

Towards an Automated, High-Throughput Identification of the Greenness and Biomass of Rice Crops

Rhett Jason C. Buzon*, Louis Timothy D. Dumlao[†], Micaela Angela C. Mangubat[‡],
Jan Robert D. Villarosa[§] and Briane Paul V. Samson[¶]

*[†][‡][§][¶]College of Computer Studies, De La Salle University - Manila

*Email: rhettbuzon@gmail.com

[†]Email: louis.dumlao@yahoo.com

[‡]Email: micaela.mangubat@gmail.com

[§]Email: janrobertvillarosa@gmail.com

[¶]Email: 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 one by one. However, 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 plants and crops which are not commonly found in the Philippines. In order to alleviate this problem, research was conducted to develop an automated, high-throughput phenotyping system that automates the identification of plant greenness and plant biomass of rice. The system was developed in order to provide an efficient way of phenotyping rice by automating the process. It implements various image processing techniques and was tested in a screen house setup containing numerous rice variants. The system's design was finalized in consultation with and tested by rice researchers. The respondents were pleased with the system's usability and remarked that it would be beneficial to their current process if used. To evaluate the system's accuracy, the generated greenness and biomass values were compared with the values obtained through the manual process. The greenness module registered a 21.9792% mean percent error in comparison to using the Leaf Color Chart. On the other hand, the biomass module yielded 206.0700% mean percent error using compressed girth measurements.

I. 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 [1]. 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 [2]. However, 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 [1].

II. RELATED WORKS

A. Related Works on Automated Phenotyping Systems

Existing technologies utilize image processing to automate plant phenotyping. The Scanalyzer 3D Phenotyping platform [3] 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 [3]. While the Scanalyzer 3D has been properly tested and deployed, the platform is proprietary and the investment on infrastructure to deploy it is expensive [4].

PHENOPSIS [5] 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 [5].

The paper published by Tsiftaris and Noutsos [4] details a setup that is low cost and easy to deploy in nature, making it the most suitable model among all the other researched related systems to Luntian. In order to achieve such characteristics, the system discussed in the paper makes use of digital cameras which are inexpensive in terms of mass purchase and satisfactory in taking the required images. Additionally, each camera has been set up to utilize an open source firmware called Canon Hack Development Kit. This firmware allows further manipulation of digital camera options which in turn, allow the researchers to tweak the settings to various factors that are present in the system environment. Examples of some manipulated settings are manipulation of the ultra intervalome-

ter for taking time lapse images and utilizing the long exposure intervalometer for taking night time photos.

B. Related Works on Image Processing Pipelines

HTPheno is an image processing pipeline specifically made for plant phenotyping and was designed by Hartmann et.al [6]. 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* [7].

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.

III. LUNTIAN

A system called Luntian, the Filipino word for "green", was conceptualized and developed which provides automated, high-throughput, phenotyping that can determine the greenness and biomass of rice crops. The manual process of phenotyping greenness and biomass takes a minimum of 24-48 hours to complete. The time taken in manual phenotyping also increases significantly as more test beds are included in the phenotyping process. As the system is automated, it is going to have a relatively higher throughput than that of the manual process. In addition, the system also provides a way to phenotype different types plant specimens in batches, without losing considerable accuracy.

Luntian is designed to work together with a data gathering hardware setup that automatically captures images of plant specimens that is fit for image processing. The captured images are automatically sent to the system for preprocessing and phenotyping.

The system utilizes numerous image processing algorithms that are used to determine the greenness and biomass of plant specimens. Image processing algorithms are also used to normalize images and reduce the impact of changing environmental conditions to the phenotyping process. The system also reduces noise from the images that might affect the phenotyping process. Luntian is built on OpenCV, an image processing library that provides functions to perform the aforementioned algorithms.

Luntian implements a database that contains the phenotypic data from the greenness and biomass modules and the file path of the raw images gathered by the cameras. The actual raw images gathered are stored in a separate directory. This implementation allows the researchers to easily access these needed data in monitoring the progress of the rice crops.

