High-Throughput Estimation of Yield for Individual Rice Plant Using Multi-angle RGB Imaging

Lingfeng Duan[1,2], Chenglong Huang[1,2], Guoxing Chen[3], Lizhong Xiong[4],
Qian Liu[5], and Wanneng Yang[1,2,4,*]

[1] Agricultural Bioinformatics Key Laboratory of Hubei Province,
Huazhong Agricultural University, Wuhan 430070, P.R. China
ywn@mail.hzau.edu.cn
[2] College of Engineering, Huazhong Agricultural University, Wuhan 430070, P.R. China
[3] MOA Key Laboratory of Crop Ecophysiology and Farming System
in the Middle Reaches of the Yangtze River,
Huazhong Agricultural University, Wuhan 430070, China
[4] National Key Laboratory of Crop Genetic Improvement
and National Center of Plant Gene Research,
Huazhong Agricultural University, Wuhan 430070, P.R. China
[5] Britton Chance Center for Biomedical Photonics,
Wuhan National Laboratory for Optoelectronics-Huazhong University of Science
and Technology, Wuhan 430074, P.R. China

Abstract. Modern breeding technologies are capable of producing hundreds of new varieties daily, so fast, simple and effective methods for screening valuable candidate plant materials are urgently needed. Final yield is a significant agricultural trait in rice breeding. In the screening and evaluation of the rice varieties, measuring and evaluating rice yield is essential. Conventional means of measuring rice yield mainly depend on manual determination, which is tedious, labor-intensive, subjective and error-prone, especially when large-scale plants were to be investigated. This paper presented an in vivo, automatic and high-throughput method to estimate the yield of individual pot-grown rice plant using multi-angle RGB imaging and image analysis. In this work, we demonstrated a new idea of estimating rice yield from projected panicle area, projected area of leaf and stem and fractal dimension. 5-fold cross validation showed that the predictive error was 7.45%. The constructed model achieved promising results on rice plants grown both in-door and out-door. The presented work has the potential of accelerating yield estimation and would be a promising impetus for plant phenomics.

Keywords: yield estimation, multi-angle imaging, individual rice plant, plant phenotyping, high-throughput.

1 Introduction

The growing population and the impact of changing climate on agriculture have put major crisis on world's food supply[1]. 70% more food is needed by 2050 according to

---

[*] Corresponding author.

the recent Declaration of World Summit on Food Security (www.fao.org/wsfs/world-summit/en/). As the staple food for over half of the world's population[2], rice production is significant for the food security of the world. Modern breeding technologies are capable of producing hundreds of new varieties daily, so fast, simple and effective methods for screening valuable candidate plant materials are urgently needed[3]. Phenotyping the interested populations is widely deemed as the most laborious, costly and technically challenging part in plant breeding. Plant phenotyping has become one of the new bottlenecks in plant science and plant breeding[4, 5]. Plant phenomics enables the acceleration of progress in linking gene function, phenotype, and environmental responses, which will eventually accelerate plant breeding to develop new germplasm to meet the future food demanding.

Digital image analysis has been proved an effective tool in plant phenomics. There has been research on the application of image analysis for phenotyping of an individual plant. Yang *et. al* adopted x-ray computed tomography to measure rice tiller number. Utilization of digital imaging and image analysis, Granier *et al.* estimate growth rates from projected leaf area for Arabidopsis plants [6]. Golzarian *et. al* [7]used the projected shoot area of the cereal plants on images at three orthogonal views and plant age in days after planting as predictors to infer the cereal biomass. Jones *et al.* quantified plant responses to water stress from thermal infrared images of crop canopies[8].

Final yield is a significant agricultural trait in rice breeding. In the screening and evaluation of the rice varieties, measuring and evaluating rice yield is essential. Conventional means of measuring rice yield mainly depend on manual determination. In this method, spikelets were threshed from the panicles and filled spikelets were weighed. Generally, manual measurement is tedious, labor-intensive, subjective and error-prone, especially when large-scale plants were to be investigated.

