ORIGINAL PAPER

Study on detection of the internal quality of pumpkin seeds based on terahertz imaging technology

Bin Li[1] · Zhao-xiang Sun[1] · A.-kun Yang[1] · Yan-de Liu[1]

Abstract

At present, the internal quality detection method of pumpkin seeds has the problems such as low efficiency and poor accuracy. Therefore, the combined terahertz time-domain spectroscopy (THz-TDS) imaging technology and K-Means image segmentation method was proposed to achieve efficient and accurate detection of the internal quality of pumpkin seeds in this paper. The samples were prepared based on national standards, and four types of samples were made broken grain 1, broken grain 2, empty shell pumpkin seeds, and whole pumpkin seeds. The terahertz images of the above four samples were acquired, respectively. The acquired terahertz images suffer from the problem of indistinguishability where the husk meets the kernel. Therefore, the K-Means algorithm was used to segment the terahertz image. By calculating the area ratio of pumpkin seed shell and kernel, the grade classification of pumpkin seeds was realized. However, the conventional terahertz image acquisition was time-consuming. In this paper, the frequency domain spectrum was obtained by the Fourier transform of the 0.1–5.0 THz time-domain spectrum of the mixture of pumpkin seed husk and pumpkin seed husk kernel. The characteristic frequency was determined by analyzing the maximum peak and characteristic peak of the frequency domain spectrum, and the single frequency image was obtained. The detection error of the single-frequency image was analyzed by calculating the ratio of the defect area between the real image and the single-frequency image. The average detection errors of single-frequency images were about 6.27% and 4.27% at spatial resolutions of 0.4 and 0.2 mm, respectively. In the quality detection of pumpkin seeds, the single-frequency image can realize the rapid detection of the quality of pumpkin seeds under the premise of ensuring accuracy.

Keywords Pumpkin seeds · Terahertz time-domain spectral imaging technology · K-Means algorithm · Fullness · Fourier transform · Single-frequency image

Introduction

Pumpkin seeds are dry and mature seeds of Cucurbitaceae pumpkin, are rich in nutrients, and have certain medicinal value [1]. With the quiet rise of the pumpkin seed deep processing industry at home and abroad and the continuous increase in the export of pumpkin seeds, the cultivation area of pumpkin is expanding year by year. The quality of pumpkin seeds has an important impact on the processing and export of related products in the later period. Due to improper planting or storage, the seeds inside the shell of pumpkin seeds are often damaged, empty shells, or other abnormal conditions. If the selected pumpkin seeds have the above abnormal particles, the quality of the products in the later stage will be affected. Therefore, it is necessary to detect the internal quality of pumpkin seeds.

Thousand-seed weight is a common method for quality inspection of pumpkin seeds, that is, the weight of one thousand seeds is generally measured by the seeds in the air-dried state. Three groups of 1000 seeds are randomly selected and they are weighed respectively to obtain the average value [2]. However, this method has the disadvantage of low efficiency and poor accuracy. Virgílio G Uarrota et al. [3] used a combination of near-infrared spectroscopy (NIR) and Stoichiometry to achieve rapid, simple, and inexpensive identification of soybean seeds with high or low vigor. Lalit Mohan

✉ Yan-de Liu
jxliuyd@163.com

1   Institute of Intelligent Electromechanical Equipment Innovation, National and Local Joint Engineering Research Center of Fruit Intelligent Photoelectric Detection Technology and Equipment, East China Jiao Tong University, Nanchang 330013, China

Kandpal et al. [4] used the near-infrared hyperspectral imaging (HSI) system combined with the PLS-DA-SR method to predict the vigor of melon seeds with the classification accuracy of the validation set of 94.6%. However, both of them can only be used to detect seed vigor and they cannot be used to visualize the inside of thin-shelled seeds.

Terahertz time-domain spectroscopy is a new approach that has emerged in recent years by combining imaging and spectroscopy information. Terahertz waves have highly transmissive, low-energy, and fingerprinted, and many molecules exhibit unique absorption or scattering spectra in the terahertz range, so the terahertz can be used as a non-ionizing alternative to generate high-resolution images from inside objects [5], and it has good penetration to many dielectric materials and non-polar substances. So, it has been greatly applied in the fields of agriculture and food safety [6, 7], security inspection [8, 9], and medical treatment [10]. Ju Xingang et al. [11] used the Maximum Between-Class Variance (Otus Method) to segment the moth-eaten area in the insect-eaten wheat image collected by THz, and a genetic algorithm was proposed to improve the efficiency of the maximum between-class variance method to solve the optimal threshold and effectively identify Out of the moth-eaten area inside the kernel. Gente et al. [12] used terahertz time-domain spectroscopy instead of x-rays to distinguish between healthy and defective seeds of sugar beet. Yew Li Hor et al. [13] used terahertz spectroscopy and imaging techniques to detect voids, cracks and defects in cork stoppers. Liu Cuiling [14] used terahertz time-domain transmission imaging technology, combined with morphological filtering and K-Means image segmentation to explore the internal quality of sunflower seeds. Those researchers have realized the detection of the internal quality of the sample, but no solution is proposed for the problem that terahertz imaging is time-consuming and cannot be used for rapid industrial detection.

Compared with existing methods, terahertz can be used to take on the internal state of pumpkin seeds without damaging the outer shell. Therefore, in this paper the terahertz time-domain spectral imaging technology was proposed to detect the internal quality of pumpkin seeds, the images of pumpkin seeds with different defect degrees at 0.1–5 THz were obtained by a terahertz imaging system, and the images in the appropriate band (1–1.5 THz) were selected based on their signal to noise ratio, The K-Means algorithm was used to achieve the segmentation of pumpkin seed kernel and pumpkin seed shell area, and the defect degree of pumpkin seeds was calculated to realize the grade classification of pumpkin seeds. However, conventional terahertz imaging was time-consuming. The time domain spectrum of pumpkin seed husk and pumpkin seed husk kernel mixture in the range of 0.1–5.0 THz was obtained by terahertz spectroscopy, the frequency domain spectrum was obtained by Fourier transform, and their spectral characteristics were analyzed to determine the characteristic frequency to obtain a single-frequency image. The defect identification error of the single-frequency image was determined by comparing the area ratio of the defect area between the single-frequency image and the physical image to verify that the feasibility of the single-frequency terahertz image can be used to achieve rapid and non-destructive detection of the quality of pumpkin seeds.