Luntian is just one of two components in the Butil system, which aims to provide researchers an automated method of plant phenotyping. The other component is Seight, which, through the use of different image processing algorithms, obtains the plant height and tiller count from the images taken by the hardware setup. Luntian and Seight share the same image gathering setup, and both interface with the same data management module. Even with the separation of components,

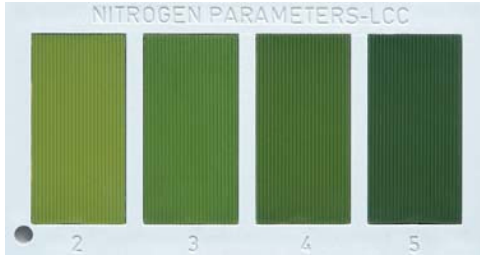


Fig. 1. The Four Panel Leaf Color Chart

the intention is for Butil to be ran and used by crop scientists as one complete system.

A. System Functions

1) *Data Capturing Module*: The Data Capturing Module is responsible for gathering the data for phenotyping. Researchers can schedule the phenotyping dates or capture appointments. These appointments will trigger the camera and a certain date and time, gather the raw images captured and store them into the database. Remotely triggering the cameras is done using the OpenCV API. The use of this library not only enables the cameras to be triggered through functions that are executed according to the scheduled camera appointment, but can also immediately send the captured images to the image processing modules since OpenCV is also used in these modules.

2) *Preprocessing Module*: The Preprocessing Module uses the raw images gathered by Data Capturing Module and prepares and processes them for automated phenotyping. Color balancing is done through a technique called Normalization. This is done through shifting the values the Saturation and Value channels in the HSV color space to produce a softer curve in the histogram. Normalization in the system is done with the OpenCV function `normalize()` which attempts to normalize the value range of the two color spaces by shifting and scaling the values. After correcting the color balance, Segmentation then isolates the plant from the whole image, making phenotyping easier and more accurate. The segmentation algorithm used in this module is Otsu Thresholding which considers the darkness intensity of pixels in grayscale. Finally, noise filtering is done on the segmented image using morphological operations.

3) *Greenness Module*: The manual method of phenotyping involves comparing the color of the plant with the Leaf Color Chart (LCC). The Four Panel LCC, shown on Figure 1 determines the chlorophyll content or greenness of the plant in four values (2,3,4,5). The Greenness Module analyzes the pre-processed images and determines the greenness of the plant by its LCC values. To determine how green the plant truly is, the greenness intensity of the plant will be looked at using the mean Hue value of the plant. Since the Hue of the plant is not reliant on the Saturation and Value spaces, it provides a more thorough conditions that will be less affected by lighting conditions. The mean Hue of the plant is compared to the Hue values of the four LCC panels. The LCC value of the

plant is estimated by using the LCC Panel's Hue value closest to the plant's mean Hue value. After determining the LCC value of the plant specimen, the phenotypic data is saved to the database, allowing researchers to retrieve the estimated phenotypic data in the future.

During the process of development, the Hue value of the LCC has to be retrieved as comparison points to the mean Hue value of the plant. The Hue values of each LCC panel were not available in any of the manufacturer's documentation. To get the values, two methods were considered for the experimentation process. The first method is to get the Hue value by taking a picture of the LCC and defining that as a Hue value for all image samples. The second method to dynamically change the LCC value for each image sample.

In the first method, the Hue values of the LCC were retrieved by capturing a photo of the LCC and determining its Hue values in Adobe Photoshop. Since OpenCV stores Hue values in integers 0 -180, and Adobe Photoshop determines Hue values in 0 - 360 degrees, the values retrieved in Photoshop is divided in half. The Hue values of each LCC panels are in I.

TABLE I
CORRESPONDING LCC AND OPENCV HUE VALUES USING THE FIRST METHOD

LCC Panel	OpenCV Hue Value
2	32
3	44
4	57
5	80

In the second method, the Hue values were retrieved by dynamically getting the Hue values through sampling the LCC attached on the board. The LCC was attached to the segmentation board as a point of reference, and this region in the image samples was isolated. After isolating the LCC in an image each panel was sampled to dynamically retrieve the Hue values of each LCC panel.