There has been work on yield-related trait measurement[9, 10]. In these works, rice panicles were harvested from the plant and spikelets were threshed manually or by the threshing machine. The major limit is that, the threshing machine had a threshing error of 3.33% for filled spikelets and 2.27% for total spikelets, and was prone to break the spikelets. Precision agriculture using remote sensing technology, in which the yield of the above-ground canopy for a large area is estimated from satellite and airborne images, is a wide-spread method for yield prediction [11-14]. However, this technology is not capable of estimating yield of an individual plant. For yield estimation of individual plant in vivo, academic publication is unavailable.

The objective of the present study is to develop an in vivo, automatic and high-throughput method to estimate the yield of individual pot-grown rice plant using multi-angle RGB imaging and image analysis. In this work, we demonstrated a model that used mixed variables of projected panicle area, projected area of leaf and stem, and fractal dimension achieved good performance in predicting rice yield. In order to improve the generalization ability of our method, we tested our method on rice plants grown both in-door and out-door and achieved promising results.

2   Experiments and Methods

2.1   Rice Sample Preparation

390 pot-grown rice plants, including 302 rice plants growing in the greenhouse (greenhouse rice) and 88 rice plants growing outdoor (outdoor rice), were used in this study. The samples were harvested at maturing stage and rice yield were measured manually. Yield varied from 7.04 g-43.10 g for the greenhouse rice and 23.68 g-58.43 g for the outdoor rice. The average yield for the greenhouse rice and outdoor rice were 27.55 g and 41.20 g, respectively. The samples were randomly split into two datasets at ratio of 2:1: a training dataset for model construction (260 plants) and a validation dataset for model validation (130 plants).

2.2   Hardware Setup and Image Acquisition

Previously, our group developed a high-throughput facility dubbed H-SMART for measuring rice tiller number[3]. The H-SMART facility used an industrial conveyor to transfer pot-grown rice plant to the imaging area for image acquisition. A rotation platform enabled the rice plants to be lifted and turned. A barcode scanner read barcode of each pot for indexing purpose. To allow measurement of rice yield, we incorporated visible light imaging into the H-SMART facility. As occlusion becomes problematic in rice plants at maturing stage, multi-angle imaging was adopted in the research. Plants were illustrated by fluorescent light tubes both from side and top. Images were taken at every 30°intervals by a Charge Coupled Device (CCD) camera (Stingray F-504C, Applied Vision Technologies, Germany) as the plant was rotated. For each rice plant, 12 images (2452×2056) at different angles were taken. Lighting conditions were constant throughout the research. Image acquisition was performed by NI-IMAQ Virtual Instruments (VI) Library for LabVIEW (National Instruments Corporation, USA). More details concerning the H-SMART system can be found in Yang et al. 2011[3].

2.3   Image Analysis and Feature Extraction

Glozarian et al. proved projected shoot area on two dimensional images was a good predictor of shoot biomass of the cereal plants[7]. And rice yield was the product of the harvest index and the biomass (more precisely, the total dry weight)[15]. It can be inferred that projected shoot area should have a good correlation with yield. Panicle was the organ where grains grow on. So projected panicle area may correlate well with yield. For this reason, we raised a hypothesis that separating the projected shoot area into two parts: projected panicle area and projected area of leaves and stems would reduce the bias in yield estimation. Gong demonstrated the feasibility of predicting yield per m2 for field rice from fractal dimension and texture features of the rice image[16]. In this study, we developed a model using projected panicle area, projected area of leaves and stems, fractal dimension and texture features.

The image analysis and feature extraction was performed in LabVIEW(National Instruments Corporation, USA). Fig. 1 shows the image analysis pipeline for this study. Firstly, the background was removed from the image to get the binary image of rice

plant, from which projected area of rice plant (A) and information dimension (IFD) were calculated. Intensity image of the RGB image was extracted according to Eq. 1. Using the binary image as the mask, intensity component of rice plant, from which texture features including differential box counting dimension (DBC) [17] and 5 histogram features were calculated, can be obtained. 5 histogram features include the mean value (M), the standard deviation (S), the third moment (mu3), the uniformity (U) and the entropy (E). After that, pixel discrimination and region recognition were performed to obtain the binary image of panicle, from which projected panicle area (PA) was calculated.

