Precision Agriculture Research Infrastructure for Sustainable Farming

Cheol-Hong Min, Chih Lai, Scott Morgan, Andrew Hafferman, Ron Chiang

School of Engineering
University of St. Thomas
St. Paul, Minnesota

A paper from the Proceedings of the
14th International Conference on Precision Agriculture
June 24 – June 27, 2018
Montreal, Quebec, Canada

Abstract. *Precision agriculture is an emerging area at the intersection of engineering and agriculture, with the goal of intelligently managing crops at a microscale to maximize yield while minimizing necessary resource. Achieving these goals requires sensors and systems with predictive models to constantly monitor crop and environment status. Large datasets from various sensors are critical in developing predictive models which can optimally manage necessary resources. Initial experiments at University of St. Thomas (UST) greenhouse have demonstrated the feasibility of sensor system for use in Precision Agriculture. Hundreds of high-value vegetables such as lettuce, beans, peas, onions, spinach and bok-choy were planted in different types of soils and nutrients. Soil moisture and pH level of each plant were monitored and collected by soil sensors. A weather station was installed to collect the air temperature, air moisture, light, and wind information in the greenhouse. A Photosynthetically Active Radiation (PAR) sensor was also installed to monitor how the plant grows responding to different wavelengths. A multi-spectral camera was also used to observe the Near-Infrared Reflection (NIR) from each plant. This information reveals the amount of photosynthesis occurring in each plant, providing an important indicator of plant's health. All information collected was time stamped such that different sensor information could be correlated. This information was then fed back to a controller to release water and nutrition to different plant groups at different times in order to meet different growth patterns.*

During the 50-day growth period, over one billion data points were generated from 6 different plant types. In order to handle and process such big data. Cloudera Hadoop Cluster and software modules to process such data is being developed. Using the information collected from our infrastructure, more sophisticated models can be developed enabling more sustainable farming.

Keywords. *Sensor Systems, Predictive Model, Machine Learning, Photosynthetically Active Radiation (PAR), Near-Infrared Reflection (NIR), Precision Agriculture, Sustainable Farming.*

Introduction

With the advance in autonomous control technology, sensors and computing capacity, a new trend is emerging which is largely known as the Precision Agriculture (PA). Among variety of tasks involved in farming, continuous seeding, weeding and harvesting is becoming a more and more difficult task for the aging farmers. According to the Census of Agriculture, about 1% of the US population is farmers with average age of 58.3 years and age of the operators over 45 years of age comprises 85% of total farmers (Census of Agriculture 2012). Therefore, such tasks are becoming a heavy burden with aging farmers. In addition, younger generation of potential farmers are leaving farms to seek less strenuous and more attractive jobs in the cities. The good news is that if technology is used properly it could assist the current farmers and may attract younger generations who could continue to be involved in agriculture industry.

Autonomous robots can perform repetitive and dangerous tasks and fill the needs of current gap and reshape the future of the agriculture industry. We will need less people performing physical labor and new jobs will be created which requires more hardware development, computer operating and software programming skills. Operators can be working on site or remotely monitoring and operating the autonomous machines. Farming can also be controlled more precisely which could improve the yield and sustainability.

Among seeding, growing/weeding and harvesting, growing/weeding which involves monitoring and controlling healthy crop growth is the most repetitive and difficult task. Proposed method of managing crop growth depends on a set of image processing algorithms which is done by taking pictures of crop, analyzing its health and identifying any weeds, providing nutrients and removing the identified weeds. Image analysis algorithm should be able to adapt to changing weather patterns, light conditions and any anomaly not seen during the algorithm test stage. Therefore, smart image analysis algorithm is needed to monitor crop health and identify weeds.

Improving crop yield is also dependent on how well the resources are managed and weeds are controlled. Therefore, the robots used in farming need to perform various tasks, such as seeding, taking pictures using camera to monitor crop growth, growing, weeding and harvesting. Such autonomous robots must be either equipped with multiple tools to handle various tasks or the tools must be interchangeable depending on the task in execution. In this paper, we describe our PA architecture and initial implementation results.

