Invited Paper

Detection method of cottonseed fullness on the basis of terahertz spectroscopy and imaging technology

Junjie Wang, Yang Li, Jingzhu Wu *, Cuiling Liu, Xiaorong Sun, and Shanzhe Zhang
Beijing Key Laboratory of Food Safety Big Data Technology,
Beijing Technology and Business University, Beijing 100048, China
* E-mail: pubwu@163.com

(Received March 2023)

Abstract: Cottonseed fullness is an important indicator in characterizing the quality of cottonseed. This paper explores the application of terahertz time-domain spectral transmission imaging technology with image refactoring image processing and other methods in cottonseed fullness detection research. In this study, the Terapulse 4000 terahertz (THz) time-domain spectroscopy system and a transmission imaging accessory are used to collect THz spectral images of cottonseed samples. Certain differences can be observed in the composition and spatial distribution of the seed kernel and the seed cavity for which terahertz images at different depths is useful in clearly distinguishing the seed kernel and the seed cavity. To explore the imaging results of the samples under different refactoring methods and use the maximum inter-class variance method (Otsu method) to remove the background to obtain complete sample images, the mathematical morphology algorithm is used to select the circular structural element with a side length of 3 as the check image for dilation corrosion and non-destructively extract the images of different tissues of cottonseed. The THz time-domain spectral images are organized, and then a fullness model of cottonseed based on the THz time-domain spectral images is established according to the enrichment quantification calculation formula. Experimental results show that the non-destructive testing of cottonseed samples can be achieved by using terahertz time-domain spectral imaging technology and is expected to provide a theoretical basis and a method reference for shelled seeds' non-destructive testing.

Keywords: Terahertz time-domain spectral transmission imaging technology, Cottonseed fullness, Image refactoring, Image processing

Doi:

1. Introduction

Cottonseed, a Gossypium plant of the Malvaceae family, is the main fibre crop grown in the world. China is the world's second largest cotton producer and largest cotton consumer[1, 2]. The excellent quality of cottonseed is an important foundation of the development of the cotton industry. Seed fullness is the proportion [3] of seed kernel in the seed cavity. Studies have shown a significant positive correlation between the germination rate and fullness of cottonseed, so the research on the fullness of cottonseed has a certain practical significance in guiding the growth and development of high-vigour cottonseed[4, 5]. Traditional methods for measuring seed fullness include expert experience, 1000-seed weight and the seed density method[6-7]. The physical properties of seeds are used to determine their fullness, but the above methods are highly subjective and cannot provide accurate measurement. Furthermore, cottonseed is a shelled seed, and the quality of the internal kernel is difficult to observe without destroying the seed shell. Meanwhile, because the seed coat is thick and cannot reflect the size of the seed kernel, the new requirements of non-destructive testing proposed by the development of modern agriculture are difficult to meet.

Terahertz pulses are in transition from macro-electronics to micro-photonics[8]. They have dual characteristics of electronics and photonics and can perform spectral analysis on the chemical properties of different substances. Meanwhile, the width of the terahertz pulse is in the order of picoseconds and can reflect the depth information of the material with high precision[9]. Therefore, optical imaging technology with terahertz radiation as the light source has advantages that make traditional optical non-destructive testing methods difficult to replace and has gradually been used extensively in the agricultural field[10, 11]. Jiang et al.[12] used a Terahertz (THz) imaging system to identify wheat grains at different stages of germination. It was found that the terahertz spectra of the main components of wheat grains, maltose and starch were significantly different. The first five images of principal components analysis (PCA) were input into the partial least squares regression (PLSR), least squares support vector machine (LS-SVM) and back propagation neural network (BPNN) model, respectively. The germination time of 7 seeds was classified from 0–48 *h*, and the prediction accuracy could reach 90%. Jiang et al.[13] used terahertz imaging technology to collect THz images of polyethylene and maltose wheat flour mixtures, obtained the regional average spectral signals, and performed PCA on the characteristics of the THz spectra. The extracted image features were then used for the predictive analysis of the maltose content of the mixture using various models. Sun et al.[14] used terahertz transmittance imaging to measure the plumpness of sunflower seeds. In the range of 0.5–2.0 *THz*, worms, defective shells and nuclei could be identified, while THz and RGB images were used to build a plumpness measurement model with a high prediction accuracy, i.e. R2 = 0.91 and RMSEP = 4%. The above research shows that terahertz imaging technology has great advantages in exploring seed quality characteristics, but most of the research uses tablet sampling that is destructive and does not meet the new development needs of modern agriculture.