Experiment

Experimental sample preparation

The pumpkin seeds used in the experiment were purchased from Huayue Trading Co., Ltd. in Bayannaoer, Inner Mongolia. In order to study the feasibility of terahertz time-domain spectroscopic imaging technology for the detection of pumpkin seed defects, three abnormal samples were prepared based on SN/T 1963–2007 "Import and Export of Pumpkin Seed Kernels and Sunflower Seed Kernel Sensory Inspection Methods", which were broken grain 1, broken grain 2, empty shell grain and whole grain for control experiments, as shown in Fig. 1 a, b, c and d. Some pumpkin seeds were taken, separate their shells and kernels were separated to make pumpkin seed shell tablets, and then some parts of the pumpkin seeds were taken to pulverize the whole grains to make tablets of the mixture of pumpkin seed shells and kernels. The thickness of the tablets was 0.8 mm, as shown in Fig. 1 e and f. The pumpkin seeds used in the experiment
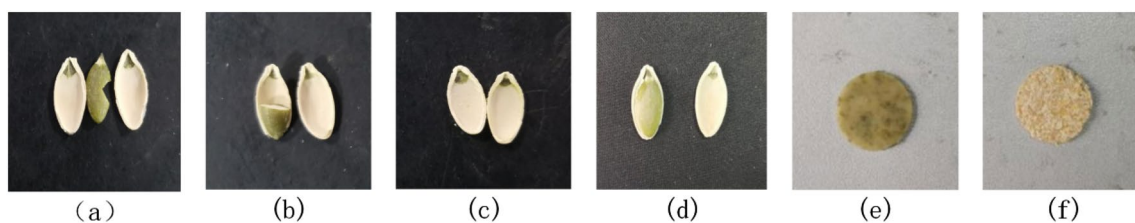

Fig. 1 Experimental sample of pumpkin seeds **a** broken grain 1 **b** broken grain 2 **c** empty shell grain **d** whole grain **e** Pumpkin seed shell 0.8 mm tablet **f** Pumpkin seed shell kernel mixture 0.8 mm tablet

were all dried at 60 °C for five hours before preparing the experimental samples to reduce the influence of the water in the pumpkin seeds.

Experimental instrument

Image acquisition was performed using the TAS7500 terahertz time-domain spectroscopy system. The principle is shown in Fig. 2: the femtosecond pulse emitted by the femtosecond laser has a wavelength of 1550 nm and a pulse width of is less than 50 fs. The femtosecond pulse is divided into pump light and probe light by an optical splitter. The pump light enters the terahertz emitter and produces terahertz radiation. The terahertz pulse emitted by the emitter is focused on the sample after passing through a set of metal parabolic mirrors, and then the terahertz wave passes through the sample and it is focused on the detector with the sample information. At the same time, the detection light is also irradiated on the detection crystal. The THz frequencies used in the experiments are between 0.1–5.0 THz, and the maximum area of the transmission scan can reach 150 mm × 150 mm. The movement of the X–Y horizontal platform is driven by a stepping motor to realize point-by-point scanning of the sampling area.

Terahertz data acquisition

The ambient temperature of the laboratory is controlled at about 23 °C, and the spectral data of 0.1–5.0 THz is measured, the resolution is set as 7.63 GHz, and the number of single-point scans is 256. During the experiment, it is necessary to use an air compressor to continuously input dry air into the experimental chamber to keep the humidity of the experimental chamber below 10% to reduce the influence

of environmental moisture on the experiment. Each sample is measured 10 times, and the average value is taken as the terahertz spectrum of this sample. In the terahertz band, the polyethylene is transparent, and polyethylene does not absorb terahertz waves, so a polyethylene plate with a thickness of 5 mm is used as a platform for placing pumpkin seeds. The terahertz spatial resolution is set as 0.4 mm, the average number of scans per point is set as 16, and the images of 0.1–5.0 THz are collected.

From the time domain waveforms in Fig. 3 a, it can be seen that the maximum peaks of the pumpkin seed shell and pumpkin seed kernel mixtures are different. The longer the scanning time, the delay times of different substances are different, and the longer the delay time of the substance signal, the worse the peak attenuation signal is. This phenomenon is caused by the absorption, reflection and dispersion of terahertz pulses on the sample surface. In addition, the wave amplitude of the sample exhibits a certain degree of attenuation due to reflection, dispersion and absorption at the sample surface. There is a considerable attenuation of the amplitude but little change in the waveform, which is caused by the reflection and scattering of the sample signal and the overall absorption of the terahertz wave signal by the sample.

