Computers and Electronics in Agriculture 106 (2014) 91–101

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Identifying blueberry fruit of different growth stages using natural outdoor color images

Han Li [a,b], Won Suk Lee [b,*], Ku Wang [a]

[a] Department of Electrical Engineering, China Agricultural University, Beijing 100083, China
[b] Department of Agricultural & Biological Engineering, University of Florida, Gainesville, FL 32611, United States

ARTICLE INFO

Article history:
Received 5 April 2013
Received in revised form 18 April 2014
Accepted 30 May 2014

Keywords:
Image classification
K-means
K-nearest neighbor
Naïve Bayesian classifier
Precision agriculture
Yield mapping

ABSTRACT

This study was conducted to identify blueberry fruit of different growth stages using natural outdoor images toward the development of a blueberry yield mapping system. As blueberries usually contain different maturity stages in a same branch, identification of blueberry fruit and their maturity stages from different background is very important for yield mapping. In this study, maturity stages of the fruit were divided into four categories: mature (m), near-mature (nm), near-young (ny) and young (y). A stepwised algorithm, termed 'color component analysis based detection (CCAD)' method, was developed and validated to identify blueberry fruit using outdoor color images. Firstly, a dataset was built using manually cropped pixels from training images. Three color components, red (R), blue (B) and hue (H), were selected using the forward feature selection algorithm (FFSA), and used to separate all fruit of four maturity stages from background through different classifiers. In this work, not only the traditional classifiers such as K-nearest neighbor (KNN), and naïve Bayesian classification (NBC) were used, but another newly introduced 'supervised K-means clustering classifier (SK-means)' was also developed and applied to the dataset. In the second step, classifiers were built to separate a group of 'mature & near-mature' fruit from a group of 'near-young & young' fruit from all fruit pixels. Finally, classifiers were developed to separate mature fruit from near-mature fruit, and near-young fruit from young fruit. The classifiers obtained from these different steps were then applied to validation images, resulting in final identification. Cross validation was conducted using these different classifiers and their results were compared. KNN classifier yielded the highest classification accuracy (85–98%) from the validation set of the prebuilt pixel dataset collected from the training images in all separations. An one-way ANOVA was used to compare the performance of the three classifies, which shows KNN performed significantly better than other methods. The newly proposed 'SK-means' classifier yielded a fairly high accuracy (90%) for the separation of mature and near-mature fruit. The newly developed 'CCAD' method for blueberry was proved to be efficient for identifying blueberry fruit of different growth stages using natural outdoor color images toward the development of a blueberry yield mapping system.

1. Introduction

Highbush blueberry is a good source of fiber, and contains antioxidants, which makes it an excellent choice for fresh market fruit production all over the world (U.S. Highbush Blueberry Council, 2012). In Florida, USA, southern highbush blueberry acreage and production have increased rapidly. Compared with the 9 million pounds in 2008, 19 million pounds of blueberry has been harvested in 2012 (Brazelton, 2013). Since fresh Florida blueberries are mostly hand-harvested, the primary production cost is harvesting labor which can exceed 50% of the total picking, grading, and packing costs. The production window of blueberry in Florida is from about April 1 until May 15, which is relatively short. The prices usually drop fast as berries enter the market from northern regions (Yang et al., 2012). Since blueberry fruit in a same branch usually do not ripen at the same time before harvesting season, it is important for farmers to estimate the quantity of blueberry fruit on the bushes at different stages of their growth, so that they can make proper arrangement for harvesting labor and its distribution to specific locations in their fields. Also early yield estimation can be used to provide feedback on how crops respond to certain soil and crop management practices and to determine recommendation rates for many crop production inputs (Arslan and Colvin, 2002).

* Corresponding author. Tel.: +1 352 392 1864; fax: +1 352 392 4092.
E-mail address: wslee@ufl.edu (W.S. Lee).

http://dx.doi.org/10.1016/j.compag.2014.05.015
0168-1699/

Automated computer vision system is one of the most commonly used methods for yield estimation of various crops, such as apple and citrus. Several studies have focused on machine vision-based, non-destructive fruit yield estimation and mapping. For apple detection, Alchanatis et al. (2007) developed a method for automatically detecting apples using hyperspectral imaging, and the overall correct detection rate was 87% with an error rate of 15%. Aggelopoulou et al. (2011) suggested that apple flowering distribution maps could be used for yield mapping, and modeled correlation between flower density as determined from image analysis and fruit yield. Their results indicated that potential yield could be predicted early in the season from flowering distribution maps and could be used for orchard management during the growing season. Zhou et al. (2012) proposed an apple recognition algorithm with color differences such as red minus blue (R−B), and green minus red (G−R). The coefficients of determination (R^2) between apples detected by the fruit counting algorithm and actual harvested yield ranged from 0.57 for young fruit to 0.70 for fruits in a fruit ripening period.

