



Luntian: An Automated, High-Throughput Phenotyping System for the Greenness and Biomass of C4 Rice

Rhett Jason C. Buzon¹, Louis Timothy D. Dumlao², Micaela Angela C. Mangubat³,
Jan Robert D. Villarosa⁴ and Briane Paul V. Samson^{5,*}
^{1, 2,3,4,5} College of Computer Studies, De La Salle University-Manila
^{5,} briane.samson@dlsu.edu.ph*

Abstract: Plant phenotyping is a vital process that helps farmers and researchers assess the growth, health, and development of a plant. In the Philippines, phenotyping is done manually, with each plant specimen measured and assessed individually, but this process is laborious, time-consuming, and prone to human error. Automated phenotyping systems have attempted to address this problem through the use of cameras and image processing, but these systems are proprietary and designed for other plants. In order to alleviate this problem, the research aims to develop a high throughput rice phenotyping system that automates the identification of plant greenness and plant biomass of C4 Rice. This research group is currently developing a system that provides an efficient way of phenotyping C4 rice by automating the process. The system is being created and tested with consultation and feedback with the International Rice Research Institute (IRRI). It is based on various image processing techniques, existing tools and screen house setup used by IRRI. The system is now able to output Leaf Color Chart values for greenness and cubic centimeters for biomass approximation. The plant images are captured based on a specified schedule then these will be processed by the system. A web application is used to view the collected phenotypic data, and change system settings. Results show that the biomass generated by the system is moderately correlated (0.674) to the manually computed biomass while the greenness values are still undergoing verification.

Key Words: Rice Phenotyping; Image Processing; Greenness; Biomass; C4 Rice



1. INTRODUCTION

Plant phenotyping is the process of gathering the observable traits of a plant in order to assess its growth, health and development, which is vital in the assessment of its more complex traits (Helmert & Lasinger, 2010). This process helps in finding out how to increase the yield and resilience of crops. In the middle of 1990s, plant breeders depend on intuition to select different traits that can increase crop yield or resiliency, but with the advent of modern genetics, scientists and farmers now have the capability to breed crops selectively with precision and accuracy (Finkel, 2009), they still need to grow, and analyze the genetic traits of these plants to aid in selective breeding making plant phenotyping vital in the selective breeding process (Helmert & Lasinger, 2010).

2. RELATED WORKS

2.1 *Related Studies on Automated Plant Phenotyping Systems*

Existing technologies utilize image processing to automate plant phenotyping. The Scanalyzer 3D Phenotyping platform (Eberius, 2014) developed by LemnaTec is an automated plant phenotyping system that can phenotype mature plants, such as corn, tomato and rice. It is capable of simultaneously phenotyping plants in large quantities by automatically moving them into a stereoscopic camera by placing all the plants in a conveyor belt. For rice, the system uses the HTS Bonit, an image processing software that analyzes the area, color, and height of the leaves (Eberius, 2014). While the Scanalyzer 3D has been properly tested and deployed, the platform is proprietary and the investment on infrastructure to deploy it is expensive (Tsaftaris & Noutsos, 2009). PHENOPSIS (Granier, et al., 2006) is an automated system by Optimalog that uses image processing for phenotyping *Arabidopsis thaliana*. It uses a mechanical arm to optimally position a displacement sensor, and a camera on each plant to collect phenotypic data such leaf area, leaf thickness, and proportions of leaf tissues. PHENOPSIS is a proprietary solution designed for the French National Institute for Agricultural Research. The system only handles phenotyping of *Arabidopsis thaliana* and it is only designed to work with screen houses used by the institute (Granier, et al., 2006).

2.2 *Related Studies on Image Processing Techniques and Tools*

HTPheno is an image processing pipeline specifically made for plant phenotyping and was designed by Hartmann et.al (Hartmann, Czauderna, Hoffman, Stein, & Schreiber, 2011). HTPheno is not designed to any specific phenotyping setup and so it is made to be flexible and highly adaptable to different plant phenotyping setups and environments. HTPheno can analyze and collect 6 different plant phenotypic traits using only the top and side view images of a plant specimen.

