Computers and Electronics in Agriculture 153 (2018) 248–255

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Application driven key wavelengths mining method for aflatoxin detection using hyperspectral data

Han Zhongzhi*, Deng Limiao

School of Science and Information Science, Qingdao Agriculture University, Qingdao, Shandong 266109, China

ARTICLE INFO

Keywords:
Aflatoxin
Peanut
Hyperspectral data
Key wavelength
Sorter design

ABSTRACT

The first step of developing Aflatoxin intelligent sorter is to determine the key wavelengths for aflatoxin detection. In order to find more accurate wavelengths, in this paper, three kinds of sensor system are built separately: the first sensor is hyper-spectrometer by ASD spectrometer, the second is the multispectral camera system based on Liquid crystal tunable filter (LCTF) and the third one is the hyperspectral camera based on Grating spectrometer module (GSM). Under 365 nm UV LED illumination, using these three systems, three hypersepctral datasets of 45, 41 and 73 peanut samples before and after aflatoxin contaminated have be collected separately. In order to select the best key wavelengths, four feature selection methods (Fisher, SPA, BestFist and Ranker) and four classifier models (KNN, SVM, BP-ANN, RandomForest) were analyzed and compared. Using all selected wavelengths based on different datasets, a weighted voting method was proposed and 10 key wavelengths (440 380 410 460 420 370 450 490 700 600 nm) were selected. Based on the best model (Random Forest), the best integrated average recognition rate is 94.5%. And then, using these key wavelengths and the best classification model, a new design system for aflatoxin sorter base on a Ploygon mirror was proposed. The structure of this system is simple, detection accuracy is high, which can be applied to online sorting of aflatoxin detection.

1. Introduction

Aflatoxin is a highly toxic carcinogen, and the toxicity is about 68 times of that of arsenic. It is the strongest chemical carcinogen found so far and is the biggest cause of liver cancer. It exists widely in peanut, corn and their products. It has been limited strictly in the national standards of China and the United States and the limited standards are respectively 20 ppb and 100 ppb in food grade and feed grade (GB 2761-2011; USDA, 2002). Aflatoxin is a metabolic product of Aspergillus and parasitic Aspergillus. It has the characteristic of fluorescence emission and aflatoxin B1, B2 emit blue fluorescence, aflatoxin G1, G2 green fluorescence under ultraviolet light. At present, aflatoxin is detected mainly by biochemical methods, including thin layer chromatography, high performance liquid chromatography, micro column method and enzyme linked immunosorbent assay (Bao et al., 2001). Although the detection accuracy is high, but the detection means and instruments are complex and the detection speed is slow. So these methods cannot be used in online detection.

In recent years, spectrum detection has become a new method for aflatoxin detection, which has drawn widespread attention (Teena et al., 2013). Hypersectral wave, multispectral images and hyperspectral images are three main kinds of tools for nondestructive exploration of agricultural product and food. The research team of Yao's mainly focuses on aflatoxin contamination in maize by hyperspectral data. They studied the spectral of single grain through artificial aflatoxin cultivation using the maximum likelihood estimation and binary encoding method based on hyperspectral image. And they found that the best fluorescence reflection peak of aflatoxin is around 437 and 537 nm, the recognition rate is 87–88%, and the fluorescence intensity decreased with the increase of the content of aflatoxin and there was a phenomenon of peak shift (Yao et al., 2013). And based on these two points, they applied for a US patent (Yao et al., 2010). But the device was still made manually instead of automatically. Ataş et al. (2012) studied the multispectral imaging analysis technique of aflatoxin contamination in the cayenne pepper bought from the market under two light sources (Halogen lamp and ultraviolet lamp) and proposed a feature selection method based on gray level histogram quantization and neural network weight optimization. The wavelength feature which plays a key role in detection was found and the recognition rate of leave-one-out method under the 420 nm reached 85%. In China, the research group of Wang Wei studied the aflatoxin detection problem using the hyperspectral images of the United State agricultural Research department (USDA, ARS) and pointed out that the aflatoxin detection rate of naturally contaminated grains was 87.5% through

* Corresponding author.
E-mail address: hanzhongzhi@qau.edu.cn (H. Zhongzhi).

https://doi.org/10.1016/j.compag.2018.08.018
Received 31 October 2017; Received in revised form 29 July 2018; Accepted 9 August 2018

common CCD imaging. They optimized five high spectral wavelengths from 700 nm to 1100 nm and the prediction accuracy rate of artificial aflatoxin contamination concentration in five highest spectral wavelengths reached 88.3% (Zhang et al., 2014).

The above literature shows that different instruments and algorithms will affect the detection effect. Which one is the best? In this paper, in order to find the accurate key wavelengths for aflatoxin detection, three kinds of datasets are collected from hyper-spectrometer, multispectral camera and hyperspectral camera separately. By analyzing the performance of four feature selection methods (Chu et al., 2014) and four classifier models (Samiappan et al., 2013), we attempt to propose a new weighted voting feature selection method to find the key feature wavelengths, and combine with the best classification model to design a new type of peanut aflatoxin sorter.

2. Materials

2.1. Materials

The peanuts were bought from a local market of Shandong, China in 2016. The variety of peanut is four red. From which, peanut kernels had been selected as samples which have a good appearance and the same size roughly (about 1 g). We divided these samples into 3 groups (45, 41 and 73 samples respectively) randomly for different experiments.

We prepared aflatoxin solution using aflatoxin B1 (form China national strain center) and acetonitrile (form market) according to 1:50. And then we dropped the aflatoxin solution onto the peanut surface using a pipette (one drop of the solution is about 1 μL) and injected 1–5 drops randomly on different samples. Different number of drops on one kernel will result in different contents of aflatoxin. If one drop, the kernel contains about 20 ppb, 2 drops means the kernel contains about 40 ppb, and so on. This means that aflatoxin content in one kernel is between 20 and 100 ppb (μg/kg) and sample kernels of each various levels is about1/5.

