Computers and Electronics in Agriculture 99 (2013) 186–193

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Application of hyperspectral imaging technology to discriminate different geographical origins of *Jatropha curcas* L. seeds

Junfeng Gao [a], Xiaoli Li [a], Fengle Zhu [a], Yong He [a,b,*]

[a] College of Biosystems Engineering and Food Science, Zhejiang University, 866 Yuhangtang Road, Hangzhou 310058, China
[b] Key Laboratory of Equipment and Informatization in Environment Controlled Agriculture, Ministry of Agriculture, China

ARTICLE INFO

Article history:
Received 21 August 2013
Accepted 19 September 2013

Keywords:
Jatropha curcas L.
Hyperspectral imaging
Image processing analysis
Effective wavelength selection
Discrimination
Least square support vector machine (LS-SVM)

ABSTRACT

A rapid and non-invasive method was investigated to identify the geographical origin of *Jatropha curcas* L. seeds in China by using near-infrared hyperspectral imaging technique on the wavelength between 874 and 1734 nm. Two hundred and forty *J. curcas* L. seed samples from four different geographical origins (Jiangsu, Sichuan, Hainan and Taiwan) in China were studied and all of them were scanned by a push-broom hyperspectral imaging system. Then the obtained data sets were analyzed by spectral and image processing technique respectively. Successive projections algorithm (SPA) was used for selecting effective wavelengths. Dimension reduction was carried out on the region of interest (ROI) image by principal component analysis (PCA). The first principal component (PC) explained over 92% of variances of all spectral bands. Gray-level co-occurrence matrix (GLCM) analysis was implemented on the principal component (PC) image to extract 5 textural feature variables in total. Moreover, 7 morphological features of samples were computed additionally. Then least squares-support vector machine (LS-SVM) classification models were built based on the extracted spectral, textural, morphological, combined spectral and textural, combined spectral and morphological, combined textural and morphological, combined spectral, textural and morphological features, respectively. The satisfactory results show the correct discrimination rate of 93.75% for the prediction samples based on spectral and morphological features. The study demonstrated that hyperspectral image technique can be a reliable tool for discriminating different geographical origins of *J. curcas* L. seeds. The above results indicated that this objective and non-destructive method can be utilized for quality control purposes and seed breeding in future.

1. Introduction

Jatropha curcas is a tree or shrub belonging to the family of Euphorbiaceae and now has become an international research subject (Annarao et al., 2008). It is chosen as an ideal biodiesel crop in China as its seed kernel has high oil content about 43–61% (Cai et al., 2010). The extracted oil is non-edible due to the existence of toxic phorbol esters. This tree can resist drought and grow on barren and marginal lands without using arable land. Moreover, *J. curcas* L. seeds quality greatly affects their breeding, crop production and industrial processing. And the quality parameters affect the value of the oil, press cake and biodiesel as well as the production efficiency of the processes involved (Montes et al., 2013). The price and quality of *J. curcas* L. seed in China strongly depends on its geographic origin, on which climate and soil conditions are crucial influencing factors. People often evaluate whether seeds quality is good or not by identifying where seeds come from. Therefore, it is necessary to develop a convenient and fast identification technique for identifying the geographic origin of *J. curcas* L. seeds.

Near infrared reflectance spectroscopy (NIRS) which is based on the absorption of electromagnetic radiation in the 780–2526 nm wavelength range can provide comprehensive structural information on the components and properties of samples at the molecular level. Its proven that this region of spectral bands arises from overtones of C—H, C—O, O—H and N—H stretching vibrations (Cen and He, 2007). NIRS had been applied to quantify and discriminate the seed quality. For example, the experiment using NIRS to estimate the oil content, fatty acid composition and protein content of seeds was investigated (Vaknin et al., 2011). Some researchers (Montes et al., 2013) also studied grain quality determination by means of near infrared spectroscopy in bulk *J. curcas* L. samples, especially for shell quality traits and phorbol eater analysis. They all concluded that NIRS has a high potential for determination of grain quality parameters. However, shells and kernels of samples were separated in the above researches, which is a drawback for breeders who require non-invasive techniques to determine the qualities of seeds.

Machine vision is a promising non-destructive method for rapid and automatic identification in many research areas (Ahmed et al., 2012; Alvarenga et al., 2010; Brescia et al., 2007; Choudhary et al.,

* Corresponding author at: College of Biosystems Engineering and Food Science, Zhejiang University, 866 Yuhangtang Road, Hangzhou 310058, China. Tel./fax: +86 0571 88982143.

E-mail address: yhe@zju.edu.cn (Y. He).

0168-1699/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.compag.2013.09.011

2008; Courtois et al., 2010; Ercisli et al., 2012; Hu et al., 2012; Wu et al., 2012b; Yonemaru and Morita, 2012). Near infrared hyperspectral imaging is a fast, simple, and non-contact technique which combines spectroscopy and digital imaging simultaneously to acquire spectral and spatial information. A spectrum for each pixel and an image for each narrow band can be acquired which enable this system to reflect componential and constructional characteristics of an object and their spatial distributions. Hyperspectral imaging technique has been successfully applied in the discrimination and quantification of many agricultural products, such as the classification of fresh and frozen-thawed porcine longissimusdorsi muscles, fish fillet and tree species (Barbin et al., 2013; Zhu et al., 2012), determination of Enterobacteriaceae on chicken fillets (Feng et al., 2013) and measurement of color distribution in salmon fillet (Wu et al., 2012a), classification of oat and groat kernels and detection of Fusarium-damaged, yellow berry and vitreous Italian durum wheat kernels (Serranti et al., 2013a, b).