$$I=(R+G+B)/3 \tag{1}$$

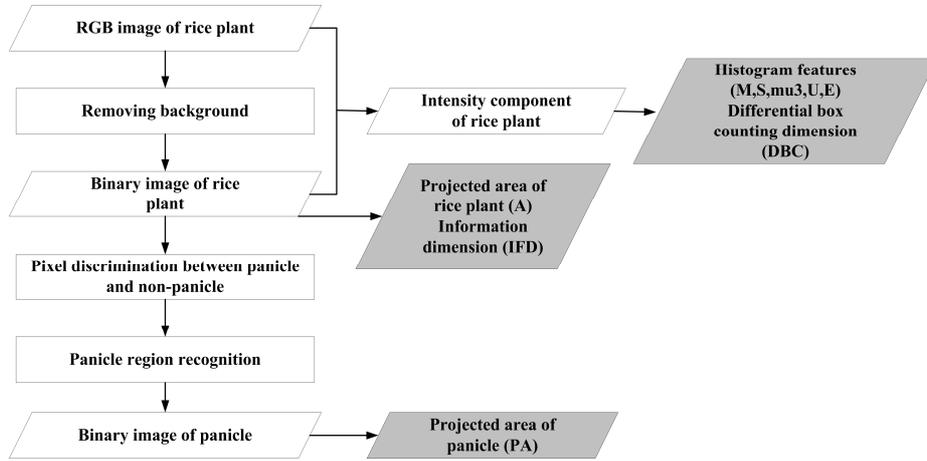where R, G, B was the R, G, B component of the RGB image.

Fig. 1. Image analysis pipeline for this study

Fig. 2 illustrates the image processing diagram for removing background. First, the ExG component and i2 component were extracted according to Eqns. 2-3. The ExG component was used to segment the green parts of the rice plant, and the i2 component was used to segment the yellow parts. The images in this study were captured in a constant illustration environment, so both the ExG component and i2 component were binarized using fixed thresholds. Then the union of the two binary images was calculated. In the end, regions whose area was less than the predefined threshold were removed. Projected area of rice plant (A) was computed by counting the number of foreground pixels in the binary image of rice plant.

$$ExG = 2Ng-Nr-Nb \tag{2}$$

$$i2= (R-B)/2 \tag{3}$$

where Ng, Nr, Nb was the normalized r,g,b component, defined by Eqns. 4-6.