Here, one drop is 1 μL, the aflatoxin in the solution is 1/50, so pure aflatoxin in one drop is $1\,\mu L * (1/50) = 1 * 10-6L * (1/50) = 20 * 10-9L$. According to the specific gravity of solution is 1, the weight of aflatoxin is $20 * 10–9$ kg, One peanut kernel is about $1g = 1 * 10–3$ kg, and one drop on one means there is $(20 * 10–9\,kg)/(1 * 10–3\,kg) = (20 * 10–6\,kg)/(1 * 10–3\,kg) = 20$ mg/kg(ppb).

Different numbers of drops represent different levels of aflatoxin contamination. The purpose of dropping different drops is to investigate the generalization ability of the models used in this paper. We just divided the test results into two categories: contaminated or uncontaminated.

2.2. System integration

The instruments in our experiment include a field spectrometer (FieldSpec 3, ASD,USA), a Liquid crystal tunable filter (LCTF, VariSpec Cri, USA), a Grating spectrometer module (GSM, V8E, Dualix Spectal Imaging, China), a SCMOS CCD (ORCA-Flash2.8 C11440-10C, HAMAMATSU, JAP), a 365 nm ultraviolet LED (HRC-UV-4A, Shenzhen Weihai lixin technology development co. LTD, CHN) and a 355 nm laser (Triple frequency Nd:YAG laser, the eleventh Institute of China Electronics Technology Group) and an electric displacement platform (HG-32-TA, Beijing UH guano Dakota Co., LTD, CHN). Beside these, some optical lens, a Mercury lamp and a Teflon panel were also used in our experiments. Based on these instruments and materials, 3 experimental systems have been established.

The first system including the ASD spectrometer and the 365 nm LED illumination is illustrated in the black dotted box of Fig. 1(a). The spectral range of the spectrometer is 350–2500 nm and 1 nm spectral bandwidth. The spectral of each sample of 45 peanuts has been collected 10 times before and after aflatoxin contamination, and there are 900 spectra totally. It should be noted that the instrument needs whiteboard calibration before obtaining the data.

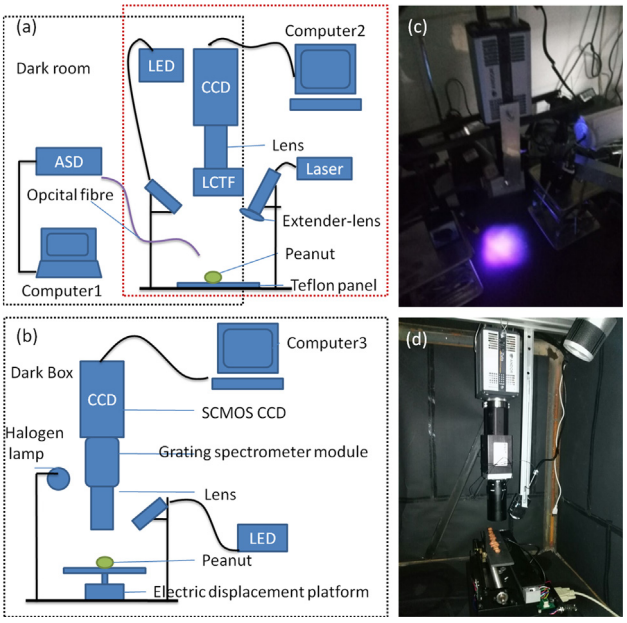

Fig. 1. The general overview figure and physical picture of 3 data collection systems.

The second hardware of the image acquisition system is composed of a SCMOS CCD camera and a Varispec liquid crystal tunable filter (LCTF), as illustrated in the red dotted box of Fig. 1(a). Using this system, multispectral images of 41 samples ranging from 400 nm to 720 nm were collected before and after aflatoxin contamination. The images of the fist 20 samples have been collected under 27 bands (400 405 410 415 420 425 430 435 440 445 450 460 470 480 490 500 520 540 560 580 600 620 640 660 680 700 720 nm) under UV 365 nm LED illumination, and the images of the last 21 sample were collected under 29 bands (above 27 band plus 530, 535 nm) and two illumination sources (UV365nm LED illumination and UV355nm laser sources).

For the convenience of the experiment, the first and second systems have been put together and the physical picture is illustrated in Fig. 1(c).

The third system contains a grating spectrometer module (292 nm-1200 nm, about 1 nm spectral bandwidth), a SCMOS CCD and a UV 365 nm LED illumination, as illustrated in Fig. 1(b) and Fig. 1(d). The electric displacement platform was used for moving the sample step by step to photo the picture one by one. An additional halogen lamp was used for illumination when we replace samples. In total, there were 73 peanut samples imaged before and after being contaminated.

3. Methods

3.1. Data preprocessing

Hyper-spectral wave data selected by system 1 is illustrated in Fig. 2(a), in which, each line represents the mean of one peanut of 10 times. The red line indicates peanuts contaminated, and the blue line indicates peanuts that are not contaminated.

The original multispectral images collected by system 2 not only include the image of peanut but also include the image of background. So, we should extract the Region of Interest (ROI) of peanut from background image firstly (Fig. 2(b)). Through this step, the original images ($2048 * 2048 * 27$ or 29 bands) were converted to the ROI images ($500 * 500 * 27$ or 29 bands). Fig. 2(c) illustrates four different band images of one sample before and after contamination. In the contaminated image of 435 nm, the area of aflatoxin can be seen faintly (position of arrow).