As shown in Fig. 3 B, the absorption curves of the pumpkin seed shell and the pumpkin seed kernel mixture are different. The small absorption coefficient of the pumpkin seed shell indicates that the cellulose and lignin contained in the pumpkin seed shell have little absorption of terahertz waves, and terahertz waves can easily penetrate the shell of pumpkin seeds; while the strong absorption coefficient of the pumpkin seed shell kernel mixture indicates that other substances such as oil in the pumpkin seed kernel have strong absorption of terahertz, and terahertz cannot easily pass
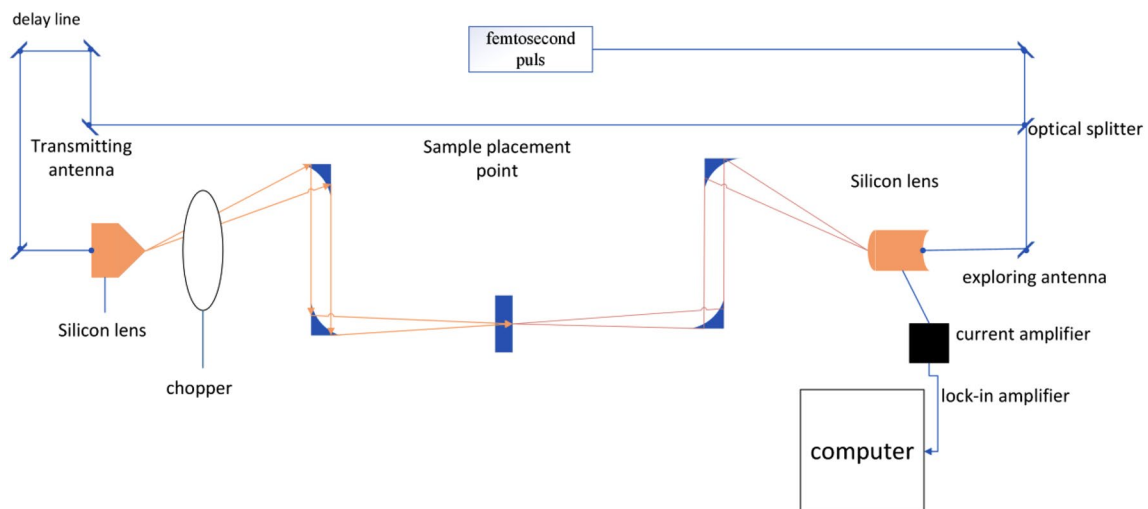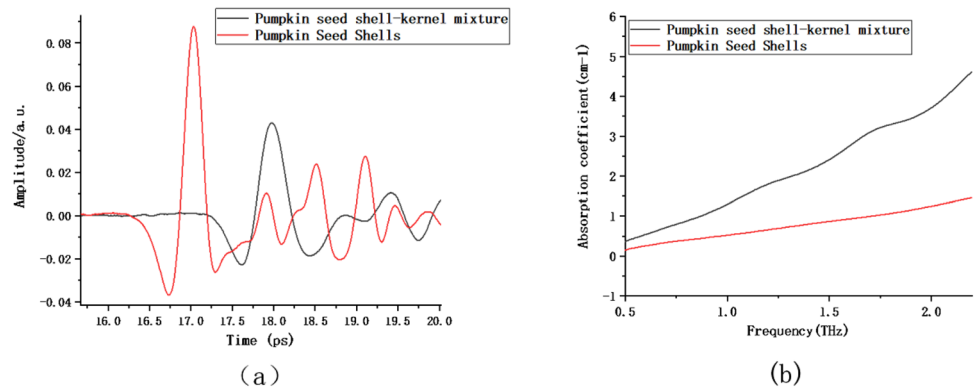

Fig. 2 Principle of THz time-domain spectroscopy system

Fig. 3 Time domain waveforms and absorption coefficients of pumpkin seeds **a** Time domain waveform **b** Absorption coefficient

through. It indicates that terahertz can be used for nondestructive detection of the internal quality of pumpkin seeds.

Data processing

Terahertz images of 0.1–5.0 THz were acquired, and the terahertz images of four frequency bands (0.1–0.5 THz, 0.5–1.0 THz, 1.0–1.5 THz, 1.5–2.0 THz) were extracted. By comparing the image information, the pumpkin seed images of 1–1.5 THz had the highest signal-to-noise ratio, and the outline of the pumpkin seed shell kernel and defects were visible. The image segmentation algorithm was used in MAT-LABR2018b to segment the 1–1.5 THz images to achieve the segmentation of seed kernels from the background. The image was processed by the image segmentation algorithm to achieve complete separation of the pumpkin seed kernel from the background and accurately present the inner kernel of the pumpkin seeds.

Threshold segmentation algorithm

The threshold segmentation algorithm uses a fixed threshold value within the whole image to segment the image, and the threshold value is determined by the gray level histogram of the target and background in the image, and when the gray level histogram has a bimodal distribution, the gray level corresponding to the valley between the two peaks is selected as the threshold value. Pixels with gray level values greater than the threshold are set as white, and pixels less than or equal to the threshold are set as black [15].

Edge detection

Edge detection is a very important method for image feature extraction in the field of computer vision. Edge detection is used to find the set of pixel points in an image where the pixel brightness changes dramatically, and these sets are expressed as contours [16]. The edge points of the image are detected and the points are connected according to a certain method to form the contour of the target.

K-means algorithm

The K-Means algorithm is an unsupervised learning and also a division-based clustering algorithm [17]. The basic idea of the K-Means algorithm: the similarity measure is used to measure the relationship between all data in the dataset, and the data that are more closely related are divided into a set. The K-Means algorithm needs to select K initialized clustering centers, and the distance from each data object to the K is calculated, Then, the mean value of the data objects in each cluster is recalculated and the mean value is used as the new cluster center, and finally the distance from each data object is calculated to get the new K initialized cluster centers and they are redivided. Later, the initialized clustering centers need to be recalculated, and the process is repeated until all data objects cannot be updated to the rest of the dataset [18].

Quality evaluation index

Pumpkin seed fullness is an important indicator of pumpkin seed grade classification. By region labeling the images obtained by K-Means algorithm, the number of pixel points of shell and kernel is calculated. The fullness of pumpkin seeds is obtained by the ratio of the pixel points of shells and kernels of pumpkin seeds.

The formula for pumpkin seed fullness is (1).

$$F = \frac{S_1}{S_2} \tag{1}$$

where: F represents pumpkin seed fullness, S_1 and S_2 represent the number of pixel points occupied by kernels and shells within pumpkin seeds on the image.

