

RESEARCH ARTICLE

# A new method for lint percentage non-destructive detection based on optical penetration imaging

Lijie Geng, Zhikun Ji, Pengji Yan, Ruiliang Zhang, Zhifeng Zhang, Yusheng Zhai, Wenyan Zhang, Kun Yang

School of Physics and Electronic Engineering, and Henan Key Laboratory of Magneto-electronic Information Functional Materials, Zhengzhou University of Light Industry, Zhengzhou 450002, Henan, China

## ABSTRACT

Lint percentage of seed cotton is one of the important bases for pricing in the trading segment. Unfortunately, the conventional methods of lint percentage are manually operated, which relies on the abundant experience of experts, and restrained by personal, physical and environmental factors. Up to date, the calculation of the lint percentage of seed cotton has not fully automated. In this paper, we proposed a non-destructive detection method for automatically obtaining lint percentage of seed cotton based on optical penetration imaging and machine vision, for the first time to our knowledge. The cotton seed image was obtained by the penetration imaging setup with a LED white backlight source. To accurately identify the number of cotton seeds, the image features of the cotton seed was studied and three key features was found, which are the circumference, area, and greyscale value, respectively. A calculation system based on the three key features was presented to process the images and then automatically calculate the lint percentage of seed cotton. The first step of the system is to segment the original image using adaptive thresholding followed by morphological operations. Afterwards, the number of cotton seed was obtained by the three key features of the cotton seed. Then, the lint percentage was achieved by a professional industry formula. The suggested lint percentage detection methods were verified by the experiments with two seed cotton varieties samples of H219 and ZHM19. The experimental results indicated that the detection average accuracy of the developed system for seed cotton varieties H219 and ZHM19 were 96.33% and 95.40%.

**Keywords:** Machine vision; Image processing; Cotton; Intelligent decision

## INTRODUCTION

Cotton, being a semi-halophyte with commercial value and is widely cultivated in Xinjiang, China (Q. Wang et al., 2021; Z. Zhang, Dong, Wang, & Pu, 2020). Over the past few years, China has emerged as a big country of cotton production, and the cotton industry is one of the most important pillars of China's economy (Alpermann, 2014). Therefore, the price of cotton attracts much more attention than before. Price of seed cotton setting is a multi-factorial decision with various determinants such as micronaire values, moisture regain, inclusion rate, short fiber rate, lint percentage and other factors. Among these factors, the lint percentage is one of the most important factors. The higher the lint percentage, the higher the price of seed cotton. Accordingly, the lint percentage of seed cotton has become a focus of attention. Unfortunately, the calculation of the lint percentage of seed cotton has not fully automated. Up to date, to obtain the lint percentage,

the operator needs to weigh the seed cotton samples, separate the lint and cotton seed with a cotton ginning machine, then weigh the separated lint again, and finally according to the professional formula, the lint percentage can be calculated by the operator. This conventional method of lint percentage obtaining suffers from time-consuming, human-intensive, complicated procedures and damaging the samples. Automation and intelligence will be a major trend in future evolution of various industries with the process of industrialization. Conventional detection methods constrain the rapid trading of seed cotton in the transaction links and impedes the industry process of automation and intelligence. To counteract the shortcomings of the conventional detection method, it's necessary to exploit a fast, simple, non-destructive, and low-budget method for lint percentage detection.

In recent years, image processing and machine vision make a fast evolution with the progress of the computer

### \*Corresponding author:

Lijie Geng, School of Physics and Electronic Engineering, and Henan Key Laboratory of Magneto-electronic Information Functional Materials, Zhengzhou University of Light Industry, Zhengzhou 450002, Henan, China. **E-mail:** 2014114@zzuli.edu.cn

**Received:** 18 November 2021; **Accepted:** 03 April 2022

and multimedia technologies. Image processing are being currently used in many different areas. Huang et al. (Huang & Chien, 2017) used a CCD color camera collect photos of three Taiwan rice seed breeds and then extracted 7 shape features using image processing. Finally, a classifier was used for classification. The classification accuracy of the resultant classifier for three rice seed varieties were 92.68%, 97.35% and 96.57%, respectively. Zhang et al. (F. Zhang & Zhang, 2011) used the image processing technology to grade tobacco leaves through their shape, color and surface texture features. The average accuracy of classification was 94%. There are many reported works in this regard (Chawgien & Kiattisin, 2021; Drewry, Luck, Willett, Rocha, & Harmon, 2019; Ghodrati, Kandi, & Mohseni, 2018; Miao, Jeon, & Park, 2020; Mondal, Rajan, & Ahmad, 2006; Sandra, Prayogi, Damayanti, & Djoyowasito, 2020; Sunoj et al., 2018; Tan et al., 2020; Yang et al., 2009). Image processing are also already widely used in solving problems related to cotton. Yang et al. (Yang et al., 2009) presented a novel way through image transformation, image enhancement and image segmentation to process images of foreign fibers. Ni et al. (Ni et al., 2020) presented a new classification method, which using the hyperspectral imaging with a spectral spectrum of 1000 - 2500 nm to online inspection the membrane of cotton. The variable weighted stacked autoworkers were used for extracting features and an artificial neural network were used for parameter optimization. The average accuracy of classification was 95%.

Penetration imaging is a well-known technique of detection. Because the non-contact, non-destructive and sensitivity properties, it is widely used for various studies. Taraghi I et al. (Taraghi, Lopato, Paszkiewicz, & Piesowicz, 2019) presented a THz imaging ways to examine the flaws in the sandwich panels. Coic, L et al. (Coic et al., 2019) proposed the hyperspectral imaging technique can as an effective method of falsified medicines analysis. Elmasry et al. (Elmasry et al., 2011) exploited a penetration imaging system to evaluate the grade of turkey. There are many reported works in this regard (Hong et al., 2012; Jaillon, Makita, Min, Lee, & Yasuno, 2011; Sumriddetchkajorn, Kamtongdee, & Chanhorm, 2015; Uchida & Okamoto, 2006; Xu, Wang, Shi, Yu, & Zhang, 2021).

In this paper, aiming at the above defects of conventional method for obtaining the lint percentage of seed cotton, we present a novel way for automatically obtaining the lint percentage based on penetration imaging and machine vision technology.