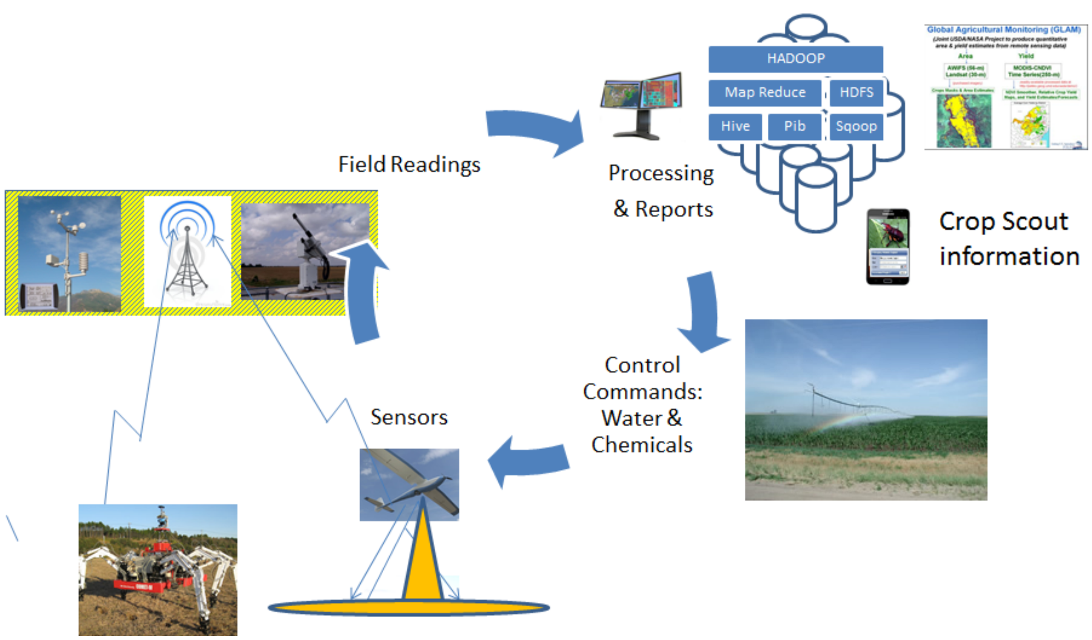


Fig 1. Precision agriculture enterprise system

Precision Agriculture Architecture at the University of St. Thomas

As shown in Fig 1 and Fig 2, University of St. Thomas (UST) has developed a system to collect sensor data, perform data analysis and utilize the analyzed information to manage crop/plant growth. Mobile platforms involve autonomous robots, which we call, a ground rover, and aerial drones to observe large scale farming area and in-situ sensors to understand soil, moisture and weather conditions to optimally manage resources.

Proposed system architecture was implemented in a smaller scale and experiments were performed in the university greenhouse facility to test the feasibility of our PA system. In lieu of autonomous ground rovers and aerial drone, they were replaced with a mechanical structure equipped with a camera which could move above the plants. Enhanced version of this structure is the robot which can be programmed to move horizontally and vertically to acquire images as shown in Fig 2. In addition, in-situ sensors were placed in the soil to determine the soil temperature, soil conductivity, and weather station was installed to monitor wind direction, wind speed, humidity, temperature, barometric pressure, light intensity and Photo-synthetically Active Radiation (PAR) values. All acquired data was reported via secure communication via WiFi tunneling to Precision Agriculture Enterprise System Hadoop Cluster. Sensor data was reported to Hadoop cluster every 10 seconds.