Meanwhile, existing domestic and foreign scholars have also conducted some studies on cottonseed on the basis of terahertz technology. Li and Tu, et al.[15, 16] used THz time-domain spectroscopy to collect non-transgenic and transgenic cottonseeds and analyzed the absorbance spectrum in each region to discuss the spectral characteristics of transgenic and non-transgenic samples for the qualitative analysis of cottonseeds, providing a basis for cottonseed gene detection based on terahertz technology. These works show that the use of terahertz technology to study cottonseeds is theoretically feasible, but the research on non-destructive testing is still in its infancy. The present study takes cottonseeds as the research object and explore the feasibility of applying terahertz time domain spectroscopy and imaging technology combined with image reconstruction and image processing to non-destructive detection of cottonseed fullness. This work provides a theoretical basis and a method reference for the further non-destructive detection of thin-shell seeds using terahertz technology.

2. Experiment and methods

A. Sample Preparation

The experimental cottonseed samples of the insect-resistant cotton variety are purchased from a seed company. The cottonseed is selected with full grain without obvious damage, as shown in Figure 1(a). To fully explore the use of terahertz time-domain spectroscopy and imaging technology to image the internal quality of the cottonseed state, the seeds are artificially damaged as shown in Figure 1(b).

Fig. 1 Cotton seed samples : (a) Whole cottonseed; (b) Broken cottonseed

B. Experimental setup

In the experiment, a TeraPulse 4000 terahertz time-domain spectrometer (Figure 2(a)) and a transmission accessory (Figure 2(b)) produced by TeraView Company in Cambridge, UK are used to collect sample images. The terahertz transmission imaging method is pulse imaging, which can emit terahertz waves with frequencies from 60 *GHz* to 4 *THz*. The parameters of the spectrometer are set as follows: the resolution is 0.953 cm^{-1}, the signal-to-noise ratio can reach 70 *dB*, the scanning area range is 20 *mm* × 20 *mm*, and the experimental environment temperature is maintained at approximately 22 ℃.

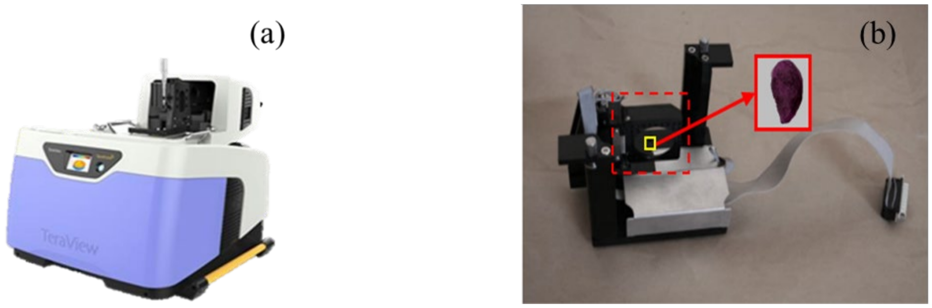

Fig. 2 Experimental instruments : (a) Terahertz pulse spectrometer; (b) Terahertz transmission imaging accessory

C. Data acquisition

First, the sample is fixed on a two-dimensional translation stage and placed at the centre of the focal plane between the terahertz emitter and the detector. The sample is moved step by step in the horizontal and vertical directions with the imaging accessory. Each point on the sample is scanned to obtain its terahertz spectral information. Then, a terahertz image is obtained by computer processing, realizing the terahertz spectral imaging of the sample.

