

Research Article

An Unsupervised Learning Method for the Detection of Genetically Modified Crops Based on Terahertz Spectral Data Analysis

Shubao Pan,¹ Binyi Qin ,^{2,3} Lvqing Bi,³ Jincun Zheng,³ Ruizhao Yang,³ Xiaofeng Yang,³ Yun Li ,⁴ and Zhi Li¹

¹School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin 541004, China

²Key Laboratory of Complex System Optimization and Big Data Processing, Guangxi Colleges and Universities, Yulin Normal University, Yulin 537000, China

³School of Physics and Telecommunication Engineering, Yulin Normal University, Yulin 537000, China

⁴College of Chemistry and Food Science, Yulin Normal University, Yulin, Guangxi 537000, China

Correspondence should be addressed to Binyi Qin; qby207@163.com and Yun Li; 675030346@qq.com

Received 10 January 2021; Revised 3 February 2021; Accepted 2 March 2021; Published 16 March 2021

Academic Editor: Liguozhang

Copyright © 2021 Shubao Pan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Genetically modified crops have been planted commercially on a large scale since 1996. However, the food safety issue of genetically modified crops remains controversial. Conventional genetically modified crops' detection methods require a plenty of detective time and complex operations that cannot rapidly identify. Previous reports show that combining terahertz time-domain spectroscopy and supervised learning has advanced to identify genetically modified crops, but supervised learning requires large data to train the model. To solve the above problem, we proposed an unsupervised learning method, PCA-mean shift, to identify genetically modified crops. Principal component analysis was employed to reduce the absorbance data dimensionality. After principal component analysis, the first three principal components were used as the input of mean shift. At last, our proposed method had 100% identification accuracy, and *K*-means had 98.75% identification accuracy. The comparison results demonstrated that PCA-mean shift outperforms *K*-means. Therefore, PCA-mean shift combined with terahertz time-domain spectroscopy is a potential identification tool for genetically modified crops' identification.

1. Introduction

A genetically modified crop (GM crop) is a crop whose DNA has been modified using genetic engineering techniques [1]. Genetically modified technique has some advantages in resisting to virus and pests and herbicides [2, 3]. Since 1996, GM crops have been planted commercially on a large scale. There are 26 countries planting GM crops. International Service for the Acquisition of Agri-biotech Applications (ISAAA) reported that the total GM crops' planted area in the world is approximately 190 million hectares in 2018. Soybean, maize, cotton, and rape are the first four planted area GM crops in the world. However, the food safety issue of GM crops remains controversial [4–6]. Polymerase chain reaction [7], southern blot [8], western blot [9], and enzyme-linked immune sorbent assay [10] are

conventional methods to identify GM crops, but these methods require plenty of detective time and complex operations. Thus, development of a practicable and effective analysis method for rapidly identifying GM crops is quite necessary.

Terahertz time-domain spectroscopy (THz-TDS) is a powerful detecting technique that operates in the frequency band from about 0.1 to 10 THz. It has been applied in the detection of biological and chemical molecules such as protein, amino acids, DNA, and harmful residues in agricultural and food products such as melamine, aflatoxin, pesticides, and antibiotics [11–19]. In recent years, some researchers reported the methods of identifying GM crops by using THz-TDS. Liu and Li [20] detected different GM cotton THz spectra by using the support vector machine (SVM). Xu et al. [21] reported discriminate

analysis (DA) and principal component analysis (PCA) have excellent performance to discriminate GM rice from non-GM rice from its parent. Chen et al. [22] proposed combining THz-TDS and chemometrics to identify GM and non-GM sugar successfully. Wei et al. [23] achieved 96.67% accuracy with discrimination of GM rice by combining the THz-TDS image and chemometrics. In our previous research, we proposed a method combining SVM and multipopulation genetic algorithm (MPGA) for identifying GM cotton seed with THz spectroscopy [24]. However, there are two problems in the above research studies. First, most research studies for identifying GM crops are based on combining THz-TDS with supervised learning. As we know, supervised learning requires large data to train the model. Fewer sample data are one of the common problems in practice. Second, the above research studies identify only one crop. How to identify different GM crops is a challenging question.

Mean shift is an unsupervised learning method, which calculates the number of clusters automatically. It does not need to know the previous knowledge about the number of clusters and does not constrain the clusters' shape. Mean shift has been applied in target detection, target tracking, and image segmentation [25–28]. In recent years, Xing et al. [29] developed a colour clustering method for Chinese traditional costume image by mean shift. Wang et al. [30] discovered common visual patterns from two images by using mean shift to group together the close transformations in the space. Ai and Xiong [31] reported that it is able to enhance activation detections by incorporating mean shift and the temporal characteristics of fMRI. As far as we know, there are no studies using mean shift in THz-TDS.