The HTPheno pipeline makes use of different image analysis algorithms such as region definition, object segmentation, morphological operation and finally, compilation of the analysis results. For the software to properly analyze the phenotypic traits of a plant, calibration is first done using image segmentation, which partitions an image into different components or segments. This is done through color image segmentation with the multidimensional histogram thresholding approach in both Red Green Blue (RGB) and Hue Saturation Value (HSV) color spaces. Segmentation is done on these two color spaces instead of just one in order to accommodate varying light conditions. One drawback to this approach is the case wherein foreign objects close to the plant have a color similar to it, as that object will be segmented into the plant as well.

After all the segments in the image have been identified, the object of interest, the plant segment, is extracted in the image. To reduce the drawback of foreign objects being included in the plant segment, morphological opening is applied in the extracted image. Morphological opening performs erosion, or reducing the segment of interest by eroding its sides, and then dilation, which expands the segment of interest by enlarging or dilating its sides. Morphological opening removes small foreign objects in the plant segment because of erosion, and it smoothes the sides of the plant segment through dilation. This technique results in a plant segment with a lower noise level.

Finally, the plant segment is analyzed for phenotypic data. The plant segment is transferred back to the original image, which now forms an outline of the plant. From the segment, the software can calculate various plant phenotypic data and outputted for analysis.



Another system used an image processing pipeline for gathering phenotypic data for *Arabidopsis thaliana* (Arvidsson, Perez-Rodriguez, & Mueller-Roeber, 2011).

To analyze and collect phenotypic data from an image, the first step is to convert it from colored to an 8-bit grayscale image, with the process assigning relatively greater pixel intensity for green pixels. Then, image segmentation is performed on the image using a binary mask. The binary mask is created by selecting pixels with a grayscale intensity greater than 130, which means the brighter areas of the image, or areas of the images that were originally greener, are selected. Afterwards, holes inside this binary mask that are caused by either particles resting on the plant or spots in the leaves are filled. The binary mask is then converted into objects, to create the plant segment. To smooth the leaf edges and reconnect some objects of the plant segment that became disconnected during the segmentation process, the objects are also dilated and eroded. From this, phenotypic parameters are calculated for the plant segment and are saved into the database. The images generated by the pipeline are shown, and the user can choose to detect and remove incorrectly identified images, as well as make annotations of the image to include additional information.

Despite the ability of the current systems to automate plant phenotyping by way of image processing, these systems are not designed for crops that can be grown in the Philippines. Additionally, these systems are expensive to deploy. Thus, there is a research opportunity to create a phenotyping system that will automate the process through image processing but will be more adaptable to the Philippine setting.

3. LUNTIAN

Luntian is an image processing system that provides automated, high-throughput, phenotyping that can determine the greenness and biomass of C4 rice. It aims to replace the manual plant phenotyping process being currently used by the International Rice Research Institute (IRRI). As the system is automated, it will have a relatively higher throughput than that of the manual process. In addition, the system will also provide a way to phenotype different types plant specimens in batches, without losing considerable accuracy.

Luntian will be designed to work together with an image gathering hardware setup that will automatically capture images of plant specimens that will be fit for image processing. The captured images will also be automatically sent to the system for pre-processing and phenotyping.

The system will utilize numerous image processing algorithms that will be used to determine the greenness and biomass of plant specimens. Image processing algorithms will also be used to normalize images and reduce the impact of changing environmental conditions to the phenotyping process. The system will also reduce noise from the images that might affect the phenotyping process.

Luntian will contain a database. The database will contain the raw images gathered by the cameras and the phenotypic data from the greenness and biomass modules. This ensures that the raw images and the phenotypic data from the system can be used by researchers to monitor the progress of the C4 Rice experiment.

4. SYSTEM FUNCTIONS

4.1 *Data Capturing Module.*

The Data Capturing Module is responsible for scheduling the phenotyping process, gathering the images, and putting the raw images into the database. The module will have a calendar or list of dates that will keep track of the next phenotyping session. It will also trigger the cameras in the screen house to capture images when the phenotyping session is about to take place. The images that will be captured by the cameras will be sent to the module and the raw images will be saved to a database.

