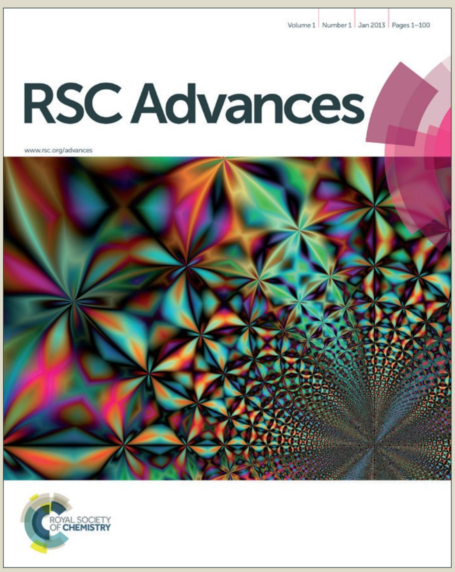
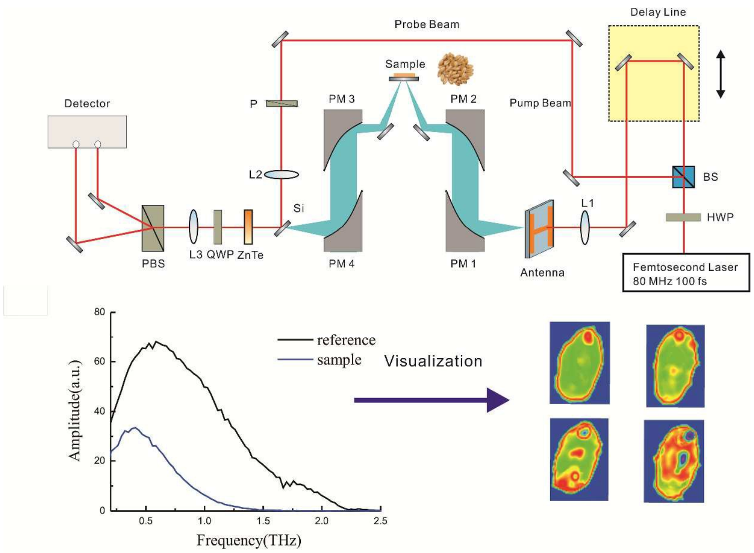
RSC Advances



Graphical abstract



Discrimination of Moldy Wheat Using Terahertz Imaging Combined with Multivariate Classification

Yuying Jiang,[a,c] Hongyi Ge,[b*] Feiyu Lian,[b] Yuan Zhang,[b] and Shanhong Xia[a]

[a]State Key Laboratory of Transducer Technology, Institute of Electronics, Chinese Academy of Sciences, Beijing 100080, China

[b]Key Laboratory of Grain Information Processing& Control, Ministry of Education, Henan University of Technology, Zhengzhou 450001, China

[c]University of Chinese Academy of Sciences, Beijing 100080, China

** Email：gehongyi@haut.edu.cn, phone: 0086-371-67756610*

Terahertz (THz) imaging was employed to develop a novel method for discriminating wheat of varying states of moldiness. Spectral data, in the range of 0.2–1.6 THz, were extracted from regions of interest (ROIs) in the THz images. Principal component analysis (PCA) was used to evaluate the spectral data and determine the cluster trend. Six optimal frequencies were selected by implementing PCA directly for each image's ROI. Classification models for moldy wheat identification were established using the support vector machine (SVM) method, a partial least-squares regression analysis, and the back propagation neural network method. The models developed from these methods were based on the full and optimal frequencies, using the top three principal components as input variables. The PCA-SVM method achieved a prediction accuracy of over 95%, and was implemented at every pixel in the images to visually demonstrate the moldy wheat classification method. Our results indicate that THz imaging combined with chemometric algorithms is efficient and practical for the discrimination of moldy wheat.

Keywords: discriminate analysis, wheat grain, moldy, terahertz imaging, spectral analysis, multivariate classification

1. Introduction

Wheat is a primary food crop worldwide, and contains high amounts of carbohydrates, proteins, fat, and vitamins (Oladunmoye, Akinoso, & Olapade, 2010). Mildew such as Aflatoxinand and Aspergillusniger are prevalent throughout all stages of wheat growth and production. When improperly stored and processed, these mildews pose a potential threat to humans and fowls (Neethirajan, Karunakaran, Jayas, & White, 2007). Recently, food quality and safety assessment have increased within the food industry. Conventional moldy grain detection methods, such as naked-eye observations, microscope inspection, liquid chromatography, and enzyme-linked immunosorbent assays, are time-consuming and labor-intensive (Turner, Subrahmanyam, & Piletsky, 2009).

To satisfy the demand for high-quality consumer products, extensive studies into grain quality via nondestructive rapid evaluations have been performed. Wang et al. (Wang, Liu, Yu, Wu & He, 2011) presented a new approach for non-invasive classification of raisins by using computer vision techniques. Eifler et al. (Eifler, Martinelli, Santonico, Capuano, Schild, & Di Natale, 2011) used an electronic nose to differentiate between infected and non-infected wheat grains. Arngren et al. (Arngren, Hansen, Eriksen, Larsen, &Larsen, 2011) used near-infrared hyperspectral imaging combined with nonlinear neural networks to identify early-stage pregermination in barley grains. ElMasry et al. (ElMasry, Wang, ElSayed, Ngadi, 2007) proposed a novel tool for nondestructive determination of moisture content, total soluble solids, and acidity in strawberry using NIR spectroscopy. However, these measurement techniques do not probe the far-infrared spectral region, which contains a wealth of physical and chemical information.