In this paper, we proposed an unsupervised learning method PCA-mean shift to identify GM crops. THz spectrum is high-dimensional data. Firstly, we use PCA to reduce the THz spectrum dimension. And then, the first three principal components are chosen as the input of mean shift. Mean shift is an unsupervised learning method. At last, we compare our proposed method with K -means. The results indicate that PCA-mean shift is better than K -means, and PCA-mean shift is a potential method to identify GM crops.

2. Materials and Methods

2.1. Experimental System. The experimental system comprises a terahertz time-domain spectrometer Z-3 (Zomega Terahertz Corp., USA) and an ultrafast fiber laser (TOP-TICA Photonics Inc., Germany). An ultrafast fiber laser generates 100 fs pulse widths with an 80 MHz repetition rate and 780 nm central wavelength. The spectral resolution is less than 5 GHz, and the peak dynamic range of the entire experimental system is better than 70 dB. The schematic diagram of the experimental system is shown in Figure 1. A laser beam is divided into two parts: a pump beam and a probe beam. The pump beam irradiates a photoconductive antenna to excite a terahertz beam. After that, the terahertz beam goes through the sample. And then, the terahertz beam meets the probe beam at electro-optic crystals ZnTe. The probe beam is modulated by the terahertz beam by the

electro-optic effect. After transmitting through a quarter-wave plate (QWP) and a Wollaston prism (WP), the modulated probe beam is then detected by a set of balanced photodiodes (PD).

All measurements were carried out at room temperature (about 295 K) under the circumstance of a dry air-purged container with the relative humidity of less than 1%. Furthermore, we used the THz-TDS system in the transmission mode.

2.2. Sample Preparation. Two types of GM maize powder (GA21 and MIR604) were purchased from Shenzhen Excellence Biotechnology Co., Ltd. Non-GM maize powder was purchased from a local supermarket. Two types of GM cotton seeds (Lumianyan No.18 and Xinqiu No. k638) were obtained from Shandong Xinqiu Agricultural Science and Technology Co., Ltd.

In the two types of GM cotton seeds, the shell was removed, and they were crushed into powder, respectively. After that, five types of powder (GA21, MIR604, non-GM maize, Lumianyan No.18, and Xinqiu No. k638) were sieved by filtering laws using 100-eye sieves. The sieved powder was dried at 323 K for 1 hour and then was pressed into circular slices about 1.0 mm thick and 13 mm in diameter under a pressure of 8 MPa with a tablet press. At last, five types of specimens were obtained: GA21, MIR604, non-GM maize, Lumianyan No.18, and Xinqiu No. k638. For each type of specimen, 16 samples were prepared.

2.3. Principal Component Analysis. PCA is a common method for data dimension reduction. The fundamental idea of PCA is to approximate an original matrix X by a product of two small matrices shown in equation (1). X is an original data matrix consisting of n rows and p columns, U is a small matrix (called the score matrix) consisting of n rows and d columns, and L is another small matrix (called the loading matrix) consisting of p rows and d columns. T is the transpose of a matrix.

$$X = UL^T. \quad (1)$$

The principal components (PCs) are determined based on the maximum variance criterion. Each subsequent PC describes a maximum of variance that is not modeled by the former components. According to this, most of the variance of the data is contained in the first PC. In the second component, there is more information than in the third one, and so on. Thus, a large fraction of the variance can be described with one, two, or three PCs, and the data can be visualized by plotting the PCs against each other. In our experiment, the original data X were THz spectra of samples.

2.4. Mean Shift. Mean shift is a density clustering algorithm without parameter estimation [32, 33]. It assumes that different clusters in a dataset accord with different probability density distributions. Mean shift can find the direction that the density of a sample point increases fastest. High density of the sample area corresponds to

