Remote Sensing of Environment 187 (2016) 91–101

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Development of methods to improve soybean yield estimation and predict plant maturity with an unmanned aerial vehicle based platform

Neil Yu [a], Liujun Li [b], Nathan Schmitz [a], Lei F. Tian [b], Jonathan A. Greenberg [c], Brian W. Diers [a,*]

[a] Department of Crop Sciences, University of Illinois at Urbana-Champaign, 1101 W. Peabody Drive, Urbana, IL 61801, USA
[b] Department of Agricultural and Biological Engineering, University of Illinois at Urbana-Champaign, 1304 W. Pennsylvania Avenue, Urbana, IL 61801, USA
[c] Department of Geography and Geographic Information Science, University of Illinois at Urbana-Champaign, 605 East Springfield Avenue, Champaign, IL, USA

ARTICLE INFO

Article history:
Received 13 May 2016
Received in revised form 19 September 2016
Accepted 2 October 2016
Available online xxxx

Keywords:
Soybean
Breeding efficiency
UAV
Multispectral image
Object classification

ABSTRACT

Advances in phenotyping technology are critical to ensure the genetic improvement of crops meet future global demands for food and fuel. Field-based phenotyping platforms are being evaluated for their ability to deliver the necessary throughput for large scale experiments and to provide an accurate depiction of trait performance in real-world environments. We developed a dual-camera high throughput phenotyping (HTP) platform on an unmanned aerial vehicle (UAV) and collected time course multispectral images for large scale soybean [*Glycine* max (L.) Merr.] breeding trials. We used a supervised machine learning model (Random Forest) to measure crop geometric features and obtained high correlations with final yield in breeding populations ($r = 0.82$). The traditional yield estimation model was significantly improved by incorporating plot row length as covariate ($p < 0.01$). We developed a binary prediction model from time-course multispectral HTP image data and achieved over 93% accuracy in classifying soybean maturity. This prediction model was validated in an independent breeding trial with a different plot type. These results show that multispectral data collected from the UAV-based HTP platform could improve yield estimation accuracy and maturity recording efficiency in a modern soybean breeding program.

1. Introduction

One of the greatest challenges of the 21st century will be to expand crop production to meet increasing demands for food, clothing, and fuel brought on by both the growing human population and its increasing affluence. The most environmentally friendly way to meet these demands is through developing and providing highly productive crop cultivars to farmers (Tester and Langridge, 2010). The recent biotech revolution is impacting the power and efficiency of this crop improvement process by increasing the capabilities of researchers to rapidly analyze large populations of plants with abundant genetic markers and we are on the brink of being able to fully sequence the genome of a large number of plants in breeding programs (Thomson, 2014). However, genetic analysis has its greatest value when it can be associated with plant phenotypes. While our ability to analyze DNA is increasing at an exponential rate, the capacity to phenotype plants in a field setting has not improved nearly as rapidly. Advances in phenotyping technology are critical to ensure the genetic improvement of crops to meet future demands.

* Corresponding author.
E-mail address: bdiers@illinois.edu (B.W. Diers).

Conventional remote sensing applications are based on satellite and manned aircraft to acquire visible (VIS), near-infrared (NIR), and short-wave infrared radiation reflected and far-infrared radiation emitted by the crop to estimate the yield potential and environmental stress for large land areas (Atzberger, 2013). The reflectance of vegetation in multiple spectral regions have been shown to be good estimators of crop biomass, yield, canopy coverage, leaf area index (LAI), chlorophyll content and plant senescence (Daughtry et al., 2000; Gitelson et al., 2003; Hatfield and Prueger, 2010; Merzlyak et al., 1999; Penuelas et al., 1994). However, the application of conventional remote sensing in plant phenotyping for breeding purposes is limited by the expensive targeted data acquisition and coarse spatial resolution (Wang et al., 2010). The ideal phenotyping system for breeding programs requires high throughput evaluation of resource capturing ability, utilization efficiency, and growth and development on a plot level while the crop develops.

High throughput phenotyping (HTP) was first implemented in controlled greenhouses and growth chambers by using automated imaging systems to assess plant growth and performance (Luis Araus and Cairns, 2014). However, environmental factors and heterogeneous conditions in the field are not replicated in the controlled facilities. In addition, controlled facilities do not have the space needed to evaluate the large amount of germplasm in breeding programs. Field-based phenotyping platforms are being considered increasingly to deliver the necessary throughput for large scale experiments and to provide an accurate depiction of trait performance in real-world environments (White et al., 2012).

Researchers have been working on high resolution sensing systems for agricultural settings. Unmanned aerial vehicles have proved to be flexible platforms for sensing crop growth conditions (Xiang and Tian, 2011) and in-field crop monitoring towers (Ahamed et al., 2012) have provided higher density data for near-real-time remote sensing. With the recent advances in multi-rotor systems, UAVs have progressed as a viable aerial platform for proximal remote sensing (Zhang and Kovacs, 2012). In particular, UAVs can operate at a low altitude to capture images with ultra-high spatial resolution of up to 1 cm per pixel, which is sufficient resolution for measuring individual field plots (Turner et al., 2012). In addition, the system can be deployed on demand with great flexibility to ensure optimal temporal resolution during the crop growing season. Finally, the costs and technical skills required to operate these platforms are becoming lower over time, which makes the UAV-based HTP a promising solution for plant breeding programs.

