the above three abnormal samples at a resolution of 0.2 mm, and a spectral image of normal sunflower seed was taken as a reference. Four sunflower seed terahertz images were reconstructed by peak-to-peak imaging. The terahertz images can preliminarily determine the shape of the seed kernel in the shell, but there were still problems such as low contrast and blurred edge information, which needed further optimization. The morphological filtering algorithm was used to filter the terahertz images of sunflower seeds. The flat diamond structure element with a side length of 3 was selected as the collation image for one expansion, and then the external gradient of the image was calculated to complete the filtering process of the image. At the same time, the morphological filtering results were compared with the median filtering results, the mean filtering results and the non-local mean filtering results. It was found that the morphological filtering not only ensured the sharpness of the image, but also preserved the edge information, and could also make an obvious boundary between the sunflower seed sample and background, which was conducive to subsequent image segmentation processing. In order to more accurately detect the shape of the sunflower seed kernel, the filtered images were segmented. The K-means clustering algorithm was used to segment the filtered terahertz images. In order to improve the accuracy of the segmentation results, the number K of different initial cluster centers was determined for the images of different samples, among which the kernel damaged $K=4$, the worm-eaten kernel $K=5$, the empty shell $K=3$, the normal grain $K=4$. The segmented images accurately showed the morphology of the kernels in the sunflower seed hull. The study showed that the terahertz time-domain transmission imaging technique combined with morphological filtering and K-means image segmentation method was feasible for the non-destructive detection of the internal quality of sunflower seeds, which laid a foundation for the establishment of non-destructive testing model for the quality of shelled sunflower seeds, and provided a new method reference for nondestructive detection of the internal quality of shelled oil crops.

Keywords Terahertz; Time-domain imaging; Morphological filtering; K-means image segmentation; Sunflower seed; Non-destructive detection

(Received Nov. 22, 2019; accepted Mar. 9, 2020)