2950 Niles Road, St. Joseph, MI 49085-9659, USA
269.429.0300 fax 269.429.3852 hq@asabe.org www.asabe.org

An ASABE Meeting Presentation

Paper Number: 131593276

Band Selection of Hyperspectral Images for detecting Blueberry Fruit with Different Growth Stages

Ce Yang, Ph.D. Candidate

Department of Agricultural and Biological Engineering, University of Florida, Rogers Hall 143, Museum Road, Gainesville, FL 32611 U.S.A. Email: ycicj@ufl.edu

Won Suk Lee, Associate Professor

Department of Agricultural and Biological Engineering, University of Florida, Rogers Hall 227, Museum Road, Gainesville, FL 32611 U.S.A. Email: wslee@ufl.edu

Paul Gader, Professor

Department of Computer & Information Science & Engineering, University of Florida, E301A CSE Building, Gainesville, FL 32611 U.S.A. Email: pgader@cise.ufl.edu

Written for presentation at the

2013 ASABE Annual International Meeting

Sponsored by ASABE

Kansas City, Missouri

July 21 – 24, 2013

Abstract. *Hyperspectral imagery deals with large volumes of data due to hundreds of spectral bands used in the images. Although hyperspectral images with higher spectral resolution usually carry more information, the processing of the images requires a significant amount of memory and is usually very slow. In addition, hyperspectral images contain considerable amount of redundant information, which does not help or even hinder the algorithm in making the correct decision. Band selection is an approach to both reduce the dimensionality of hyperspectral images and save calculation time for further applications, such as detection of blueberry fruit with different maturity stages. Hyperspectral images of blueberry fruit were taken in a commercial blueberry field. Mature fruit, intermediate fruit, young fruit and background were the four classes to be studied. A supervised band selection method was proposed using Kullback-Leibler divergence (KLD). Wider bands were made by combining 20 hyperspectral bands so that the selected bands could be used in a blueberry yield mapping system using a lower-cost multispectral camera. Based on the analysis, six combined bands were selected:*

543.1 – 572.6 nm, 627.4 – 658.8 nm, 663.6 – 695.2 nm, 725.4 – 757.4 nm, 773.5 – 805.6 nm and 838 – 870.5 nm. The test result showed that the proposed band selection method worked well for the task of blueberry growth stages detection.

Keywords. Band selection, blueberry, Kullback-Leibler divergence, precision agriculture, yield mapping

Introduction

Florida is ideal for producing early-season blueberries because of its warm weather across the year. Berries from Florida mainly supply the fresh markets from early April to late May. Therefore, all Florida blueberries are hand-picked by manual labor. Insufficiency and high cost of labor are the major concerns. Therefore, efficient assignment of harvesting labor based on spatial variability of yield in a large-scale blueberry field is of great necessity for Florida blueberry growers.

In order to estimate blueberry yield, different stages of fruit growth need to be identified and the fruit amount of each growth stage should be estimated. Remote sensing techniques do not physically contact the objects, which is suitable for estimating blueberry yield of different growth stages. It is easy to distinguish mature blueberries in regular color images because of its dark-blue color. Zaman et al. (2008) estimated the wild blueberry yield by simply collecting the blue pixels in color images. However, color images do not show much difference for young fruit and leaf, especially when leaves are in the well-illuminated condition.

Hyperspectral imagery is useful for distinguishing visually similar materials based on their different spectral properties. However, they contain a large amount of redundancy bands, which is not helpful or even hinder the performance of useful bands. Yang et al. (2012) analyzed the spectral signatures of blueberry fruit and leaves in a laboratory. However, they selected useful bands based only on laboratory measurements, which could be very different from outdoor conditions.

Many scholars worked on selecting wavebands from hyperspectral images during the last few years (Yang and Lee, 2011; Zare and Gader, et al. 2008; Martinez-Uso, et al. 2007; Sotoca and Pla, et al. 2006; Chang and Wang, et al. 2006). However, they focused on reducing to a limited number of bands to reduce the calculating complexity. In addition, most of the methods were unsupervised and required one whole image as input.

The objective of this paper was to select important bands using Kullback-Leibler divergence based on a hyperspectral blueberry image set collected from the field. The band selection result would be used in the development of a blueberry yield mapping system using a multispectral camera, which is cheaper, faster, and thus will work in real-time.