Terahertz (THz) radiation (with frequencies from 0.3 to 10 THz and wavelengths from 3.3 to 333 cm^{-1}) occupies the region between the microwave and infrared bands; it can be used for non-destructive and non-invasive analyses, and possesses attractive features such as extremely low-energy levels, broad spectral bandwidth, transparency, and good penetration through various materials (Ferguson, & Zhang, 2002; Tonouchi, 2007). THz spectroscopy and imaging are rapidly becoming novel techniques in the field of optics research. The new techniques are widely used as solutions in art conservation (Fukunaga, & Hosako, 2010), security problems (Melinger, Laman, & Grischkowsky, 2008), biomedical applications (Oh. et. al., 2014; Siegel, 2004), agricultural quality control (Gowen, O'Sullivan, & O'Donnell, 2012; Ge, Jiang, Xu, Lian, Zhang, & Xia, 2014), and other fields (Guillet. et. al., 2014). THz imaging is performed both by the transmission and reflection of THz waves. In reflectance imaging, THz waves reflect not only from the surface of samples, but also from interfaces present in the samples within the penetration depth of the radiation (Safrai, Ben Ishai, Polsman, Einav, & Feldman, 2014). Thus, both surface and depth information can be obtained from the timing and amplitude of the reflected waves. Time- and frequency-domain structural images can be acquired from detected THz waves associated with various parameters at each pixel in the measured sample area (Reid, Pickwell-MacPherson, Laufer, Gibson, Hebden, & Wallace, 2010). Owing to the absorption, reflection, scattering, and phase-shifting of the imaged material, measured parameters can change due to differing wave delay and attenuation.

The aim of this study was to evaluate the validity and feasibility of identifying different moldy states of wheat using THz imaging and multivariate data analysis methods. THz spectra of wheat grains with different moldy statuses were extracted in the range of 0.2–1.6 THz from regions of interest (ROIs) in each THz image. Principal component analysis (PCA) was used to explore features of the spectral data and select the optimal frequencies. Support vector machine (SVM), partial least-squares regression (PLSR), and back propagation neural network (BPNN) models were established based on the full frequencies and optimal frequencies for discriminating between the four stages of moldy wheat. Finally, THz images of wheat with different moldy states were investigated using the optimal classification method (i.e., PCA-SVM).

2. Materials and methods

2.1 Experimental setup

A standard THz-TDS laboratory setup, using reflection geometry as developed by Zomega Terahertz Corporation in USA, was used in our experiment. A schematic of the THz-TDS reflection imaging system is shown in Fig. 1. The THz imaging system employed an externally pulsed femtosecond laser Ti-sapphire with a pulse width, central wavelength, and repetition frequency of 100 fs, 800 nm, and 80 MHz, respectively. The beam produced by the laser was split into a pump and a probe using a polarizing beam splitter. The pump beam was irradiated on a photoconductive dipole antenna fabricated on a LT-GaAs wafer for generation of the THz waves, and the probe beam was focused onto an electro-optic ZnTe crystal for detection of the THz waves (Taylor. et. al., 2008). The THz pulses emitted by the generator were focused on the sample via two metal parabolic mirrors, and the THz pulses reflected by the sample via two additional parabolic mirrors were guided to the detection antenna. The system measures far-infrared spectra between 0.1 THz and 3.0 THz. The sample was scanned by moving the two-dimensional motorized stage, and the obtained image data were saved and analyzed using a computer. Details about the principles of the system are explained elsewhere (Kim. et. al., 2012). The experiment was performed at room temperature, and the humidity was maintained at approximately zero by purging the system with dry nitrogen to avoid absorption of vapor.

2.2 Sample preparation

Wheat used in the experiment was collected from the School of Food Science and Technology, Henan University of Technology, Zhengzhou, China. The wheat was of the same variety and produced in 2013. Wheat grains were moistened at a humidity of 28% and were evenly distributed in a circular Petri dish. The Petri dish was put into an incubator box that was maintained at a constant temperature of 25°C, where it remained for eight days. Wheat with different stages of mold growth (none, slight, moderate, and serious) where then selected (as shown in Fig.2) and individually imaged by the THz imaging system with a spatial resolution of 0.25 mm. For each degree of mold contamination, 50 samples were used without further processing.

2.3 Multivariate Analysis Methods

2.3.1 Principal component analysis

PCA (Lin, Zhao, Sun, Chen, & Zhou, 2011; Noori, Sabahi, Karbassi, Baghvand, & Zadeh,

2010) is a multivariate statistical and dimensional reduction method that can be used to reduce the complexity of input variables when dealing with large datasets. In this method, a large volume of data is transformed into a small number of principal components (PCs). PCs can be expressed as:

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \cdots + a_{in}X_n \quad (1)$$

where Z_i represents the PCs, a_i represents the related eigenvectors, and X_i represents the input variables. This information can be acquired by solving following equation.

$$\left|R - I\lambda\right| = 0 \quad (2)$$

where R is the variance-covariance matrix, I is the unit matrix, and λ is the eigenvector.