The first and second algorithms of determining greenness were tested and compared with each other. The discussion of the results of both algorithms will be detailed in the Results section. Since the first algorithm obtained the lesser percentage error between the two, it was then used as the final algorithm in the system.

4) *Biomass Module*: The Biomass Module analyzes the pre-processed images and approximates the biomass of the plant by estimate its plant volume. Plant radius is first determined by counting the back pixels per row of the resulting binary image. The pixel counts are then averaged to get the mean pixel count. This mean count is converted from pixels to centimeters and used as the plant radius. Approximation of biomass is done by estimating the plant's volume when packed inside a cylinder. The formula for the volume of a cylinder is used to compute the approximate biomass of the plant.

$$volume = \Pi radius^2 height \quad (1)$$

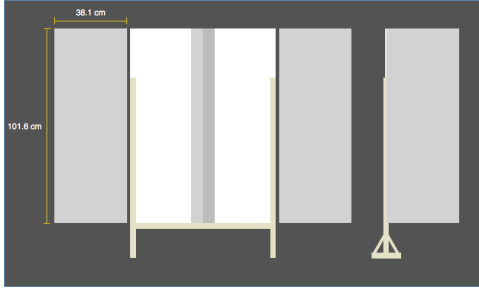


Fig. 2. Separation Board specifications

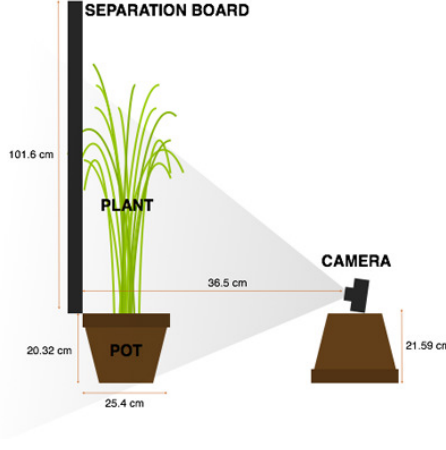


Fig. 3. Measurements and Specifications for the Data Gathering Setup

The height will be retrieved from the *Seight* system which approximates the height of the plant.

5) *Data Management Module*: The Data Management Module is the main interface of the system. The user can use this module to view the collected phenotypic data, and change system settings. This component is in the form of a web application in order to make the system accessible to the researchers in any location.

subsectionPhysical Environment and Resources

Luntian works alongside a hardware setup that will capture the images needed for phenotyping. As a proof of concept, the data capturing setup was tested on one test bed in a screen house. The test bed has two plants placed in front of the camera. These two plants serve as the representatives of the whole test bed and are used for phenotyping. A separation board was used to isolate the plant specimen from the background. Specifications of the separation board is detail in Figure 2

IP cameras were used in order to make use of the remote capturing functionality available in the OpenCV library. The cameras must be positioned in the screen house properly for optimal data gathering. The measurement and specifications for the placement of the data gathering setup are shown in Figure 3

The IP camera that is used in the setup during the time of development and testing is D-Link DCS-932L. It has a

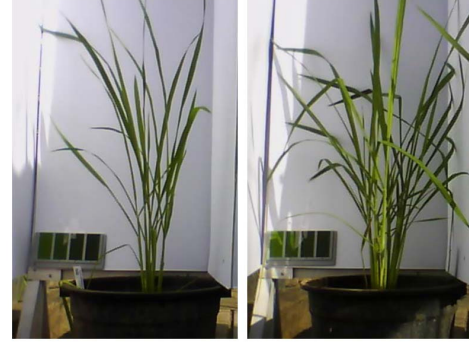


Fig. 4. Mid-tillering growth stage (left) and Tillering growth stage (right)

resolution 640x480 Video Graphics Array (VGA) resolution and is compatible with the OpenCV framework.