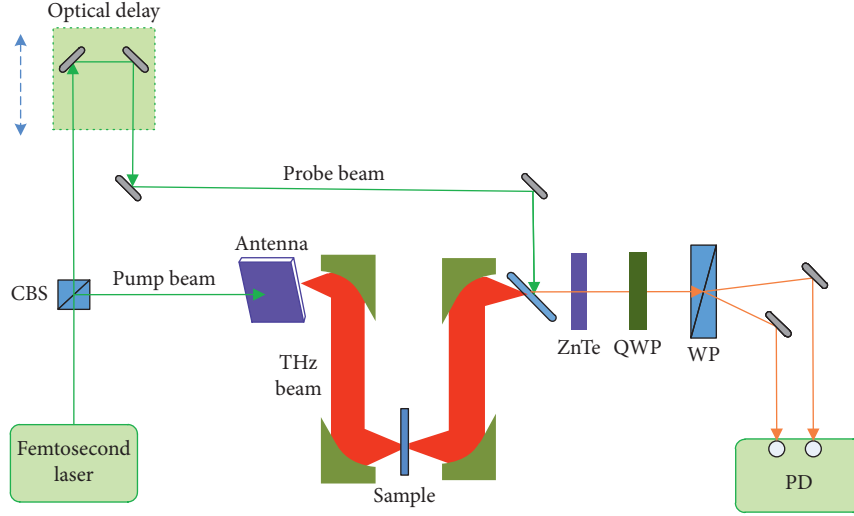


FIGURE 1: Schematic of the experimental apparatus.

the distribution of the maximum value of the sample points. These sample points will end up in a maximum density of local convergence. And the convergence to the same local maximum point is considered to be the same cluster member of the class.

Let $x_i, i = 1, 2, \dots, n$, be a set of d -dimensional points in the space R^d . For a sample point x , the mean shift vector is defined as follows:

$$M_h(x) = \frac{1}{k} \sum_{x_i \in S_h} (x_i - x), \quad (2)$$

where k is the number of points that lie in S_h . S_h is a high-dimensional space with h radius. It is defined as

$$S_h(x) = \{y | (y - x)^T (y - x) \leq h^2\}. \quad (3)$$

The procedure of the mean shift algorithm is as follows:

- Step 1: calculate the mean shift vector of each sample $m_h(x), m_h(x) = M_h(x) + x$
- Step 2: move each sample with $m_h(x)$, such as $x_i = x_i + m_h(x_i)$
- Step 3: repeat Step 1 until the sample point converges, ($m_h(x) = 0$)
- Step 4: samples that converge to the same point are considered to be members of the same cluster

3. Results and Discussion

3.1. Spectroscopic Analysis. There are five types of specimens: GA21, MIR604, non-GM maize, Lumianyan No.18, and Xinqiu No. k638. For each type of the specimen, we measured 16 samples. Figure 2(a) displays the time-domain waveforms of five different specimens. Due to absorption and the refractive index difference between five specimens, the pulse amplitude and time delay are different. To compare the five types of specimens' time-domain

waveforms, a novel type of plot-colour contour mapping of time-domain waveforms in terms of time is employed, as shown in Figure 2(b). Yellow means the pulse peak, and blue means the pulse valley. In Figure 2(b), the location of the pulse peak and valley between GA21 and non-GM maize is similar. Thus, it is impossible to immediately identify GA21 and non-GM maize by THz time-domain waveforms.

To obtain the absorbance, we translate the time-domain waveform of the sample and reference (air) into the frequency domain and then calculate it as follows [15]:

$$\text{absorbance}(\omega) = -\lg \left| \frac{E_s(\omega)}{E_{\text{ref}}(\omega)} \right|^2, \quad (4)$$

where ω is the frequency. And $E_s(\omega)$ and $E_{\text{ref}}(\omega)$ are the amplitude of the sample and reference signal in the frequency domain, respectively.

Figure 3(a) displays the terahertz absorbance spectra of five types of specimens in the frequency range of 0.4 THz to 1.4 THz. Figure 3(b) is the plot-colour contour mapping of absorbance spectra in terms of the frequency. Yellow represents the absorbance is strong, and blue represents the absorbance is weak. In Figure 3(b), the absorbance of MIR604 has strong absorption in the range of 1 THz to 1.4 THz. Lumianyan No.18 and Xinqiu No. k638 both have an absorption peak nearby 1.2 THz and 1.3 THz. For GA21 and non-GM maize, there is no sharp absorption peak in the range of 0.4 THz–1.4 THz. Comparing Figures 2(b) and 3(b), we find that absorbance spectra have more discrimination than time-domain waveforms. So, we selected absorbance spectra for further identification study.