## MATERIALS AND METHODS

### Lint percentage of seed cotton

The lint percentage of seed cotton refers to the ratio of lint mass to cotton seed mass. Lint percentage is one of the principal factors that can affect price. The higher the lint percentage, the higher the price of cotton. Lint percentage  $P$  was calculated with the following formula.

$$P = \frac{M_C - M_F}{M_C} \quad (1)$$

Where  $M_C$  is the total seed cotton mass,  $M_F$  is the total cotton seed mass,  $P$  value stands for lint percentage of seed cotton. Total seed cotton mass  $M_C$  was obtained by a weighing sensor, then the  $M_C$  was passed into the developed system for calculating the lint percentage of seed cotton.  $M_F$  is given by

$$M_F = N \times M_f \quad (2)$$

Where  $N$  is overall number of cotton seed,  $M_f$  is the statistical mass of an individual cotton seed. Cotton seed total number  $N$  was the focus of our study. We detailly described how we detect cotton seed total number in this paper.  $M_f$  was obtained from the dataset. The Equation (1) and Equation (2) were written into the algorithms to achieve calculation of the lint percentage automatically.

### Image processing system

An image processing system is constructed as shown in Fig. 1, which is consisted of a camera (Hikvision MV-CE100-30GC 0.8 Lux 3840 × 2748), an PCI Bus Frame Grabber, a quartz glass plate (315 × 315 × 5mm), a LED white backlight source (32.3 × 32.3 × 7mm), a monitor, a computer and a printer.

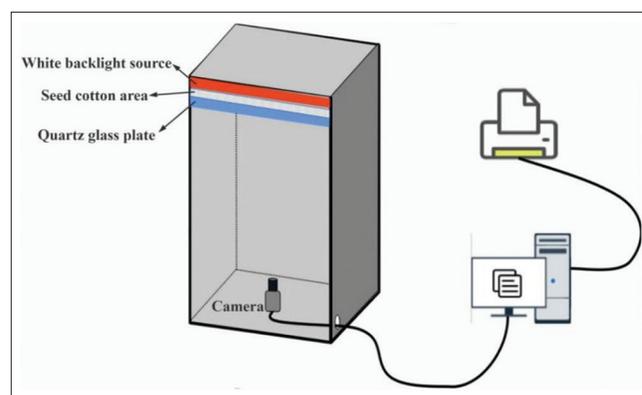


Fig 1. Image processing system

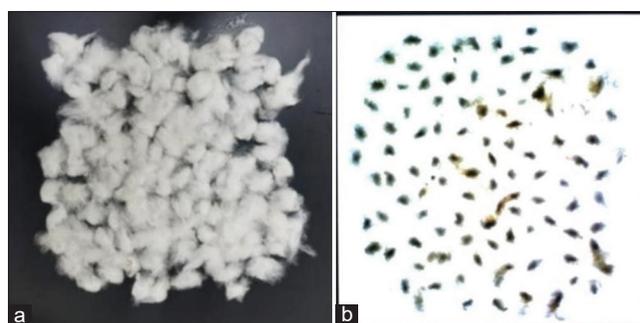
The color camera is fixed on the bottom center of the cabinet, which could move up-down, left-right, and forward-back. The quartz glass plate was placed above the color camera. The distance between the quartz glass plate and color camera was about 52 cm to ensure clear imaging. The seed cotton samples are placed evenly on the quartz glass plate, and the LED white backlight source was placed above the quartz glass plate. The distance between the LED white backlight source and quartz glass plate was about 7 mm to prevent the stacking of seed cotton layers. A computer is connected to the color camera for transmitting image information, and a printer was used to print the testing results. Python programming language is used to accomplish all algorithms.

### Image acquisition and preprocessing

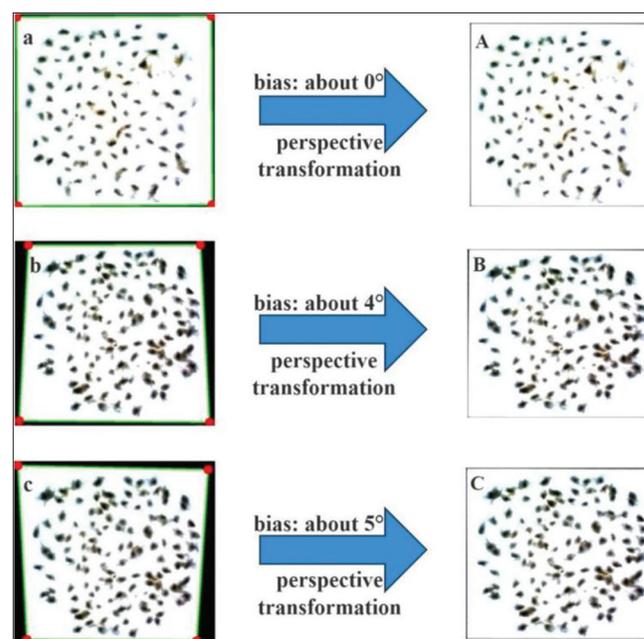
Image acquisition is one of the most basic steps during the lint percentage of seed cotton detection. The reflection images and the corresponding transmittance image of the seed cotton samples were obtained in experiment as shown in Fig. 2, one can see that there are considerably different between the two images. Transmittance image of seed cotton holds more abundant information, which included cotton seed, nep and impurity. Cotton knots are knots of cotton fibers tightly twisted together, which are denser and harder for white light to penetrate. Cotton seed had higher gray value compared to the nep in terms of gray value. Morphologically, cotton seed and impurity have similar shapes and different sizes. Although, they also have similar gray values. Cotton seed, however, which the margin gray value is changed gradually. Therefore, depending on the features of grayscale value and morphology (e.g., circumference and area), it was possible to distinguish cotton seed, nep and impurity from transmittance image. Additionally, a single cotton seed thickness was between 4 and 6 mm. The seed cotton was placed the sample region between the LED white backlight source and quartz glass plate during image acquisition. The spatial height of sample region is rigorously set at 7 mm to prevent the stacking of seed cotton layers. To summarize, we adopted penetration imaging method to obtain transmittance image of seed cotton to identify cotton seed, statistics the number of cotton seed and calculate the lint percentage of seed cotton.

Image preprocessing of the transmittance image is one of the most critical steps during the lint percentage of seed cotton detection. The perspective transformation (Haralick, 1980) and gaussian filtering (Pedder, 1993) were used to correct the image distortion and improve the image quality. In this experiment, in order to capture a photo of distortion-free, the optical axis of the CMOS camera should keep vertical to the quartz glass plate plane. However, it is difficult to achieve perpendicular between

the CMOS camera and quartz glass plate plane. In case of the CMOS camera and quartz glass plate plane were non-vertical to each other, the obtained image of seed cotton appears different level distortions. We used perspective transformation method to eliminate effects of image distortion and obtain corrected image. Four corner points in the transmittance image were used to constitute a polygon bounding the region of interest. The area included in this polygon for each transmittance image was then rotated into a corrected image using a perspective transformation. In this study, when the angular bias between the CMOS camera and quartz glass plate plane is no greater than 7 degrees, the transmittance image can be correct with the perspective transformation method. As shown in Fig. 3.