Terahertz time-domain spectroscopy and imaging technology can obtain rich sample information, including not only the spatial information of the sample but also the information on the time axis of the sample. Each pixel in the THz image corresponds to a spectral signal, the amplitude or phase imaging at any moment can be obtained from the time-domain spectrum, and the distribution of the sample spatial density, the refractive index and the sample thickness can be reconstructed using the time-domain spectrum.

The main components of the cottonseed shell are cellulose and lignin, accounting for more than 70% of the seed shell, while the seed kernel is rich in adipose and protein, accounting for 32.2% and 39% of the seed kernel, respectively[17]. Figure 3(a) shows the differential imaging of the damaged sample, and Figure 3(b) shows the time-domain spectra at different positions. The THz image information indicates that different tissue components have different absorption intensities of terahertz light. The intensities of the pixel values in different tissues vary and combined with the spectral information at different positions reveal that the spectra of the seed cavity and the reference signal are roughly similar in spectral line trends, and the signal intensity of the seed cavity is significantly lower than that of the reference signal, indicating that the seed shell has a certain absorption effect on the terahertz waves, resulting in accelerated signal attenuation. The time-domain spectrum of the seed kernel is clearly distinguishable from the seed cavity signal, and the characteristic peak of the seed kernel time-domain spectrum shifts to the right relative to the reference signal, indicating that a significant delay occurred after the terahertz wave penetrates the seed shell and kernel. Meanwhile, at -3*ps*, obvious absorption peaks can be observed, providing important sample information for the subsequent image refactoring.

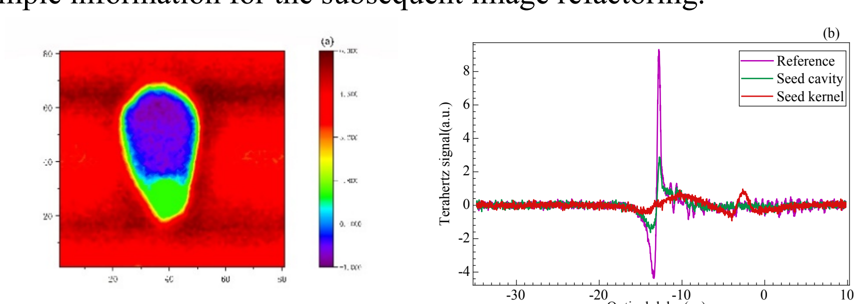

Fig. 3 Differential time slice and spectral comparison of samples : (a) Differential time slice of damaged sample; (b) Time-domain spectra at different locations

3. Calculations

A. Image refactoring

Each pixel in the THz image corresponds to a spectral signal, and image refactoring is performed by imaging the spectra of all the pixel positions of the sample in different ways. Different refactoring methods highlight different sample information. Commonly used image refactoring methods include Time slice imaging, Time difference imaging, Peak–peak Image[18] and correlation coefficient imaging.

1) Time slice imaging (Time Slice): This method obtains information by imaging the measured signal intensity of the sample at a certain time delay. Different time delays are selected to obtain the single-layer transmission images formed by terahertz waves, reaching different depths of the sample at a specific time point.

2) Time difference imaging (Differential Time Slice): This approach selects two delayed moments, subtracts the amplitude of the signal at the start delay time from the signal amplitude at the end delay time, and performs imaging through the signal difference.

3) Peak–peak Image: In this method, each pixel corresponds to the difference between the maximum and minimum values on the time-domain spectrum and subjected to two-dimensional imaging. This refactoring method can present the contour features of the sample to the greatest extent.