Materials and Methods

Hyperspectral imaging system

The hyperspectral imaging system contained a digital CCD camera (MV-D1312, Photonfocus AG, Lachen SZ, Switzerland), a line scanning spectrometer (V10E, Specim, Oulu, Finland), a wide-angle lens (CNG 1.8/4.8-1302, Schneider Optics, North Hollywood, CA, USA), an image grabber (NI-PCIe 6430, National Instruments Inc. Austin, TX, USA), a DAQ card (NI-6036E, National Instruments Inc. Austin, TX, USA), an encoder (Omron-E6B2, OMRON cooperation, Kyoto, Japan), a tilting head (PT785S, ServoCity, Winfield, KS, USA) and a laptop (DELL Latitude E6500). Image acquisition program was written in LabVIEW (National Instruments Corporation, Austin, TA, USA). The tilting head carries the camera body to move vertically in the range of view. The encoder is mounted on top of the tilting head to measure the tilting speed. A trigger signal is generated by the LabVIEW program when it counts the pulses from the encoder. The camera obtains one line image once it receives a trigger signal.

Figure 1 is the schematic diagram of the hyperspectral imaging system. Figure 2 shows the hyperspectral imaging system used in a blueberry field.

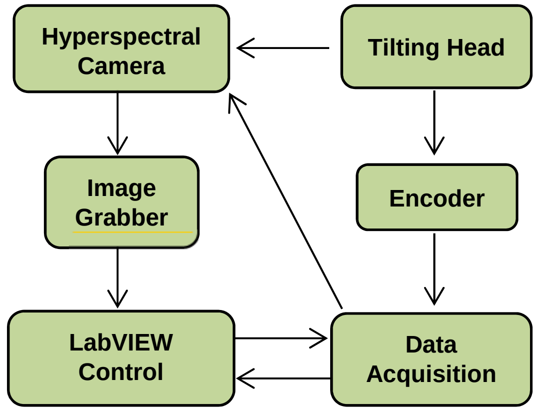

Figure 1. Schematic diagram of the hyperspectral camera system.

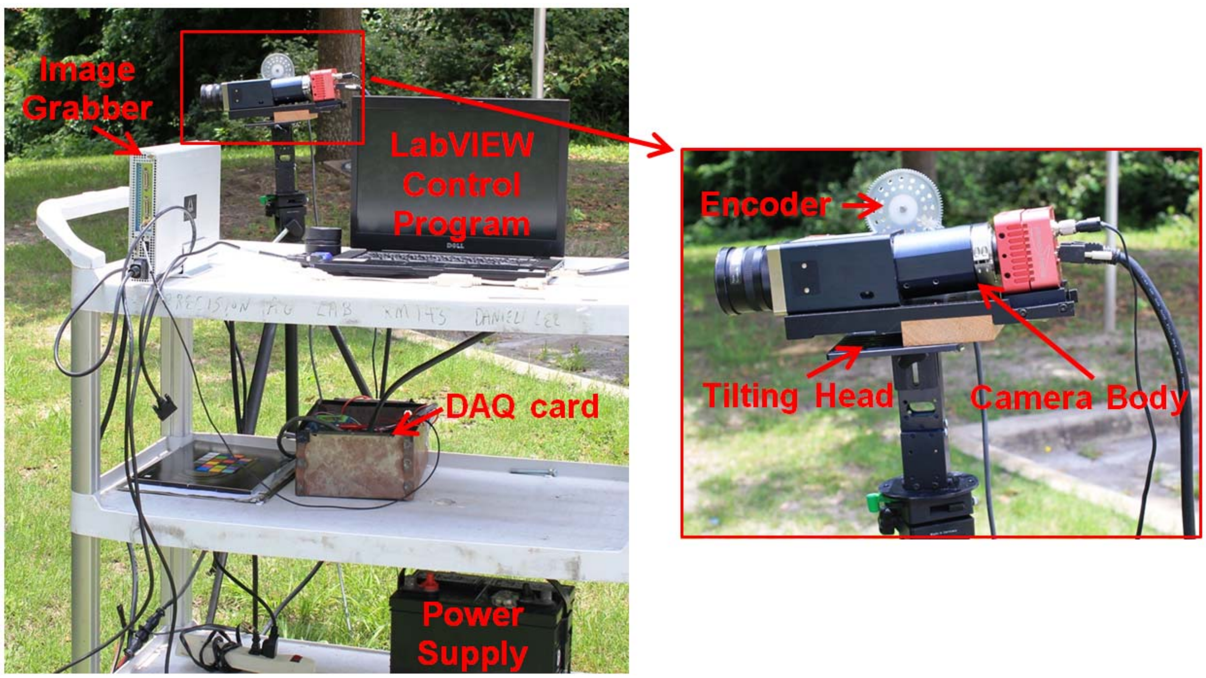

Figure 2. Hyperspectral camera system in the blueberry field.

