Computers and Electronics in Agriculture 152 (2018) 281–289

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Assessment of spinach seedling health status and chlorophyll content by multivariate data analysis and multiple linear regression of leaf image features

Avinash Agarwal*, Snehasish Dutta Gupta

Agricultural and Food Engineering Department, Indian Institute of Technology Kharagpur, Kharagpur 721302, India

ARTICLE INFO

Keywords:
Leaf spectral reflectance
Plant health status
Image analysis
Multivariate data analysis
Multiple linear regression

ABSTRACT

Plant health and physiological status significantly influence chlorophyll content and photosynthetic capacity. Analysis of leaf reflectance information from digitized leaf images allows high-throughput, non-invasive and real-time estimation of chlorophyll content in a cost-effective manner. In the present study the application of multivariate data analysis tools, viz. principal component analysis (PCA) and agglomerative hierarchical clustering analysis (AHCA), has been discussed for distinguishing between spinach seedlings having high and low chlorophyll contents by simultaneously using the information provided by various image features. Further, leaf color information contained within different color spaces, viz. RGB (red, green and blue), *rgb* (normalized red, green and blue), HSI (hue, saturation and intensity), CIE (Commission Internationale de l'Eclairage) L*a*b*, CIE-XYZ, and CIE-xyY color spaces, has been used to predict chlorophyll content in terms of SPAD (Soil Plant Analysis Development) chlorophyll meter values by multiple linear regression. It was observed that the color indices R, G, R + G, R−B, G−B, R + G−B, Y (luminance) and DGCI (dark-green color index) exhibited high correlation ($R^2 > 0.8$) with the SPAD values. Further, subjecting the leaf reflectance information provided by these color indices to PCA and AHCA enabled a clear segregation of seedlings with high and low chlorophyll contents. SPAD values predicted by the L*a*b* color space information yielded the lowest RMSE (root mean square error) and the highest R^2 (coefficient of determination) amongst the six color space features assessed. The findings of the present study indicate that concatenation of leaf reflectance information provided by different color indices may be more useful than individual color indices for assessing plant health status and predicting chlorophyll content using machine vision.

1. Introduction

Environmental stress and nutrient deficiencies faced by plants are manifested as changes in the pigment composition of leaves (Carter, 1993; Bacci et al., 1998). Decrease in chlorophyll content under conditions of stress results in reduced light absorption and poor photosynthetic activity (Gitelson et al., 2003). Unhealthy and senescent leaves typically exhibit reduced greenness due to lower chlorophyll content (Carter and Knapp, 2001). Reduction in leaf water content is also perceptible through changes in leaf spectral reflectance (Ahmad and Reid, 1996). Since nitrogen is a key constituent of the chlorophyll molecule, plant nitrogen status is also reflected through leaf coloration (Pagola et al., 2009; Rorie et al., 2011). Hence, canopy reflectance is considered as a reliable indicator of plant health status and physiological stress (Kawashima and Nakatani, 1998; Carter and Knapp, 2001; Raun et al., 2001). Assessment of leaf spectral reflectance yields information about the photosynthetic potential and allows timely detection of symptoms of stress and nutrient deficiencies in plants (Wood et al., 1993; Carter and Knapp, 2001; Xu et al., 2011; Riccardi et al., 2014).

Although the use of chlorophyll meters as well as chlorophyll fluorometry, thermal imaging and hyperspectral imaging have proven to be reliable for studying leaf spectral properties and for rapid and non-destructive assessment of plant health status (Wood et al., 1993; Mahlein et al., 2012), the techniques require expensive specialized equipments and are thus out of reach for most cultivators. Advances in digital image acquisition and processing techniques over the past decades have equipped us with tools for rapid, non-destructive, high-throughput and cost-effective real-time quantification of leaf spectral properties using just a digital camera and a computer (Dutta Gupta

* Corresponding author.
E-mail addresses: avinash.agfe@iitkgp.ac.in, avinash.agarwal.1612@gmail.com (A. Agarwal), sdg@agfe.iitkgp.ac.in (S. Dutta Gupta).

https://doi.org/10.1016/j.compag.2018.06.048
Received 5 June 2017; Received in revised form 14 August 2017; Accepted 28 June 2018