4) Correlation coefficient imaging: This method is characterized by a known spectrogram, and the known distribution information of each sample tissue in the image analysis area is provided by the correlation coefficient of each point. That is, taking a certain spectrum as the standard, the correlation coefficient between the spectrum of each pixel on the image and the standard spectrum is calculated and uses for imaging[19]. In this study, the centre position spectrum of the sample image is selected, that is, the seed kernel spectrum, and the correlation coefficient between the spectrum of each pixel point of the image and the seed kernel spectrum is calculated and used for visualization. The correlation coefficient is calculated using Equation (1).

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{1}$$

The value range of correlation coefficient r is 0 to 1, and its size can reflect the degree of correlation between the spectrum of each pixel and the spectrum of the seed kernel. When $r > 0.7$, a strong correlation exists between the spectrum at the pixel and the spectrum of the seed kernel, that is the seed kernel area. The smaller the r value is, the lower the similarity of the response spectrum is, and that is the background area or the seed cavity area.

B. Quantification of cottonseed fullness

Terahertz waves have a good penetrating ability, and the radiation of terahertz waves does not cause photoionization to destroy the substance to be tested, so the internal quality information of the sample can be obtained safely[20]. The cottonseed fullness degree model in this paper is characterized by the ratio of the kernel area to the seed shell area, as shown in Formula (2):

$$P = \frac{s_1}{s_2} \times 100\% \tag{2}$$

where s_1 is the kernel area, s_2 is the seed shell area, and P is the seed fullness. The area is obtained by counting the number of pixels of the seed kernel and shell.

4. Results and discussion

Terahertz pulse imaging is formed by the two-dimensional point-by-point linear scanning of the sample to be tested on the focal plane of the terahertz wave driven by a stepping motor. Jaggedness[21] can be observed around the sample image. To obtain a clear image to eliminate the edge jaggedness, the odd-numbered lines in the image are moved by minimizing the mean square error of the time domain signal between the odd-numbered line and its two adjacent even-numbered lines.

A. Analysis of image refactoring

Given different tissue components, the absorption intensity of terahertz light varies, so the final detected signal intensity is reflected in the obvious colour difference in different positions of the image. The imaging results of different refactoring methods are discussed with the whole grain of insect-resistant cotton. Figures 4(a), (b) and (c) show the Time Slice of cottonseed at -10*ps*, -12*ps* and -14*ps*, respectively. Figure 4(b) shows that the same time delay is affected by physical and chemical factors, such as composition and density, and the colour depth on the image is different. The comparison of Figures 4(a), (b) and (c) reveals that at different time delays, the samples at the same pixel position also show different images and signal intensity information, indicating that the absorption effects of terahertz signals through the samples vary at different depths. As shown in Figure 3(b), kernels have absorption peaks at -10*ps*, and imaging at the peak may provide more sample information, as shown in Figure 4(c). Complete kernels imaging information can be seen at -10*ps*, but the exact location information of the seed shells cannotbe found. Meanwhile, the approximate contour information of the seed shells can be observed at -12*ps* in Figure 4(b).

However, the kernel image information is fuzzy.

Figure 4(d) and (e) show the time difference imaging of the samples in the intervals of -9*ps*–2*ps* and -13*ps*–-11*ps*. The most complete cottonseed image can be obtained through time difference imaging. The clear boundary can clearly distinguish the seed shell, kernel and cavity and compensate for the fuzzy boundary of the tissues of different samples in time slice imaging at a single time delay position. Figure 4(f) shows the Peak–peak results of the cottonseed samples. Compared with time-difference imaging, Peak–peak Image refactoring can effectively eliminate background noise and the influence of sample thickness and density, thereby distinguishing the sample contour and the background information clearly.