Hyperspectral image acquisition

Hyperspectral images were obtained from a commercial blueberry field in Waldo, FL, United States in July 2012. The research area has ten rows with approximately 20 trees per row. Four sample images were taken from each row. Every image corresponded to an actual area of 100 cm^2. Forty hyperspectral images were acquired. The spatial resolution is 1 mm. The spectral range of the images is 398 - 1010 nm. After binning, the spectral resolution is 1.59 nm. The line-scan image size is 388 (spectral) × 1312 (spatial). The line scan data was saved in 16-bit binary files and processed later to create image cubes containing spatial and spectral data. Reflectance was obtained by using a universal white standard (Spectralon, Labsphere Inc., North Sutton, NH, USA) for all images. The RGB channels of a hyperspectral image are shown in Figure 3 (a) with blue: 440 nm, green: 550 nm and red: 650 nm. Figure 3 (b) is the corresponding digital camera image. There are four classes in the image: mature fruit, intermediate fruit, young fruit and background. Background class mainly consists of leaf. Branch, soil, sky and all other man-made objects (polyvinyl chloride pipes, ribbons, etc.) in the view are also considered as background. In order to make the selected bands suitable for working in a yield mapping system using a multispectral camera, every 20 neighboring bands were combined into one wider band.

(a) (b)

Figure 3. (a) RGB channels of an image taken by the hyperspectral imaging system; (b) Corresponding color image taken by a digital camera.

Band selection

MATLAB R2012a (The mathematic software MATLAB version 7.5.0 (The MathWorks Inc., Natick, MA, USA) was used to carry out the band selection in this study. The KLD-based method calculates the variation of classes at each combined band. Normalized histograms of every class and combined band were plotted. The symmetric KLD of the normalized histograms of every two classes with every combined band were calculated based on the following equation:

$$D_{KL}(x,y) = \sum_{i=1}^{N} P_i(x)\log\frac{P_i(x)}{P_i(y)} + \sum_{i=1}^{N} P_i(y)\log\frac{P_i(y)}{P_i(x)} \quad (1)$$

where: (x, y) is a class pair and $D_{KL}(x,y)$ is the KLD of (x, y) in one combined band. $P_i(x)$ and $P_i(y)$ are the probability distributions simulated by the normalized histogram of class x and y. The histograms were discretized, and N is the number of bins for discretization. With a larger difference between two classes in a specific combined band, $P_i(x)$ and $P_i(y)$ would be very different from each other and thus have higher KLD value. Therefore, the combined band is considered to be more useful than those with lower KLD values.

Six hundred pixels were collected from the image set as the training and validation data. The pixels were labeled manually since the spatial resolution of the images is high. The pixel set contains equal number of pixels of mature fruit, intermediate fruit, young fruit and background classes. Leaf pixels took the largest part of the background pixels. The pixels in every class were randomly divided into training and validation sets with equal size.

After the band selection, K-Nearest neighbor classifier (KNN) was used for classifying pixels of mature fruit, intermediate fruit, young fruit and background. In KNN, if the instance's K nearest neighbors contains mostly the training samples from one class than any other classes, the instance is classified into this class.

Results and discussion

Blueberry spectra

The four classes have significantly different spectra, which are shown in Figure 4. The spectra of a leaf pixel were used to represent the background spectra. Mature fruit have lower reflectance across the whole spectral range. Intermediate berries have higher reflectance than mature fruit in the red band and the near-infrared (NIR) range. Young berries have higher reflectance in the green band. The leaves have the highest reflectance in the NIR range mostly because of its high content of chlorophyll.

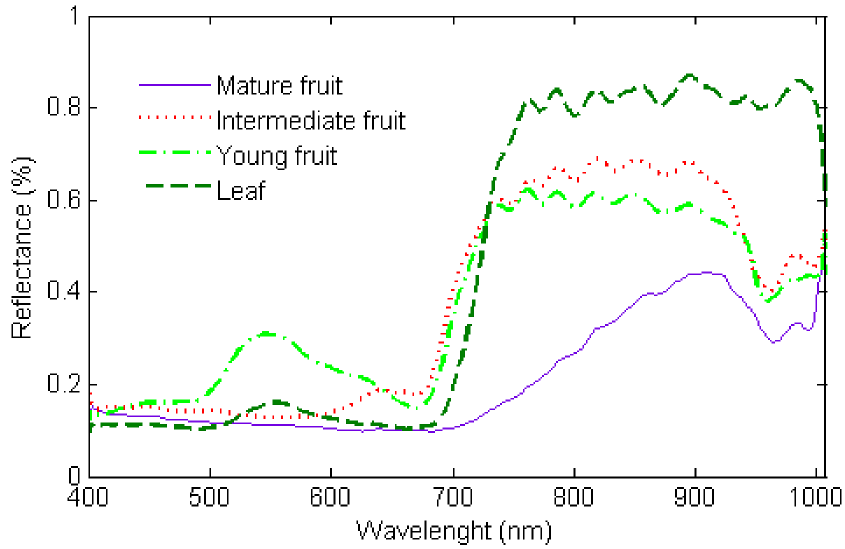

Figure 4. Spectra of the blueberry fruit growth stages and leaf as background.