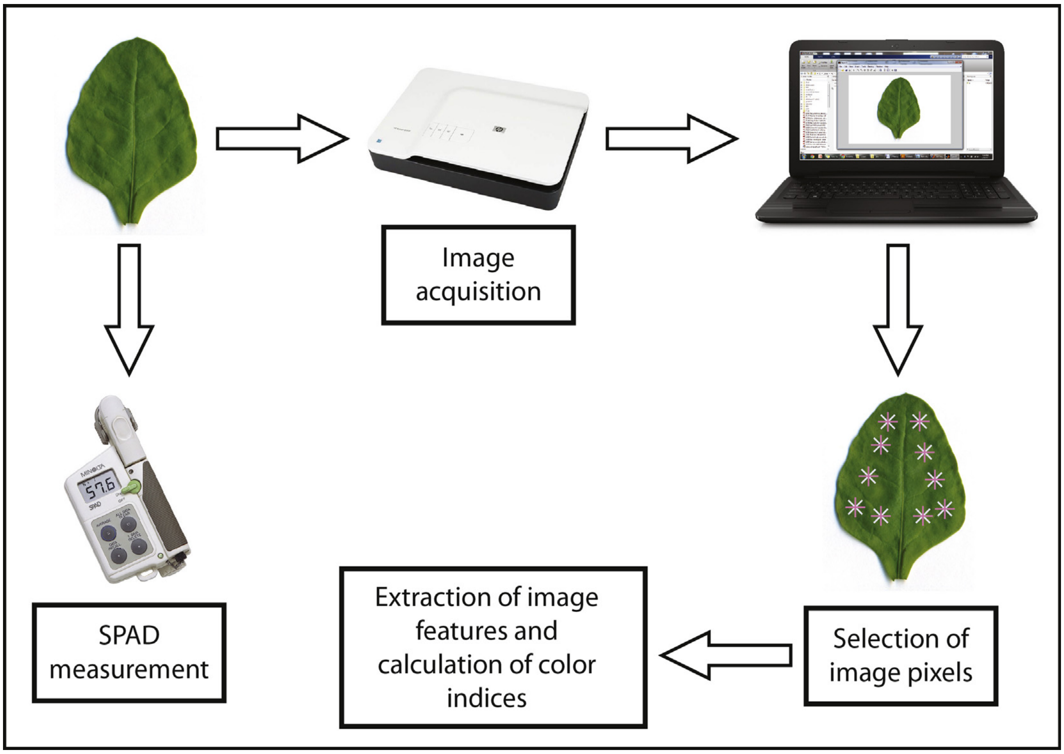

Fig. 1. Steps for acquisition of SPAD values and image features from spinach leaves.

et al., 2014). Leaf digital-image analysis aims at assessing the health of plants via colorimetric characteristics of the leaves. The close association of chlorophyll content with plant health, leaf greenness, photosynthetic efficiency and plant nitrogen status forms the basis of image-based plant health assessment (Rorie et al., 2011). The technique enables the quantification of leaf reflectance information in terms of various color spaces and numerous chromatic transformations derived thereof. The color spaces commonly used for digital quantification of colorimetric information include RGB (red, green and blue), HSI (hue, saturation and intensity), CIE (Commission Internationale de l'Eclairage) $L^*a^*b^*$, CIE-XYZ and CIE-xyY. Characteristic features of these color spaces have been discussed at length in earlier publications (Palus, 1998; Gonzalez and Woods, 2002; Ohta and Robertson, 2005; Schanda, 2007).

RGB color space data contained within digital leaf images has been used for assessing chlorophyll content (Kawashima and Nakatani, 1998; Yadav et al., 2010; Hu et al., 2013), plant stress (Bacci et al., 1998; Ahmad and Reid, 1996), nitrogen content (Pagola et al., 2009; Tewari et al., 2013; Wang et al., 2013; Saberioon et al., 2014) and macronutrient deficiencies (Wiwart et al., 2009; Xu et al., 2011). The *rgb* color features derived from the RGB values have also been used in many such studies (Ahmad and Reid, 1996; Dutta Gupta et al., 2013; Lee and Lee, 2013; Wang et al., 2013, 2014; Dey et al., 2016). Indicating leaf color features in terms of hue (H) and saturation (S) in combination with brightness (B), luminosity (L) and intensity (I) parameters has been discussed in a few reports (Ahmad and Reid, 1996; Wiwart et al., 2009; Mata-donjuan et al., 2012; Lee and Lee, 2013; Vesali et al., 2015; Rigon et al., 2016). The application of $L^*a^*b^*$ color space values for assessing plant nutrition status has also been demonstrated earlier (Graeff et al., 2008; Hu et al., 2010; Wang et al., 2014). Assessment of plant stress using *xy* chromaticity data obtained from spectrometer (Ruth et al., 1991) and colorimeter (Bacci et al., 1998) readings has been reported earlier. However, earlier reports have not addressed the use of CIE-XYZ tristimulus color space or the derived CIE-xyY color model for evaluating chlorophyll content and assessing plant health using digitized leaf images.

Most of the previous reports have credibly correlated leaf chlorophyll content with color indices such as '(R − B)/(R + B)' (Kawashima and Nakatani, 1998), 'R + G' (Hu et al., 2010) and 'G' (Yadav et al., 2010). Dark green color index (DGCI), a colorimetric transformation of the HSV based image features, has been applied for estimating leaf chlorophyll content by many scientists (Rorie et al., 2011; Raper et al., 2012; Saberioon et al., 2014; Vesali et al., 2015; Rigon et al., 2016). The studies have focused on comparing the potential of individual color indices in predicting plant health, chlorophyll content and nutrient status from leaf images. We propose the concatenation of information acquired from different color indices for segregating the healthy and stressed plants by multivariate data analysis and predicting chlorophyll content on the basis of various trichromatic leaf image features.

The objective of this study was to explore the potential of multivariate data analysis tools, viz. principal component analysis (PCA) and agglomerative hierarchical cluster analysis (AHCA), for assessing seedling health status by distinguishing between spinach seedlings with high and low chlorophyll contents using digitized leaf image features. Another aim of the study was to assess the capacity of different image feature trichromatic values or triplets, viz. RGB, *rgb*, HSI, CIE-$L^*a^*b^*$, CIE-XYZ tristimulus values and the CIE-xyY color features, to predict chlorophyll content in terms of SPAD values by a multiple linear regression approach.

2. Materials and methods

2.1. Plant material and growth conditions

Seeds of spinach (*Spinacia oleracea* L. cv. "All Greens") were purchased from Sutton and Sons India Pvt. Ltd., Kolkata, India. The seeds were sown in six plastic trays containing soilrite mix (perlite + vermiculite + sphagnum peat moss 1:1:1). The plastic trays were kept in a controlled environment chamber for a period of 35 days under cool white fluorescent lamps (FL), blue LEDs (BL, 470 nm), red LEDs (RL, 630 nm) and different combinations of blue + red LEDs (3:1, B3RL; 1:1, BRL; 1:3, BR3L). Variations in leaf chlorophyll content under different