**Fig 2.** The (a) and (b) indicate seed cotton images of the same group under different imaging methods. The (a) is reflection image of seed cotton and the (b) is transmittance image of seed cotton



**Fig 3.** Transformation of transmittance image based on 4 corner points. The red points are corner points and the green polygon is the boundary for the perspective transformation. a, b and c indicating that the angular bias between the CMOS camera and quartz glass plate plane is about 0°, 4° and 5° respectively. A, B and C represents the perspective transformation images of a, b, and c, respectively

The CMOS camera for picture collection is subjected to miscellaneous sources of noise, including thermal noise and shot noise (Moroni, Mei, Leonardi, Lupo, & La Marca, 2014), which also caused the obtained transmittance image to be impressionable to noise. So far, about noise removing ways, median filtering and Gaussian filtering are normally used (Jiang, Zhang, Han, Liu, & Liu, 2013). The gaussian-filter eliminates noise by using a fixed size Gaussian kernel (Weber, Pauling, List, & Baumbach, 2020). Compared with the other filtering algorithm, Gaussian filtering has better smoothing effect and flexible filtering adjustment scale (G, L, K, & Y, 2019). In view of the white Gaussian noise was most widespread in transmittance image, the Gaussian filter was selected for noise decrease.

Gaussian smoothing is an image smoothing way based on the idea of neighborhood weighted averaging, where diverse positions of the pixels are given varying weights. (Meylan & Susstrunk, 2006). The Gaussian function with a one-dimensional zero-mean (Antolik, Sabatier, Galle, Frégnac, & Benosman, 2021) is as follows:

$$G(x) = e^{-\frac{x^2}{2\sigma^2}} \quad (3)$$

Where the width of the Gaussian filter is determined by the Gaussian distribution parameter  $\sigma$ . Among the image processing, a discrete Gaussian function with a 2-dimensional 0-mean is always chosen as the smoothing filter. The discrete Gaussian function that we used is as follows:

$$G[i, j] = e^{-\frac{(i^2 + j^2)}{2\sigma^2}} = e^{-\frac{r^2}{2\sigma^2}} \quad (4)$$

There are two methods to create a Gaussian filter: the convolution approach and the Gaussian template method (Cai et al., 2021; Kittisopikul, Virtanen, Taimen, & Goldman, 2019). In this study, the Gaussian template method was chosen. We design a Gaussian filter by calculating the template weights of a discrete Gaussian distribution. For computational simplicity, the filter weights are always integers and the weights of the filter templates must be normalised so that the uniform grey areas of the image are not to be influenced. Here a Gaussian template has been chosen for the equation with  $\sigma^2 = 1/2$ ,  $m = 3$ . The Gaussian template we have been selecting is shown as follows:

$$G = \frac{1}{16} \times \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \quad (5)$$

Fig. 4 shows the variance between the Gaussian filtered fractionally anisotropic image and the image containing

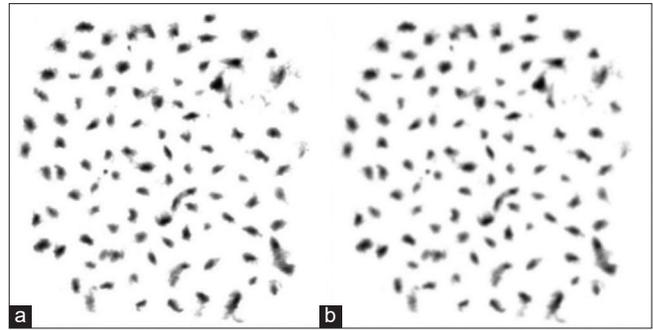


Fig 4. (a and b) Performance comparison of Gaussian filtering

noise. It is evident that the white Gaussian noise was largely removed and the image detail was well retained.

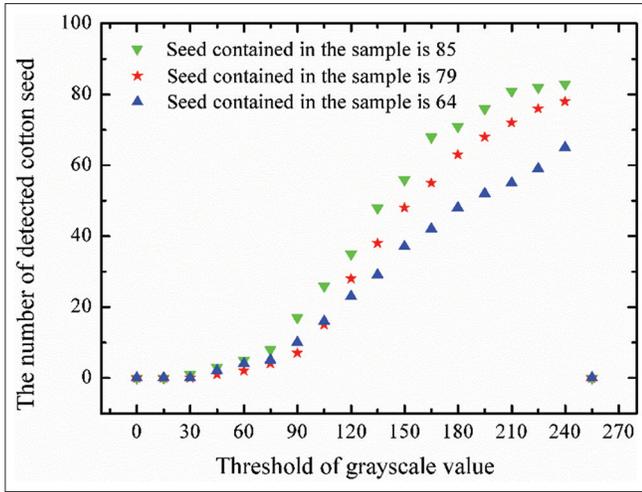
### Features extraction

#### Choice of optimal threshold based on the gray value

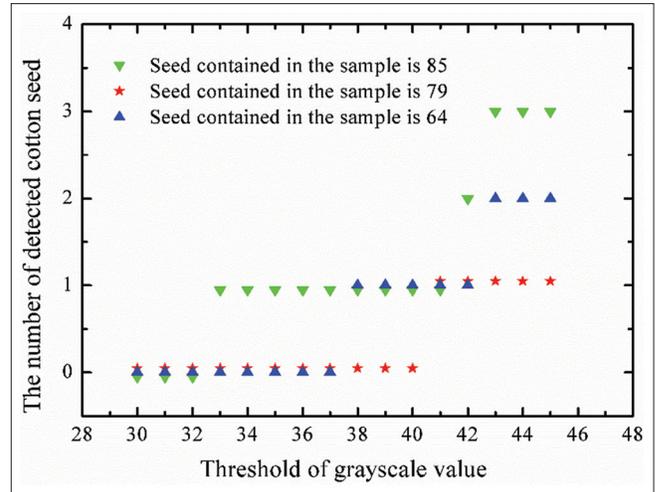
In the grayscale image, the grayscale value refers to the brightness of single pixel points, and the larger the grayscale value indicates that the individual pixel is brighter. These values are displayed in 256 grayscales, from 0 to 256 (Shimomura et al., 2021). The gray values have been widely used as the features of the image in many studies. For the purpose of obtaining the average surface ruggedness  $R$  of operational conductors. Liu et al. (Liu, Bian, Cao, & Zhuansun, 2015) presented a feature of two-dimensional surface morphology based on a matrix of grey values  $G$ . They found a linear relationship between the variance  $F$  and  $R$  within a wide range of  $R$ . As a result,  $R$  can be obtained from the images and  $G$  without the need to dismantle the conductor being used, and it is more economical and time efficient for the measurement.