The discrimination of different geographic origins of agricultural products had been published. Lin et al. (2009) used Visible and Near-Infrared Spectroscopy technique to identify the geographical origin of olive oil. Hwang et al. (2012) discriminated the geographical origins of rice samples using transmission spectra with Enhanced Raman spectroscopic. The origins of pistachio samples were authenticated by Raffaele et al. with NIR spectroscopy and Chemometrics (Vitale et al., 2013). Besides, the texture features extracted using image processing analysis were utilized in some other works, such as characterization of different typical Italian breads (Brescia et al., 2007), sorting of tea categories (Li et al., 2011; Wu et al., 2008), fruits detection based on texture analysis and multiview geometry (Rakun et al., 2011) and classification among wheat varieties (Zapotoczny, 2011). Furthermore, morphological parameters based on a series of images processing analysis can be applied to the discrimination or characterization (Saad et al., 2011; González-Velasco et al., 2011; Alvarenga et al., 2010).

The purpose of this work was to develop a rapid, non-destructive and effective analytical procedure to identify the geographical origins of *J. curcas* L. seeds. This experiment was carried out directly on the seeds without any sample pre-treatment in order to solve problem in previous researches (Vaknin et al., 2011; Montes et al., 2013). Some effective spectral wavelengths were extracted by SPA algorithm, and the textural and morphological features were extracted from imaging processing analysis for highlighting differences and enabling discrimination. Lastly, we explored the optimal combined features to discriminate the geographical origin according to the correct discrimination rate of least squares-support vector machine (LS-SVM) classification models.

2. Materials and methods

2.1. Seed samples

The *J. curcas* L. seeds were collected from local farmers in Jiangsu, Sichuan, Hainan and Taiwan of China. All the seeds were sun dried and stored at room temperature in plastic bags before used as samples for the experiment. 60 samples from each origin were randomly picked and scanned by the NIR hyperspectral imaging system. Three replicate measurements were made on each seed sample to avoid errors. An example of samples from different origins is shown in Fig. 1.

2.2. Near infrared hyperspectral imaging system

2.2.1. Main components and configuration of the system

A line-scanning NIR hyperspectral imaging system was employed in this work for the acquisition of hyperspectral images of each seed sample. The main components of this system include a spectrograph (InSpector N17E, Specim, Finland), a Te-cooled InGaAs photodiode array sensor with pixel resolution of 12 bits (Xenics, Belgium), two 150 W tungsten halogen lamps (Fiber-Lite DC950 Illuminator, Dolan Jenner Industries Inc., USA) for illumination, a conveyer belt (Isuzu Optics Corp., Taiwan, China) driven by a stepping motor, and a computer with data acquisition and preprocessing software (Spectral Image-N17E, Isuzu Optics Corp., Taiwan, China). The spectral resolution is 5 nm in 874–1734 nm. The camera, fitting with a C-mount 22.5 mm OLES22 lens, has 320 × 256 (spatial × spectral) pixels. The schematic diagram of main components of the hyperspectral imaging system is illustrated in Fig. 2. The speed of the conveyer belt was adjusted to 19 mm/s to synchronize with the scanning of camera, whose exposure time was set to 5 ms, in order to achieve a square pixel. The illumination unit was fixed above the seed sample from both sides at an angle of 45° to reduce the shadowing effects, and the vertical distance between sample and lens was 230 mm. Each hyperspectral seed image consisted of 256 congruent images at 256 contiguous spectral bands. The dimension of the acquired hyperspectral image was 320 pixels in *x*-direction, *n*-pixels in *y*-direction and 256 bands in *z*-direction. The images were stored in raw format and exported to the Environment for Visualizing Images (ENVI) V4.6 software (ITT Visual Information Solutions, Boulder, USA) for following processing.

2.2.2. Image calibration and denoising

All raw hyperspectral seed images ($\boldsymbol{I}_{raw}$) were calibrated for reflectance to minimize differences among samples due to sensor response and illumination. The dark reference image ($\boldsymbol{I}_b$) of approximately 0% reflectance was acquired for removing the influence of dark current in the camera when the light source was turned off together with the camera lens completely covered with its opaque cap. The image of a white Teflon tile with about 100% reflectance was used as the white reference image ($\boldsymbol{I}_w$). The acquired images ($\boldsymbol{I}_{raw}$) from the hyperspectral imaging system were corrected by using the following equation:

$$\boldsymbol{I} = \frac{\boldsymbol{I}_{raw} - \boldsymbol{I}_b}{\boldsymbol{I}_w - \boldsymbol{I}_b} \times 100 \quad (1)$$

Because of the low signal-to-noise ratio in the first 15 and the last 12 bands of calibrated images, these bands were eliminated, and spectral region from band 15 to 244, corresponding to the wavelength range of 921–1693 nm, was employed. Minimum

Fig. 1. The four different geographical origins seed samples.