Development of distributed computing architecture for spatio-temporal mapping of cotton

Vaishnavi Thesma, Glen C. Rains, and Javad Mohammadpour Velni

University of Georgia

College of Engineering
UNIVERSITY OF GEORGIA

Introduction

Farmers are facing many challenges to produce sufficient food and fiber for the rapidly growing world population, including climate change, reduced natural resources, and an unskilled workforce [2, 6].

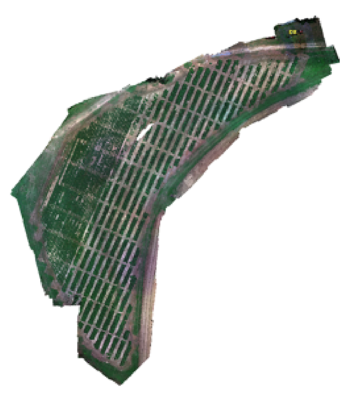
Cotton farmers specifically experience challenges to properly manage and maximize yield and quality. However, it is difficult to manually count cotton bolls and blooms [3, 4, 5, 6]. Several works have addressed these challenges in detecting and counting cotton bolls and blooms to predict yield using traditional and modern computer vision methods. These methods use ground vehicles and collect vast amounts of cotton image data [5, 6].

This data must be systematically and autonomously processed to provide for cotton boll and bloom counts, yield estimation, and potential early harvesting. 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 contribution of this work 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 using Hadoop and deep learning.

Data Collection and Preparation

The cotton bloom data was collected using an autonomous ground vehicle equipped with stereo camera from the University of Georgia Tifton campus in Tifton, GA, USA [6]. An aerial view of the farm is shown in Figure 1 as well as the ground vehicle used and an example of the extracted data.



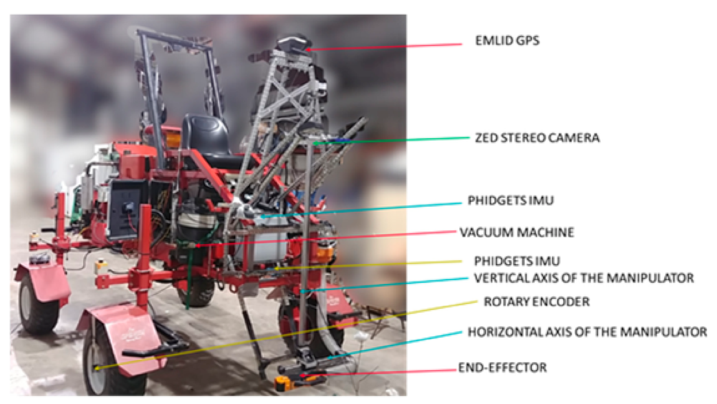


Fig. 1: (Top Left) Aerial view of cotton farm in Tifton, GA; (Top Right) West Texas Lee Corp. ground vehicle to capture cotton plant data; (Bottom) An example of the extracted data.

Hadoop

Apache Hadoop is an open-source framework that handles processing big data in a distributed architecture [1]. Hadoop contains four main parts, namely

1. **Hadoop Common** which contains the core files to run Hadoop,
2. **Hadoop Distributed File System (HDFS)** that serves as the storage component for a cluster,
3. **YARN** that serves as the resource manager for Hadoop, and
4. **MapReduce programming model** for parallel processing data.

Distributed Computing Architecture

Our distributed cluster consists of a single master node and five slave nodes. Each node is a virtual machine created on Oracle's VirtualBox, each with 2GB of RAM, 2 CPUs, 25 GB of disk space, a flavor of Ubuntu operating system, and Hadoop version 3.3.4. The master and slave nodes communicate with each other over the network.

