Computers and Electronics in Agriculture 122 (2016) 139–145

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Classification of maize seeds of different years based on hyperspectral imaging and model updating

Min Huang [a,*], Jinya Tang [a], Bao Yang [b], Qibing Zhu [a]

[a] Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi 214122, China
[b] Department of Electrical & Computer Engineering, Oakland University, USA

ARTICLE INFO

Article history:
Received 3 August 2015
Received in revised form 26 January 2016
Accepted 27 January 2016
Available online 9 February 2016

Keywords:
Classification
Seeds
Hyperspectral imaging
Model updating
Incremental support vector data description
Least squares support vector machine

ABSTRACT

Seed classification and identification exhibit potential for detecting seed purity and increasing crop yield. In this study, hyperspectral imaging was employed to develop classification methods for maize seeds. A total of 2000 seeds, including four varieties of maize seeds of different years, were evaluated. Hyperspectral reflectance images were acquired between 400 nm and 1000 nm. Classification models based on the mean spectral features of seeds were developed using least squares support vector machine (LSSVM). Model updating using incremental support vector data description was also applied to update the LSSVM model online and ensure accurate identification of maize seeds of different years. The classification accuracy of the LSSVM model combined with model updating reached 94.4% and was 10.3% higher than that of other non-updated models. This study showed that combined hyperspectral imaging and model updating could be an effective method for classification of seeds of different years.

1. Introduction

Seeds are the foundation of agricultural production. Severe degradation of seed purity because of the development of hybrids from different seed varieties leads to decreased crop yield. Research indicates that a 1% decrease in the purity of maize seeds can decrease the maize yield to 9 kg per 667 m^2 (Liu and Wang, 2000). Therefore, seed classification and identification play an important role in seed quality detection.

Over the past 20 years, researchers have developed several methods, such as morphology method, protein electrophoresis, and DNA molecular marker technology, for accurate seed identification. However, these methods are time consuming and require experienced operators and costly equipment (Mahesh et al., 2015). These methods are also inapplicable for single seed recognition. Therefore, scholars have focused on developing rapid, objective, and nondestructive detection techniques for seed classification and identification.

Machine vision, near infrared (NIR) spectroscopy, and hyperspectral imaging are potential nondestructive techniques for agricultural products. Machine vision evaluates image features, such as color, size, shape, and surface texture (Jog et al., 2011; Tsai et al., 2011), but is unsuitable for predicting the chemical composition of food because of limited spectral information (Xiong et al., 2015). NIR spectroscopy is used to obtain spectral information by using a single spot and is influenced by the uniformity of sample distribution (Reed et al., 2013; Ji et al., 2015). Hyperspectral imaging integrates machine vision and NIR technology into one system. This technique obtains both image and spectral information, as well as collects spectral information not only from a single point but at each pixel of an image, thereby overcoming the limitations of machine vision and NIR technology, respectively (Ferrari et al., 2015). Several studies have widely recognized the advantages of hyperspectral imaging in nondestructive tests for seed classification and identification (Mendoza et al., 2011; Huang et al., 2013; Rodríguez-Pulido et al., 2013; Wang et al., 2014; Ravikanth et al., 2015). However, in most applications, samples in training and test sets are obtained from the same years; hence, the accuracy of identifying seeds of different years cannot be guaranteed using hyperspectral information. In this regard, model updating can be applied to improve the generality of the developed model; the use of this technique can yield accurate identification of seeds of different years and reduce production cost.

Model updating has been extensively studied (Farrell et al., 2012; Wu et al., 2008; Xie and Ying, 2012; Wen et al., 2012). This technique is traditionally implemented based on prior knowledge of the spectral information of test samples. However, obtaining test

* Corresponding author at: School of Internet of Things, Jiangnan University, 1800 Lihu Avenue, Wuxi, Jiangsu Province 214122, China. Tel.: +86 510 85910635, mobile: +86 15861596626.

E-mail address: huangmzqb@163.com (M. Huang).

http://dx.doi.org/10.1016/j.compag.2016.01.029
0168-1699/

seeds in a batch is difficult. As such, an online model updating strategy, where classification model can be updated in an incremental learning manner, must be established in actual production.

This study proposes the combination of online model updating based on incremental support vector data description (ISVDD) algorithm with hyperspectral imaging for classification of maize seeds. This study specifically aims to: (1) extract the mean spectral features of maize seeds of different years, (2) establish a maize seed classification model using least squares support vector machine (LSSVM) based on the extracted mean spectral characteristics, and (3) update the classification model online by the ISVDD algorithm.