Based on SN/T 1963–2007 "Import and export of pumpkin seed kernels, sunflower seed kernels sensory inspection

methods", the grade of pumpkin seeds is classified. The number of kernels whose cotyledons are broken and mutilated and whose lost part reaches or is greater than one-half of the volume of the complete particles of this product is regarded as broken kernels. And the broken kernels accounting for less than 3% are AA grade pumpkin seeds; those less than 5% are A grade pumpkin seeds. The nondestructive detection of fullness of pumpkin seeds by terahertz time-domain transmission imaging technology is of great significance to achieve the classification of pumpkin seeds by grade.

Results and discussion

Image segmentation

The inner kernel of pumpkin seeds is encased in a shell. Normally, the physical and chemical properties of the inner
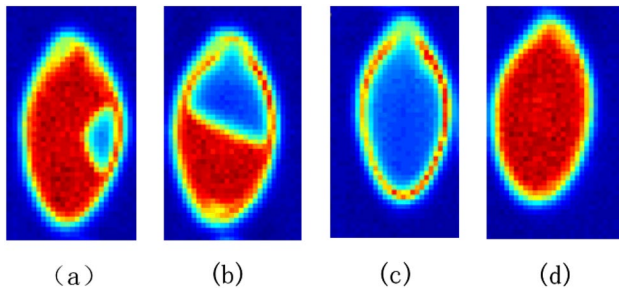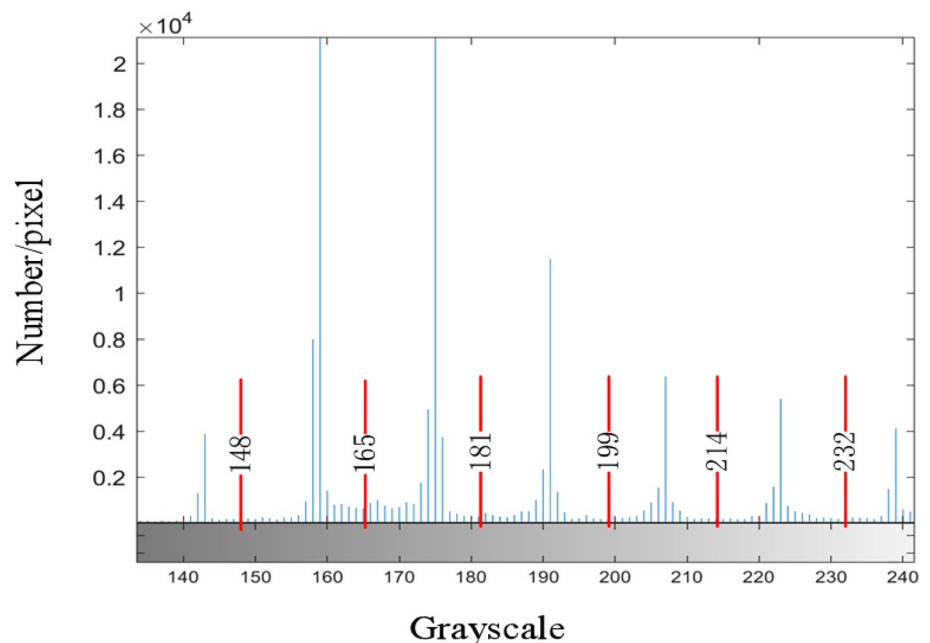

Fig. 4 THz images of pumpkin seeds at 1–1.5 THz **a** broken grain 1 **b** broken grain 2 **c** empty shell grain **d** whole grain

kernel of pumpkin seeds without damaging the shell cannot be obtained. The terahertz waves can pass well through the shell of pumpkin seeds and the information about the inner kernel of pumpkin seeds can be obtained. The terahertz image of pumpkin seeds obtained by terahertz time-domain transmission imaging is shown in Fig. 4, it can be seen from that the morphology of the inner kernel of pumpkin seeds is clearly presented, but the edges are blurred and the distinction between shell and kernel is not obvious.

Image segmentation is one of the important prerequisites of image processing, and segmenting precisely the part you need from a complex image background is beneficial to achieving accurate judgment of image features and it is important for the direction of crop quality detection [19]. Four commonly used algorithms for image segmentation of crops are threshold-based, clustering-based, edge-based and deep learning-based image segmentation algorithms, respectively. To be able to accurately identify the state of the inner kernel of pumpkin seeds, several segmentation algorithms are used to segment the terahertz images of pumpkin seeds and the algorithm with the best segmentation effect by observing the image features is selected.

The grayscale histogram of the image is shown in Fig. 5, and the values of 148, 165, 181, 199, 214, and 232, which are the valleys between the bimodal peaks, are selected as the threshold values of the threshold segmentation algorithm to segment the terahertz image of pumpkin seeds. In thresholding, the threshold value is the key to the image segmentation effect, and the manual threshold segmentation based on the gray histogram of the image has a great randomness for the threshold selection of the image due to the lack of the typical bimodal structure of the gray histogram of pumpkin