3.2. Detection Results. In order to evaluate the performance of PCA-mean shift, confusion matrix and average accuracy were employed:

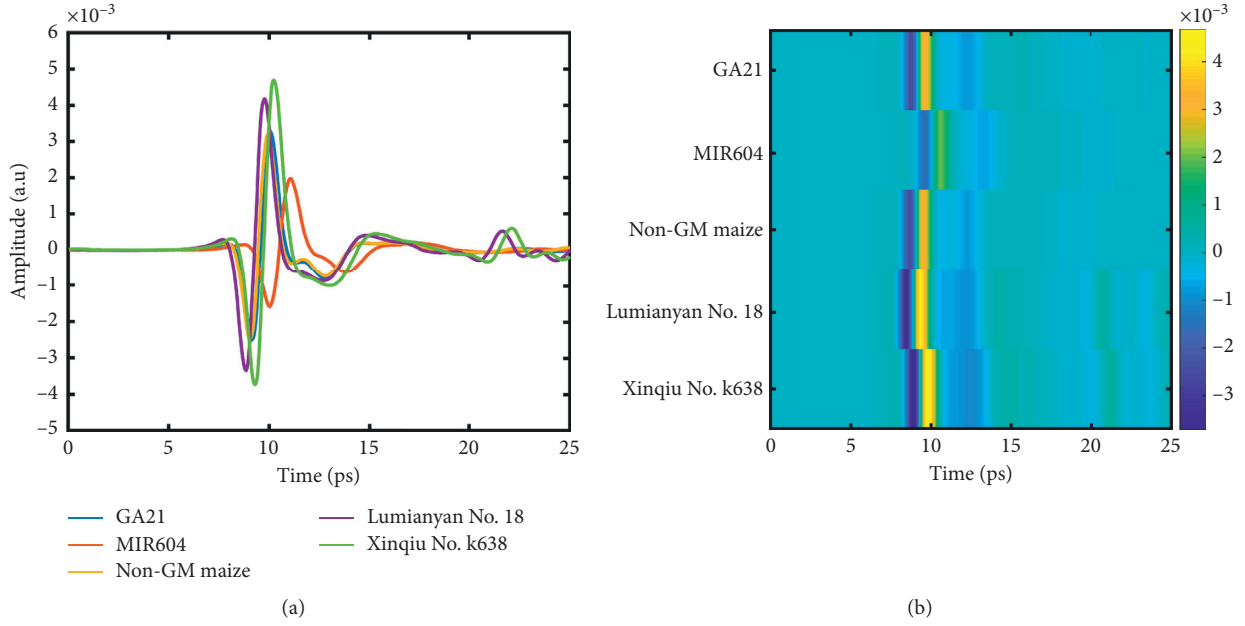


FIGURE 2: Time-domain waveforms of five types of specimens. (a) Time-domain waveforms. (b) Plot-colour contour mapping of time-domain waveforms.

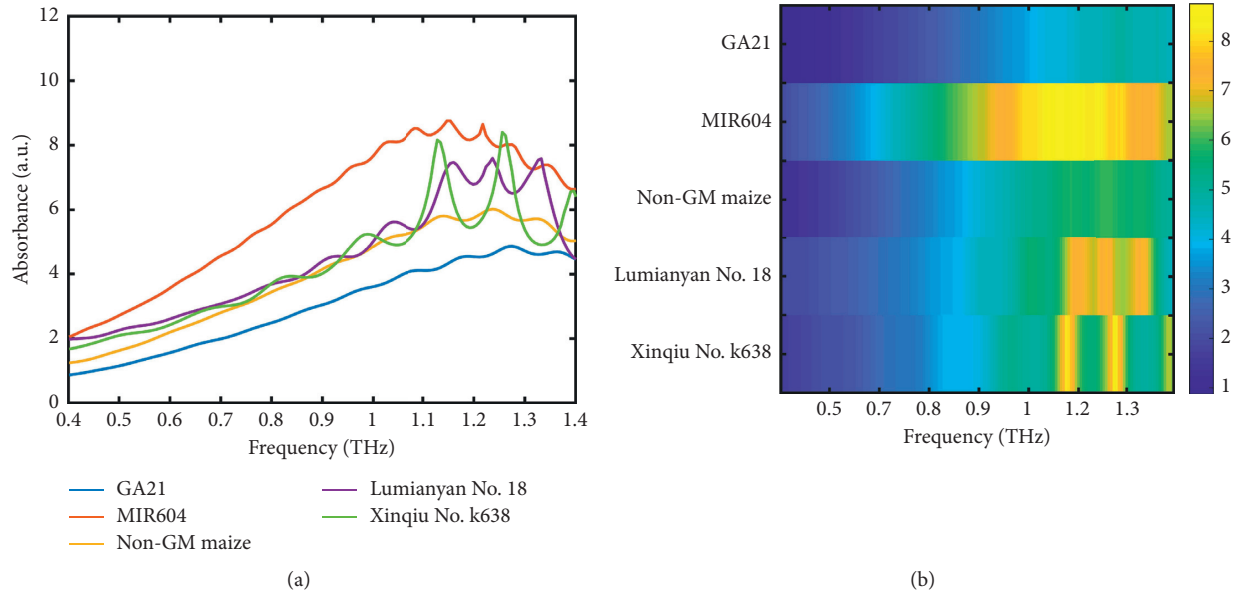


FIGURE 3: Absorbance of five types of specimens. (a) Absorbance. (b) Plot-colour contour mapping of five types of specimens' absorbance.