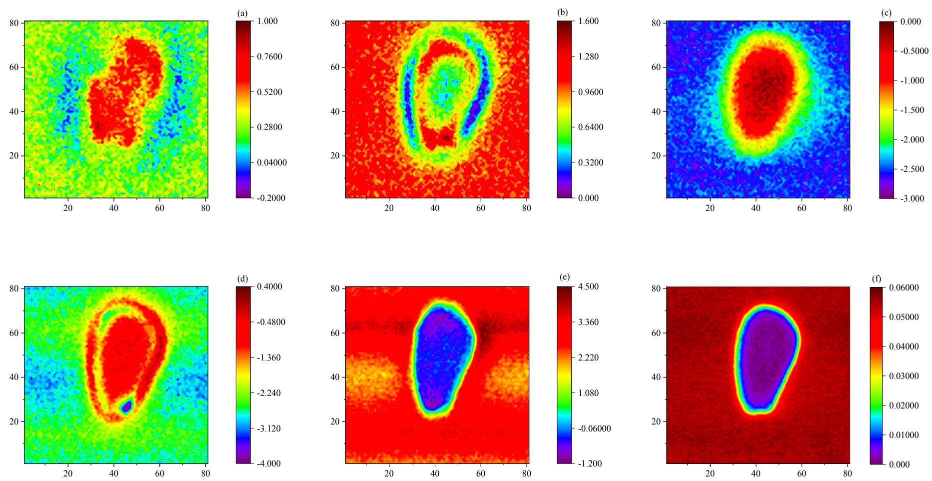

Fig. 4 Different image refactoring methods in Terapulse : (a)-14ps Time Slice; (b)-12*ps* Time Slice; (c)-10*ps* Time Slice; (d) Time Slice of -9*ps*–2*ps*; (e) Time Slice of -13*ps*–-11*ps*; (f) Peak–peak Image.

By extracting the full pixel spectrum of the sample transmission imaging, the kernel spectrum is selected at the imaging centre, the correlation coefficient between the kernel spectrum and each pixel is calculated, and the correlation coefficient is visualized according to the location information of each pixel. The correlation coefficient value ranges from [0, 1]. Different cottonseed tissues can be distinguished according through correlation coefficient imaging, as shown in Figure 5.

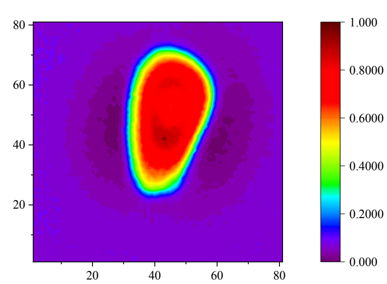

Fig. 5 correlation coefficient imaging result

In summary, Different refactoring methods show differences in the composition and spatial distribution of seed kernels and hulls. Under temporal difference imaging, the contour information of complete seed hulls and kernels are presented, supplementing single time slice imaging. Meanwhile, in the peak-to-peak refactoring mode, maximum and minimum value difference imaging can be used on the time-domain waveform of each pixel point to display the most complete contour information and eliminate the influence of sample thickness. At the same time, the correlation coefficient imaging method based on the characteristic spectrum extracted in this study, considering the correlation degree between the spectra of different tissues of the sample, the correlation coefficient imaging results can clearly distinguish the seed cavity and the seed kernel area, facilitating the accurate determination of the fullness. Therefore, the time slice of -13*ps*–-11*ps* and the Peak–peak and correlation coefficient imaging results to quantify the subsequent enrichment degree.

B. Image processing results

The image processing first considers the background segmentation of the image. In this study, the maximum inter-class variance method (Otsu method) is used for background segmentation. The Otsu method[22] is an adaptive threshold segmentation method. Through statistical methods, the grayscale variance between the target and background areas is maximized so that the image is divided into two parts, exhibiting the greatest separation and facilitating the separation of the cottonseed from the background. The time slice of -13*ps*–-11*ps* is taken as an example. Figure 6(a) shows the segmentation result of the Otsu method. After obtaining the grayscale image of the seed shell, the complete sample image is obtained by using the grayscale binarized image for the mask operation, as shown in Figure 6(b). The part of the seed cavity between the seed kernel and the seed shell is small. Consequently, the seed kernel image is difficult to obtain through simple threshold segmentation, and a large amount of sample contour information is also lost.