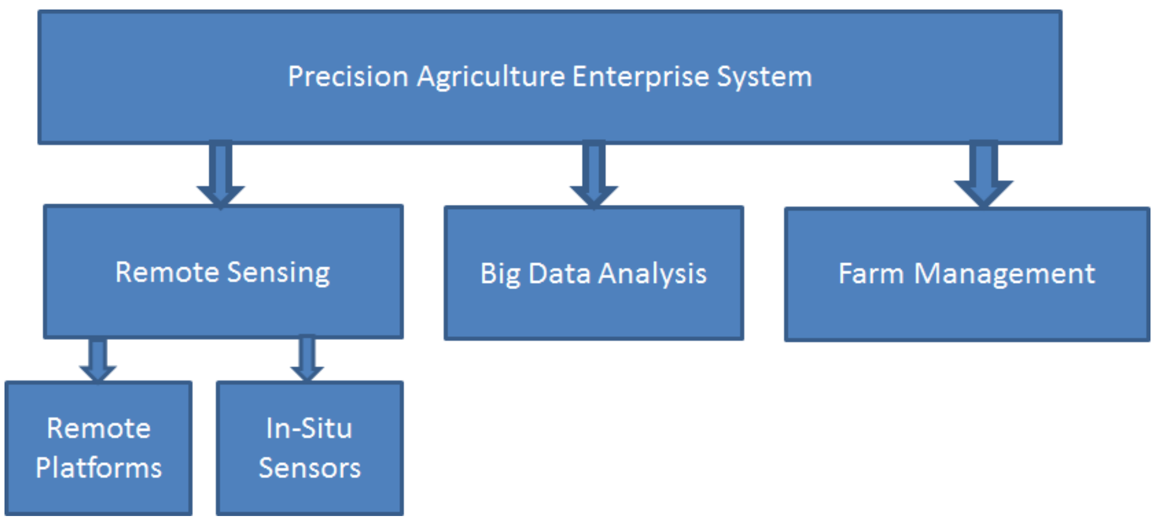


Fig 2. Proposed intelligent precision agriculture system architecture

Data Collection and Analysis via Machine Learning

In this section, we demonstrate the use of predictive modeling within PA in order to conserve precious resources, such as water, reduce environment impact, and promote more efficient growth (Lai et al. 2014; Zhang 2015; Anderson 2016). We develop predictive models that utilizes images as well as sensor inputs such as soil moisture, nutrition, temperature, sun light, and soil-type to identify and predict the location, size and spread of stressed areas of vegetation.

Data Collection Methods

Initial experiment was conducted at UST to demonstrate the feasibility of PA in the. Hundreds of plants from different types of high-value vegetables (lettuce, beans, peas, onions, spinach and bok choy) in different types of soil. The soil moisture and pH level of each plant are monitored and collected by sensors every second. A weather station was also installed to collect the air

(a) (b) (c)

Fig 3. Data collection from proposed system (a) Greenhouse robot for image acquisition, (b) Greenhouse infrastructure with in-situ sensors and weather station to collect temperature, moisture, and multiple sensor data. (c) Plant health monitoring via image analysis

temperature, air moisture, light level, and wind information in the greenhouse. This information is processed and fed back to a computer based controller to release individualized amount of water and nutrients to different group of plants. Data collected from the controlled environment can be used to understand the relationship between environment and their growth. The variety of growth patterns of different plants are also used to train our predictive models. In addition to RGB images obtained, a specialized multi-spectral camera was also used to observe the Near-Infrared Reflection (NIR) from each plant. This reflection reveals the amount of photosynthesis occurring in each plant which provides an important indicator of plant's health status. The multi-spectral camera was installed on a moving structure so images for different plants can be taken over the growth period can be properly indexed.

We have collected a large amount of data over the 50-day growth period. When such experiment is extended to a farm, amount of data collected may be extremely large. While storing all big datasets alone poses a large obstacle, analyzing these datasets to extract meaningful growth patterns may be harder. This is because growth patterns of different plants under a dynamic environment may contain complex and changing quantitative relationships among combined factors like amount of water, nutrition, light, etc. This is another reason that very few "smart" software packages have been integrated into precision agriculture systems (Bakhtiari and Hematian 2013; Shearer 2014; Anderson 2016). Therefore, it is our goal to develop an efficient model to identify the growth patterns of different plants under variety of growth conditions.