In this work, the maximal gray value and the minimum gray value of the cotton seed in the transmittance image were defined as  $G_{max}$  and  $G_{min}$ , respectively. Then, when the gray value of pixel point was between  $G_{max}$  and  $G_{min}$ , it may be the cotton seed. Therefore, the gray scale value of the image can be used as a parameter to distinguish cotton seed, and we also determined experimentally the values of  $G_{max}$  and  $G_{min}$ . Under the circumstance of other experimental conditions be equal, three seed cotton samples were used to investigate the relation between the threshold of grayscale value from 0 to 255 and the number of cotton seed, as shown in Fig. 5. The actually number of seed contained in the three seed cotton samples were 85, 79 and 64, respectively.

As one can see in Fig. 5, for the three groups of samples, when the threshold of grayscale value less than 30, the numbers of cotton seed detected in each group was 0, when the threshold of grayscale value greater than 30 and less than or equal 45, the cotton seed were detected in the each group, so the lowest grayscale value of cotton seed is  $G_{min}$ ,



**Fig 5.** The relation between the threshold of grayscale value from 0 to 255 and the number of detected cotton seed

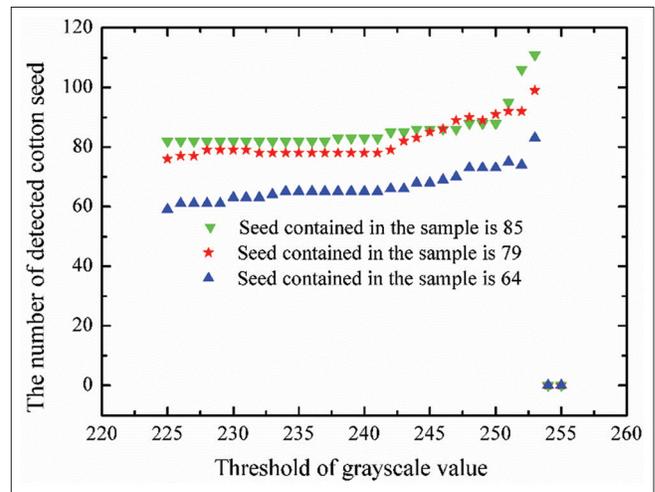


**Fig 6.** The relation between the threshold of grayscale value from 30 to 35 and the number of detected cotton seed

which values is between 30 and 45. The numbers of cotton seed detected in each group increases with increase in the threshold of grayscale value. However, we found that when the threshold of grayscale value changed from 240 to 255, the numbers of cotton seed detected in each group was 0, this is because there were no differences for the cotton seed and background in the transmittance image after binarization as the threshold of grayscale value increases to a certain value, and the features of cotton seed could not be extracted. In order to get an accurate  $G_{min}$  and  $G_{max}$ , in the range [30, 45] and [225, 255], we further investigated the relation between the threshold of grayscale value and number of cotton seed. As shown in Fig. 6 and Fig. 7.

It can be seen from Fig. 6, for the three groups of samples, the lowest threshold of grayscale value in each group is 33, 38 and 41, respectively. Therefore, we have chosen  $G_{min}=33$  to ensure all cotton seed were detected.

As can be seen from Fig. 7, the numbers of cotton seed detected in each group around the true value as the threshold of grayscale value gradually changed from 230 to 241, which indicated that the threshold of grayscale value has reached the  $G_{max}$  and all cotton seed in each group of samples have been detected. When threshold of grayscale value gradually changed from 241 to 253, the numbers of cotton seed detected in each group rose rapidly, which indicated that the threshold of grayscale value has reached the minimum gray value for lint and some of the lint has been misidentified as cotton seed. When the threshold of grayscale value increases to 254, the numbers of cotton seed detected in each group was 0, which indicated that the threshold of grayscale value has reached the gray value of background and there were few features in the transmittance image after binarization. In order to ensure all cotton seed were detected as far as



**Fig 7.** The relation between the threshold of grayscale value from 225 to 255 and the number of detected cotton seed

possible and to avoid misidentifying lint as cotton seed. We have chosen  $G_{max}=236$ .

**Classification of perimeter based on the number of pixels**

The sum number of all pixels on the boundary of the connected area of the image is the perimeter (Venegas et al., 2020). In some studies, an image with perimeter features plays an important role. Y. Wang and McClung (Y. Wang & McClung, 2018) presented an assessment of the surface area of exposed grains in a 2D polished section analysis of porphyry copper ores based on corresponding HRXMT data. The results show that the polished cross-sectional analysis based on perimeter gives a good approximation of the surface area of the exposed grains as determined by HRXMT. Stachowiak et al. (Stachowiak, Stachowiak, & Podsiadlo, 2008) used image analysis software to determine the perimeter and area parameters of each wear particle and

performed a statistical analysis of these parameters. Each particle class was then evaluated according to the features of perimeter and area.

There are cotton seed, nep and impurities in a transmittance image of seed cotton. Fortunately, they differed significantly by circumferences. The circumference of cotton seed is larger, followed by the impurities and nep, as shown in Fig. 8.

We can obtain the relationship between the circumference and the numbers of cotton seed by image analysis, as shown in Table 1. We will employ the data to count the numbers of cotton seed in the next Section.

#### Screening of area based on the number of pixels

The sum number of all pixels in the connected area of the image is the area (Harris et al., 2018). Similar to the circumference, the area features of an image are also important in some studies. Tong et al. (Tong, Li, & Jiang, 2013) used the area parameter of seedling leaves to measure the quality of seedlings. They developed a vision system to detect seedling leaves in the growing stage, which enables real-time detection of seedling quality. Leaf area is also used by Silva et al. (Silva, Reis, Nascimento, Nascimento, & Filho, 2021) to explore the effects of N, K and Mg elements on the growth of arugula plants.

In a similar manner, the cotton seed could be differentiated from impurity and nep based on the size of area, as shown in Fig. 9.

We can obtain the relationship of the cotton seed and the area similarly, as shown in Table 2. These data will be used to count the numbers of cotton seed in the next Section.