4.2 *Pre-processing Module.*

The Pre-processing Module will use the raw images from the Data Capturing Module in order to prepare the images and process them for phenotyping. Image normalization is done first in order to fix the contrast and color-balance of the image. This is a crucial step, especially in determining the greenness of the plant, wherein a change in contrast will also change the color intensity of the plant specimen in the image. Image segmentation is then employed to isolate the plant specimen from the background image. Noise filtering is done on the segmented image by using Morphological opening and closing. Dilation

is applied on the white background to effectively erode the plant segment, and then erosion is applied on the white background which will dilate the plant segment. The main flow of the Pre-processing module is shown on Figure 1.

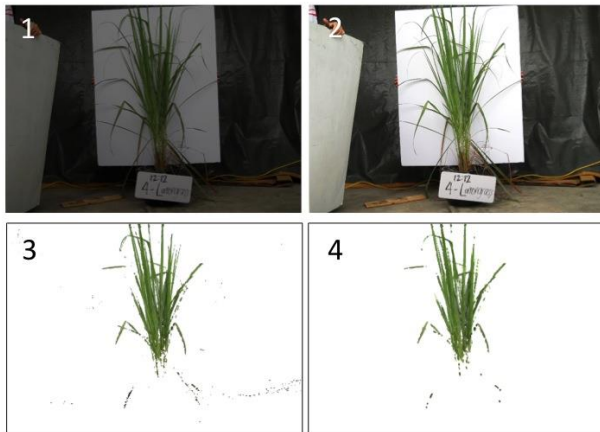


Fig. 1. The process of preparing the raw image for phenotyping through the pre-processing module. Image 1 is the raw image, image 2 is after normalization, image 3 is after image segmentation, and image 4 is after noise filtering.

4.3 Greenness Module.

The Greenness Module will analyze the pre-processed images and determine the greenness of the plant by its LCC values. To determine the greenness of the plant specimen, the green intensity of the plant will be looked at Hue value of the HSV color space. Since different parts of the plant segment will produce a different green intensity because of lighting, and plant characteristics, the overall greenness of the plant will be done by getting the average Hue value. The average Hue value of the plant will be then compared to the different Hue values in the LCC panels. The LCC values of the plant specimen will be determined by the LCC color values closest to the average value of the plant segment. After determining the LCC value of the plant specimen, the phenotypic data will be saved to the database that can be retrieved by the users in the future.

4.4 Biomass Module.

The Biomass Module will analyze the pre-processed images and determine the biomass of the plant. The biomass can be approximated by

determining the volume of the plant. This can be computed by getting the plant's height and width. This computation relies on the assumption that when the plant is compacted into a tight cylinder, its approximated volume will be close to the cylinder it's contained in. Plant width is determined by first, approximating where the approximate tiller base should be. The tiller base is approximated by counting the number of pixels that intersects two vertical lines representing the sides of the supposed cylinder; the more pixels of the plant that intersects the lines, the more likely that the tiller base is located in the vertical coordinate. This approximation technique is shown in Figure 2, with the black segment being the product of the technique. Once the two lines are fixed into place, the black pixels are counted to get the plant width in pixels, which is then converted to centimeter. After determining the approximate biomass of the plant specimen, the phenotypic data will be saved to the database that can be retrieved by the users in the future.



Fig. 2. The process of determining plant width in order to approximate the plant biomass. The left side of the image shows the approximated tiller base superimposed on the plant while the right side shows the enlarged version.

4.5 Data Management Module.

The Data Management Module will serve as the main interface of the system. The user can view use this module to view the collected phenotypic data, and change system settings. This component will be a web application in order to make the phenotypic data accessible to IRRI researchers in any location. Users can changes settings in the Data Capturing Module



such as adding and editing phenotyping appointments in the calendar, and configuring test beds and the data capturing setup in order to have control over the components deployed inside the screen house. Raw images and images produced by the system will also be stored in the database and will be accessible to the user for research related purposes.

5. DEVELOPMENT PROGRESS

Currently, the system is able to determine the greenness and biomass of the plant from an image manually uploaded by the user, but the system can be switched to accept inputs from the data capturing module during the actual system deployment.