Fig. 5 Histogram of pumpkin seeds grayscale

Fig. 6 Segmentation results of THz images of pumpkin seeds under different thresholds **a** Threshold of 148 **b** Threshold of 165 **c** Threshold of 181 **d** Threshold of 199 **e** Threshold of 214 **f** Threshold of 232
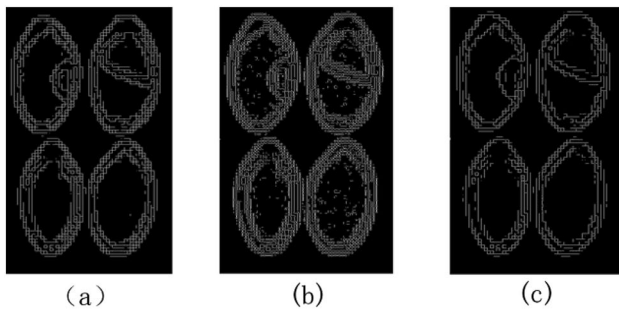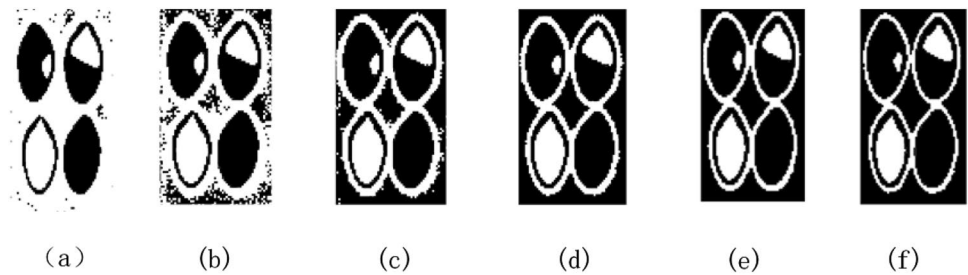
（a）　（b）　（c）　（d）　（e）　（f）

Fig. 7 Pumpkin seed segmentation images based on different operators **a** Sobel operator **b** Laplacian operator **c** Roberts operator

（a）　（b）　（c）

seeds. As shown in Fig. 6, the images segmented by different thresholds have indistinguishable shell-kernel connections and are grouped together.

The edge detection Sobel operator is often used for the edge detection task of images with more noise and gradual grayscale; the Laplacian operator is often used to normalize the detection results and perform variance calculation, which can be used on the blur detection task; the Roberts operator is often used for the edge detection task of images with obvious vertical edges or with steep low noise. In this paper, the edge segmentation of pumpkin seed images is performed by these three operators, and the results are shown in Fig. 7. In the edge-based image segmentation algorithm, because the terahertz image of pumpkin seeds has both the edge information of the inner kernel and the edge information of the outer shell, the edge segmentation cannot guarantee the continuity and closure of the edges. There are a large number of broken edges in the area where the shells and kernels meet, and it is difficult to form a complete region. So, the segmentation effect is poor.

The K-Means algorithm iteratively divides the samples by presetting the initialized clustering centroids K [20]. The value of the initialized clustering centroid K has a direct impact on the image segmentation results. By trying different k-values, the obtained results are shown in Fig. 8. When K = 5, the segmented images have indistinguishable shell and kernel connections and are lumped together. When K = 6, the distinction between the shell and kernel of pumpkin seeds is obvious, the edges are clear, and the inner kernel
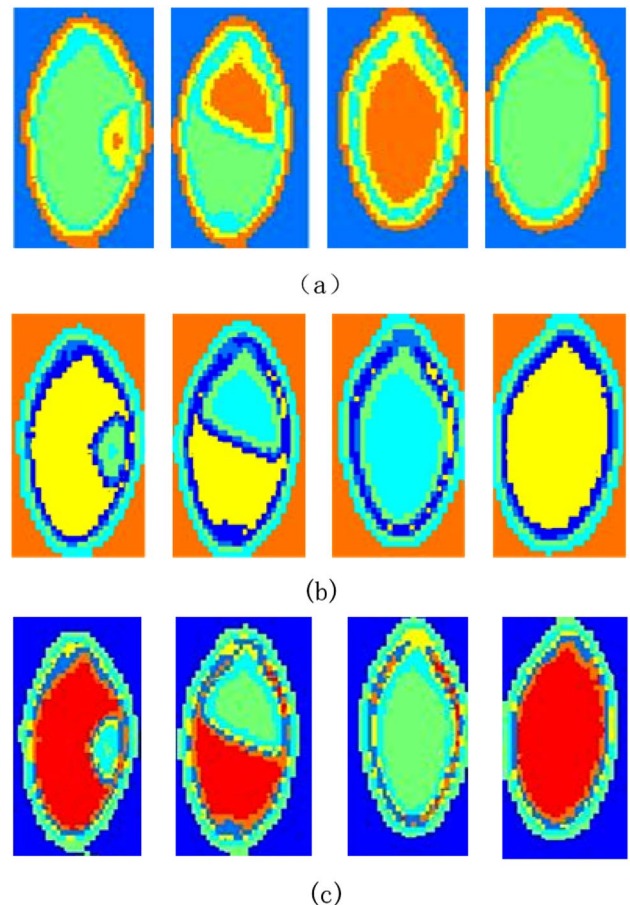
（a）

（b）

（c）

Fig. 8 K-Means image segmentation results for different states of pumpkin seed images **a** K = 5 **b** K = 6 **c** K = 7

morphology is completely expressed. When K = 7, the inner kernel morphology can be seen, and the distinction between the shell and kernel is obvious, but the image is cluttered with too much information in color. Therefore, the initialized clustering center K = 6. It provides great convenience for the subsequent fullness calculation.

Analysis

In the image segmentation of terahertz images of pumpkin seeds, the threshold segmentation algorithm, edge detection

algorithm and K-Means algorithm were utilized, due to the lack of typical bimodal structure of the gray histogram of the threshold segmentation algorithm, the threshold value of the image is difficult to select, so the result of image segmentation is poor. In the edge detection algorithm of the complex edge information of pumpkin seeds, the continuity and closure of the edges cannot be guaranteed, resulting in the existence of a large number of broken edges, which makes it difficult to achieve effective segmentation. The K-Means algorithm can be used to accurately segment the shell and kernel information of pumpkin seeds, and the segmentation effect is good at the shell and kernel junction, which improves the fullness calculation accuracy.

Single frequency images

Conventional terahertz image acquisition is a full-band image, which has a long acquisition time and cannot be used to achieve rapid detection the quality of pumpkin seeds. In contrast, single-frequency images of terahertz have a short acquisition time and low acquisition cost, and have the potential to replace terahertz full-band data acquisition for rapid nondestructive detection of pumpkin seed quality [21].