Band selection result

The training pixel set was used for calculating the variation of probability distribution, which was represented by normalized histograms. Six bands were selected based on KLD for separating every two classes. Band 201 to 220 was selected for separating mature fruit and intermediate fruit pixels. Mature fruit and young fruit pixels can be separated the best by band 91 to 110. Band 160 to 179 separates mature fruit and background pixels. Band 68 to 87 performs the best for separating intermediate fruit and young fruit pixels. Band 130 to 149 separates intermediate fruit and background pixels. Band 12 to 31 is the most suitable for separating the young fruit and background pixels. Table 1 shows the corresponding wavelengths for the selected combined bands. Three selected bands are from the visible range and the other three combined bands are from the NIR range. This result shows that the spectral properties of the classes in the near-infrared range are useful for the separation of the blueberry growth stage and background classes. The performance of the combined bands is shown in Figure 5.

Table 1. Selected bands for classifying blueberry growth stages and background.

Combined band number	Wavelength range (nm)	Separation
12 - 31	543.1 – 572.6	Young fruit versus background
68 - 87	627.4 – 658.8	Intermediate fruit versus young fruit
91 - 110	663.6 – 695.2	Mature fruit versus. young fruit
130 - 149	725.4 – 757.4	Intermediate fruit versus background
160 - 179	773.5 – 805.6	Mature fruit versus background
201 - 220	838 – 870.5	Mature fruit versus intermediate fruit

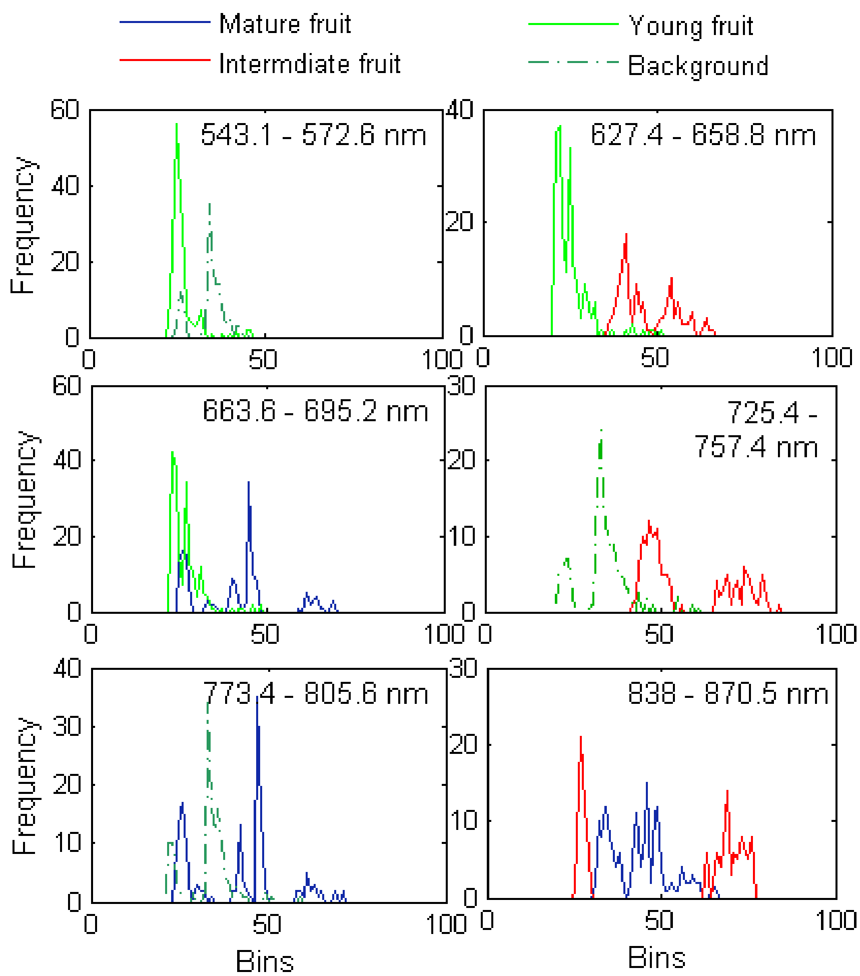

Figure 5. Separation ability of the combined bands.