2.3.2 Support vector machines

SVM is a widely used, supervised statistical learning method for analyzing data and recognizing patterns (He, Yang, & Xie, 2013; He, Wu, & Sun, 2014). SVM demonstrates advantage over other methods when dealing with small samples, and high-dimensional and non-linear data. In the multi-class SVM method, $k(k-1)/2$ classifiers are constructed, where k is the class number of the data. The following two-class classification problem was implemented by training the ith and jth data classes:

$$\min_{w^{ij}, b^{ij}, \xi^{ij}} \frac{1}{2}(w^{ij})^T w^{ij} + c(\sum_t (\xi^{ij})t) \quad (3)$$

subject to

$$\begin{aligned} &(w^{ij})^T \phi(x_t) + b^{ij} \geq 1 - \xi^{ij}, \text{ if } x_t \text{ in } ith \text{ class} \\ &(w^{ij})^T \phi(x_t) + b^{ij} \leq -1 + \xi^{ij}, \text{ if } x_t \text{ in } jth \text{ class} \\ &\xi^{ij} \geq 0 \end{aligned} \quad (4)$$

where w and b define the optimal hyperplane, ξ represents the slack variable, c is the penalty factor, and $\phi(x)$ is the sample set. Selection of the kernel function in SVM models significantly affects model performance. In this paper, the commonly used radial bias function (RBF) $k(x_i, y_i) = \exp(-\frac{\|x_i - y_i\|^2}{\gamma^2})$ was used. The adjustable kernel function parameter C controls the trade-off between the minimum model complexity and minimum training error, while γ represents the degree of generalization and the width of the kernel function. A grid-search procedure was employed to find the optimal parameters of the model (Maali, & Al-Jumaily, 2013).

The root mean square error (RMSE) was used to evaluate the performance of the established model (Zhang. et. al., 2008). The RMSE is calculated as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i^{pre} - y_i)^2}{N}} \tag{5}$$

where y_i represents actual value of the *i*th sample in the data set, y_i^{pre} is the predicted weight ratio value of the *i*th sample in the developed model, and N is the sample size.

2.3.3 Partial least squares regression

PLSR is one of most robust and reliable multivariate-data analysis methods, and is particularly suitable for use in situations where there is a linear relation between the spectra and properties of the considered objects (Brereton, 2000). A PLSR analysis was performed to establish a regression model for the prediction of target chemical concentrations (variable matrix Y) based on the corresponding spectra data (variable matrix X). The underlying PLSR model is expressed as:

$$X = TP^T + E$$
$$Y = UQ^T + F \quad (6)$$

where T and U are the feature matrices of the variable matrix of X and Y respectively, P and Q represent the orthogonal loading matrices, and E and F are the error terms.

2.3.4 Back propagation neural network

BPNN is a type of nonlinear multi-layer network, and it has been used extensively to solve a variety of classification and regression problems (Dubey, Bhagwat, Shouche, & Sainis, 2006). A BPNN is based on an algorithm that rectifies the weights within each layer in proportion to the error obtained from the previous layer. In this study, an input layer, a hidden layer, and an output layer were used. By optimizing the hidden nodes from the input variables by "trial and error," BPNN was used to classify samples into predefined varieties, and a new output layer that provided a more precise discrimination of a sample's variety was obtained. Details of the BPNN method are discussed extensively elsewhere (Marengo, Bobba, Robotti, & Lenti, 2004). The whole experiment procedure by using THz imaging technique, as illustrated in Fig. 3, is made from three steps to prepare the data structure for mold statuses wheat identification.

3 Results and Discussion

3.1 Spectral Analysis

3.1.1 Moldy wheat spectra

After THz images of wheat with different stages of mold growth were acquired, the only wheat grain areas are segmented as the ROIs to exclude the interfering information origin from the background in each image. The spectra of each pixel within the ROI were extracted and averaged at each frequency to generate a mean value, which is then expressed as the ROI spectrum. The average frequency domain spectra of each degree of mold growth, in the range of 0.1–2.0 THz, are shown in Fig. 4. It is seen that an intense trough is present at around 1.67 THz, which is related to the absorption of water within the grain. And the spectral curves of these four mold statuses wheat are quite similar at the beginning. Hence, spectral frequency range from 0.2-1.6 THz is employed for further identification study. Meanwhile, the general trends of the four spectral curves show no obvious differences, which indicated that mold statuses of the wheat could not be identified from spectral curves directly.

To solve this problem, more sophisticated computational analysis methods were employed to differentiate between the mold statuses of the wheat. Therefore, a dataset with 512 spectral features and 200 wheat samples was selected in order to construct a classification model to discriminate between the different degrees of moldiness. A dataset consisting of 200 samples was randomly split into a calibration set (120 samples) and a prediction set (80 samples). The classification errors would clearly decrease when training more samples. Hence each wheat sample leaves fewer samples to analyze and obtains higher prediction accuracy. But when more training number, redundant information (existed in the large number of input variable) would affect the robust and ability of the classification models. Meanwhile, the less input simplify the classification models and accelerate the calculated speed.

3.1.2 PCA Analysis

PCA was performed on all of the spectral data (with a frequency range of 0.2–1.6 THz) obtained from the normal, slightly moldy, moderately moldy, and seriously moldy wheat samples to reduce the high dimensionality of the problem and qualitatively identify the samples. The explained variance rate for the top four PCs extracted from the original THz spectra data are 93.22%, 3.61%, 1.24%, and 0.21%, respectively. The top four PCs explain 98.25% of the total contribution to the original data. It is shown that the cumulative reliabilities of the top four PCs represent 98% of the total information to the original data. Thus, they contain the maximum information across all the wheat samples and reduce the dimensions from 512 spectral measurements for classification of different mold statuses of wheat to only three components. Figure 5 shows the three-dimensional scores plotted for the first three PCs for all of the samples. As we can see, the different mold statuses are distributed separately in the three-dimensional area. However, some sample points near the boundaries of normal and slightly moldy wheat are mixed although their sample points are clustered. Therefore, it is necessary to employ an adequate classification model based on the PCA process for further discrimination.

3.1.3Optimal Frequency Selection

A PCA was used for each ROI image to select the optimal frequencies. PC loadings were employed to identify sensitive frequencies that were highly correlated with each PC. The x-loading weights of the first four PCs were used to select each frequency in the full spectral range. Strong peaks and troughs for the top four PCs were selected as the optimum frequencies. As seen in Fig. 6, six frequencies with the values of 0.32 THz, 0.59 THz, 0.87 THz, 1.0 THz, 1.29 THz, and 1.58 THz were selected as discriminators of different moldy statuses. The reduced number of frequencies decreased the time to acquire and process each image.