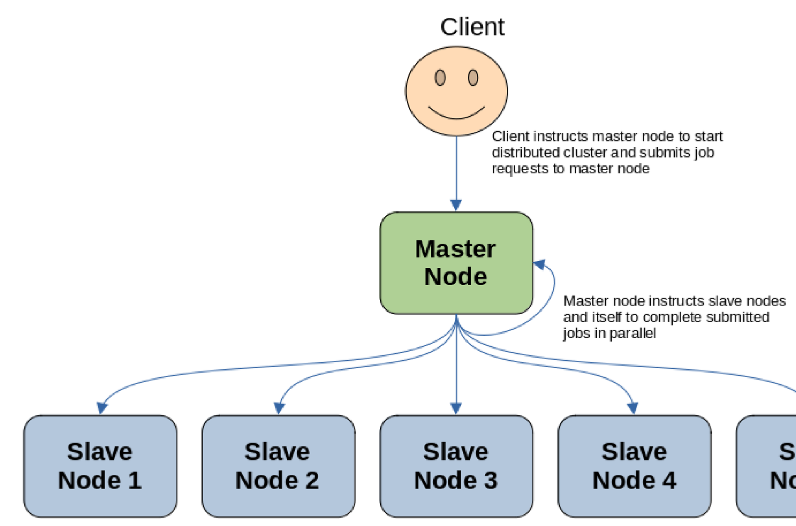


Fig. 2: Schematic of our distributed computing architecture consisting of a client, a single master node, and five slave nodes.

Methodology

Once the data is ingested into HDFS, the data is preprocessed in parallel. The images are split and sliced to increase the size of the cotton blooms with respect to the overall image. The sliced images are stored locally for cotton bloom detection using tiny-YOLOv4 and the detection results are stored across all nodes in the cluster. Next, slices that contain at least 1 detected cotton bloom are inputted into HDFS for spatio-temporal map creation. The spatio-temporal map slices are then merged together to create a comprehensive frame showing the spatial distribution of cotton blooms at different times throughout a season.

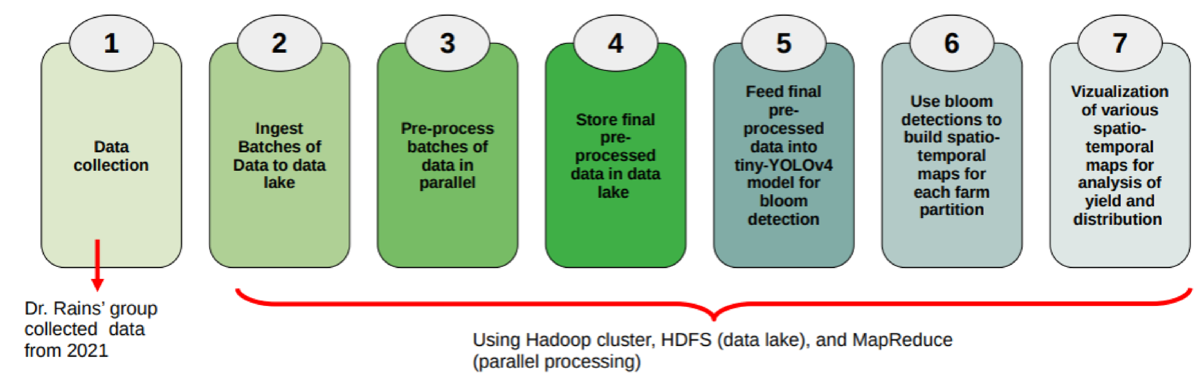


Fig. 3: Our distributed computing methodology to process the cotton image data.

Results

As an example of creating a spatio-temporal map, we select 5 dates from the same row in the 2021 farm. Namely, July 26, August 4, August 11, August 23, and August 26. We process each of the row's collected images as per our aforementioned methodology. The total number of raw extracted images processed was 628 images. The average processing time per row is seen in the table below.