For citrus fruit detection, Annamalai et al. (2004) developed a color vision system for estimating citrus yield. Chinchuluun and Lee (2006) developed a fruit identification method using the Watershed transform to separate and split touching fruit for accurate fruit counts and yield estimation. Kane and Lee (2007) equipped a monochromatic near-infrared camera with interchangeable optical band pass filters to capture images of citrus fruit for in-field immature green citrus identification, and their results revealed the potential of a multispectral imaging system to be used in an automated early season yield mapping system. Okamoto and Lee (2009) developed a hyperspectral image processing method to detect immature green citrus fruit in individual trees, and the fruit detection tests revealed that 80–89% of the fruit in the foreground of the validation set were correctly identified, though many occluded or highly contrasted fruit were identified incorrectly. More recently, Kurtulmus et al. (2011) developed a new 'eigenfruit' approach to detect immature green citrus fruit from color images acquired under natural outdoor conditions and reported a detection accuracy of 75%. Also Bansal et al. (2013) implemented a novel technique utilizing the fast Fourier transform leakage values for detecting immature green citrus, and obtained an accuracy of 82% for correct fruit identification.

For blueberry fruit detection, Zaman et al. (2008) and Zaman et al. (2010) reported that there was a real potential to estimate fruit yield and detect wild blueberry plants and bare spots from digital photography within the fields in Canada, because there was significant correlation between percentage of blue pixels and actual fruit yield in field 1 ($R^2 = 0.90$) and field 2 ($R^2 = 0.97$). Swain et al. (2010) presented the development of an automated yield mapping system for real-time fruit yield estimation for wild blueberry. The experiment showed a significant correlation between percentage of blue pixels representing ripe fruit in the field of view and actual fruit yield. Chang et al. (2012) developed an automated yield monitoring system consisting of two color cameras, a real time kinematic-GPS receiver, and custom software. Highly significant relationship between the percentage of blue pixels and actual fruit yield in two fields were shown through linear regression results. Farooque et al. (2013) evaluated the performance of an integrated automated system in two chosen fields. Plant height, fruit yield, slope and elevation were measured in real-time simultaneously with harvesting. The results suggested that this system was reliable for mapping such information in real-time. Wild blueberry plants are very close to the ground, and so it is easier to take a picture and estimate yield. However, southern highbush blueberry plants in Florida, which was studied in this paper, are taller and bear fruit in different branches. Therefore, new methods need to be developed to estimate fruit yield of southern highbush blueberry plants.

The overall objective of this study was to identify blueberry fruit of different growth stages using red, green, and blue (RGB) color images acquired outdoors using a regular digital camera, which would be a first step for developing a blueberry yield mapping system. Specific objectives were to:

1. develop an efficient method to identify blueberry fruit of four different growth stages in color images acquired under natural outdoor conditions,
2. compare the results obtained from different classifiers and choose one classifier with best performance to be used in blueberry fruit identification.

2. Materials and methods

2.1. Image acquisition

A total of 46 images were acquired from a commercial blueberry farm (Straughn Farm) in Waldo, Florida, USA from April 12th to April 22nd, 2011, for the variety of Jewel, which was one of the representative varieties in Florida. A digital SLR camera (EOS Rebel T2i, Canon Inc., Japan) with an 18–55 mm lens was used, and the camera was set as a full auto mode. The images were saved in JPEG/Exif format with 3648 × 2736 pixels, corresponding to an approximately 13 cm × 10 cm actual scene, and the compression rate was 10:1. Image acquisition was performed between 11:00 AM and 2:00 PM local time. The illumination was affected by the sun light change, wind, and cloud in the sky when capturing the images.

2.2. Building data library

Among the 46 images, 23 images were randomly selected as training images, and the other 23 were used as validation images. A pixel data library including ten classes of different objects mainly appearing in the images was built by manually cropping from the 23 training images using an image editing software (Corel Painter Photo Essentials 4, Ottawa, Ontario, Canada). The 10 different classes mostly found in the images are: mature (*m*), near-mature (*nm*), near-young (*ny*), young (*y*), leaf (*leaf*), burgundy branch (*bran_b*), green branch (*bran_g*), older branch (*bran_older*), other background objects (*bg*), and sky (*sky*). An example image of these classes is shown in Fig. 1, and a brief description is listed in Table 1. As the fruit do not ripen at the same time, maturity of the fruit was divided into four stages: *m*, *nm*, *ny* and *y* for yield prediction.

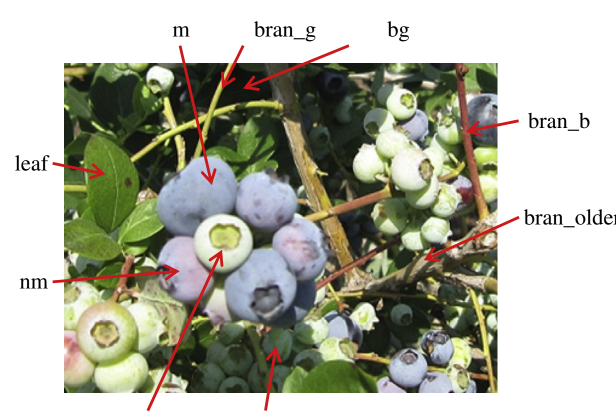

Fig. 1. Example image of blueberry fruit with different growth stages and other objects.