3.2 Multivariate Data Analysis

3.2.1 Multivariate Data Analysis Based on Full spectra

SVM, PLSR, and BPNN classification models were used to predict the degree of moldiness using the entire spectral dataset. Within the SVM models, the optimization values for the regularization parameter γ and the RBF kernel function parameter C were selected when the smallest RMSE was obtained. The optimal parameters γ and C were set at 3.6 and 1.8, respectively, which were determined by using the grid search algorithm. For the BPNN model, after several attempts to optimize the parameters, the learning rate factor, momentum factor, initial weight, permitted training error, and maximal training times were set at 0.1, 0.1, 0.6, 0.00001, and 1,000, respectively.

The SVM, PLSR, and BPNN models were constructed using the top four PCs as inputs. The discrimination results of normal, slightly moldy, moderately moldy, and seriously moldy wheat in the calibration set and prediction set using these models are presented in Table 1.

Table 1 Results of the classification models based on full spectra (Cal. represents the calibration set of the samples and Pre. represents the prediction set of the samples.)

Model	Accuracy per type (%)				Overall prediction
	Normal	Slightly moldy	Moderately	Seriously	

					moldy		moldy		accuracy
	Cal.	Pre.	Cal.	Pre.	Cal.	Pre.	Cal.	Pre.	(%)
PCA-SVM	100%	100%	100%	86.67%	100%	84%	100%	100%	96.5%
PCA-PLSR	100%	95%	91.43%	86.67%	88%	84%	100%	95%	93%
PCA-BPNN	93.33%	90%	88.57%	80%	84%	76%	93.33%	90%	87%

As the table shows, the performance of the SVM model was, in general, better than those of the PLSR and BPNN models, and achieved a prediction accuracy of 96.5%. The SVM model achieved a classification rate of the normal and serious moldy statuses of 100% in both the calibration and prediction sets; however, the classification rates of the prediction sets of slightly moldy and seriously moldy wheat were relatively lower. Moreover, the PLSR and BPNN models misclassified some statuses, with an overall prediction accuracy of 93% and 87%, respectively. The results indicate that PLSR and SVM models can be used as effective methods for moldy wheat identification, with the SVM model considered the optimum method.

3.2.2 Multivariate Data Analysis Based on Optimal frequencies

Although the classification models have good moldy wheat prediction performances, the large number of frequency variables resulted in complicated and time-consuming data processing. Instead, the use of optimal-frequency selection can reduce the complexity and time required for model establishment. As a consequence of optimal frequency selection, the top four PCs and the selected six frequencies (0.32 THz, 0.59 THz, 0.87 THz, 1.0 THz, 1.29 THz, and 1.58 THz) were used as inputs to the SVM, PLSR, and BPNN models. The performance of the optimized models based only on the optimal frequencies is presented in Table 2.

Table 2 Results of the classification models based on their optimal spectra (Cal. represents the calibration set of the samples and Pre. represents the prediction set of the samples.)

Model	Accuracy per type (%)								Overall prediction accuracy (%)
	Normal		Slightly moldy		Moderately moldy		Seriously moldy		
	Cal.	Pre.	Cal.	Pre.	Cal.	Pre.	Cal.	Pre.	
PCA-SVM	100%	100%	97.14%	86.67%	92%	84%	100%	95%	95%
PCA-PLSR	100%	95%	91.43%	80%	92%	84%	96.67%	95%	92.5%
PCA-BPNN	93.33%	85%	88.57%	73.33%	84%	76%	93.33%	90%	86%

As shown in Table 2, the BPNN model had the worst prediction result, with a classification accuracy of 86%. The classification rates of the SVM and PLSR models in both the calibration and the prediction sets were all over 80%. The SVM model obtained the highest overall prediction accuracy, 95%, and a classification accuracy of 100% for normal and seriously moldy wheat in the calibration set. The slightly moldy and moderately moldy wheat showed poorer prediction accuracy in all models, compare with the normal wheat and seriously moldy wheat.

The plots of the actual values compared to the predicted values using the PCA-SVM models based on the full spectra and selected optimal frequencies are shown Fig. 7. A threshold value (dummy variable ±0.5) was set to define the class limits. Subintervals from 0.5–1.5, 1.5–2.5, 2.5–3.5, and 3.5–4.5 represent normal, slightly moldy, moderately moldy, and seriously moldy wheat samples, respectively. It can be seen in Figs. 6 (a) and (b) that a similar distribution of points between the full spectrum and the optimal frequencies was obtained. The experimental results demonstrate the feasibility of using selected optimal frequencies for the discrimination of wheat grains with different mold statuses.

3.3 THz Images of Moldy Wheat

The implementation of a visualization process is helpful for determining the degree of moldiness of a wheat grain, which can be difficult when observed by just the naked eye. In this study, the PCA-SVM model acquired the best classification accuracy and therefore was used to generate THz moldy wheat images. Training of the SVM model was done using the optimal frequencies selected by the PCA. The reduced spectral data were then used as input to the SVM model. The output value of the model was the reflectivity of each pixel, which corresponds to a different component within each wheat grain. When the values of all pixels within the wheat grain were calculated, an image was generated based on the spatial positions of each pixel.

Figure 8 shows the THz images of normal, slightly moldy, moderately moldy, and seriously moldy wheat. Regions (1), (2), and (3) represent the embryo of each wheat grain. Except for the embryo structure, the inner structures of the wheat sample in Fig. 7(a) and 7(b) are evenly distributed. However, in Fig. 7(b) the embryo and edge structure have changed, indicating that the wheat is in its moldy infancy, while it is seen that the wheat in Fig. 7(a) is not contaminated with mold. In Fig. 7(c), the embryo area and small range of inner structures are damaged, indicating that the sample has a moderate degree of mold growth. Finally, in 7(d), the red area

(5) indicates that the inner structures of this wheat sample are totally damaged, and the embryonic area is absent.