Data Storage and Transfer Methods

Data sets derived from sensors that are continuously monitoring the conditions that impact the farm's products are so large and complex that traditional data processing applications are inadequate to deal with them. In addition, farm data sets are of such a diverse nature and constantly evolving so planning a traditional for searching, analysis, sharing, querying and updating becomes a messy and complicated business. Big Data computing is required to handle these problems. Big data computing requires a specific infrastructure. The capability is typically rented from a company like Amazon, but this presents a problem for hosting developmental applications in H2O, Scala, R or Python. It was decided to create a Cloudera Hadoop Cluster from 38 Sun Fire Servers. The cluster hosts 76 processors with 2 cores each, 640 GB of RAM and 60 TB of hard disk storage.

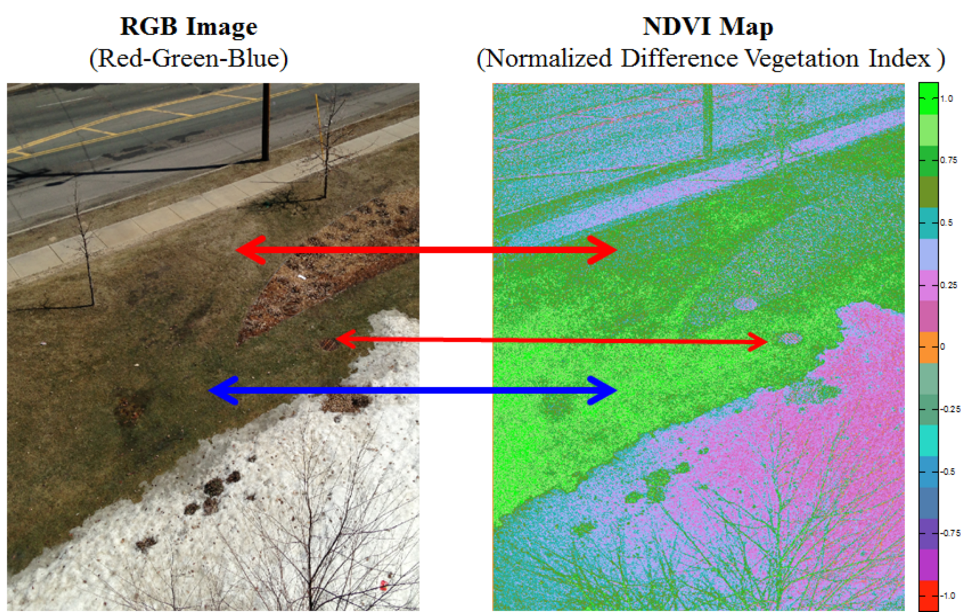
Remote sensor data transfer to the Precision Agriculture Cluster is not as simple as pasting a file in a document and sending email. Sensor data for example from the remote weather system is captured as textual data that is time stamped with latitude and longitude as part of the file header. These small files are streamed using a custom application created in Python to Kafka Producer. Kafka Producer takes care of the Security related to SSH on the external router tunneling to the Master node of the Cluster where Kafka Topic messages (channels) are received by Kafka

Customer. Kafka Customer then convers the Kafka Topic messages to text files that are stored on the master node of the Precision Agriculture Cluster. We plan to continue the work to take the text files stored on the master node using a custom application written in Python H2O or Scala and store the files on the slave nodes as Hadoop Distributed File System (HDFS) files. The HDFS files can then be accessed and analyzed by MapReduce and custom applications written in R. to produce actionable tasks, indicate possible risk vectors to the farm and discover other trends.