#### Establishment of dataset

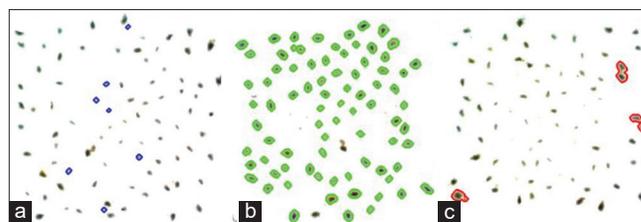
In order to assess the reliability and universality of the lint percentage detection method, setting up a dataset of individual cotton seed mass about H219 (varieties of seed cotton) and ZHM19 (varieties of seed cotton) to obtain individual cotton seed mass for calculating the lint percentage. In this study, we employed H219 and ZHM19 as experimental subjects purchased from breeding organizations. The detailed process of dataset establishment about individual cotton seed mass of H219 and ZHM19 is described below.

Step1 Five groups seed cotton sample (H219 or ZHM19) was obtained.

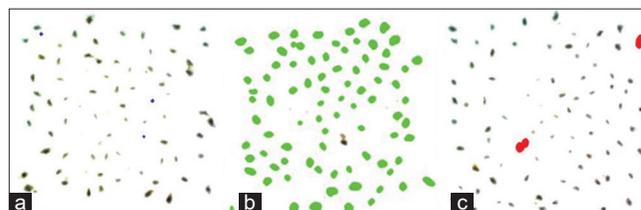
Step2 Five groups of seed cotton was separated into lint and cotton seed.

Step3 Five groups of cotton seed was weighed.

Step4 Five groups of cotton seed was counted.



**Fig 8.** The features of cotton seed, impurity and nep. (a) each of the area surrounded by the blue curve is the impurity or nep, which represent area with an absence of cotton seed. (b) each of the area surrounded by the green curve represent 1 cotton seed. (c) each of the area surrounded by the red curve represent 2 cotton seed



**Fig 9.** The features of cotton seed, impurity and nep. (a) Each of the blue area is the impurity or nep, which represent area with an absence of cotton seed. (b) Each of the green area represent 1 cotton seed. (c) Each of the red area represent 2 cotton seed

**Table 1: Circumference and the corresponding numbers of cotton seeds**

Circumference (pixels)	Numbers of cotton seeds
85-180	2
20-85	1
0-20	0

**Table 2: Area and the corresponding numbers of cotton seeds**

Area (pixels)	Numbers of cotton seeds
500-1200	2
27-500	1
0-27	0

Step5 The mass of individual cotton seed was calculated.

The dataset of individual cotton seed mass about H219 and ZHM19 was established according to these steps, as shown in Fig. 10.

Based on the number and mass of cotton seed in the samples, the average mass of each group cotton seed was calculated, then the average mass of each variety five groups cotton seed was calculated. The results of calculated as the dataset of individual cotton seed mass, as shown in Table 3.

## RESULTS

The presented algorithm is experimented with using a PC i7-7700HQ, GeForce 950M 4 GB, 16 GB RAM, 512 GB HDD. The programming language used Python 3.7 in the

PyCharm integrated development environment, with some packages namely numpy, imutils, matplotlib, argparse, and opencv-python. Experiments on Windows 10 and 64-bit operating systems. The presented algorithm is tested against 10 groups of seed cotton samples. It includes 5 groups seed cotton of variety H219 and 5 groups seed cotton of variety ZHM19. The details related to the processes for counting the number of cotton seed are given in Fig.11 and experiment results are provided in Table 4.

**Statistics the numbers of cotton seed**

In this work, a large numbers of transmittance images were collected. In brief, statistical the number of cotton seed embraces the three main steps: The first step was to preprocess the transmittance image to improve the image

quality. The pre-processing contains two main processes which are perspective transformation and Gaussian denoising. Perspective transformation is used to correct the image which was slightly distorted, and Gaussian filter is used for smoothing the image. The second step was binarize the grayscale image by setting a suitable threshold, the image is changed from gray scale status to binary image with the help of the threshold technique which separates the cotton seed from the whole image. The next step is the finding of the contours of the image. Finally, the filtering contours way is applied for cotton seed detection and count.

The numbers of cotton seed were auto-counted according to these steps as shown in Fig. 11.

**Table 3: Dataset of individual cotton seed quality**

Variety	Breed origins	Origin	Numbers of cotton seed	Mass	$M_i$
H219	461—3×ZM—4	Xinjiang	2623	278.8g	0.10630g
ZHM19	Xiang Z201×H101	Hunan	2565	277.4g	0.10814g

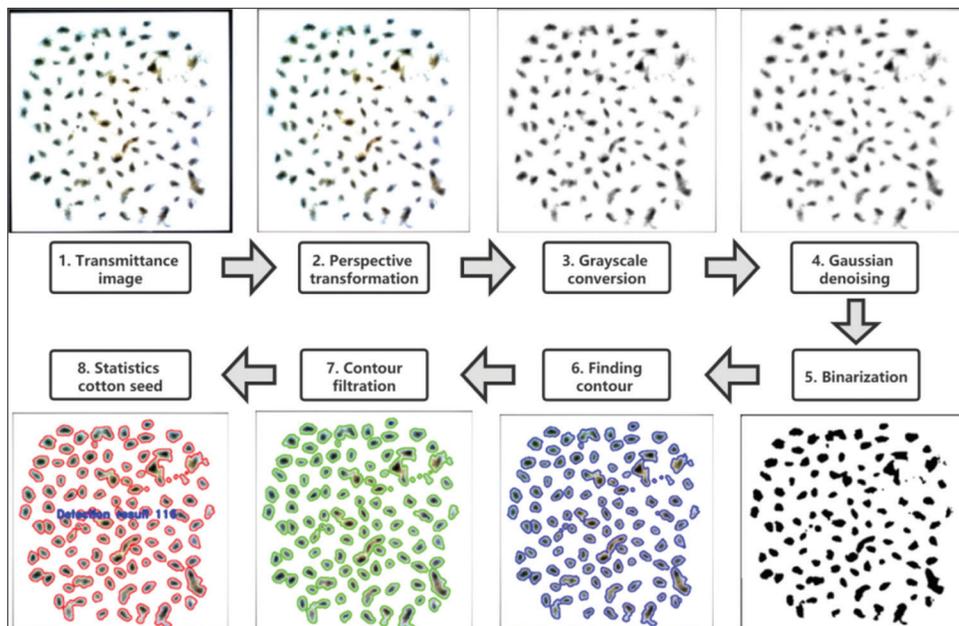
INPUT: Transmittance image, which was captured by a color camera.