Single frequency image analysis in different terahertz frequencies

The frequency domain spectraof the terahertz time domain spectral signals of the pumpkin seed shell and pumpkin seed shell kernel mixtures were obtained by Fourier transformation, and it is shown in Fig. 9. In this paper, the frequencies of 0.89 THz and 1.51 THz at the maximum peak and the frequencies of 1.78 THz and 2.31 THz at the two characteristic peaks of the pumpkin seed hull kernel mixture and pumpkin seed hull are selected as the imaging frequencies
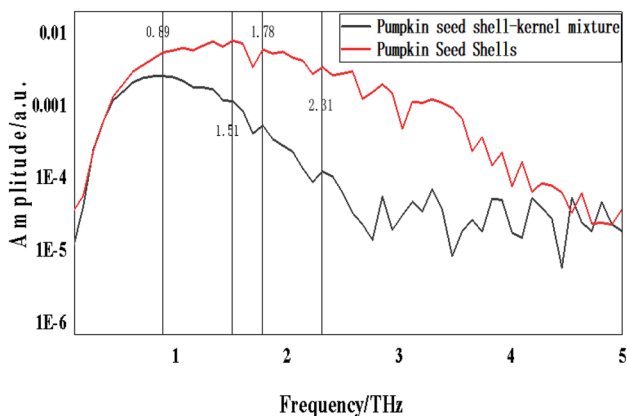

Fig. 9 Frequency domain spectrum

of the single-frequency images. The images after 2.31 THz are too noisy to be used.

Figure 10 shows the single frequency images of pumpkin seeds at different terahertz frequencies. At 0.89 THz, the inner kernel of the pumpkin seeds is visible, but the edges of the pumpkin seeds are blurred. At 1.51 THz, the defective pumpkin seeds are visible, and the outline of the pumpkin seed shell edge is visible with good identifiability. However, when the frequencies are 1.78 THz and 2.31 THz, the terahertz images of pumpkin seeds become blurred again and the shell-kernel contours are not distinguishable. When the frequency is 2.50 THz, the image is distorted and the internal information of pumpkin seeds cannot be effectively identified. Therefore, the single-frequency image at 1.51 THz is chosen for the subsequent processing and calculation.

Defective degree analysis

Although the single-frequency image can be used to distinguish the edge contour of pumpkin seeds, it cannot be used to effectively distinguish at the junction of kernel and shell. However, the defective degree of pumpkin seeds is visible, so the quality of pumpkin seeds can be reflected by the segmentation of the defective area and the calculation of the area of pumpkin seeds.

Table 1 shows the magnitude of the defect area of different pumpkin seeds of single frequency images and 1–1.5 THz images. The comparison of the defect area on the single-frequency image and the 1–1.5 THz image show that the accuracy of the single-frequency image is reduced. Taking the 1–1.5 THz image as the standard, the detection error of a single-frequency image is 8.59% in the broken grain 1, 5.38% in the broken grain 20.65% in the empty shell grain, and 0 in the intact grain, respectively.

Experimental validation

Based on SN/T 1963–2007 "Sensory Inspection Methods for Import and Export Pumpkin Seed Kernels and Sunflower Seed Kernels", 14 sets of experimental samples of pumpkin seeds with different defect degrees were made as validation sets to verify the accuracy problem of single-frequency images for the detection of defects of pumpkin seeds. The physical images of 14 sets of pumpkin seeds and the single frequency images at 1.51 THz were collected. The spatial resolution of the single-frequency images is set as 0.4 mm, the resolution is 7.63 GHz, and the average number of scans at a single point is set as 16. By image segmentation algorithm, the defective area of pumpkin seeds is segmented and the percentage of the defective area is calculated. The detection accuracy of the single-frequency image is examined by using the percentage of defect area of the physical image as the standard.

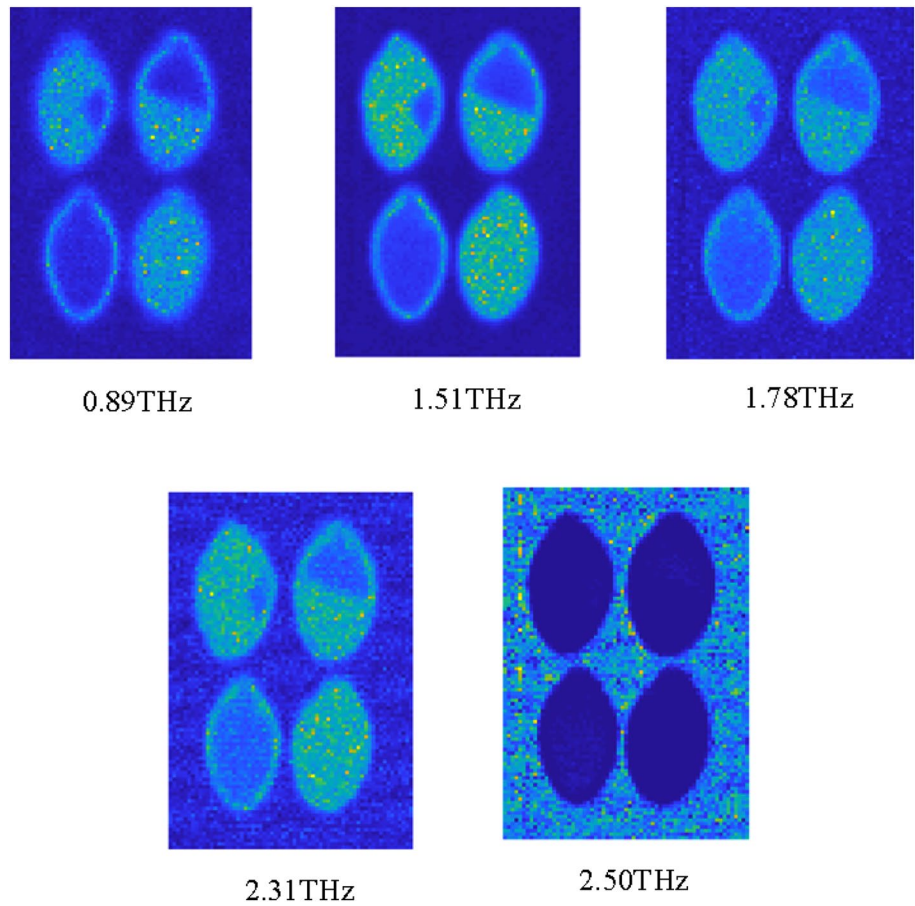**Fig. 10** Single frequency images of differentfrequencies