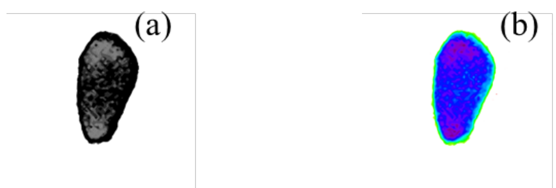

Fig. 6 Otsu result figure : (a) Grayscale image after removing background; (b) Sample image

The image of the seed kernel should retain the edge information in the transmission imaging, and the image information of the seed kernel must be extracted from the image after background segmentation using mathematical morphological operations. The image is processed, using the best processing method that is selected according to the contour edge extraction effect to provide pixel statistics for the subsequent establishment of a substantial quantitative model. This study uses the seed shell image in Figure 6(b) to ensure the effective preservation of image edge information, the morphological circular structure element with a size 3 is used to remove the seed cavity part, and the result is shown in Figure 7(a). Some seed cavity pixels remain after morphological processing, and the seed kernel image require filtering. By extracting the edge contour of the seed kernel, the result obtained by filtering the edge image of the seed kernel with the median filter and then using the edge position information after filtering to obtain the seed kernel tissue area is shown in Figure 7(b), which is the result obtained by using 3 × 3 median filtering. The part of the seed cavity around the seed kernel is removed after using the median filter image, and the edge information is clearly visible, probably because the median filter uses the median grey value of the neighbourhood of the pixel to replace the pixel, to make the surrounding pixel values nearer the real value and eliminate abnormal points.

Fig. 7 Kernel image after morphological and filtering processing : (a) Seed kernel image; (b) 3 × 3 median filter

C. Quantitative results of cottonseed fullness

Formula (2) is used to calculate the cottonseed fullness degree, and the pixel areas of the seed kernel and shell after image processing are used for calculation. The fullness degrees of the insect-resistant cotton samples are obtained and are shown in Table 1. After calculation, the fullness of the cottonseed under different image refactoring methods can exceed 75%. Table 1 shows that compared with Difference Time Slice and correlation coefficient imaging, Peak–peak Image can provide the most contour information. Meanwhile, correlation coefficient imaging can best show the relationship between the different tissue spectra of cotton samples and provide literature references for the in-depth research on cottonseed.

Tab. 1 Statistical table of cottonseed fullness degree under different imaging methods

Image refactoring method	Pixels of the Seed Shell	Pixels of Seed kernel	Cottonseed fullness
Time Slice of -13ps–-11ps	1106	838	75.77%
Peak–peak Image	1407	1130	80.3%
correlation coefficient imaging	1319	993	75.28%

5. Conclusions

The measurement method for cottonseed fullness is studied using terahertz transmission imaging technology. This study has found that the time-domain signals between the seed cavity and the seed kernel are significantly different in the range of -15*ps*–0*ps*. By using a variety of common image reconstruction methods, there exists a visual imaging method based on the inter-spectral correlation coefficient. By comparing the cottonseed THz imaging results, three refactoring results are selected for image processing, and a terahertz time-domain spectral image-based imaging method is established. The cottonseed fullness model obtains the cottonseed fullness using different image refactoring methods. The experimental results show that terahertz transmission imaging technology can be used to measure the fullness of cottonseed.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61807001) and the National Key Research and Development Program of China (2018YFD0101004-03).

References

[1] C. Y. Wang. "Cotton industry status analysis and countermeasure analysis". *Agricultural Development & Equipments.*, 5, 32–33 (2021).

[2] S. A. Tung, Y. Huang, S. Ali, et al. "Mepiquat chloride effects on potassium acquisition and functional leaf physiology as well as lint yield in highly dense late-sown cotton". *Ind. Crop. Prod*., 129, 142–155 (2019).

[3] Shanghai Academy of Agricultural Sciences. " Growth and tissue anatomy of cotton". *Shanghai Scientific & Technical Publishers* [in China], 13–20 (1988).