IV. RESULTS

A. Testing Methodology

For the purposes of testing the accuracy of the system's greenness and biomass algorithms, 28 plant specimen of C4 rice varying in size, greenness and plant stages were chosen for data capturing. These plants were chosen at random to give variety to different plants. At the time of capturing, 14 plants were on the mid-tillering growth stage and 14 other plants were on the tillering growth stage. The two growth stages give different plant structure and color to the plants and this gives a more thorough characteristic in the testing data. This can be seen in Figure 4.

Following the planned data gathering setup, 28 images were captured by the system, one for each plant. After data capturing, crops were manually phenotyped for plant greenness (using the LCC), girth of the plant when tightly compressed, and girth of the plant when loosely compressed. In order to measure the girth when tightly compressed, the tillers are compacted as close as possible to each other to reduce the gaps in between then a tape measure is used for determining the circumference or girth. For the loosely compressed girth, the tape measure is placed around the edges of the plant without compacting the tillers. After assessing the measurements, the compressed girth was used in comparing the results since it is more appropriate to the formula used for computing the biomass which was defined in the previous chapters. The manual phenotyping process was done by the researchers to ensure the accuracy of the manual phenotypic data. The data taken from manual phenotyping were used to assess the accuracy of the system.

The mean error difference was computed by getting the mean of the difference between the manual phenotyping results and the automated phenotyping results. The percentage error for each specimen was computed using the formula below. The mean percentage error was computed by averaging the percentage error for each specimen.

$$percenterror = \frac{automatedval - manualval}{manualval} \times 100 \quad (2)$$



Fig. 5. An image segmentation of the plant in an image with a resolution of 2000x3008

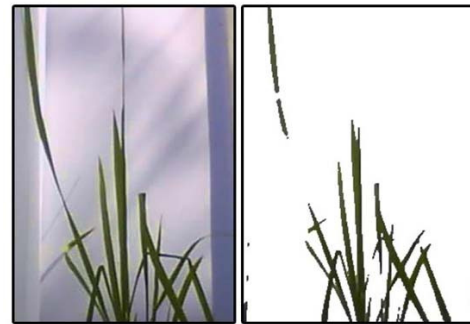


Fig. 6. An image segmentation of the plant using the 640x480 resolution image captured by the system

B. Greenness

In the development of greenness, two algorithms were tested in parallel so results can be compared with one another. The first algorithm is to rely on static Hue values of the LCC for all images. The other algorithm is to rely on dynamic Hue values of the LCC that changes for every image. Two results of the two algorithms will be discussed and compared in detail.

After processing the 28 images through the first algorithm, the system has registered a mean percentage error of 21.9792%. While the accuracy displayed by the system seems fairly high, the small range of values (2-5 LCC values) means the percentage error in the system can result in discrepancies between the actual value and approximated value significant enough for the greenness algorithm to become unreliable. The mean error difference of the LCC value generated by the system compared to the Manual Phenotyping method is 0.7083, so on average the LCC Value generated by the system can have an error up to more or less .71 the original value.

The second algorithm however, yielded a much higher mean percentage error of 33.2066%. This means that using the second algorithm with the dynamic Hue values has a less accurate estimation than using the first algorithm, although it has the advantage of being more adaptable to different conditions. The mean error difference for the second algorithm is 1.0833, which is also higher than the first one. In the second algorithm, the error difference can mean it can over or underpredict by 1 LCC value.

Further analysis show that a number of factors that may affect the accuracy of the system. One major factor that affects results is the resolution of the camera. The 640 x 480 pixel resolution results in less pixels being used by the algorithm for determining greenness. The resolution also affects the segmentation of the plant (seen in Figures 5 and 6) and this is because the delineation and edges of the plant are clearer and JPEG compression is less evident on larger images. Segmentation in lower resolution images can result in loss of data with the upper tillers not being segmented as part of the plant.