Task	**Proccesing time** (min per row)
Image Splitting	1.23
Image Slicing	2.23
Cotton bloom detection	15.02
Spatio-temporal map creation	8.02
Image merging	1.85
Total time per row	30.2

Spatio-temporal Map Creation

We use July 26 image data as our base date and the following dates serve as future time-points to when and where new cotton blooms appear and their distribution. Figure 4 shows an example of our spatio-temporal map implementation using these dates. The left image is a cropped image frame from July 26 and several cotton blooms are visible. The right image shows our spatio-temporal map implementation result with the colored boxes representing the future bloom detection after July 26.

Blooms on July 26, August 4, August 11, August 23, and August 26 are shown by colors dark blue, green, pink, yellow, and orange, respectively. There are fewer blue and red boxes as mid- to late-July are when blooms first appear. There are more blue, and pink blooms as these are when most blooms appear in higher frequency.

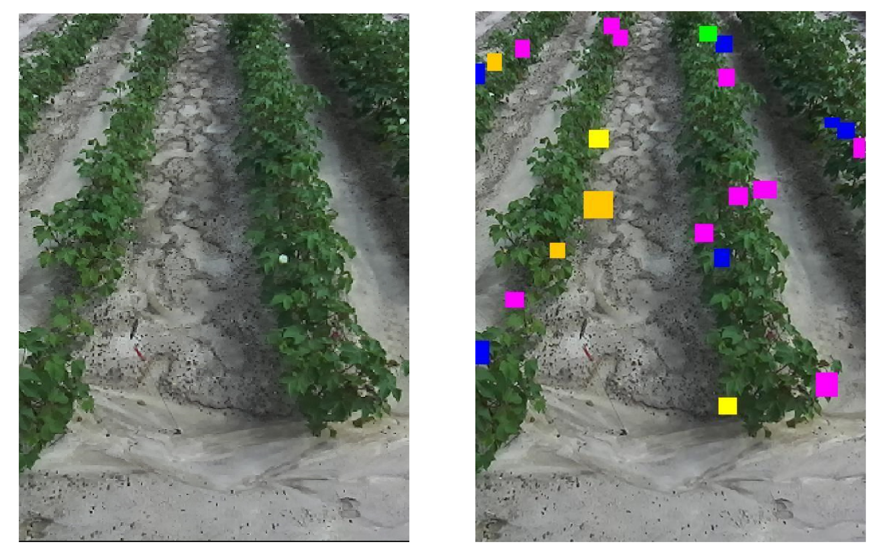


Fig. 4: Our spatio-temporal map implementaion using a cropped image frame from the right lens of the stereo camera collected on July 26 with several blooms present (Left) and 4 dates to overlay future blooms (Right).

Future Works

We plan to extend this work by processing all the cotton farm data collected in Tifton, GA and create spatio-temporal maps for each row. We plan to analyze the spatio-temporal distributions of the cotton blooms appearance.

Acknowledgements

This work was supported by the United States National Institute of Food and Agriculture (NIFA) under award no. 2020-67021-32461.

References

[1] Apache Software Foundation. *Hadoop.* Version 3.3.4. URL: `https://hadoop.apache.org`.

[2] Michael Buzzy et al. "Real-time plant leaf counting using deep object detection networks". In: *Sensors* 20.23 (2020), p. 6896.

[3] Ziyun Huang, Yanan Li, and Haihui Wang. "In-field cotton boll counting based on a deep neural network of density level classification". In: *Journal of Electronic Imaging* 29.5 (2020), p. 053009.

[4] Yu Jiang et al. "DeepFlower: a deep learning-based approach to characterize flowering patterns of cotton plants in the field". In: *Plant methods* 16.1 (2020), pp. 1–17.

[5] F Kadeghe, R Glen, and P Wesley. "Real-time 3-D measurement of cotton boll positions using machine vision under field conditions". In: (2018).

[6] Vaishnavi Thesma et al. "Spatio-temporal Mapping of Cotton Blooms Appearance Using Deep Learning". In: *IFAC-PapersOnLine* 55.32 (2022), pp. 36–41.