$$Nr=R/(R+G+B) \qquad (4)$$

$$Ng=G/(R+G+B) \qquad (5)$$

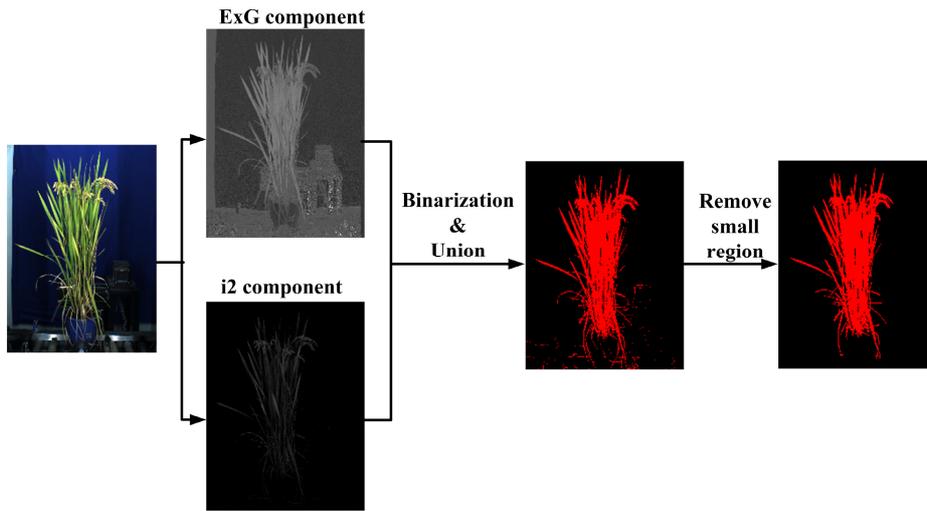$$Nb=B/(R+G+B) \qquad (6)$$

Fig. 2. Image processing diagram for removing background

Fractal dimension can be used to describe shape complexity[18]. The most commonly used method for calculating fractal dimension is box-counting method as it is simple to compute. In this research, the authors adopted information dimension, as it provides a more precise estimate of the fractal dimension than the box-counting method and meanwhile is easy to compute. IFD is calculated by plotting information I(ϵ) (defined as Eq. 7) against the natural logarithm of box size ϵ and IFD is computed as the slope of the regression line.

$$I_\varepsilon = \sum_{i=1}^{N_\varepsilon} p_i \ln \frac{1}{p_i} \qquad (7)$$

Where N_ε is the number of boxes, ϵ is the box size, p_i is the probability of foreground pixels falling into the ith box.

Fig. 3 indicates the image processing diagram for generating binary image of panicle. Pixel discrimination between panicle and stem/leaf (non-panicle) based on discriminant analysis was performed to segment the panicle region[19]. There may be non-panicle region in the extracted panicle regions after pixel discrimination. So, region recognition was used to remove non-panicle regions. Rice yield is mainly contributed by effective panicles, which generally locates in the uppermost one third region of the rice plant. Therefore, in this study, a simple location discrimination method was used to recognize the panicle region. More specifically, the image was

equally divided into three sub-regions according to the plant height. Regions that were located in the uppermost region were regarded as panicle region and other regions were treated as non-panicle regions. Projected panicle area (PA) was computed by counting the number of foreground pixels in the binary image of panicle. Projected area of leaf and stem (LSA) was computed by A minus PA.
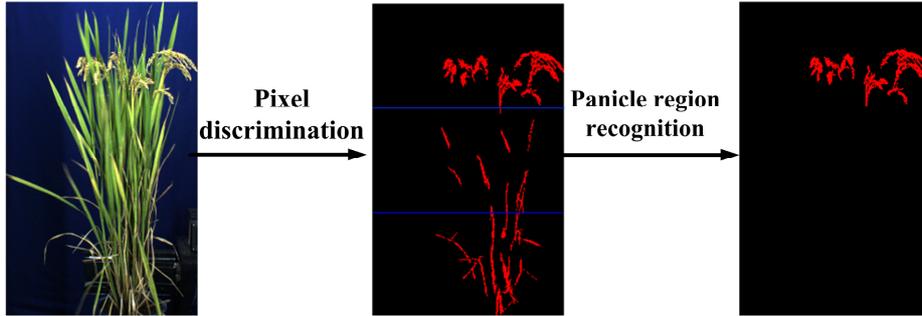

Fig. 3. Image processing diagram for removing background

2.4   Model Construction and Validation

In this study, 12 images at different angles were captured. For each image, 9 features were extracted, including PA, LSA, IFD, 5 histogram features and DBC. For yield prediction, the features were averaged among the 12 angles.

Linear regression analysis is one of the most widely used methods in agriculture [7, 20, 21]. In this research, linear regression was used to predict rice yield. Before modeling, all the data were log-transformed in order to meet normality assumptions and increase model performance[22]. For variable selection and optimal model identification, all-possible regression method was performed, using criteria of Akaike's information (AIC)[23], adjusted coefficient of determination (adjusted R2), and prediction error sum of squares (PRESS statistic)[24]. Adjusted R2 indicates the proportion of variation explained by the predictors and increases on adding a variable only when the variable decreases the residual mean square. Taking the complexity of the model into account, AIC is generally regarded as a measure of the goodness of fit of the model. PRESS statistic indicates the prediction ability of the model. A model with a highest adjusted R2 and lowest AIC and PRESS statistic was selected as the final model. Other assumptions requisite to meet the specification of the multiple regression analysis, such as significance of the regression, significance of individual regression coefficients and multi-collinearity, were also checked. Model construction was performed in SAS (version 9.2, SAS Institute Inc., USA).