OUTPUT: Output image with number and contour of the cotton seed.

Step1: Transmittance image was captured by a color camera.



**Fig 10.** The process of dataset establishment about individual cotton seed mass of H219 and ZHM19



**Fig 11.** The processes for counting the number of cotton seed

**Table 4: Experiment results**

Variety	Group	Mass ( $M_c$ )	Real number of cotton seed	Detection result(N)	Lint percentage (P)	Accuracy
H219	①	12.9g	65	60	50.39%	93.44%
H219	②	14.1g	71	74	43.97%	98.41%
H219	③	15.8g	82	79	46.84%	97.38%
H219	④	10.5g	56	57	41.90%	93.61%
H219	⑤	18.5g	90	95	45.41%	98.82%
ZHM19	⑥	13.2g	66	65	46.97%	91.22%
ZHM19	⑦	20.9g	115	113	41.63%	95.16%
ZHM19	⑧	20.0g	109	116	37.50%	96.15%
ZHM19	⑨	14.7g	91	81	40.14%	98.33%
ZHM19	⑩	13.2g	77	72	40.91%	96.14%

Step2: By applying perspective transformation for correcting the image.

Step3: The corrected image was converted into the grayscale image.

Step4: By using Gaussian filter to remove noise from the photo.

Step5: The image was binarized, set threshold (33, 236). The pixels with gray values below the 33 and above 236 were set to 255, and gray values between 33 and 236 were set to 0.

Step6: Extracting the contours information from the binarized image.

Step7: Filtering the contours.

Step7.1: If the circumference range is between 0-20 or area range is between 0-27 is considered as 0 cotton seed.

Step7.2: If the circumference range is between 20-85 or area range is between 27- 500 is considered as 1 cotton seed.

Step7.3: If the circumference range is between 85-180 or area range is between 500- 1200 is considered as 2 cotton seed.

Step8: The number of cotton seed was counted.

In this research, several groups of seed cotton samples have been used to verified the method of statistics the numbers of cotton seed, and the results were shown in Fig. 12.

### The calculated results of lint percentage

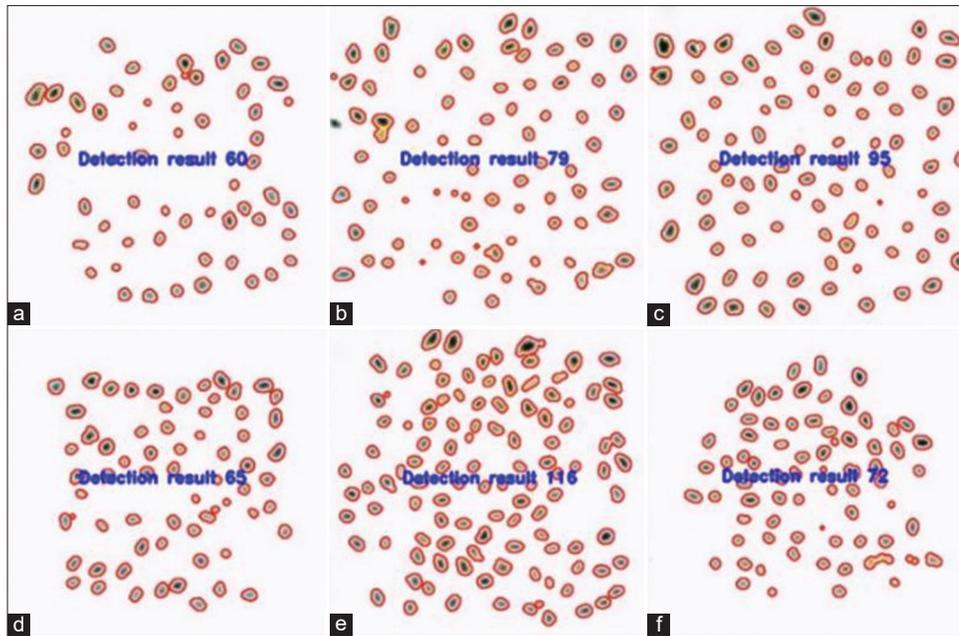
The functions of software include file operations (gather, load, and save photo), image preprocessing (greyscale-transformed, binarization, perspective transformation, gaussian denoising), statistics cotton seed, and the calculation of lint percentage. Five groups seed cotton of variety H219 and five groups seed cotton of variety ZHM19 were used as experimental samples. The mass of each group sample is  $M_c$ , which is weighed and recorded by the load cell, and then automatically transferred to the calculation software. The dataset of individual cotton seed mass has been stored in the calculation software, which recalls the corresponding individual cotton seed mass

$M_f$  automatically depending on the input seed cotton variety. The number of cotton seed  $N$  is obtained with the calculation software. Based on equations (1) and (2) within the software, the software can calculate the lint percentage of seed cotton accurately and rapidly. Several group samples have been used to verified the software, the results were shown in Table 4.

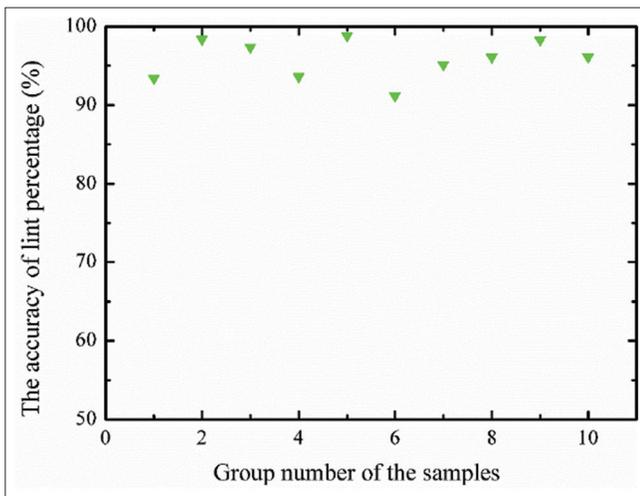
According to the Table 4, the accuracy of lint percentage was plotted for each group, as shown in Fig. 13. From Fig. 13, it can be seen that the accuracy of the algorithm is between 91.22% and 98.82%. The average accuracy for lint percentage of variety H219 and ZHM19 was calculated, which was 96.33% and 95.40%, respectively. The experiments show that the method has high feasibility and effectiveness.