2. Materials and methods

2.1. Sample preparation

This study utilized 2000 maize seeds, including four varieties from the same classes (JIDAN7, JUNDAN18, JUNDAN20, and LUNDAN818) harvested in different years (Table 1). The seeds were provided by a commercial seed company in China (Fengle Seed Inc., Hefei, China) and stored in bottlenecks closed by paraffin wax in a refrigerator at 4 °C. Selection of representative training samples is important for developing a classification model but is difficult in practice; as such, model updating is generally applied to improve the generality of the classification model. To simulate the case and verify the model performance through model updating, this study employed a training set that consisted of the following varieties: JIDAN7 (collected in 2012), JUDAN18 (collected in 2007), JUDAN20 (collected in 2009), and LUDAN818 (collected in 2012). A test set was also established using each variety of maize seeds collected from three different years.

2.2. Hyperspectral image acquisition and correction

2.2.1. Hyperspectral imaging system

A hyperspectral imaging system in the visible/NIR range (400–1000 nm) was used to acquire images of maize seeds in reflectance mode (Fig. 1). The system consisted of a charge-coupled device (CCD) camera (Pixelfly QE IC/285AL, Cooke, USA), imaging spectrograph (1003A-10140 Hyperspec™ VNIR C-Series, Headwall Photonics Inc., Fitchburg, USA), zoom lens (10004A-21226 Lens, F/1.4 FL23 mm, Standard Barrel, C-Mount., USA), illumination unit with 150 W tungsten halogen lamps (Halogen lamp, EKE, 3250 K, Techniquip, USA) connected with a single optic fiber, translation stage, and computer equipped with data acquisition and control software (Hyperspectral Scanning and Image Rendering Software, Rev A.2.1.3, Headwall Photonics Inc., Fitchburg, USA). The CCD area detector of the camera comprised 1392 × 1024 pixels and an imaging spectrograph with an effective spectral range of 400–1000 nm and a 25 μm slit. The hyperspectral imaging system exhibited a spatial resolution of 0.15 mm/pixel and a default spectral resolution of 0.64 nm/pixel.

Table 1
Maize seeds including four varieties from different years.

Variety	Harvest year	Training set	Test set	Total
JTDAN7	2012	200	100	
	2013	0	100	
	2014	0	100	
JUNDAN18	2007	200	100	
	2011	0	100	
	2014	0	100	2000
JUNDAN20	2009	200	100	
	2010	0	100	
	2011	0	100	
LUDAN818	2012	200	100	
	2013	0	100	
	2014	0	100	

2.2.2. Hyperspectral image acquisition and correction

In a push-broom system, each seed was placed on an objective table and scanned at a constant speed line by line at 250 ms exposure time to create a hyperspectral image. A total of 750 scanning lines were obtained under 60 mm scan length and 80 μm step size. The hyperspectral images were obtained within the spectral range of 400–1000 nm with 6.38 nm interval among contiguous bands (total of 94 bands). Finally, a hypercube with a size of 1392 × 750 × 94 was generated, and the images were saved in a band-interleaved-by-line format. The entire acquisition process was conducted in a closed black box to weaken the interference of the external light source. An image should be corrected after four images were collected to minimize changes in light and system noise. The procedure was conducted based on the raw acquired images (R_o) and two reference images (R_b, R_w) by using the following equation:

$$R_c = \frac{R_o - R_b}{R_w - R_b} \quad (1)$$

where R_c and R_o are the corrected and raw spectral images of the sample, respectively; R_b is the dark reference image acquired by completely blocking the lens with an opaque cap; and R_w is the white reference image obtained from 99% diffuse reflectance white standard.

2.3. Image segmentation and feature extraction

Selection of region of interest (ROI) in hyperspectral images is crucial because it significantly affects the extraction of spectral features. Image segmentation plays a key role in ROI selection. Threshold segmentation is one of the most widely applied methods in engineering practices because of its simplicity and high efficiency (Karasulu and Korukoglu, 2011; Guo et al., 2014; Liu et al., 2015). In the present study, the reflectance image of maize seeds at 782.59 nm was selected because the seeds exhibited the clearest outline in this image (step I). Prior to image segmentation, the obtained gray image was subjected to pre-processing operations, such as image filtering and enhancement, to remove noise (step II). A threshold based on the gray image was then established to build a binary image (step III). The selected ROI is presented in red lines in Fig. 2. Finally, the mean spectral features of the 2000 maize seeds were extracted in the ROI of 94 bands to characterize the seeds (step IV). Schematic of image segmentation and feature extraction is shown in Fig. 2. In this study, all programs were implemented in MATLAB 2009b (MathWorks, Inc., USA) with LSSVM toolboxes.

2.4. Classification model

Support vector machine (SVM) is an important classification method for pattern recognition and machine learning (Vapnik, 1998). In SVM, data input space is mapped into a high-dimensional feature space through a kernel function by using minimal training data. To make SVM suitable for engineering applications, researchers developed LSSVM by applying least squares errors in the training error function and solving the problem by transforming quadratic programming (QP) into a linear equation; this modification simplifies the training process of SVM and reduces computation time and cost (Su et al., 2015). LSSVM is widely used for nondestructive detection in agricultural production (Wu et al., 2008; Khanmohammadi et al., 2014). In the present