For model validation, the final model was tested on the validation dataset. In addition, cross validation technique was used to evaluate the prediction error of the model[7]. In the cross validation method, the samples are randomly partitioned into K approximately equal-sized K groups. In each iteration, the i-th (i=1,2…K) group is treated as validation set, and the remaining K-1 groups are used to fit the model. This process are repeated for K iterations, and the prediction errors, such as root mean square error (RMSE) and mean absolute percentage error (MAPE), are averaged over the iterations. Model validation was performed in Matlab ((Mathworks Inc., USA).

3 Results and Discussion

3.1 Image Analysis

Rice yield is mainly contributed by effective panicles, which generally locates in the uppermost one third region of the rice plant. In this study, the candidate panicle regions were firstly segmented using discriminant analysis. Then, the panicle regions were discriminated from non-panicle regions according to their location. After that, projected panicle area (PA) was computed. This method would be fit for most rice varieties. However, some rice varieties may have effective panicle that locates in the middle of the plant. In this occasion, some panicle regions may be mistaken treated as non-panicle regions and consequently the computed projected panicle area may be biased.

This paper mainly focuses on demonstrating the feasibility of estimating rice yield from projected panicle area, projected area of leaf and stem and fractal dimension. Improvement of projected panicle area extraction would be our future orientation.

3.2 Model Development and Validation

In variable selection and optimal model identification, we noticed that when the number of variables in the model was fixed, the model with the highest R2 generally had the lowest AIC and PRESS statistic. Therefore, among models with the same number of variables, the model with the highest R2 was selected for further consideration to accelerate the optimal model identification. This process reduced the number of candidate models to nine. Table 1 shows the performance of the nine candidate models.

When using alone, IFD was the most powerful variable for predicting yield, explaining 62.8% of the variation in the 260 training samples. Combination of PA and LSA outperformed other models with two variables, increasing adjusted R2 to 0.87, and dramatically reducing AIC and PRESS. When the number of variables in the model was more than 4 (excluding constant), adding more variables to the model did not produce less AIC and PRESS. The model with four variables IFD, E, PA and LSA had the highest Adjusted R2 and lowest AIC and PRESS. However, multi-collinearity was noticed in this model (VIF>5).Therefore, the model with three variables (Eq.8 and Table 2) was selected as the final model. Model diagnostics suggested that underlying assumptions of the regression model were met.

$$\text{Model 1: } \ln(Y) = a_0 + a_1 \cdot \ln(PA) + a_2 \cdot \ln(LSA) + a_3 \cdot \ln(IFD) \qquad (8)$$

Table 1. Performance of the nine candidate model based on training samples

Number of variables	Adjusted R^2	AIC	PRESS	Variables in the model
1	0.6284	-940.80	7.01	IFD
2	0.8725	-1218.04	2.40	PA, LSA
3	0.8904	-1256.28	2.08	IFD, PA, LSA
4	0.8917	-1258.37	2.06	IFD, E, PA, LSA
5	0.8916	-1257.08	2.07	DBC, IFD, E, PA, LSA
6	0.8917	-1256.44	2.08	IFD, M, S, mu3, PA, LSA
7	0.8915	-1255.06	2.10	IFD, E, M, S, mu3, PA, LSA
8	0.8915	-1253.94	2.11	IFD, E, M, S, U, mu3, PA, LSA
9	0.8910	-1251.98	2.13	DBC, IFD, E, M, S, U, mu3, PA, LSA

Table 2 shows the summary of the final model based on the training samples. F-test value and its significance (P>F) showed that the overall fit of the model was significant. The t-test values and its significance (P>|t|) indicated that the individual independent variables were statistically significant. Variance inflation factor (VIF) for each individual independent variable showed that there was no multi-collinearity between the variables.