3.4 Discussion

The excellent discrimination results demonstrate that the THz reflection imaging technique combined with PCA feature extraction and a SVM classification model can be used to identify wheat grains with different mold statuses. Six optical frequencies (0.32 THz, 0.59 THz, 0.87 THz, 1.0 THz, 1.29 THz, and 1.58 THz) were selected according to the top four PC loading weights. The overall prediction accuracy of the PCA-SVM model based on the selected optimal frequencies was 95%, which is higher than that achieved with the PCA-PLS and PCA-BPNN models. The optimal frequency-based models used six frequencies instead of 159 frequencies, indicating a decrease of 96.49%. The performance of each classification model showed only a slight decline from full spectra to optimal frequencies, implying that the optimal frequencies were effective, and as such, we encourage further study of them. Furthermore, the fewer input variables accelerated the data calculation speed and simplified the model complexity. In further studies, different frequency selection methods and different classification models will be applied to improve the prediction accuracy and explore the optimal frequency for moldy wheat identification.

Additionally, the PCA-SVM model was used to classify the THz image data and determine the degree of mold contamination as normal, slightly moldy, moderately moldy, and seriously moldy. The THz images provided information regarding the spatial distribution of different components within the wheat grain, and were helpful for detecting changes in a grain's inner structure due to varying mold status. Our results show that THz imaging can be used to recognize the wheat when it is in its early moldy stage, which cannot be done with conventional imaging and spectroscopy, and thus provides an early warning technique for mold contamination. The THz imaging technique has the potential to be an effective tool for agriculture quality and safety control. Therefore, it is essential to expand the sample variety number and optimize the image classification algorithm in further studies to assist in discriminating the multiple statuses of wheat mold en masse and for practical applications.

4 Conclusion

THz imaging combined with multivariate data analyses was employed to discriminate wheat grains with different mold statuses. Spectral information was extracted from the THz images, in the range of 0.2–1.6 THz, for each wheat sample. The feature data of each spectrum were analyzed and six optimal frequencies were selected using PCA. In addition, the SVM, PLSR, and BPNN models were constructed based on the full spectra and optimal frequencies to help discriminate between different moldy wheat samples. The prediction accuracies of the full spectra were similar to those obtained using only the optimal frequencies. The PCA-SVM model was considered to be the optimal model, and the prediction accuracies reached 95%. The PCA-SVM model was also used on THz images as a visual demonstration of the classification technique. Our experimental results demonstrate that THz imaging is a potential tool for the classification of moldy wheat.

ACKNOWLEDGMENTS

This work is supported by the National High-tech R&D Program of China (863 Program) (Grant No. 2012AA101608), the National Natural Science Foundation of China (Grant No. 61071197), the Key Project of Educational Committee of Henan Province of China (Grant No. 14B413011), the Key Science and Technology Program of Henan Province of China (Grant No. 122102210217), the Plan of Nature Science Fundamental Research in Henan University of Technology (11JCY07), and the High-Level Personnel Funds (2012BS047) in Henan University of Technology. Finally, the authors are grateful to the reviewers for their helpful comments and constructive suggestions.

REFERENCES

Arngren, M., Hansen, P. W., Eriksen, B., Larsen, J., & Larsen, R. (2011). Analysis of Pregerminated Barley Using Hyperspectral Image Analysis. *Journal of Agricultural and Food Chemistry*, 59(21), 11385-11394.

Brereton, R. G. (2000). Introduction to multivariate calibration in analytical chemistry. *Analyst*, 125(11), 2125-2154.

Dubey, B. P., Bhagwat, S. G., Shouche, S. P., & Sainis, J. K. (2006). Potential of artificial neural networks in varietal identification using morphometry of wheat grains. *Biosystems Engineering,* 95(1), 61-67.

Eifler, J., Martinelli, E., Santonico, M., Capuano, R., Schild, D., & Di Natale, C. (2011). Differential Detection of Potentially Hazardous Fusarium Species in Wheat Grains by an Electronic Nose. *Plos One*, 6(6).

EIMasry, G., Wang, N., EISayed, A., Ngadi, M., 2007. Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry. Journal of Food Engineering, 81 (1), 98–107.

Ferguson, B., & Zhang, X. C. (2002). Materials for terahertz science and technology. *Nature Materials,* 1(1), 26-33.

Fukunaga, K., & Hosako, I. (2010). Innovative non-invasive analysis techniques for cultural heritage using terahertz technology. *Comptes Rendus Physique*, 11(7-8), 519-526.

Ge, H. Y., Jiang, Y. Y., Xu, Z. H., Lian, F. Y., Zhang, Y., & Xia, S. H. (2014). Identification of wheat quality using THz spectrum. *Optics Express*, 22(10), 12533-12544.

Gowen, A. A., O'Sullivan, C., & O'Donnell, C. P. (2012). Terahertz time domain spectroscopy and imaging: Emerging techniques for food process monitoring and quality control. *Trends in Food Science & Technology,* 25(1), 40-46.

Guillet, J. P., Recur, B., Frederique, L., Bousquet, B., Canioni, L., Manek-Honninger, I., Desbarats, P., & Mounaix, P. (2014). Review of Terahertz Tomography Techniques. *Journal of Infrared Millimeter and Terahertz Waves*, 35(4), 382-411.

He, M., Yang, G. L., & Xie, H. Y. (2013). A hybrid method to recognize 3D object. *Optics Express*, 21(5), 6346-6352.