$$\text{average accuracy} = \frac{\text{the number of correct detection samples}}{\text{total of samples}}. \quad (5)$$

In addition, PCA-mean shift was compared with the common unsupervised learning method K -means. Firstly, we construct a dataset called Dataset1 by using absorbance spectra of five different specimens. Details of Dataset1 are shown in Table 1. And then, we use PCA to reduce the dimension of Dataset1.

By using PCA, absorption spectra were reduced from 80 dimensions to 3 dimensions. The variance contribution rate and cumulative variance contribution rate are listed in Table 2. Usually, as the cumulative variance contribution rate is large enough (typically $\geq 85\%$), the original dataset can be replaced approximately [34]. The cumulative variance contribution rate of the first three PCs is 90.43%. It means that the first three PCs contain major information of the original absorption spectra. Figure 4(a) shows the two-dimensional score of the first two PCs. After performing PCA, all the cotton seed samples are

TABLE 1: Composition of Dataset1.

Specimens	Number of samples	Region of frequency (THz)
GA21	16	0.4-1.4
MIR604	16	0.4-1.4
Non-GM maize	16	0.4-1.4
Lumianyan No. 18	16	0.4-1.4
Xinqiu No. k638	16	0.4-1.4

TABLE 2: The variance contribution rate and cumulative variance contribution rate.

Component	Variance (%)	Cumulative rate (%)
PC1	61.14	61.14
PC2	25.41	86.55
PC3	3.88	90.43