0.89THz   1.51THz   1.78THz

2.31THz   2.50THz

Table 1 Comparison of defective area of pumpkin seeds in different images

Different quality pumpkin seeds	1–1.5THz image/ pixel	Single fre- quency image/ pixel	Error
Broken grain 1	2585	2363	8.59%
Broken grain 2	14,330	13,559	5.38%
Empty shell grain whole grain	30,308	30,112	0.65%
	0	0	0%
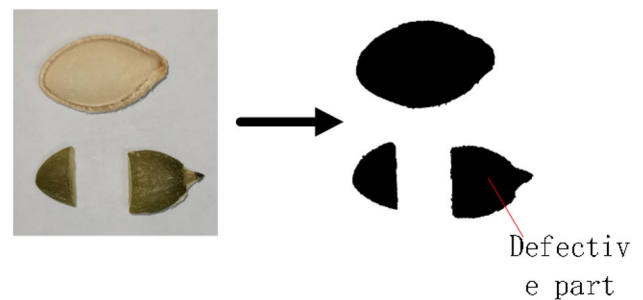
A threshold segmentation algorithm is applied to the physical image to obtain the binarized image of the physical image, the area of the defective region and the intact grain is calculated, and the defective area percentage of the physical image is achieved. The single-channel image for the terahertz single-frequency image is extracted (the second channel is selected), and then the single-channel image is threshold segmented to obtain the binarized image, and the area of each region is calculated to realize the calculation of the percentage of the defected area for the single-frequency image, as shown in Fig. 11. The accuracy and feasibility of the single-frequency image for pumpkin seed detection are examined by compari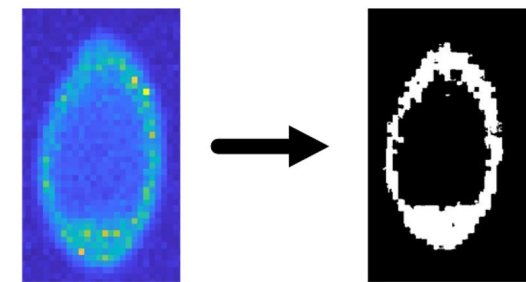ng the defect degree of the physical image with that of the single-frequency image, and the experimental results are shown in Table 2.

Fig. 11 Image segmentation

Table 2 Percentage of defective area of pumpkin seeds in 14 groups with different degrees of defects

	Physical image	Single frequency images(0.4 mm)		Physical image	Single frequency images(0.4 mm)
1	36.52%	29.86%	8	38.70%	31.87%
2	65.63%	58.13%	9	60.77%	53.89%
3	37.73%	30.55%	10	22.34%	13.98%
4	21.32%	14.40%	11	24.07%	19.75%
5	47.25%	42.44%	12	35.69%	29.74%
6	62.40%	56.59%	13	60.73%	54.87%
7	66.69%	61.50%	14	37.08%	31.58%

Table 2 shows the comparison of the percentage of the defective area of pumpkin seeds between the physical image and the single-frequency image (spatial resolution 0.4 mm). Compared with the physical image, the single-frequency image has a certain error in detection accuracy, and the average detection error of the single-frequency image of 14 groups of pumpkin seeds is about 6.27%.

It is speculated that the spatial resolution of the single-frequency images may be one of the factors affecting the detection error, and seven experimental samples of pumpkin seeds with defects closest to 50% are selected, namely, the 3rd, 5th, 6th, 8th, 9th, 12th, and 13th groups of pumpkin seeds. The spatial resolution is set as 0.2 mm, other conditions are kept constant, and the image data of pumpkin seeds are collected. The images are segmented, and the percentage of defective area is calculated, and the results are shown in Table 3. With the spatial resolution of 0.2 mm, the average detection accuracy of single-frequency images of seven groups of samples is about 4.27%, and the detection accuracy of single-frequency images is significantly improved. It can be seen that the single-frequency images can be used to achieve rapid detection with guaranteed detection accuracy. It provides a theoretical basis and experimental reference for achieving rapid, nondestructive and accurate detection of the quality of pumpkin seeds.

Conclusion

In this study, the problem of pumpkin seed fullness was studied by terahertz time-domain transmission imaging. Firstly, the transmission terahertz images of pumpkin seeds with different defective degrees were obtained, and then the K-Means image segmentation algorithm was used to segment them, the detection accuracy of the inner kernel morphology of pumpkin seeds was improved, and the non-destructive detection of pumpkin seed fullness was achieved, thus it was verified that the combined terahertz time-domain spectroscopy imaging technology and K-Means image segmentation method could be used for nondestructive detection of pumpkin seed grade. However, the normal terahertz imaging time was time-consuming. So, the single-frequency terahertz time-domain transmission image was used to detect the quality of pumpkin seeds the average detection error was about 6.27% at 0.4 mm spatial resolution and about 4.27% at 0.2 mm spatial resolution when comparing the percentage of the defective area of single-frequency images and physical images of pumpkin seeds. It is verified that the terahertz imaging technology in single-frequency image can be used for fast, nondestructive and accurate detection of pumpkin seeds quality with guaranteed accuracy.