Plant breeding efficiency is limited by spatial field variability and phenotyping capacity, which could be potentially improved using UAV-based HTP platforms and the application of UAVs in field crop phenotyping was summarized by Sankaran et al. (2015). Spatial field variability usually results from agriculture management, soil heterogeneity and variability in field topography. Such variation in a breeding trial decreases the repeatability of the phenotypic traits evaluated and the precision of the trait mean estimation of any given experimental entry. Soil, nutrition and water spatial variation have been reported to lead to large plot residual variance in multiple studies (Cairns et al., 2013; Masuka et al., 2012; Robertson et al., 2008). The UAV-based HTP was shown to capture field stress variation which can be used to assist crop genetic improvement (Zaman-Allah et al., 2015). In addition, non-pattern variation due to planting issues, which is directly reflected as plot row length, is challenging to explain using traditional biometric models, such as a randomized complete block design (RCBD) or nearest neighbor analysis. Therefore, characterization of such random variability on a plot level and removal of these effects from treatment variation are critical to increasing the genetic effect to noise ratio.

Modern breeders take individual visual ratings and measurements of plants to estimate important agronomic traits in much the same way they did decades ago. Both biases among individuals taking the data and subjective criteria create imprecision in phenotypic data collection and the visual ratings are limited by what a human can visualize at the ground level. In addition, it is difficult to take visual ratings of the tens of thousands of lines evaluated in modern breeding programs. Such large populations are needed to achieve successful selection of multiple loci, which can be in linkage disequilibrium with unfavorable traits or to dissect the genetic architecture of complex traits (Dinka et al., 2007). Aerial spectral imaging has been shown to deliver plant density estimations, physiological condition assessments and stress detection in different crops (Gonzalez-Dugo et al., 2015; Hunt et al., 2010; Liebisch et al., 2015; Thorp et al., 2008). Therefore, a UAV-based HTP platform could be customized as a quick and low-cost approach for agronomic trait evaluation to improve the efficiency of crop breeding programs.

Random Forest (RF), a machine learning algorithm develops multiple classification and regression trees (CART) based on a random subset of the input variables using randomly selected bootstrap samples and the ensemble of these trees are combined to ensure the prediction accuracy (Breiman, 2001). Random Forest has been found to be superior to other machine learning algorithms because it is not sensitive to data skewness, can easily handle a high number of model parameters, and has fewer issues with overfitting (Horning, 2010). The applications of RF in image analysis have been increasingly reported in remote sensing classification studies (Guo et al., 2015; Peters et al., 2007; van Beijma et al., 2014; Wiesmeier et al., 2011).

Our research aimed to establish a HTP platform based on a UAV equipped with a multispectral sensor system through acquiring high resolution image data followed by RF machine learning analysis to improve soybean breeding efficiency. The first objective was to measure the plot-based canopy geometric features and test the significance of plot row length as a covariate in yield estimation models. The second objective was to develop a machine learning model for binary soybean plot maturity prediction using multispectral data.

2. Materials and methods

2.1. Experimental setup

The HTP remote sensing study was conducted at a three ha soybean field located at Urbana, IL (40.053602 N, 88.235721 W) in 2014. Two different breeding research trials were planted in this field and these trials were designated as GS and NAM. The GS trial represented a genomic selection (GS) study containing 2980 plots from 26 populations and two check cultivars ('IA2102' and 'IA3023'). Each population consisted of approximately 110 recombinant inbred lines (RILs) developed from different breeding crosses. The NAM trial included two experiments that each contained 60 breeding lines from a nested association mapping (NAM) study. These lines were replicated twice using a randomized completed block design and there were 240 plots in the NAM trial. The GS trial was planted on June 9, 2014 and the plots were single rows with a 0.76 m row spacing and a 1.2 m length. The plots were not replicated and were grown in blocks of 20 entries and each block included check cultivars. The NAM trial was planted on June 7, 2014 and the plots were four rows wide with a 0.76 m row spacing and a 3.6 m length.

To ensure the geo-referencing accuracy of images collected by the UAV HTP platform, 12 flat 61-by-61 cm cross-patterned wooden panels were mounted on metal poles sunk into the soil to act as permanent ground control points (GCPs) (Fig. 1A). The center position of each panel was measured using the Trimble R8 GNSS system (Sunnyvale, CA), which provided cm-level accuracy. In addition, a 122-by-122 cm wooden board painted white and black in a four-cell checkerboard pattern with paint-sand mix to create a near-Lambertian reflection surface (Fig. 1B). This board was placed in the center of the field and used as a reference object to account for different irradiance conditions at each data collection time.

2.2. Development of the HTP platform

An autonomously flying octocopter, "X8" (3D Robotics, San Diego, CA) was purchased as a ready-to-fly kit. Two Canon S110 point-and-shoot digital cameras (Canon Inc., Lake Success, NY) were mounted under the UAV to point at the nadir, one of which had been converted for capturing NIR only. These digital cameras were equipped with complementary metal-oxide semiconductor (CMOS) sensors, which are sensitive to wavelengths between 350 nm and 1100 nm. Nijland et al. (2014) reported when the infrared (IR) rejection filter in the digital camera was replaced by 590 nm or greater long-pass filters, the Blue-channel recorded IR only and was the second most sensitive to IR after the Red-channel. According to the IR photography professional company Kolari Vision (Raritan, NJ), more chromatic aberrations and softer focus, which might decrease image quality, would occur with a camera converted with a 590 nm or 665 nm filter compared to 720 nm or higher filters (kolarivision.com/articles/choosing-a-filter/). Therefore, one of our cameras was modified into a standard NIR camera using a 720 nm long pass filter by Kolari Vision, and the blue-channel of the images taken by this modified camera was used to extract the NIR information.