He, H. J., Wu, D., & Sun, D. W. (2014). Potential of hyperspectral imaging combined with chem ometric analysis for assessing and visualising tenderness distribution in raw farmed salmon fillet s. Journal of Food Engineering, 126, 156-164.

Kim, K. W., Kim, K. S., Kim, H., Lee, S. H., Park, J. H., Han, J. H., Seok, S. H., Park, J., Choi, Y., Kim, Y. I., Han, J. K., & Son, J. H. (2012). Terahertz dynamic imaging of skin drug absorption. *Optics Express,* 20(9), 9476-9484.

Lin, H., Zhao, J. W., Sun, L., Chen, Q. S., & Zhou, F. (2011). Freshness measurement of eggs using near infrared (NIR) spectroscopy and multivariate data analysis. *Innovative Food Science & Emerging Technologies*, 12(2), 182-186.

Maali, Y., & Al-Jumaily, A. (2013). Self-advising support vector machine. *Knowledge-Based Systems*, 52, 214-222.

Marengo, E., Bobba, M., Robotti, E., & Lenti, M. (2004). Hydroxyl and acid number prediction in polyester resins by near infrared spectroscopy and artificial neural networks. *Analytica Chimica Acta*, 511(2), 313-322.

Melinger, J. S., Laman, N., & Grischkowsky, D. (2008). The underlying terahertz vibrational spectrum of explosives solids. *Applied Physics Letters,* 93(1).

M.Tonouchi. (2007). Cutting-edge terahertz technology. *Nature Photonics*, 1(2), 97-105.

Neethirajan, S., Karunakaran, C., Jayas, D. S., & White, N. D. G. (2007). Detection techniques for stored-product insects in grain. *Food Control*, 18(2), 157-162.

Noori, R., Sabahi, M. S., Karbassi, A. R., Baghvand, A., & Zadeh, H. T. (2010). Multivariate statistical analysis of surface water quality based on correlations and variations in the data set. *Desalination*, 260(1-3), 129-136.

Oh, S. J., Kim, S. H., Ji, Y. B., Jeong, K., Park, Y., Yang, J., Park, D. W., Noh, S. K., Kang, S. G., Huh, Y. M., Son, J. H., & Suh, J. S. (2014). Study of freshly excised brain tissues using terahertz imaging. Biomedical *Optics Express*, 5(8), 2837-2842.

Oladunmoye, O. O., Akinoso, R., & Olapade, A. A. (2010). Evaluation of Some Physical-Chemical Properties of Wheat, Cassava, Maize and Cowpea Flours for Bread Making. *Journal of Food Quality*, 33(6), 693-708.

Reid, C. B., Pickwell-MacPherson, E., Laufer, J. G., Gibson, A. P., Hebden, J. C., & Wallace, V. P. (2010). Accuracy and resolution of THz reflection spectroscopy for medical imaging. *Physics in Medicine and Biology*, 55(16), 4825-4838.

Siegel, P. H. (2004). Terahertz technology in biology and medicine. *Ieee Transactions on Microwave Theory and Techniques,* 52(10), 2438-2447.

Safrai, E., Ben Ishai, P., Polsman, A., Einav, S., & Feldman, Y. (2014). The Correlation of ECG Parameters to the Sub-THz Reflection Coefficient of Human Skin. *Ieee Transactions on Terahertz Science and Technology*, 4(5), 624-630.

Taylor, Z. D., Singh, R. S., Culjat, M. O., Suen, J. Y., Grundfest, W. S., Lee, H., & Brown, E. R. (2008). Reflective terahertz imaging of porcine skin burns. *Optics Letters*, 33(11), 1258-1260.

Turner, N. W., Subrahmanyam, S., & Piletsky, S. A. (2009). Analytical methods for determination of mycotoxins: A review. *Analytica Chimica Acta*, 632(2), 168-180.

Wang, S. J., Liu, K. S. , Yu, X. J. , Wu, D., & He, Y. 2011. Application of hybrid image features for fast and non-invasive classification of raisin. Journal of Food Engineering, 109 (3), 531-537.

Zhang, Y., Peng, X. H., Chen, Y., Chen, J., Curioni, A., Andreoni, W., Nayak, S. K., & Zhang, X. C. (2008). A first principle study of terahertz (THz) spectra of acephate. *Chemical Physics Letters,* 452(1-3), 59-66.

Figure captions

Fig.1 THz reflectance imaging experimental setup.

Fig.2 Wheat samples with different stages of mold contamination: (a) normal; (b) slightly; (c) moderately; (d) seriously.

Fig. 3 Flowchart of the procedure of discrimination moldy wheat by using THz imaging: (a) Imaging pre-processing; (b) Spectral analysis; (c) Imaging visualization.

Fig. 4 Frequency-domain THz spectra of the moldy wheat samples

Fig.5 Scores scatter plot of PC1, PC2, and PC3 for each moldy wheat sample

Fig.6 Loading weights of the top four PCs used for selecting the optimal frequencies

Fig.7 Scatter plots of the actual value versus the predicted value using the PCA-SVM model based on (a) the full spectrum and (b) the optimal frequencies for different moldy wheat samples.

Fig. 8 THz images of four wheat grains with different mold statuses: (a) normal; (b) slightly moldy; (c) moderately moldy; (d) seriously moldy.