Table 3 Percentage of defective area of pumpkin seeds in 7 groups with different degrees of defects

	Physical image	Single frequency images(0.2 mm)		Physical image	Single frequency images(0.2 mm)
3	37.73%	35.44%	9	60.77%	56.94%
5	47.25%	43.20%	12	35.69%	30.78%
6	62.40%	59.51%	13	60.73%	56.75%
8	38.70%	32.74%			

Declarations

Conflict of interest The authors declare no competing interests.

References

1. A. Shaban, R.P. Sahu, Pumpkin seed oil: an alternative medicine. J Pharmacogn Phytochem **9**(2), 11 (2017)
2. R. Jevtić, V. Župunski, M. Lalošević, M. Novica, B. Orbović, The combined effects of multiple diseases and climatic conditions on thousand kernel weight losses in winter wheat. Eur J Plant Pathol **152**(2), 469–477 (2018)
3. V.G. Uarrota, C. Segatto, A.P.L. Voytena, M. Marcelo, L.N.V. Avila, D.C.D.S. Kazama, C.M.M. Coelho, C.A. Souza, Metabolic fingerprinting of water-stressed soybean cultivars by gas chromatography, near-infrared and UV-visible spectroscopy combined with chemometrics. J Agron Crop Sci **205**(2), 141–156 (2019)
4. L.M. Kandpal, S. Lohumi, M.S. Kim, J.S. Kang, B.K. Cho, Near-infrared hyperspectral imaging system coupled with multivariate methods to predict viability and vigor in muskmelon seeds. Sens Actuat B-Chem **229**(229), 534–544 (2016)
5. F. Blanchard, L. Razzari, H.C. Bandulet, G. Sharma, R. Morandotti, J.C. Kieffer, T. Ozaki, M. Reid, H.F. Tiedje, H.K. Haugen, F.A. Hegmann, Generation of 1.5 μJ single-cycle terahertz pulses by optical rectification from a large aperture ZnTe crystal. Opt Express **15**(20), 13212–13220 (2007)
6. Y. Liu, X. Du, B. Li, Y. Zheng, J. Hu, X. Li, J. Xu, Detection of purple rice adulteration by terahertz time domain spectroscopy. Spectrosc Spect Anal **40**(08), 2382–2387 (2020)
7. C. Wang, R. Zhou, Y. Huang, L.J. Xie, Y.B. Ying, Terahertz spectroscopic imaging with discriminant analysis for detecting foreign materials among sausages. Food Control **97**, 100–104 (2019)
8. J. Ma, R. Shrestha, J. Adelberg, C.Y. Yeh, H. Zahed, E. Knightly, J.J. Miquel, D.M. Mittleman, Security and eavesdropping in terahertz wireless links. Nature **563**(7729), 89–93 (2018)
9. M. Kato, S.R. Tripathi, K. Murate, I. Kazuki, K. Kodo, Nondestructive drug inspection in covering materials using a terahertz spectral imaging system with injection-seeded terahertz parametric generation and detection. Opt Express **24**(6), 6425–6432 (2016)
10. Y. Liu, H. Liu, M.Q. Tang, J.Q. Huang, W. Liu, J.Y. Dong, X.P. Chen, W.L. Fu, Y. Zhang, The medical application of terahertz technology in non-invasive detection of cells and tissues: opportunities and challenges. RSC Adv **9**(17), 9354–9363 (2019)
11. X. Ju, F. Lian, Y. Zhang, H. Ge, Y. Jiang, The detection of vermiculate grain of wheat using terahertz imaging. J Chin Cereal Oil Ass **33**(08), 106–111 (2018)
12. R. Gente, S.F. Busch, E.M. Stübling, L.M. Schneider, C.B. Hirschmann, J.C. Balzer, M. Koch, Quality control of sugar beet seeds with THz time-domain spectroscopy. IEEE T Thz Sci Technol **6**(5), 754–756 (2016)
13. Y.L. Hor, J.F. Federici, R.L. Wample, "Nondestructive evaluation of cork enclosures using terahertz/millimeter wave spectroscopy and imaging. Appl Optics **47**(1), 72–78 (2008)
14. C.L. Liu, S.M. Wang, J.Z. Wu, X.R. Sun, Study on internal quality nondestructive detection of sunflower seed based on terahertz time domain transmission imaging technology. Spectrosc Spect Anal **40**(11), 3384–3389 (2020)
15. Z.D. Niu, H.D. Li, Research and analysis of threshold segmentation algorithms in image processing. J Phys Conf Ser **1237**(2), 022122 (2019)
16. M. Mittal, A. Verma, I. Kaur, B. Kaur, M. Sharma, G.L. Mohan, R. Sudipta, K. Tai-Hoon, An efficient edge detection approach to provide better edge connectivity for image analysis. IEEE Access **7**, 33240–33255 (2019)
17. Saroj and Kavita, Review: study on simple k-mean and modified K-mean clustering technique. CSSE **6**(7), 279–281 (2016)
18. A. Fahim, K and starting means for K-Means algorithm. J Comput SCI-Neth **55**, 101445 (2021)
19. X.N. Zheng, F. Yang, F.Z. Li, Overview of crop image segmentation algorithms. Mod Comput **2020**(19), 72–75 (2020)
20. H.L. He, B. Sun, Y. Yang, J. Chen, A K-means optimization algorithm suitable for fast clustering of WebGIS massive data. J Phys Conf Ser **2171**(1), 012069 (2022)
21. Y. Shen, Y.X. Yin, B. Li, C.J. Zhao, G.L. Li, Detection of impurities in wheat using terahertz spectral imaging and convolutional neural networks. Comput Electron Agr **181**, 105931 (2021)