Computers and Electronics in Agriculture 109 (2014) 23–31

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Hyperspectral band selection for detecting different blueberry fruit maturity stages

Ce Yang [a,1], Won Suk Lee [b,*], Paul Gader [c,2]

[a] Bioproducts and Biosystems Engineering Department, University of Minnesota, St. Paul, MN 55108, United States
[b] Agricultural and Biological Engineering Department, University of Florida, Gainesville, FL 32611, United States
[c] Computer Information Science and Engineering Department, University of Florida, Gainesville, FL 32611, United States

ARTICLE INFO

Article history:
Received 10 March 2014
Received in revised form 19 August 2014
Accepted 20 August 2014

Keywords:
Band selection
Blueberry
Hyperspectral imagery
Kullback–Leibler divergence
Precision agriculture
Yield mapping

ABSTRACT

Hyperspectral imagery divides spectrum into many bands with very narrow bandwidth. It is more capable to detect or classify objects, where visible information is not sufficient for the task. However, hyperspectral image contains a large amount of redundant information, which eliminates its discriminability. Band selection is used to both reduce the dimensionality of hyperspectral images and save useful bands for further application. This study explores the feasibility of hyperspectral imaging for the task of classifying blueberry fruit growth stages and background. Three information theory based band selection methods using Kullback–Leibler divergence: pair-wise class discriminability, hierarchical dimensionality reduction and non-Gaussianity measures were applied. Three classifiers, *K*-nearest neighbor, support vector machine and AdaBoost were used to test the performance of the selected bands by the three methods. The selected bands achieved classification accuracies of 88% and higher. Therefore, the band selection methods are very useful in reducing the volume of the hyperspectral data, and constructing a multispectral imaging system for detecting blueberry fruit maturity stages.

1. Introduction

Labor expense of handpicked blueberries for fresh markets is increasing due to severe shortages of available farm workers. Management cost of Florida's commercial blueberry field excluding harvesting labor is approximately $9884 ha (Williamson et al., 2012). The average blueberry yield in Florida is 6310 kg/ha (USDA, 2012). Morgan et al. (2011) estimated that the hand harvest cost is $1.59 kg. Therefore, the cost of harvesting labor takes higher than $10,000 ha, which is more than half of the total management cost of the blueberry production. Efficient harvesting labor assignment in a large blueberry field can reduce much of the harvesting cost. Furthermore, yield estimation prior to harvest helps growers to find problems in their fields as early as possible. It is useful for growers to make further decisions such as irrigation, pest control, and weed control. Therefore, yield estimation of blueberry field prior to harvesting is beneficial for the growers. During the harvest season, individual blueberries in a fruit cluster usually mature at different times. A cluster may contain all growth stages including young fruit (green color), intermediate fruit (red color) and mature fruit (dark blue/purple) at the same time. Fig. 1 is an example picture taken from a blueberry field during the blueberry harvest season in 2013.

Efficient labor deployment based on yield monitoring requires that the yield be estimated in advance of berry ripening. Remote sensing is a method of detecting objects without physically touching or breaking them. Therefore, it is logical to use remote sensing for the yield estimation of fruit amount of different growth stages. Wild blueberry fruit estimation was carried out by digital image processing (Zaman et al., 2008) and high prediction accuracy was obtained. The color images of wild blueberry in the study contained only mature fruit, which was easily distinguishable because of its significant color contrast in the blue band. However, as shown in Fig. 1, a southern highbush blueberry cluster has all growth stages at the same time. It is difficult to distinguish young fruits and intermediate fruits from the noisy background in the visible range. To estimate the blueberry yield in advance of harvesting, all growth stages should be detected so that all fruits on the bushes are considered. Hyperspectral imaging has been used in detecting fruit and vegetable quality such as maturity, firmness, starch content, soluble solid contents for over a decade. Lu and Peng (2006) investigated peach fruit firmness using hyperspectral

* Corresponding author. Tel.: +1 352 392 1864x227.
E-mail addresses: ceyang@umn.edu (C. Yang), wslee@ufl.edu (W.S. Lee), pgader@ufl.edu (P. Gader).
[1] Tel.: +1 612 626 6419.
[2] Tel.: +1 352 392 1527.

http://dx.doi.org/10.1016/j.compag.2014.08.009
0168-1699/

Fig. 1. A blueberry fruit bunch that contains all three growth stages: young, intermediate and mature.

scattering. They selected 10 or 11 wavelengths with r^2 of 0.77 and 0.58 for two peach cultivars. Nagata et al. (2004) estimated strawberry maturity by measuring the soluble solids content of 'Akihime' strawberries and had a correlation coefficient of 0.784 using five-predictor firmness model. However, all the five predictors they chose are in the visible range. Rajkumar et al. (2012) studied banana maturity stages at different temperatures using hyperspectral imaging and obtained coefficient of determinant of 0.85, 0.87 and 0.91 for total soluble solids, moisture and firmness, respectively. There are also studies on maturity estimation of mango, peanut pod, etc. (Sivakumar et al., 2011; Carley, 2006). However, all of these research were carried out as a post-harvest step in the packing house or lab, which have more ideal condition compared to the outdoor environments. The samples for these experiments were placed in the lighting house and images were taken with even illumination, and without wind or shadow factors. These methods are not applicable for on-site early crop maturity detection.