Table 2. Summary of the final model based on training samples

	Coefficients		t	P>|t|	VIF	F	P>F	R^2	RMSE
	Value	Standard error							
a_0	-10.613	0.39223	-27.06	<.0001	0				
a_1	0.4181	0.01892	22.1	<.0001	1.36293	702.32	<.0001	0.89	2.71 g
a_2	0.64369	0.0439	14.66	<.0001	2.09423				
a_3	3.67453	0.56133	6.55	<.0001	2.59348				

Fig. 4 shows the plot of residuals versus fitted values $\ln(\hat{Y})$ of the final model. It indicated that the residuals were evenly distributed at the two sides of y=0 and can be contained in a horizontal band, suggesting no obvious model defects were existed. In addition, no significant differences were presented among outdoor rice and greenhouse rice. This meant that the model performed well on both outdoor rice and greenhouse rice.
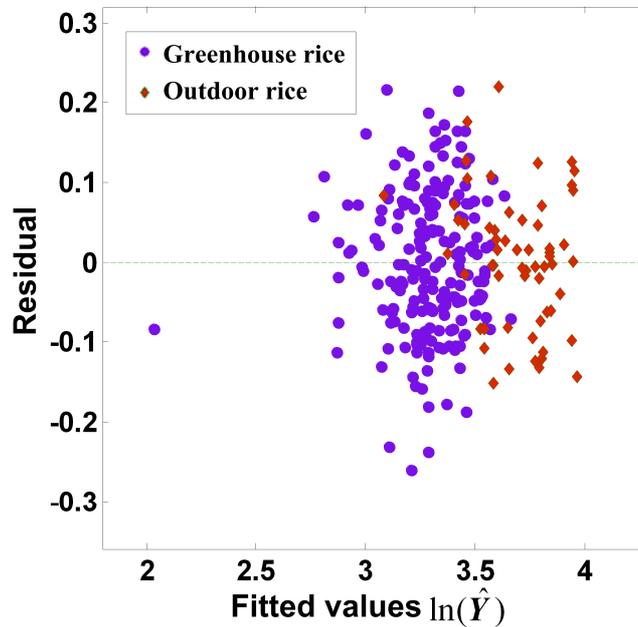

Fig. 4. Plot of residuals against the fitted values $\ln(\hat{Y})$ of the final model

The final model with three variables was then applied to the validation dataset and performed well (Fig. 5). The RMSE and MAPE for the validation dataset were 3.04 g and 7.74%.

5-fold cross validation showed that the RMSE and MAPE of the final model were 2.83 g and 7.45%.
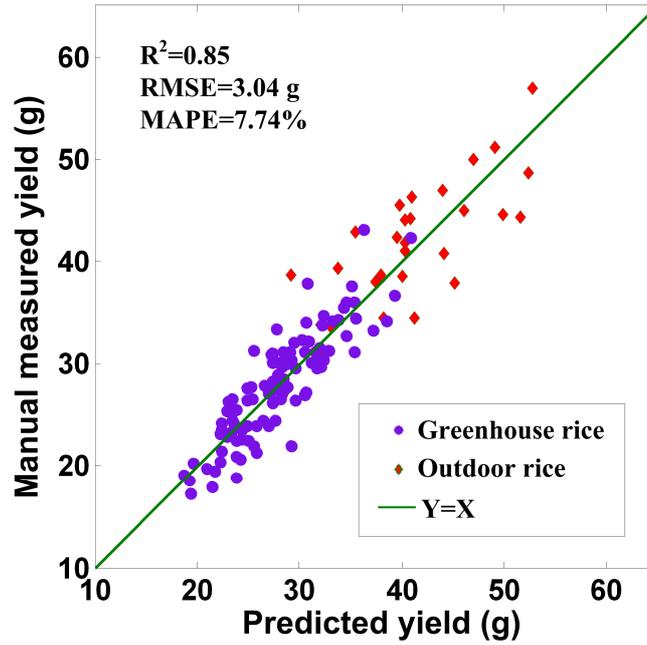

Fig. 5. Prediction for testing set using the final model

3.3 Performance Comparison of the Different Models

For comparison, we also tested the model using projected shoot area (Eq.9) and the model using projected panicle area and projected area of leaf and stem (Eq.10). And compared the two models with the model used in this study (Table 3). Separating A into PA and LSA significantly reduces the estimation error in yield prediction, which proved our hypothesis. And Adding fractal dimension into the model had a slight promotion on model performance. This was because that fractal dimension describes shape complexity of rice and can compensate the influence of organ occlusion problem at a certain degree.