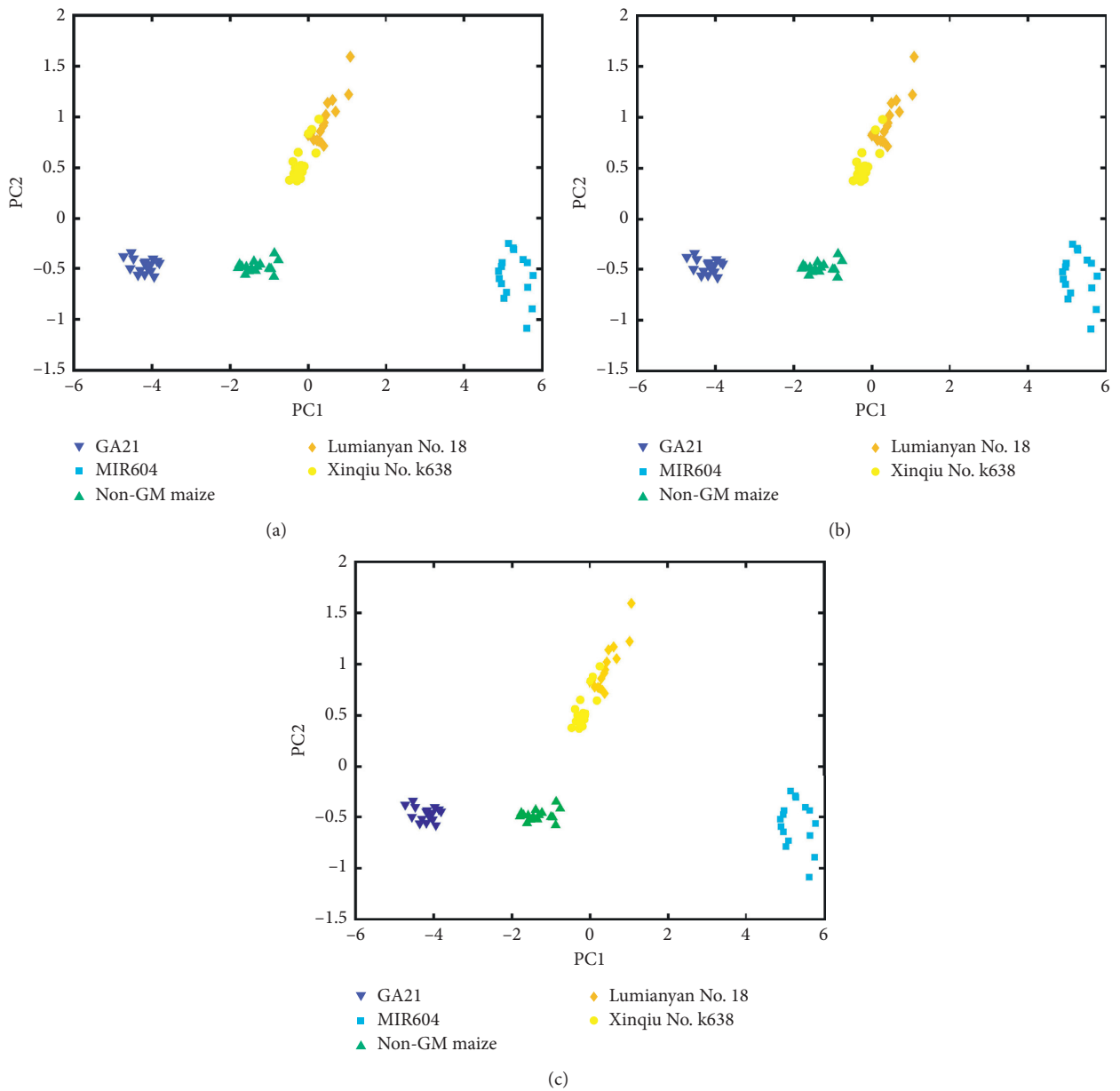


FIGURE 4: Results of detection. (a) PCA. (b) K-means. (c) PCA-mean shift.

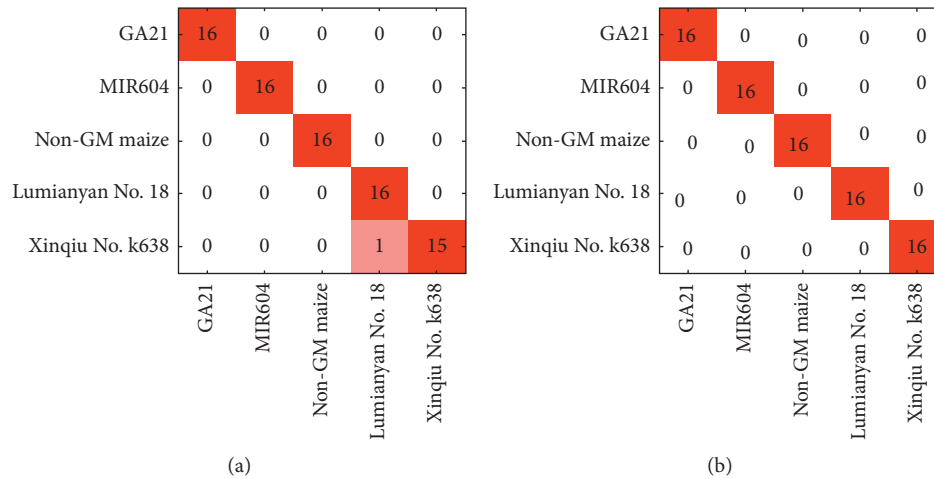


FIGURE 5: Confusion matrix. (a) *K*-means. (b) PCA-mean shift.

located on the upper half-plane, and all maize samples are located on the lower half-plane. It is easy to identify cotton and maize by their location. In the lower half-plane, GA21 is distributed in the left, non-GM maize is distributed in the middle, and MIR604 is distributed in the right. It is consistent with the absorption strength that GA21 has weak absorbance and MIR604 has strong absorbance. These three types of maize can be classified successfully. In Figure 4(a), these two types of cotton specimens are partly overlapped. It is due to that the absorbance of these two cotton specimens is similar. Thus, PCA is unable to identify Lumianyan No. 18 and Xinqiu No. k638 correctly.

After PCA, we used the first three PCs as the input of *K*-means and mean shift. In Figures 4(b) and 4(c), cotton and maize samples are divided into upper and lower half-planes. Thus, both methods are able to distinguish between cotton and maize. Because GA21, MIR604, and non-GM maize located on the lower half-plane have a large gap between each other, both *K*-means and PCA-mean shift can differentiate between the three types of maize. Unlike the sample points' distribution of the three types of maize, two types of cotton sample points overlap each other. The performance of classifying two types of cotton between *K*-means and PCA-mean shift is different. We employed the confusion matrix to evaluate the performance of these two methods, as shown in Figure 5. Figure 5(a) displays that one Xinqiu No. k638 sample is recognized as Lumianyan No. 18 by using *K*-means. Figure 5(b) shows that all the samples can be identified correctly by using PCA-mean shift. From the confusion matrix, the average accuracy of *K*-means and PCA-mean shift is 98.75% and 100%.