In a previous study, blueberry spectral property was analyzed based on laboratory measured spectral data by Yang et al. (2012). The analysis showed that hyperspectral property would be helpful in classifying different growth stages of blueberry fruit. While blueberry spectral properties have been analyzed in a laboratory, it cannot be coupled with field measurement directly because of their different measurement conditions. The laboratory is a more ideal environment because of its stable indoor light source. In addition, the samples were well prepared without much noisy background. However, field measurement uses the sunlight as its illumination source, and the background contains not only leaves, but also soil, sky, and man-made objects such as PVC irrigation pipes. A portable spectrometer can only measure either a spot or a small area as one spectrum, which cannot provide sufficient information. A color image is not easy to detect all the fruit maturity stages because of the similar colors of young fruit and leaves. On the other hand, hyperspectral images obtained from field conditions have both high spatial and spectral resolution. Therefore, hyperspectral imagery can be used for the detection of blueberry of different growth stages in the field with complicated background objects.

Due to the high spectral resolution, hyperspectral images contain considerable amount of redundancy. The images usually have several hundred bands, but some bands are useless or even hinder the discriminability of useful bands. Adjacent bands in the spectrum tend to be highly correlated (Cai et al., 2007). Band extraction methods such as principal component analysis (PCA), and maximum noise fraction (MNF) reduce dimensionality by projecting the original bands into new dimensions. However, the projected features combine the original information in these methods and do not have physical meaning. In contrast, band selection methods choose original features, which have physical information. Some selected original bands can be used for yield estimation using a multispectral camera system. A multispectral camera is of lower cost and higher processing speed compared to a hyperspectral camera system. Therefore, a multispectral imaging system with selected bands is more suitable for the task of blueberry yield prediction.

During the last decade, many band selection methods have been developed as preprocessing of hyperspectral image analysis. Some methods used different criteria to measure the importance of bands. The separability of bands may be measured with transformed divergence, Bhattacharyya distance, and Jeffries-Matusita distance (Yang et al., 2011). Other methods employed a criterion to prioritize bands, and then bands with the highest rankings in dissimilar band clusters are selected. The band ranking criterion contains variance, correlation, signal-to-noise ratio (SNR), etc. Information measures have also been used for hyperspectral band selection using mutual information or information divergence (Martinez-Uso et al., 2007). However, the purpose of these band selection methods was to reduce data volume and calculation complexity. They did not focus on what specific selected bands were.

The objectives of this study were to explore the feasibility of hyperspectral imagery in classifying different blueberry growth stages, and to select useful bands that are suitable for a multispectral imaging system, which is of lower cost and higher processing speed. The selected bands are supposed to yield a high accuracy of classification. A supervised band selection method based on the Kullback–Leibler divergence (KLD) was proposed, which measures pair-wise discriminability of spectral bands. The proposed method was compared with two other band selection methods: hierarchical dimensionality reduction and non-Gaussianity measures. The band selection methods were tested by *K*-nearest neighbor (KNN), support vector machine (SVM) (Martinez-Uso et al., 2007; Chang and Wang, 2006) and AdaBoost (Freund and Schapire, 1995) by their performance in classifying the blueberry growth stages and background.

2. Materials and methods

2.1. Hyperspectral image acquisition

Hyperspectral images were obtained from a blueberry research and demonstration farm at the University of Georgia cooperative extension in Alma, GA, United States (31.534°N, 82.510°W) in July, 2012. There were ten rows with 20 trees per row. In each row, four trees were randomly selected for hyperspectral image acquisition. Therefore, a total of 40 images were obtained. In each image, an area of $15.2 \times 15.2\ cm^2$ of the view was acquired. A hyperspectral imaging system was used for image acquisition, consisting of a line scanning spectrometer (V10E, Specim, Oulu, Finland), a digital CCD camera (MV-D1312, Photonfocus AG, Lachen SZ, Switzerland), a lens (CNG 1.8/4.8-1302, Schneider Optics, North Hollywood, CA, USA), an encoder (Omron-E6B2, Omron Cooperation, Kyoto, Japan), a tilting head (PT785S, ServoCity, Winfield, KS, USA), an image grabber (NI-PCIe 6430, National Instruments Inc., Austin, TX, USA), a data acquisition card (NI-6036E, National Instruments Inc. Austin, TX, USA), and a laptop (DELL Latitude E6500) with a control and vision acquisition program written in LabVIEW (National Instruments Corporation, Austin, TA, USA). The tilting head carried the camera to rotate vertically. When the camera rotated, the encoder generated pulses, which was sent to the program for generating a trigger signal. The camera acquired one line