Quantify Health Statuses of Plants from Images

Since healthy vegetation strongly reflect near infrared (NIR) light and strongly absorb red light (less visible light reflection) in the process of photosynthesis, we can use multispectral camera to collect both visible and NIR light reflected from vegetation and other objects. If we compute the ratio of these light reflections, it indicates the wellbeing of vegetation. In other words, the ratio of the NIR and red light determines the strength of the photosynthesis activity in the vegetation (Herring 2000). This ratio is referred to as Normalized Difference Vegetation Index (NDVI). NDVI has a range of values from +1.0 to -1.0. Strong photosynthetic activity will lead to a strong NDVI value, indicating healthy vegetation. Ideal photosynthetic activity would reflect 100% NIR and absorb 100% red giving an NDVI value of +1.0, indicating a healthy vegetation. Conversely, a dying plant that absorbs 100% NIR and reflects 100% red would produce a NDVI value of -1.0. Fig 3 (best viewed in color) shows an example in which healthy and unhealthy vegetation areas can be marked by a calculated NDVI map.

After the overall distribution of NDVI in each image is computed (see Fig 4 and Fig 5), we need to associate the health status of each plant with environmental factors such as soil moisture. Unfortunately, the location of each plant is not fixed in various images that are taken over different time. Hence, another image processing algorithm needs to be developed so the center of each plant can be identified and the NDVI (i.e. health status) of each plant can be correctly associated with its soil moisture and soil pH level. This process requires large amounts of complex and iterative computations.



(a) (b)

Fig 4. NDVI map: (a) show that healthier (greener) grass areas in the RGB picture (left). Areas with healthier vegetation will have larger NDVI values (brighter green) in the NDVI map while stressed grass areas in the RGB picture are

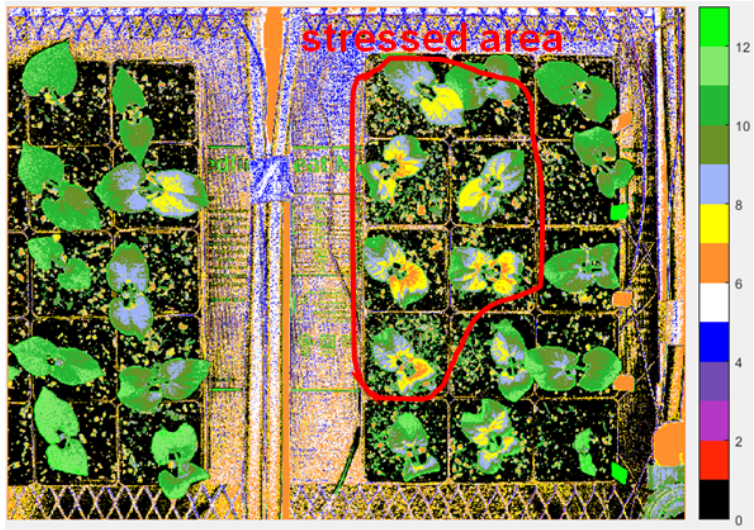
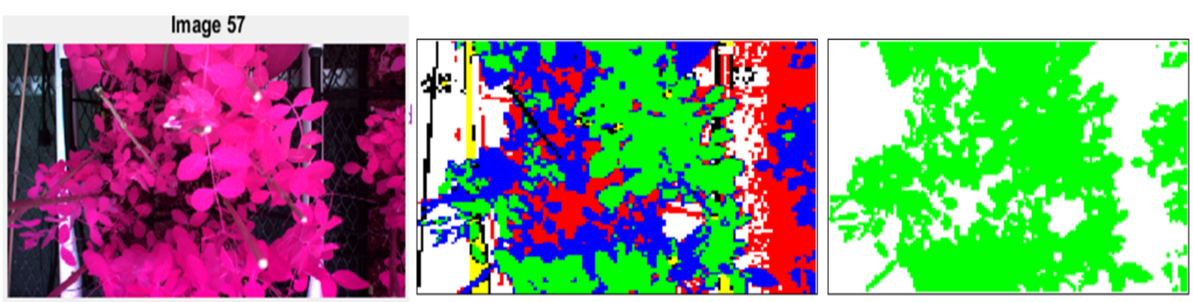


Fig 5. One NDVI map calculated from one image taken in our greenhouse experiment. The color bar on far right shows the range of NDVI indices. Higher NDVI indices (i.e. 12) indicate healthier vegetation. As we can see, the plants near the center of the figure (circled by red) are less healthy than the other plants.