4. Conclusions

In this paper, an unsupervised learning method, PCA-mean shift, was proposed to identify two types of cotton and three types of maize with absorbance spectra in THz frequency. PCA was used to reduce the dimensionality of THz absorbance spectra. To compare our proposed method with *K*-means,

confusion matrix and average accuracy were employed. The results indicated that PCA-mean shift had a higher average accuracy than *K*-means. Thus, PCA-mean shift combined with THz-TDS is a potential identification tool for GM crops' identification.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (Grant no. 62041111), Guangxi Natural Science Foundation (Grant no. 2019GXNSFBA245076), Guangxi Key Laboratory of Automatic Detecting Technology and Instruments Foundation (Grant nos. YQ19208 and YQ20207), Opening Foundation of Yulin Research Institute of Big Data (Grant no. 2020YJKY04), Major Cooperative Project between Yulin Municipal Government and Yulin Normal University (Grant no. YLSXZD2019015), Doctoral Scientific Research Foundation of Yulin Normal University (Grant no. G2019K02), and Yulin Normal University Research Grant (Grant no. 2015YJYB06).

References

- [1] A. Repellin, M. Baga, P. Jauhar, and R. Chibbar, "Genetic enrichment of cereal crops via alien gene transfer: new challenges," *Plant Cell Tissue and Organ Culture*, vol. 64, no. 2-3, pp. 159-183, 2001.
- [2] X. K. Morin, "Genetically modified food from crops: progress, pawns, and possibilities," *Analytical and Bioanalytical Chemistry*, vol. 392, no. 3, pp. 333-340, 2008.
- [3] H. Azadi, M. Ghanian, O. M. Ghoochani et al., "Genetically modified crops: towards agricultural growth, agricultural