Classification using selected bands

KNN classifier with different number of neighbors and distance metrics were used for testing the selected bands. Figure 6 shows the prediction accuracies of three distance metrics: Euclidean, Mahalanobis and Cosine. The classifier achieved the highest prediction accuracy when K=1 using Mahalanobis and K = 2 using Euclidean. The prediction accuracy is over 98%, which proves that these selected bands worked well for distinguishing the blueberry growth stages out of the background. The performance went down when using more than two neighbors. The main reason is that this preliminary study used only 600 pixels, while KNN achieves the best performance when there are many instances. The prediction accuracy decreased much slower when using the Cosine metric than using the other distance metrics. Therefore, it is of high possibility that the Cosine metric will perform better than the other metrics when there are many more pixels for training and testing.

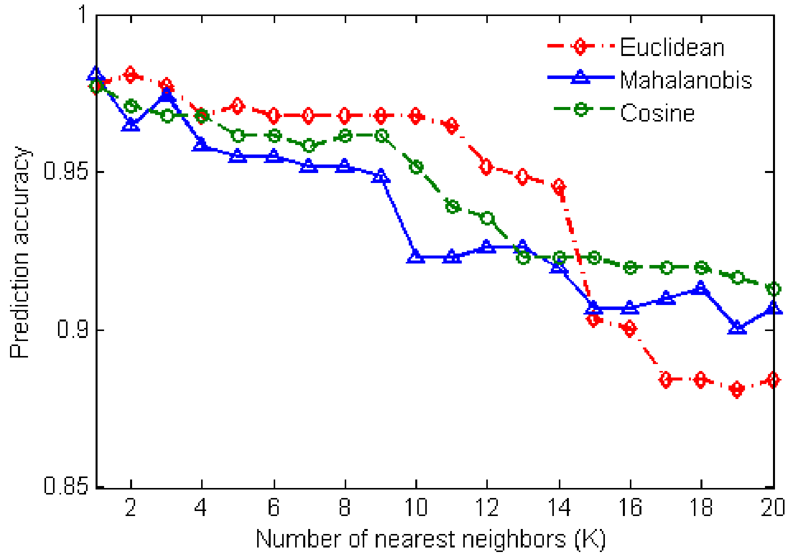

Figure 6. Performance of KNN classifier with different distance metrics and number of neighbors.

Conclusion

Blueberry hyperspectral images were taken in the outdoor conditions. Four classes: mature fruit, intermediate fruit, young fruit and background were taken into consideration for blueberry fruit detection. A KLD-based supervised band selection method was proposed. Hyperspectral bands were combined to be wider range so that these bands can be used in a blueberry yield mapping system using a lower-cost multispectral camera. KL divergence was calculated for every combined band for separating two classes. Six combined bands were selected: 543.1 – 572.6 nm, 627.4 – 658.8 nm, 663.6 – 695.2 nm, 725.4 – 757.4 nm, 773.5 – 805.6 nm and 838 – 870.5 nm. KNN was used for the construction of a classification model for distinguishing the fruit growth stages and the background. The result showed that the proposed band selection method worked well for the task of blueberry growth stages detection.

Reference

Chang C.-I., and Wang S., “Constrained band selection for hyperspectral imagery,” IEEE T Geosci. Remote: 44 (6), 1575–1585, 2006.

Martinez-Uso A., Pla F., Sotoca J.M. & Garcia-Sevilla P., “Clustering-based hyperspectral band selection using information measures,” IEEE T Geosci. Remote: 45(12), 4158–4171, 2007.

Sotoca J.M. and Pla F., “Hyperspectral data selection from mutual information between image bands,” Proc. SSPR & SPR: 853-861, 2006.

Yang C., Lee W.S. & Williamson J., “Classification of Blueberry Fruit and Leaves based on spectral signatures,” Biosys. Engi.: 113(4), 351–362, 2012

Yang H., Du Q., Su H., & Sheng Y., “An Efficient Method for Supervised Hyperspectral Band Selection,” IEEE T Geosci. Remote: 8(1), 138-142. 2011

Zaman Q.U., Schumann A.W., Percival D.C. & Gordon J., “Estimation of wild blueberry fruit yield using digital color photography,” Trans. of the ASABE: 51(5), 1539–1544, 2008.

Zare A. and Gader P., “Hyperspectral Band Selection and Endmember Detection Using Sparsity Promoting Priors,” IEEE T Geosci. Remote: 5(2), 256–260, 2008