2023 International Conference on Integrative Precision Agriculture – Local Solutions Through Global Advances

Development of a distributed computing architecture for spatio-temporal mapping of cotton using Hadoop and deep learning

V. Thesma[1*], G.C. Rains[2], J.M. Velni[3]

[1]School of Electrical and Computer Engineering, University of Georgia, Athens, GA, USA
[2]Department of Entomology, University of Georgia – Tifton Campus, Tifton, GA, USA
[3]Department of Mechanical Engineering, Clemson University, Clemson, SC, USA
[*]Corresponding author: vst16532@uga.edu

Abstract

Farmers are facing many challenges to produce enough food and fiber for a rapidly growing world population. The use of state-of-the-art computer vision methods and big data tools can help farmers analyze farm data faster and more accurately. Also, these solutions result in farmers having less reliance on manual labor, give remote insight to farm status, provide farm management and decision making recommendations, and forecast plant and animal production when using these management recommendation tools.

Cotton farmers specifically experience challenges to properly manage and maximize yield and quality. Yield prediction can be used to improve management and increase yield by keeping track of cotton boll count and their locations. However, it is difficult to manually count cotton bolls because they grow in high density and under heavy canopy. On the other hand, cotton blooms typically determine the number of cotton bolls much earlier. The cotton blooms each only last a few days and are continually produced for approximately 1 to 2 months during the summer. Thus, counting their frequency manually and everyday is also very difficult, especially for large-scale farms. Several studies

have addressed these challenges in detecting and counting cotton bolls and blooms to predict yield using traditional pixel-based methods, deep learning models, and unmanned vehicles. Also, these methods use ground vehicles that are typically equipped with cameras, on-board processors, and navigation systems to collect vast amounts of cotton image data. This data must be systematically and autonomously processed to provide for cotton boll and bloom counts, yield estimation, and potential intelligent harvesting techniques under investigation. However, current image processing methods are linear and cannot adequately process the ingested data in a timely manner. Thus, it is imperative to develop a big data architecture to process vast amounts of cotton image data in parallel.

The purpose of this research is to develop an open source, big data architecture to create spatio-temporal maps of cotton blooms appearance over time for an entire field and season. These maps serve as a visual understanding of how many blooms grow and their location prior to harvesting. This way, farmers have insight about the quantity of blooms that grow at what frequency over time. This research presents the development of a distributed computing architecture to process cotton image data using Hadoop and deep learning. A master-slave distributed computing architecture is created to ingest batches of cotton image data, preprocess them, detect and count the number of blooms per image, and create a spatio-temporal map for each day using Hadoop. The Hadoop cluster consists of a single master node and five slave nodes. The client starts the cluster via the master node and the master node submits jobs for the slave nodes and itself to process data in parallel. This research uses cotton plant video data collected from a field in Tifton, GA with a ground vehicle equipped with an on-board processor and stereo camera. Images are extracted from the video frames to be used as input to a pretrained tiny-YOLOv4 model for accurate cotton bloom prediction. The detection results from past and future bloom locations are used to build the spatio-temporal map onto mid-stage cotton plants in parallel

Keywords: Big-data, crop monitoring, computer vision