- development, or agricultural sustainability?" *Food Reviews International*, vol. 31, no. 3, pp. 195–221, 2015.
- [4] A. Bakshi, "Potential adverse health effects of genetically modified crops," *Journal of Toxicology and Environmental Health, Part B*, vol. 6, no. 3, pp. 211–225, 2003.
- [5] G. Peterson, S. Cunningham, L. Deutsch et al., "The risks and benefits of genetically modified crops: a multidisciplinary perspective," *Conservation Ecology*, vol. 4, no. 1, pp. 1–4, 2000.
- [6] M. Kramkowska, T. Grzelak, and K. Czyzewska, "Benefits and risks associated with genetically modified food products," *Annals of Agricultural and Environmental Medicine*, vol. 20, no. 3, pp. 413–419, 2013.
- [7] H. Cheng, W. Jin, H. Wu et al., "Isolation and PCR detection of foreign DNA sequences in bee honey raised on genetically modified Bt (Cry1Ac) cotton," *Food and Bioproducts Processing*, vol. 85, no. C2, pp. 141–145, 2007.
- [8] M. S. McCabe, J. B. Power, A. M. M. de Laat, and M. R. Davey, "Detection of single-copy genes in DNA from transgenic plants by nonradioactive southern blot analysis," *Molecular Biotechnology*, vol. 7, no. 1, pp. 79–84, 1997.
- [9] X. Xiao, H. Wu, X. Zhou et al., "The combination of quantitative PCR and western blot detecting CP4-EPSPS component in roundup ready soy plant tissues and commercial soy-related foodstuffs," *Journal of Food Science*, vol. 77, no. 6, pp. 603–608, 2012.
- [10] G. Liu, W. Su, Q. Xu, M. Long, J. Zhou, and S. Song, "Liquid-phase hybridization based PCR-ELISA for detection of genetically modified organisms in food," *Food Control*, vol. 15, no. 4, pp. 303–306, 2004.
- [11] M. D. King, P. M. Hakey, and T. M. Kortner, "Discrimination of chiral solids: a terahertz spectroscopic investigation of l- and d- serine," *The Journal of Physical Chemistry A*, vol. 114, no. 8, pp. 2945–2953, 2010.
- [12] Y. Ueno, K. Ajito, N. Kukutsu, and E. Tamechika, "Quantitative analysis of amino acids in dietary supplements using terahertz time-domain spectroscopy," *Analytical Sciences*, vol. 27, no. 4, p. 351, 2011.
- [13] I. Maeng, S. H. Baek, H. Y. Kim, G.-S. Ok, S.-W. Choi, and H. S. Chun, "Feasibility of using terahertz spectroscopy to detect seven different pesticides in wheat flour," *Journal of Food Protection*, vol. 77, no. 12, pp. 2081–2087, 2014.
- [14] J. El Haddad, B. Bousquet, L. Canioni, and P. Mounaix, "Review in terahertz spectral analysis," *TrAC Trends in Analytical Chemistry*, vol. 44, pp. 98–105, 2013.
- [15] B. Qin, Z. Li, Z. Luo, H. Zhang, and Y. Li, "Feasibility of terahertz time-domain spectroscopy to detect carbendazim mixtures wrapped in paper," *Journal of Spectroscopy*, vol. 2017, no. 7, 8 pages, Article ID 6302868, 2017.
- [16] B. Qin, Z. Li, F. Hu et al., "Highly sensitive detection of carbendazim by using terahertz time-domain spectroscopy combined with metamaterial," *IEEE Transactions on Terahertz Science and Technology*, vol. 8, no. 2, pp. 149–154, 2018.
- [17] B. Cao, H. Li, E. Cai, and M. Fan, "Determination of pesticides in flour by terahertz time-domain spectroscopy (thz-tds) with voigt function fitting and partial least squares (pls) analysis," *Analytical Letters*, vol. 54, pp. 1–12, 2020.
- [18] B. Cao, H. Li, M. Fan, W. Wang, and M. Wang, "Determination of pesticides in a flour substrate by chemometric methods using terahertz spectroscopy," *Analytical Methods*, vol. 10, no. 42, pp. 5097–5104, 2018.
- [19] W. Liu, P. Zhao, C. Wu, C. Liu, J. Yang, and L. Zheng, "Rapid determination of aflatoxin B1 concentration in soybean oil using terahertz spectroscopy with chemometric methods," *Food Chemistry*, vol. 293, pp. 213–219, 2019.
- [20] J. Liu and Z. Li, "The terahertz spectrum detection of transgenic food," *Optik*, vol. 125, no. 23, pp. 6867–6869, 2014.
- [21] W. Xu, L. Xie, Z. Ye et al., "Discrimination of transgenic rice containing the Cry1Ab protein using terahertz spectroscopy and chemometrics," *Scientific Reports*, vol. 5, p. 1115, 2015.
- [22] T. Chen, Z. Li, X. Yin, F. Hu, and C. Hu, "Discrimination of genetically modified sugar beets based on terahertz spectroscopy," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 153, pp. 586–590, 2016.
- [23] W. Liu, C. Liu, X. Hu, J. Yang, and L. Zheng, "Application of terahertz spectroscopy imaging for discrimination of transgenic rice seeds with chemometrics," *Food Chemistry*, vol. 210, pp. 415–421, 2016.
- [24] B. Qin, Z. Li, T. Chen, and Y. Chen, "Identification of genetically modified cotton seeds by terahertz spectroscopy with mpga-svm," *Optik*, vol. 142, pp. 576–582, 2017.
- [25] J. Michel, D. Youssefi, and M. Grizonnet, "Stable mean-shift algorithm and its application to the segmentation of arbitrarily large remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 2, pp. 952–964, 2014.
- [26] J. Kim, S. Lee, G. Lee, Y. Park, and Y. Hong, "Using a method based on a modified k-means clustering and mean shift segmentation to reduce file sizes and detect brain tumors from magnetic resonance (mri) images," *Wireless Personal Communications*, vol. 89, no. 3, pp. 993–1008, 2016.
- [27] Q. Wei, T. Dai, T. Ma, Y. Liu, and Y. Gu, "Crystal identification in dual-layer-offset doi-pet detectors using stratified peak tracking based on svd and mean-shift algorithm," *IEEE Transactions on Nuclear Science*, vol. 63, no. 5, pp. 2502–2508, 2016.
- [28] H. Cho, S. J. Kang, and Y. H. Kim, "Image segmentation using linked mean-shift vectors and global/local attributes," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 10, p. 1, 2017.
- [29] L. Xing, J. Zhang, H. Liang, and Z. Li, "Intelligent recognition of dominant colors for Chinese traditional costumes based on a mean shift clustering method," *The Journal of the Textile Institute*, vol. 109, no. 10, pp. 1304–1314, 2018.
- [30] L. Wang, D. Tang, Y. Guo, and M. N. Do, "Common visual pattern discovery via nonlinear mean shift clustering," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5442–5454, 2015.
- [31] L. Ai and J. Xiong, "Temporal-spatial mean-shift clustering analysis to improve functional MRI activation detection," *Magnetic Resonance Imaging*, vol. 34, no. 9, pp. 1283–1291, 2016.
- [32] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 790–799, 1995.
- [33] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.
- [34] T. Chen, Z. Li, and W. Mo, "Identification of biomolecules by terahertz spectroscopy and fuzzy pattern recognition," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 106, pp. 48–53, 2013.