[4] C. X. Yang. "How to improve cotton seed rate".*China Seed Industry.*, 36, 34–34 (2009).

[5] S. C. Mao. "Chinese cotton cultivation". *Shanghai Scientific & Technical Publishers.*, 34–56 (2013).

[6] Z. H. Yan, W. Z. Xin, Z. Y. Jiu. "Studies On Correlation Between Peel Hardness and Grain Plumpness of Seed-used Zucchini". *China Vegetables*., 1, 69–73 (2021).

[7] X. S. Yuan, W. Y. Ping, Z. Wei, et al. "Determination of Seed Plumpness of Panax ginseng C.A. Meyer". *Special Wild Economic Animal and Plant Research*., 35, No.4, 25–29(2013).

[8] H. J. Guang, Z. Y. Ming, Z. Y. Xin. "'Terahertz Science and Technology' - Introduction". *Acta Optica Sinica*., 46, 9–10 (2019).

[9] W. W. Bin, S. J. Yu, W. X. Ke. “Summerize of Terahertz(THz) Imaging Technology in Dimension Characterization”. *Chin. J. Lasers*., 26, 98–102 (2015).

[10] L. A. Hejri, E. Akbari, A. Toudeshki, et al. “Terahertz spectroscopy and imaging: A review on agricultural applications”. *Comput. Electron. Agric.*, 177 (2020).

[11] Y. Zhang. “Applications of THz Spectral Imaging in the Detection of Agricultural Products”. *Photonics*., 8 (2021).

[12] Y. Jiang, H. Ge, F. Lian, et al. “Early detection of germinated wheat grains using terahertz image and chemometrics”. *Sci Rep*., 6, 21299 (2016).

[13] Y. Jiang, H. Ge, Y. Zhang. “Quantitative analysis of wheat maltose by combined terahertz spectroscopy and imaging based on Boosting ensemble learning”. *Food Chem*., 307, 125533 (2019).

[14] X. Sun, J. Liu. “Measurement of Plumpness for Intact Sunflower Seed Using Terahertz Transmittance Imaging”. *Infrared Millim. Waves*., 41, 307–321 (2020).

[15] B. Li, X.C. Shen. “Preliminary study on discrimination of transgenic cotton seeds using terahertz time domain spectroscopy”. *Food Sci. Nutr*., 8, 5426–5433 (2020).

[16] T. Shan, Z. W. Tao, X. X. Ming, et al. “Principal Component Analysis for Transgenic Cotton Seeds Based on Terahertz Time Domain Spectroscopy System”. *Acta Photonica Sinica*., 44, 176-181(2015).

[17] G. X. Xin. “Study the Spectroscopy and Grain Traits Quickly Identify Cotton Variety Purity and Authenticity”. *Yangtze University* [in China], 9–11 (2014).

[18] G. T. Tian. “Terahertz Pulsed Scanning Imaging Algorithms and Applicaion”. *University of Electronic Science and Technology of China* [in China], 11–25 (2016).

[19] X. Yang, W. S. Bo, G. S. Rong, et al. “Study on Near-Infrared Spectrum Features and Identification Methods of Typical Coal-Rock in Dust Environment”. *Spectrosc. Spectr. Anal.*, 40, 3430–3437, (2020).

[20] Y. Jun, Q. L. Mei, Wu L. Qin, et al. “Research Progress of Terahertz Metamaterial Biosensors”. *Spectrosc. Spectr. Anal*., 41, 1669–1677 (2021).

[21] X. Fei, L. Dong, Y. Yang. “Application of Hard Example Mining to Terahertz Imaging Object Detection”. *Computer Technology and Development*., 31, 124–130, 136 (2021).

[22] Y. M, Z. Ping, W. Sheng, et al. “Research on Image Segmentation Method Based on Top-hat Transformation and Otsu”. *Science Technology and Engineering*.,14, 60–64 (2014).