Figures

Fig. 1

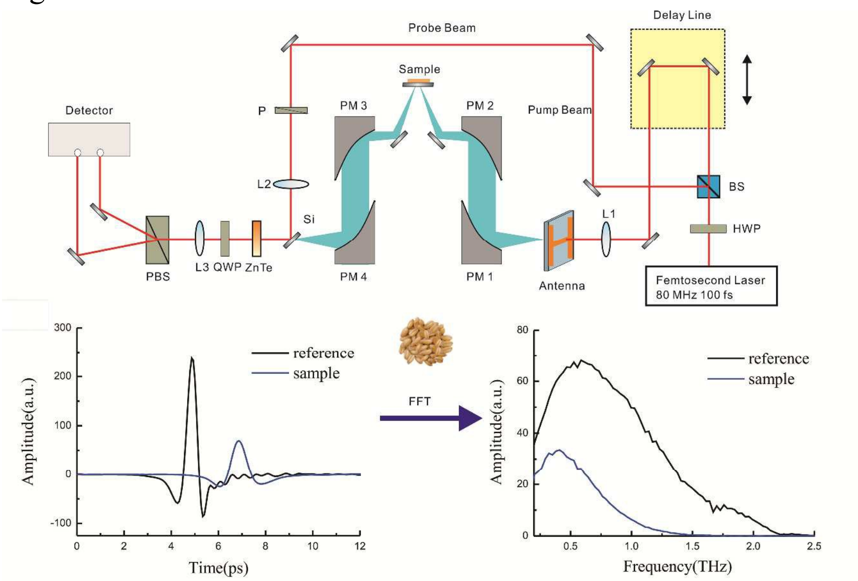


Fig. 2

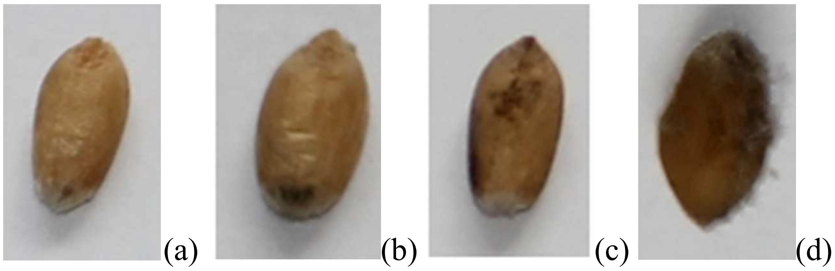


Fig. 3

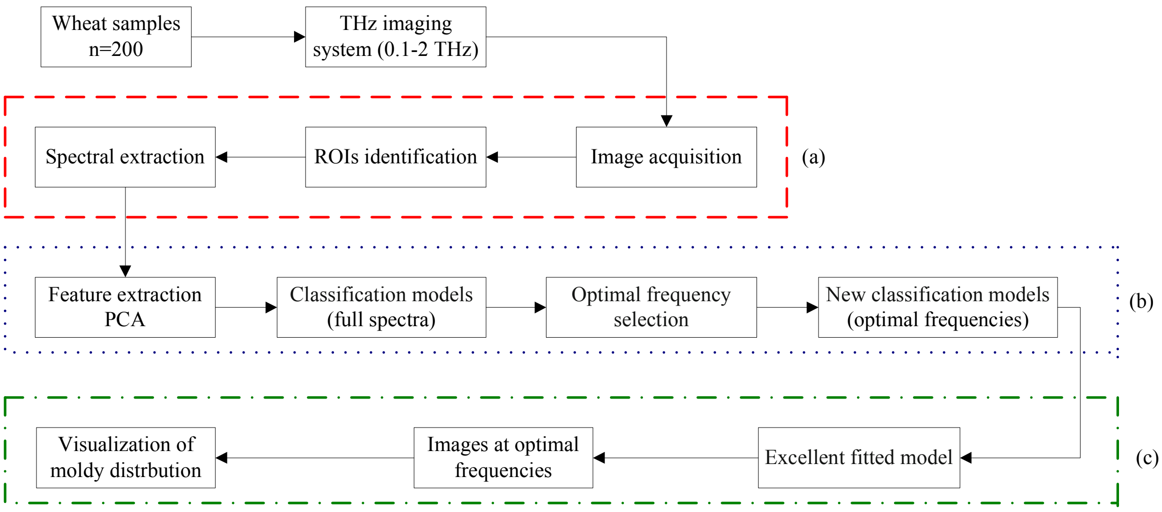

Wheat samples
n=200
THz imaging
system (0.1-2 THz)
Spectral extraction
ROIs identification
Image acquisition
(a)
Feature extraction
PCA
Classification models
(full spectra)
Optimal frequency
selection
New classification models
(optimal frequencies)
(b)
Visualization of
moldy distrbution
Images at optimal
frequencies
Excellent fitted model
(c)

Fig. 4

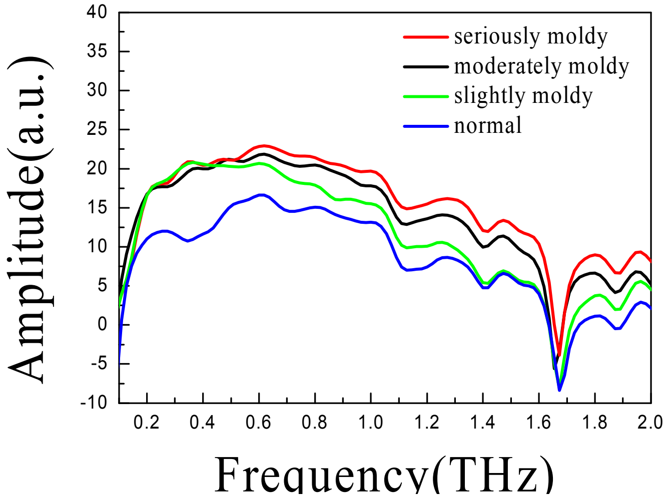

seriously moldy
moderately moldy
slightly moldy
normal
Amplitude(a.u.)
Frequency(THz)
40
35
30
25
20
15
10
5
0
-5
-10
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0