5.1 *Preprocessing*

Pre-processing raw images consists of white balancing, image segmentation, and noise filtering. For white balancing, two algorithms were tested: Histogram Equalization (`equalizeHist()`) and Minmax Normalization (`normalize()`). Histogram Equalization attempts to improve the contrast and lighting situation in an image by stretching out the intensity range of the histogram. Unfortunately, applying this algorithm on the image produced an overcorrected color balanced image with dark parts of the image becoming too dark. Minmax normalization, on the other hand, was applied to the Saturation and Value of the HSV color space utilizing the full range of values (0 – 255). Normalization scales and shifts the original values in the Saturation and Value space to produce a more natural color balanced image without over-correction.

For the image segmentation, the HSV thresholding algorithm was used instead of RGB thresholding in order to account for changing light conditions. RGB thresholding was found to have a problem with plants in the flowering stage, which had a more yellowish hue, because the green value of these plants is not higher than the red and blue values, which is interpreted by the algorithm as the object not being a plant. Using the HSV thresholding algorithm did not have these problems and an HSV range (22 – 75% for Hue 23 – 100% for Saturation, 11 – 100% for Value) was used to segment the plant out of the background.

Another algorithm for image segmentation was used mainly for processing Biomass. Binary Thresholding with a range of 170 – 255 was applied to

the white balanced image to isolate the plant from the background. Binary thresholding was more effective in getting the plant segment for Biomass because it can get the darker parts of the plant, which might not be included in the HSV thresholding algorithm.

5.2 *Greenness*

For getting the greenness of the plant, the system gets the mean Hue value of the plant and compared to the Hue values of the colors in the LCC. Currently, another algorithm is being tested, which is getting the mode Hue value of the plant segment and comparing that the Hue values of the LCC.

5.3 *Biomass*

The radius of the plant is determined and at the moment, the actual height is taken from the database in order to estimate the biomass of the plant using its volume. Different techniques were implemented and tested in order to determine the radius. First, each row of pixels in the lower half of the plant were counted and averaged. The second approach is by taking two green endpoints for each row then computing the distance in between. The results from the two techniques were unsatisfactory due to the absence of binary conversion so another approach was tested. Since the image was not converted to binary and the width depends on the pixels which are not white (image pointer value is not equal to 1), a lot of noise influenced the values. After including a binary image conversion, the image is cropped at a specific size in order to lessen the noise. The width is then measured by finding the tiller base of the plant and counting the pixels. The pixel count is then converted to centimeters by using the size of the pot as a reference.

Other modules that are integral to the system are in progress and scheduled to be tested and validated. The Data Capturing Module is currently being implemented along with the database for saving images and phenotypic data. The Data Management Module, which will serve as the front-end user interface for the system, is also currently being implemented in consultation with IRRI.

6. Initial Results and Analysis

Initial results show that there is moderate correlation between the system-generated estimated values and the manually collected measurements done by the IRRI researchers. The correlation was

generated using RapidMiner, a tool used for data analytics. The estimated values and the manually collected values were imported to the tool then the data analytic computation needed, correlation, is chosen as output. After processing, the result was a correlation value of 0.674. A visual comparison is shown in Figure 3. Based on the graph, there is a possibility that the system can be able to predict the biomass. Although there are outliers, it can be seen that there is a positive correlation between the two values.

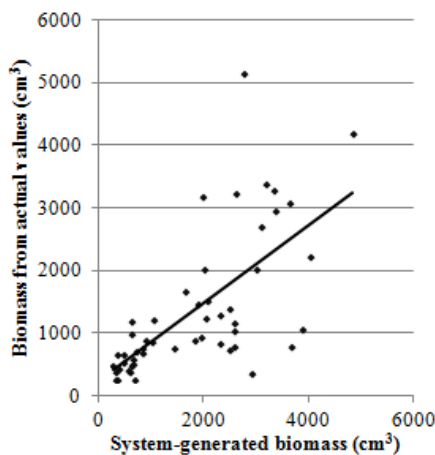


Fig. 3. The comparison of the results using the actual measurement values and system-generated values.

It can be observed that the differences between the values are inconsistent. There are overestimates, underestimates and best estimates. In order to determine the cause of the discrepancy, samples were taken from each type of estimate.