Another factor that can affect the accuracy of the greenness algorithm and its result is the effect of the daylight conditions inside the screen house. Shadows and uneven lighting can

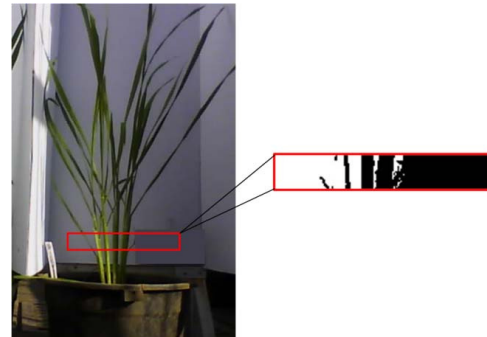


Fig. 7. Sample of removed image because of improper thresholding

cause the color of the plant to vary when compared to actual inspection of the plant. Plant images were taken in daylight conditions at 2:45 PM. External lighting equipment was not used in the hardware setup, so daylight and shadows caused un-optimal lighting conditions.

C. Biomass

After processing the images with the algorithms, some images were removed due to excessive noise that will affect the results. A sample of removed images is shown in Figure 7. It can be seen that even if the area is white, it was not successfully removed and was included as part of the plant. The error is the same for all removed images. The final dataset contains 23 samples.

Biomass results show that there is no exact trend in the differences of the measurements from the manual method with compressed radius, and the automated method values. There are also large differences between the values. This is caused by already having differences in the estimated radius values from the system, which greatly affected the estimation of biomass. Biomass values computed using compressed girth have a mean percentage error of 206.0700% while the mean percentage error of the radius for compressed girth is 66.7573%. Since there are already large percentage errors in estimating the radius, the values highly affected the results. Another factor that affected the results is the conversion of pixels to cm. Using the average number of pixels from all images is not sufficient to provide accurate results. Since the conversion was not exact,

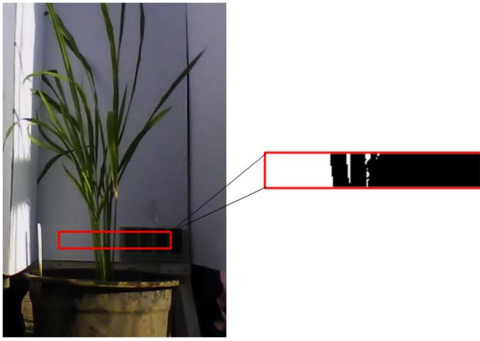


Fig. 8. Image sample with preprocessing issue

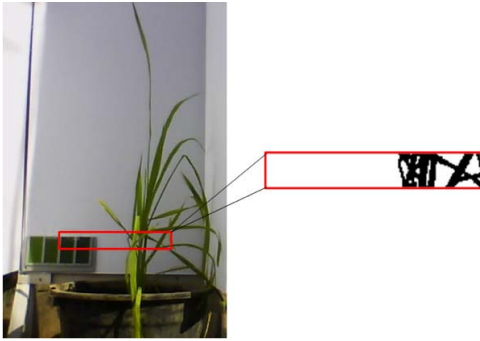


Fig. 9. Image sample with plant structure issue

there was no exact trend in the differences of the results of the automation with the manually measured values.

The results show that there are outliers in the dataset. After further analyzing some of these outliers, it can be inferred that the presence of dark portions affect the measurements since these were not completely preprocessed in the binary conversion of the pictures. Despite the fact that the system was already improved by cropping the image first before converting to binary in order to reduce the lighter shadows, there were still dark portions which the system was not able to successfully differentiate from the plant. Figure 8 shows a sample image with its corresponding binary equivalent generated by the system. In order to address this issue, the images with excessive noise were removed as mentioned earlier. For further studies, the lighting setup should be improved for better results.

It can be seen that in the right portion of the cropped binary image, the dark portion behind the plant was not successfully preprocessed. Since this image is the one used for estimating the radius, the result was a greater radius value than the manually measured radius of the compressed and loose plant.

Another issue is that plants have different structures. There are some which have leaves in the region of interest which also became noise as shown in Figure 9. In addition, there are also plants that appear wide in the front view but are actually linearly distributed. Automated results from such images may be larger than the actual since the tillers are not compressed. It could be possible to tie the tillers together before taking the picture in order to have more accurate results.