$$\text{Model 2: } \ln(Y)=a_0+a_1 \cdot \ln(A) \tag{9}$$

$$\text{Model 3: } \ln(Y)=a_0+a_1 \cdot \ln(PA)+a_2 \cdot \ln(LSA) \tag{10}$$

Table 3. Performance comparison of the three models

	Training set		Validation set		Cross validation	
	RMSE (g)	MAPE	RMSE (g)	MAPE	RMSE (g)	MAPE
Model 1	2.71	7.16%	3.04	7.74%	2.83	7.45%
Model 2	4.79	12.03%	5.48	13.82%	5.04	12.74%
Model 3	2.94	7.76%	3.35	8.48%	3.09	8.08%

3.4   Performance Comparison of the Models Using Different Numbers of Side-View Images

By increasing number of side-view images, more hidden organs were available to be seen and more complete information on plant architecture would be provided. To find out the appropriate number of side-view images, the models that used average prediction variables from 1, 2, 4 and 6 side-view images were studied. The four models were listed as follows.

Model 4: a single image at angle of 0
Model 5: two images at the angle of 0 and 90
Model 6: four images at the angle of 0, 90, 180 and 270
Model 7: six images at the angle of 0, 60, 120, 180, 240 and 300

Table 4 shows the performance comparison of the models using different numbers of side-view images. 5-fold cross validation indicated that increasing number of side-view images would decrease prediction error of the model. However, the prediction error reduction was not significant when the number of side-view images were larger than 4. In this study, the images were captured by the visible light imaging unit that incorporated into our previous H-SMART system. The processing time of the H-SMART system was 20 seconds. This time was sufficient for capturing and processing 12 side-view images. For this reason, 12 side-view images were used in this study. To promote the efficiency of the yield estimation, two perpendicular side-view images would generate acceptable results.

Table 4. Performance comparison of the models using different numbers of side-view images

Model	Number of side-view images	5-fold Cross validation	
		RMSE (g)	MAPE
Model 4	1	3.80	10.08%
Model 5	2	3.13	8.15%
Model 6	4	3.01	7.87 %
Model 7	6	2.92	7.77%
Model 1	12	2.83	7.45%

4   Conclusions

This paper presented a non-destructive and high-throughput method to estimate the yield of individual pot-grown rice plant using photonics-based technology. And we demonstrated a new idea of estimating rice yield from projected panicle area, projected

area of leaf and stem and fractal dimension. 5-fold cross validation showed that the predictive error was 7.45% with testing rice plants grown both in-door and out-door. The presented work has the potential of accelerating yield estimation and would be a promising impetus for plant phenomics and high-throughput phenotyping.

Additional equations: Definition of the five histogram features.

$$M = \sum_{i=0}^{L-1} G_i p(G_i) \tag{1}$$

$$S = \sqrt{\sum_{i=0}^{L-1} (G_i - m)^2 p(G_i)} \tag{2}$$

$$mu3 = \sum_{i=0}^{L-1} (G_i - m)^3 p(G_i) \tag{3}$$

$$U = \sum_{i=0}^{L-1} p^2(G_i) \tag{4}$$

$$E = \sum_{i=0}^{L-1} p(G_i) \log_2 p(G_i) \tag{5}$$

Where G_i was the i-th graylevel, and $p(G_i)$ was the probability of G_i.