Identify vegetation area in images

A highly accurate predictive model relies on high quality training data. Hence, one essential task before developing predictive models is to accurately identify the real vegetation area in images while excluding non-vegetation (i.e. water pipe, soil, debris, machines, etc.) area.

To automate this process, we have trained another machine learning model to correctly recognize the vegetation area without human intervention. Few major steps in this process include: Step 1. Use machine learning methods to cluster each image into few segments based on the color and infrared similarity within each image. (See Fig 6(a) and Fig 6(b) for an example). Step 2. Manually examine segments that were generated from hundreds of images. If an image segment contains vegetation, the segment is manually marked. For example, we mark the blue and the green areas in Fig 6(b) as areas that contain vegetation. Step 3. Finally, we ask machine to learn the rules to classify the marked (vegetation) and unmarked (non-vegetation) image segments. For example, Fig 6(c) shows the vegetation area begin correctly identified by our machine learning method. So far our model has achieved 93% of accuracy in recognizing vegetation area.



(a) (b) (c)

Fig 6. (a) Raw inferred image. (b) Segments generated by applying clustering algorithm to (a). (c) Vegetation area classified from (b) by machine learning methods.

Conclusion

We believe that the use of reliable predictive models in precision agriculture will enable farmers to conserve precious resources (pesticides, water, fuel, time, etc.) for more resource efficient farming. By conserving resources, we can also reduce carbon footprint, preserve more arable land, and promote more sustainable agriculture. Such objective is not easy to acquire as it requires a large amount of data collection and computational power to process and develop reliable predictive models for optimal plant growth prediction.

Beyond the computing resource, other key factors in developing reliable predictive models include validating and adjusting our models against data. As mentioned in the previous sections, our ultimate goal is to develop an automated system that can predict optimal growth conditions to autonomously grow healthy plants by controlling the right amount of water and nutrition to different plants under different environment (i.e. light / temperature, moisture in different seasons). To achieve this goal, our predictive models must be validated against big datasets with lots of fluctuating and missing values. This task definitely require iteratively examining the dataset over the course of the project. Potential machine learning algorithms that can be used to build predictive models must be able to detect non-linear relationships among environment / growth factors (i.e. moisture, light) and the health status of plants (i.e. NDVIs of plants).

Finally, since the conditions in the growing environment may continue to change (i.e. increasing temperature due change of season, global warming, or gradual loss of soil nutrition), we also plan to study how to extend static machine learning models so they can dynamically update themselves in real-time based on drifting data to gain improvement in classifier performance.

References

Census of Agriculture, US Department of Agriculture, published 2012, https://www.agcensus.usda.gov/Publications/2012/#rankings

Lai, C., Belsky, C., Holub, B., Ellingson, J., Greene, C., and Morgan, S. (2014). Building proactive predictive models with Big Data technology for precision agriculture, In Proceeding of the 12th International Conference on Precision Agriculture, Sacramento, CA, USA.

Zhang, Q. (2015). *Precision agriculture technology for crop farming*, CRC Press.

Anderson, M., (2016). MIT's Food Computer: The future of urban agriculture, *IEEE Spectrum*, 53(2),.12-13.

Herring, J. W. D (2000). *NASA earth observatory – measuring vegetation (NDVI & EVI)*, NASA earth observatory, 30(8).

Bakhtiari, A. A., and Hematian, A. (2013). Precision farming technology, opportunities and difficulty, *International Journal for Science and Emerging Technologies with Latest Trends*, 5(1), 1-14.