## DISCUSSION

In the cotton market, the lint percentage is important as it is known to influence cotton prices. The traditional method of calculating the lint percentage could not be avoided interference from human factors due to individuals' involvement. The legal rights of empor will be infringed if the calculated lint percentage is high during cotton transaction. Conversely, the legal rights of seller will be infringed. As a result, in this research a new compute the lint percentage approach is proposed, which the seed cotton image was obtained by the penetration imaging setup with a LED white backlight source. Then through the perspective transformation to obtain corrected image. The features of corrected image were studied and three key features was been found, which are the circumference, area, and greyscale value, respectively. Two morphological features (circumference, area) and one color (grayscale) features of each corrected image were extracted. These features were used to identify cotton seed and count the number of cotton seed. The mass of seed cotton is weighed and recorded by the load cell, and then automatically transferred to the developed calculation system. The dataset of individual cotton seed mass has been stored in



**Fig 12.** Statistical results of cotton seed. The seed cotton variety of a, b, c and d, e, f is H219 and ZHM19, respectively. The a, b, c, d, e and f actually contain the number of cotton seed are 65, 82, 90, 66, 109 and 77, respectively. Blue colored font indicates the result of detection with the proposed method



**Fig 13.** The accuracy of lint percentage

the calculation system, which recalls the corresponding individual cotton seed mass automatically depending on the input seed cotton variety. Afterwards, the lint percentage was calculated by a professional industry formula that written in developed calculation system. The proposed method can calculate the lint percentage and indicates favorable performance.

Fig.13 shows the experimental results. The performance of the proposed method showed that this technique can replace the traditional manual method. However, as can be seen in Table 4, the difference between the highest and lowest lint percentage for the varieties H219 and ZHM19

were 8.49% and 9.47%, respectively, although the accuracy rate of the results was very high. Theoretically, the lint percentage of the same breed was essentially equivalent. However, resulting in the difference of lint percentage between same varieties due to the segmentation algorithms failure to accurately identify numbers of cotton seed and the sample size was small in this experiment. In response to above issues, we will optimize the segmentation algorithm to improve its performance and design more reasonable experiments to carry out in-depth mechanism research. Moreover, the penetration of white light is unable to completely penetrate the nep and impurities, which also increased the difficulty of identifying cotton seed and spurred us to seek for some other alternatives.

Our contributions in this paper are summarized as following. First, to the best of our knowledge, research studies of lint percentage are scarce; thus, this study attempts to fill this research gap. Second, white light could be used as a transmitted light source for the obtaining of penetration image. Thus, we will try to generalize this method to other areas for solving various research and practical issues. Thirdly, through the experimentation, we verified the gray value, perimeter and area can be used as a feature to distinguish cotton seed, nep and impurities, which were used to separate the cotton seed achieved a better result. Finally, we have succeeded in making a data set with mass of an individual cotton seed about H219 and ZHM19, and made it publicly available for use by other researchers.

In the future, algorithms to enhance the segmentation process will be explored and used to avoid bridging problems. In addition, we will try other algorithms to solve our problem, such as deep learning algorithms. In fact, whatever algorithm we employed, it is conceivable that the whole system could be integrated into an industrial robot, but this would become cost prohibitive.

## CONCLUSIONS

In this study, we proposed a non-destructive detection method for automatically obtaining lint percentage of seed cotton based on optical penetration imaging and machine vision. The method uses the light penetration to image the cotton seeds, which are difficult to distinguish with the human eye. Cotton seeds then were identified and counted based on the three image features, which are gray value, perimeter and area, to calculate the lint percentage of seed cotton. The test results show that the calculation average accuracy of the new way for lint percentage of varieties H219 and ZHM19 were 96.33% and 95.40%, respectively and the new method is a viable way for the automatic calculating of lint percentage. This method will be one of promising alternatives, with potential advantages of fast, simple, non-destructive, and low-cost for development of a novel solution to improve the level of cotton industry automation and intelligence.

### Funding

This research was funded by the National Natural Science Foundation of China grant Number (11904327, 61905223, 62105296), the Henan science and technology development plan project (222102110279, 202102210318).

### Conflicts of interest

The authors declare no conflict of interest.

## REFERENCES

Alpermann, B. 2014. China's cotton market as a strategic action field. In: L. Augustin-Jean and B. Alpermann (Eds.), *The Political Economy of Agro-Food Markets in China: The Social Construction of the Markets in an Era of Globalization*. Palgrave Macmillan UK, London. pp. 183-209.

Antolik, J., Q. Sabatier, C. Galle, Y. Frégnac and R. Benosman. 2021. Assessment of optogenetically-driven strategies for prosthetic restoration of cortical vision in large-scale neural simulation of V1. *Sci. Rep.* 11: 10783-10783.

Cai, C., J. Friedrich, A. Singh, M. H. Eybposh, E. A. Pnevmatikakis, K. Podgorski and A. Giovannucci. 2021. VolPy: Automated and scalable analysis pipelines for voltage imaging datasets. *PLoS Comput. Biol.* 17: e1008806.

Chawgien, K. and S. Kiattisin. 2021. Machine learning techniques for classifying the sweetness of watermelon using acoustic signal and image processing. *Comput. Electron. Agric.* 181: 105938.

Coic, L., P. Y. Sacré, A. Dispas, A. K. Sakira, M. Fillet, R. D. Marini and E. Ziemons. 2019. Comparison of hyperspectral imaging techniques for the elucidation of falsified medicines composition. *Talanta.* 198: 457-463.

Drewry, J. L., B. D. Luck, R. M. Willett, E. M. C. Rocha and J. D. Harmon. 2019. Predicting kernel processing score of harvested and processed corn silage via image processing techniques. *Comput. Electron. Agric.* 160: 144-152.

Elmasry, G., P. Y. Sacré, A. Dispas, A. K. Sakira, M. Fillet, R. D. Marini and E. Ziemons. 2011. Quality classification of cooked, sliced turkey hams using NIR hyperspectral imaging system. *J. Food Eng.* 103: 333-344.

Parthasarathy, G., L. Ramanathan, K. Anitha and Y. Justindhas. 2019. Predicting source and age of brain tumor using canny edge detection algorithm and threshold technique. *Asian Pac. J. Cancer Prev.* 20: 1409-1414.

Ghodrati, S., S. G. Kandi and M. Mohseni. 2018. Nondestructive, fast, and cost-effective image processing method for roughness measurement of randomly rough metallic surfaces. *J. Opt. Soc. Am. A.* 35: 998-1013.

Haralick, R. M. 1980. Using perspective transformations in scene analysis. *Comput. Graph. Image Proc.* 13: 191-221.

Harris, A. N., H. W. Lee, G. Osis, L. Fang, K. L. Webster, J. W. Verlander and I. D. Weiner. 2018. Differences in renal ammonia metabolism in male and female kidney. *Am. J. Physiol. Renal Physiol.* 315: F211-F222.