Fig. 5

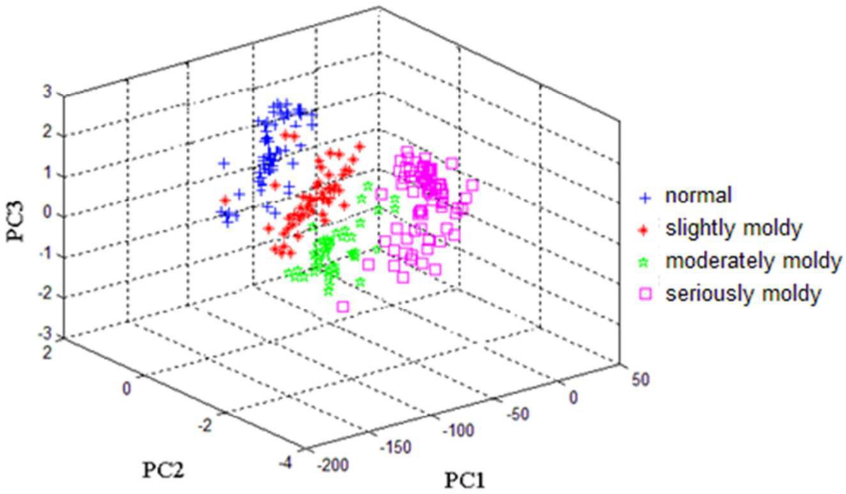

PC3
PC2
PC1
3
2
1
0
-1
-2
-3
2
0
-2
-4
-200
-150
-100
-50
0
50
normal
slightly moldy
moderately moldy
seriously moldy

Fig. 6

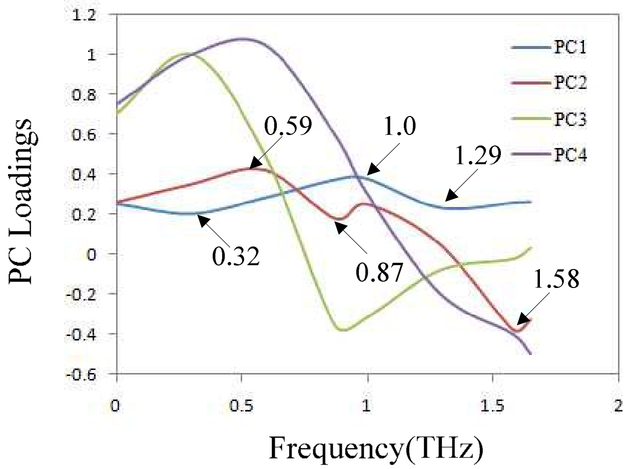

PC Loadings
Frequency(THz)
PC1
PC2
PC3
PC4
0.59
1.0
1.29
0.32
0.87
1.58
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
0
0.5
1
1.5
2

Fig. 7

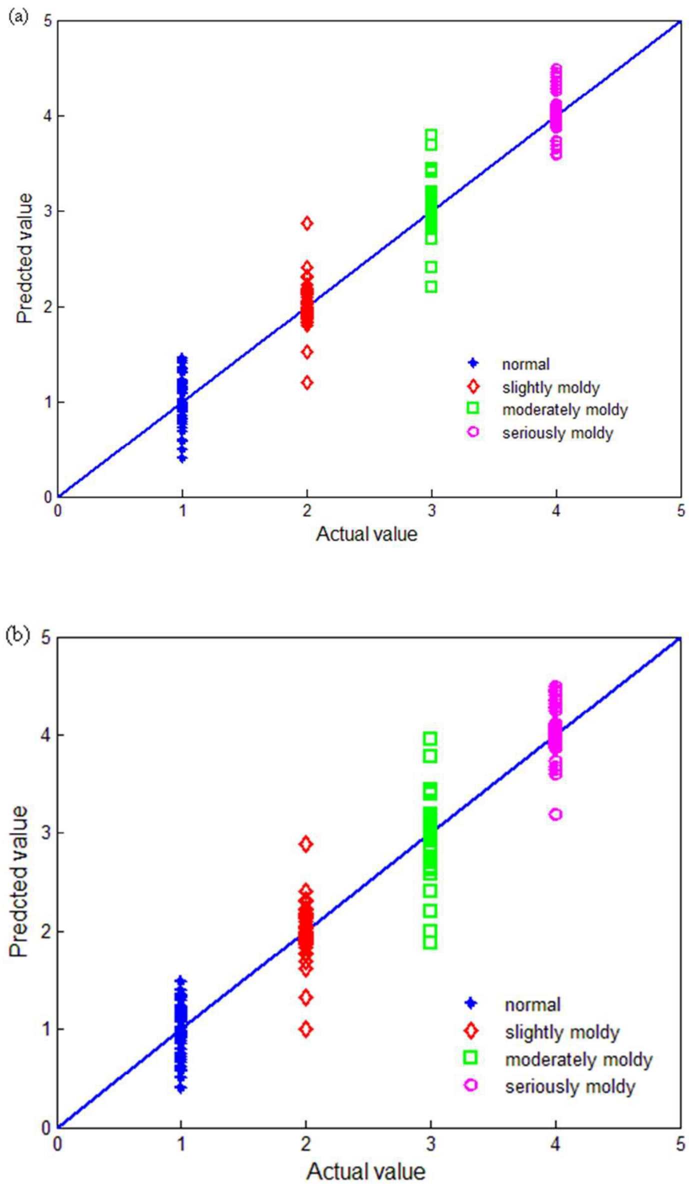

(a)
Predcted value
Actual value
normal
slightly moldy
moderately moldy
seriously moldy
(b)
Predcted value
Actual value
normal
slightly moldy
moderately moldy
seriously moldy

Fig. 8