V. CONCLUSION

The Luntian system was created for the benefit of experimental crop rice researchers to address their needs of an automated system that can speed up the phenotyping process. Throughout the study and the development of the system, consultations were made with the C4 rice researchers to match their manual phenotyping process and translate it to the automated system. Data from their manual phenotyping process were used as comparison data and metrics for determining the accuracy of the completed system.

Two greenness algorithms were tested in the research. The first algorithm, which relied on static LCC hue values for all images, yielded the most accurate results between the two in estimating the actual LCC values of the plant. The second algorithm, which relied on dynamic LCC values that is newly sampled for every image, yielded a less accurate result even though it has the advantage of being more robust than the first algorithm. While the determining greenness is shown to be accurate based on the results, both algorithms can prove to be unreliable to the researchers because of the relatively large average percent error especially when considering the level of accuracy needed for their research. The relative inaccuracy and unreliability of the algorithm is mainly rooted in how the image is captured, with factors such as uneven lighting and the low resolution of the camera used all playing a significant role in making both algorithms unreliable.

Based on the results, there are significant percentage errors in the biomass and radius estimations and there is no definite trend when compared with the compressed girth values. Further studies on the values may have to be done in order to discover a trend and use it to improve the algorithm. One recommendation is to use distribution fitting in order to find whether there are relationships between the data in order to improve the accuracy.

Overall, results yielded by the study can serve as a step towards automating plant phenotyping, which is a significant help for the researchers in developing their crops. In the future, effective automation would speed up the data gathering process, which would be more evident in larger set-ups. By curtailing the extent of human intervention in the phenotyping process, inconsistencies brought about by human error would consequently be reduced.

ACKNOWLEDGMENT

The authors would like to thank Mr. Briane Samson, Dr. Florante Salvador and Dr. Joel Ilao from the College of Computer Studies of De La Salle University for their guidance throughout the duration of writing the research. They would also like to thank Mr. Alexis Pantola, for providing the Internet Protocol camera used in the development and testing of the system. Finally, they would also like extend their gratitude to the International Rice Research Institute especially C4 Rice Researcher Mr. Albert De Luna for working with us through consultations about plant phenotyping and the C4 Rice project, as well as giving feedback in the development of the Luntian system.

REFERENCES

- [1] M. Helmert and H. Lasinger, "The scanalyzer domain: Greenhouse logistics as a planning problem," in *In International Conference on Automated Planning and Scheduling*, May 2010.
- [2] E. Finkel, "With phenomics, plant scientists hope to shift breeding into overdrive," vol. 325, no. 5939, pp. 380–381, 2009.
- [3] M. Eberius, "Lemnatec hts bonit: Image analysis for the quantification of rice in 2-d and 3-d assays," White Paper.
- [4] S. Tsafaris and C. Noustos, "Plant phenotyping with low cost digital and image analytics," in *Information Technologies in Environmental Engineering: Proceedings of the 4th International ICSC Symposium*, vol. 4, 2009, p. 239.
- [5] C. Granier, L. Aguirrezabal, K. Chenu, S. J. Cookson, M. Dauzat, P. Hamard, J.-J. Thioux, G. Rolland, S. Bouchier-Combaud, A. Lebaudy, B. Muller, T. Simonneau, and F. Tardieu, "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," vol. 169, no. 3, pp. 623–635, 2006.
- [6] A. Hartmann, T. Czuderna, R. Hoffman, N. Stein, and F. Schreiber, "Ht-pheno: an image analysis pipeline for high-throughput plant phenotyping. bmc bioinformatics," vol. 12, no. 1, p. 148, 2011.
- [7] S. Arvidsson, P. Perez-Rodriguez, and B. Mueller-Roeber, "A growth phenotyping pipeline for arabidopsis thaliana integrating image analysis and rosette area modeling for robust quantification of genotype effects," vol. 191, no. 3, pp. 895–907, 2011.