Hong, Y. J., S. Makita, F. Jaillon, M. J. Ju, E. J. Min, B. H. Lee and Y. Yasuno. 2012. High-penetration swept source Doppler optical coherence angiography by fully numerical phase stabilization. *Opt. Express.* 20: 2740-2760.

Huang, K. Y. and M. C. Chien. 2017. A novel method of identifying paddy seed varieties. *Sensors (Basel).* 17(4): 809.

Jaillon, F., S. Makita, E. J. Min, B. H. Lee and Y. Yasuno. 2011. Enhanced imaging of choroidal vasculature by high-penetration and dual-velocity optical coherence angiography. *Biomed. Opt. Express.* 2: 1147-1158.

Jiang, S., P. Zhang, T. Han, W. Liu and M. Liu. 2013. Tri-linear interpolation-based cerebral white matter fiber imaging. *Neural Regen. Res.* 8: 2155-2164.

Kittisopikul, M., L. Virtanen, P. Taimen and R. D. Goldman. 2019. Quantitative analysis of nuclear lamins imaged by super-resolution light microscopy. *Cells.* 8: 361.

Liu, L., X. Bian, L. Cao and X. Zhuansun. 2015. The evaluation method of surface roughness degree for long-term operating conductors based on gray value matrix. *IEEE Trans. Power Deliv.* 30: 1641-1643.

Meylan, L. and S. Susstrunk. 2006. High dynamic range image rendering with a retinex-based adaptive filter. *IEEE Trans. Image Proc.* 15: 2820-2830.

Miao, Y., J. Y. Jeon and G. Park. 2020. An image processing-based crack detection technique for pressed panel products. *J. Manuf. Syst.* 57: 287-297.

Mondal, P. P., K. Rajan and I. Ahmad. 2006. Filter for biomedical imaging and image processing. *J. Opt. Soc. Am. A.* 23: 1678-1686.

Moroni, M., A. Mei, A. Leonardi, E. Lupo and F. La Marca. 2014. PET and PVC separation with hyperspectral imaging. *Int. Electron. Conf. Sens. Appl.* 15: g011.

Ni, C., Z. Li, X. Zhang, X. Sun, Y. Huang, L. Zhao and D. Wang. 2020. Online sorting of the film on cotton based on deep learning and hyperspectral imaging. *IEEE Access.* 8: 93028-93038.

Pedder, M. 1993. Interpolation and filtering of spatial observations

- using successive corrections and gaussian filters. *Mon. Weather Rev.* 121: 2889-2902.
- Sandra, I. Prayogi, R. Damayanti and G. Djoyowasito. 2020. Design to prediction tools for banana maturity based on image processing. *IOP Conf. Ser. Earth Environ. Sci.* 475: 012010.
- Shimomura, M., A. Yumoto, N. Ota-Murakami, T. Kudo, M. Shirakawa, S. Takahashi and D. Shiba. 2021. Study of mouse behavior in different gravity environments. *Sci. Rep.* 11: 2665-2665.
- Stachowiak, G. P., G. W. Stachowiak and P. Podsiadlo. 2008. Automated classification of wear particles based on their surface texture and shape features. *Tribol. Int.* 41: 34-43.
- Sumriddetchkajorn, S., C. Kamtongdee and S. Chanhorm. 2015. Fault-tolerant optical-penetration-based silkworm gender identification. *Comput. Electron. Agric.* 119: 201-208.
- Sunoj, S., S. N. Subhashree, S. Dharani, C. Igathinathane, J. G. Franco, R. E. Mallinger and D. Archer. 2018. Sunflower floral dimension measurements using digital image processing. *Comput. Electron. Agric.* 151: 403-415.
- Tan, C., P. Zhang, Y. Zhang, X. Zhou, Z. Wang, Y. Du and W. Guo. 2020. Rapid recognition of field-grown wheat spikes based on a superpixel segmentation algorithm using digital images. *Front. Plant Sci.* 11: 259.
- Taraghi, I., P. Lopato, S. Paszkiewicz and E. Piesowicz. 2019. X-ray and terahertz imaging as non-destructive techniques for defects detection in nanocomposites foam-core sandwich panels containing carbon nanotubes. *Polym. Test.* 79: 106084.
- Tong, J. H., J. B. Li and H. Y. Jiang. 2013. Machine vision techniques for the evaluation of seedling quality based on leaf area. *Biosyst. Eng.* 115: 369-379.
- Uchida, K. and K. Okamoto. 2006. Measurement of powder flow in a screw feeder by x-ray penetration image analysis. *Meas. Sci. Technol.* 17: 419.
- Venegas, P., N. Pérez, S. Zapata, J. D. Mosquera, D. Augot, J. L. Rojo-Álvarez and D. Benítez. 2020. An approach to automatic classification of *Culicoides* species by learning the wing morphology. *PLoS One.* 15: e0241798.
- Wang, Q., X. Lu, X. Chen, W. A. Malik, D. Wang, L. Zhao and W. Ye. 2021. Transcriptome analysis of upland cotton revealed novel pathways to scavenge reactive oxygen species (ROS) responding to Na<sub>2</sub>SO<sub>4</sub> tolerance. *Sci. Rep.* 11: 8670.
- Wang, Y. and C. McClung. 2018. Stereological correction of perimeter based estimates of exposed grain surface area. *Min. Eng.* 126: 64-73.
- Weber, P., J. K. Pauling, M. List and J. Baumbach. 2020. BALSAM-An Interactive online platform for breath analysis, visualization and classification. *Metabolites.* 10: 393.
- Xu, S., M. Wang, X. Shi, Q. Yu and Z. Zhang. 2021. Integrating hyperspectral imaging with machine learning techniques for the high-resolution mapping of soil nitrogen fractions in soil profiles. *Sci. Total Environ.* 754: 142135.
- Yang, W., P. Y. Sacré, A. Dispas, A. K. Sakira, M. Fillet, R. D. Marini and E. Ziemons. 2009. A new approach for image processing in foreign fiber detection. *Comput. Electron. Agri.* 68: 68-77.
- Zhang, F. and X. Zhang. 2011. Classification and quality evaluation of tobacco leaves based on image processing and fuzzy comprehensive evaluation. *Sensors (Basel).* 11: 2369-2384.
- Zhang, Z., X. Dong, S. Wang and X. Pu. 2020. Benefits of organic manure combined with biochar amendments to cotton root growth and yield under continuous cropping systems in Xinjiang, China. *Sci. Rep.* 10: 4718.