References

[1] Furbank, R.T.: Plant phenomics: from gene to form and function. Functional Plant Biology 36(10), 5–6 (2009)

[2] Zhou, Y., Zhu, J., Li, Z., et al.: Deletion in a quantitative trait gene qPE9-1 associated with panicle erectness improves plant architecture during rice domestication. Genetics 183(1), 315–324 (2009)

[3] Yang, W., Xu, X., Duan, L., et al.: High-throughput measurement of rice tillers using a conveyor equipped with x-ray computed tomography. Review of Scientific Instruments 82(2), 025102–025107 (2011)

[4] Cabrera-Bosquet, L., Crossa, J., von Zitzewitz, J., et al.: High-throughput phenotyping and genomic selection: The frontiers of crop breeding converge. Journal of Integrative Plant Biology 54(5), 312–320 (2012)

[5] Furbank, R.T., Tester, M.: Phenomics–technologies to relieve the phenotyping bottleneck. Trends in Plant Science 16(12), 635–644 (2011)

[6] Christine, G., Luis, A., Karine, C., et al.: 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. New Phytologist 169(3), 623–635 (2005)

[7] Golzarian, M.R., Frick, R.A., Rajendran, K., et al.: Accurate inference of shoot biomass from high-throughput images of cereal plants. Plant Methods 7(1), 2 (2011)

[8] Jones, H.G., Serraj, R., Loveys, B.R., et al.: Thermal infrared imaging of crop canopies for the remote diagnosis and quantification of plant responses to water stress in the field. Functional Plant Biology 36(11), 978–989 (2009)

[9] Duan, L., Yang, W., Bi, K., et al.: Fast discrimination and counting of filled/unfilled rice spikelets based on bi-modal imaging. Computers and Electronics in Agriculture 75(1), 196–203 (2011)

[10] Duan, L., Yang, W., Huang, C., et al.: A novel machine-vision-based facility for the automatic evaluation of yield-related traits in rice. Plant Methods 7(1), 44 (2011)

[11] Wang, Y.-P., Chang, K.-W., Chen, R.-K., et al.: Large-area rice yield forecasting using satellite imageries. International Journal of Applied Earth Observation and Geoinformation 12(1), 27–35 (2010)

[12] Wang, P., Sun, R., Zhang, J., et al.: Yield estimation of winter wheat in the North China Plain using the remote-sensing–photosynthesis–yield estimation for crops (RS–P–YEC) model. International Journal of Remote Sensing 32(21), 6335–6348 (2011)

[13] Jégo, G., Pattey, E., Liu, J.: Using Leaf Area Index, retrieved from optical imagery, in the STICS crop model for predicting yield and biomass of field crops. Field Crops Research 131, 63–74 (2012)

[14] Gitelson, A.A., Kaufman, Y.J., Merzlyak, M.N.: Use of a green channel in remote sensing of global vegetation from EOS-MODIS. Remote Sensing of Environment 58(3), 289–298 (1996)

[15] Shouichi, Y.: Fundamentals of rice crop science. International Rice Research Institute (1981)

[16] Hongju, G.: Estimating paddy yield based on fractal and image texture analysis. Nanjing Agricultural University (2008)

[17] Sarkar, N., Chaudhuri, B.B.: An efficient differential box-counting approach to compute fractal dimension of image. IEEE Transactions on Systems, Man and Cybernetics 24(1), 115–120 (1994)

[18] Bruno, O.M., De Oliveira Plotze, P., Falvo, M., et al.: Fractal dimension applied to plant identification. Information Sciences 178(12), 2722–2733 (2008)

[19] Okamoto, H., Lee, W.S.: Green citrus detection using hyperspectral imaging. Computers and Electronics in Agriculture 66(2), 201–208 (2009)

[20] Overman, J.P.M., Witte, H.J.L., Saldarriaga, J.G.: Evaluation of regression models for above-ground biomass determination in Amazon rainforest. Journal of Tropical Ecology 10(2), 207–218 (1994)

[21] Pompelli, M.F., Antunes, W.C., Ferreira, D.T.R.G., et al.: Allometric models for non-destructive leaf area estimation of Jatropha curcas. Biomass and Bioenergy 36, 77–85 (2012)

[22] Cristian, G., David, B., Kristin, S., et al.: Temperature-controlled organic carbon mineralization in lake sediments. Nature 466(7305), 478–481 (2010)

[23] Hirotugu, A.: A new look at the statistical model identification. IEEE Transactions on Automatic Control 19(6), 716–723 (1974)

[24] Montgomery, D.C., Peck, E.A., Vining, G.G.: Introduction to linear regression analysis. Wiley (2012)