Fig. 4. Overestimate sample. Binary (left) and preprocessed (right) images of plant IR64-038.

For the overestimated value, it can be observed that the size of the cropped image is adjusted to fit the tiller base of the plant. The image, as seen in Figure 4, still contains leaves that are considered

noise and these affected the value taken from the plant and increased the pixel count

Another factor to be considered is the color of the plant. Figure 5 shows the trend of the differences between the actual and system generated radius. The plants numbered 1 to 20 are from the tillering stage, 21-40 are from the mid-tillering stage and 41-60 are from the flowering stage. The plants in each stage have different shades of green. As the plant shifts to another stage, the color becomes more yellowish. Based on the graph, it can be inferred that the closer the shade of the plant to yellow, the more under-estimated result becomes. This is reflected further in the individual samples in Figure 6 and Figure 7.

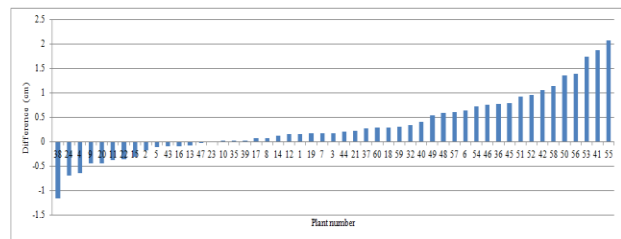


Fig. 5. Difference between actual and system generated radius.

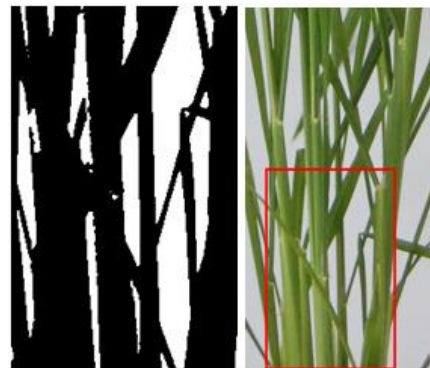


Fig. 6. Best estimate sample. Binary (left) and preprocessed (right) images of plant IR64-010.





Fig. 7. Under-estimate sample. Binary (left) and preprocessed (right) images of plant IR64-055.

It is possible that this is due to the green thresholding which is done in order to separate the plant from the background. Therefore, further improvements in color thresholding should be implemented in order to include more shades of yellow in the binary conversion and eventually in the pixel count. Another issue is the vertical threshold in cropping the tiller base. The threshold value should be modified in order to avoid overestimates by removing noise such as leaves from the image as shown in Figure 4.

7. REFERENCES

- Arvidsson, S., Perez-Rodriguez, P., & Mueller-Roeber, B. (2011). A growth phenotyping pipeline for *Arabidopsis thaliana* integrating image analysis and rosette area modeling for robust quantification of genotype effects. *New Phytologist*, *191*(3), 895-907.
- Eberius, M. (2014). Lemnatec HTS Bonit: Image Analysis for the Quantification of Rice in 2-D and 3-D Assays. Lemnatec.
- Finkel, E. (2009). With 'Phenomics,' Plant Scientists Hope to Shift Breeding Into Overdrive. *Science*, *325*(5939), 380-381.
- Granier, C., Aguirrezabal, L., Chenu, K., Cookson, S. J., Dauzat, M., Hamard, P., . . . Tardieu, F. (2006). PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in *Arabidopsis thaliana* permitted the identification of an accession with low sensitivity to soil water deficit. *169*(3), 623-635.
- Hartmann, A., Czuderna, T., Hoffman, R., Stein, N., & Schreiber, F. (2011). HTPheno: an image analysis pipeline for high-throughput plant phenotyping. *BMC Bioinformatics*, *12*, 148.
- Helmert, M., & Lasinger, H. (2010). The Scanalyzer Domain: Greenhouse Logistics as a Planning

Problem. *IAAA Publications*. AAAI Publications. From <https://www.aaai.org/ocs/index.php/icaps/icaps10/paper/view/1423>

Tsaftaris, S., & Noutsos, C. (2009). Plant Phenotyping with Low Cost Digital Cameras and Image Analytics. *Environmental Science and Engineering*.