Phenotypic Parameters Prediction of Lettuce based on Computer Vision and Regression

Minjuan Wang[1], Maowei Li[1], Wanwan Li[1], Jinsong Li[1], Ying Wang[1], Yu Zhang[1], Xiyue Guo[1], and Minjuan Wang[1]

[1]China Agricultural University

November 23, 2022

Abstract

To obtain phenotypic parameters by means of lossy measurement, we proposed a comprehensive and integrated approach to predict different parameters of four varieties of lettuces. By building different prediction models, we required predicted value of five phenotypic parameters of lettuce. Test results indicate that prediction models we have constructed are reliable and feasible. In addition,our methods can be better transferred to the research of other crops, and producers can adjust the growing environment of crops in time, so as to obtain higher yield.

Phenotypic Parameters Prediction of Lettuce based on Computer Vision and Regression

Maowei Li[a], Wanwan Li[a], Jinsong Li[a], Ying Wang[a], Yu Zhang[a], Xiyue Guo[a], and Minjuan Wang[a]

[a]Key Laboratory of Modern Precision Agriculture System Integration Research, China Agricultural University, Beijing, China

ABSTRACT

To obtain phenotypic parameters by means of lossy measurement, we proposed a comprehensive and integrated approach to predict different parameters of four varieties of lettuces. By building different prediction models, we required predicted value of five phenotypic parameters of lettuce. Test results indicate that prediction models we have constructed are reliable and feasible. In addition,our methods can be better transferred to the research of other crops, and producers can adjust the growing environment of crops in time, so as to obtain higher yield.

Keywords: Lettuce, Computer Vision, Regression, Phenotyping Parameters

1. INTRODUCTION

The traditional phenotypic parameters are usually measured by manual method, which is difficult to meet the requirements of automation and high precision[1,2] . Meanwhile, it is time-wasting and labor-consuming. More importantly, it is unacceptable to obtain phenotypic parameters by means of lossy measurement in practical agricultural production. Therefore, we proposed a nondestructive method to measure phenotypic parameters of lettuce in this research, aiming to monitor the growth and development of lettuce dynamically and efficiently. The predicted phenotypic parameters are plant height(PH), plant diameter(PD), leaf area(LA), fresh weight(FW) and dry weight(DW), respectively. The five types of phenotypic parameters obtained by this method will help producers adjust the growing environment of crops in time, which have count for much meaning to increase yield.

2. PROPOSED METHOD

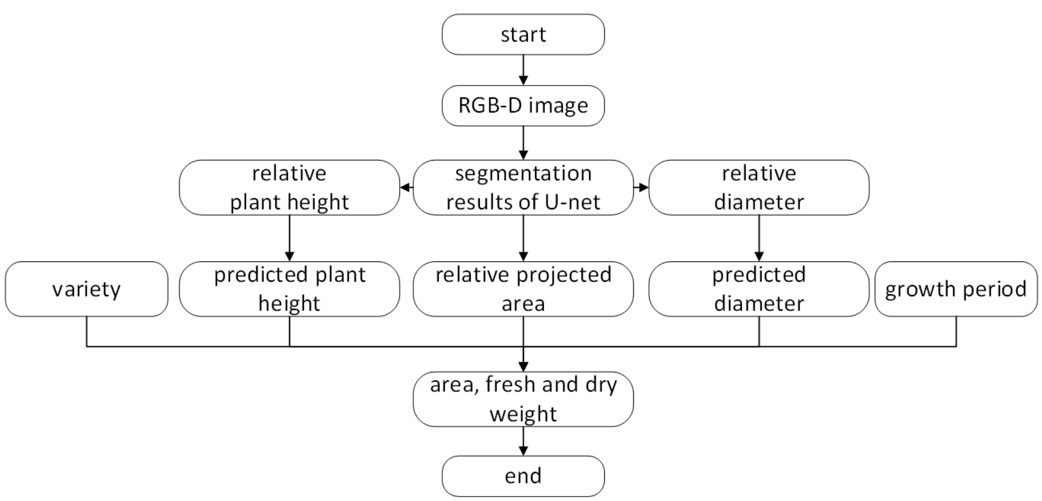

Figure 1. The overall research flowchart.

Further author information: (Send correspondence to Minjuan Wang)
Minjuan Wang: E-mail: minjuan@cau.edu.cn, Telephone: +86 136 0111 7404
Maowei Li: E-mail: sy20203081578@cau.edu.cn, Telephone: +86 178 5331 7947

The dataset in this experiment includes four types of lettuce(Lugano, Aphylion, Salanova and Satine) based on the 3rd Greenhouse Autonomous Challenge organized by Wageningen University and Tencent Lab, which were taken by Intel depth camera, containing 391 sets of depth images and color images. Among them, fifty sets of data are used to test the performance of the model. The overall research flowchart is shown in Figure. 1.

Before building prediction models, we define some evaluation indexes, including R^2, Adjusted $R^2(AR^2)$, RMSE, RMSE Radio(RR), NMSE, MAPE and Total Error(TE). Finally, R^2, NMSE, MAPE and TE are used to evaluate verification results on test data. AR^2 is corrected determination coefficient, which offsets the influence of sample size and the number of independent variables in regression on R^2. RR is the proportion of RMSE in actual average. The calculation formula of the above indicators is as follows.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - f_i)^2}{\sum_{i=1}^{n}(y_i - \overline{f_i})^2} \tag{1}$$

$$Adjusted\ R^2 = 1 - \frac{(1-R^2)(n-1)}{n-p-1} \tag{2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - f_i)^2}{n}} \tag{3}$$

$$RMSE\ Ratio = \frac{RMSE}{\frac{1}{n}\sum_{i=1}^{n} y_i} \tag{4}$$

$$NMSE = \frac{\sum_{i=1}^{n}(y_i - f_i)^2}{\sum_{i=1}^{n} y_i^2} \tag{5}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}|\frac{y_i - f_i}{y_i}| \times 100\% \tag{6}$$

$$Total\ Error = \sum_{i=1}^{m}\frac{\sum_{j=1}^{n}(y_{ij} - f_{ij})^2}{\sum_{j=1}^{n}(y_{ij})^2} \tag{7}$$

Our approach is divided into the following two parts. The goal of the first step is predict PH and PD, as is shown in Table 1. Firstly, U-net network[3] is used to perform color image segmentation to obtain binary image, and then it is projected to depth image. Through projection results, the pixel point with the smallest depth is obtained, which corresponds to the highest point of lettuce. The relative plant height (PH1) can be defined by making the difference between mininum of depth and the camera height. Secondly, the number of pixels in binary image is counted as relative projection area (RPA). Based on segmentation results, we apply contour detection function of OpenCV library to obtain all contour pixels, among which point pair with farthest distance are found to calculate the relative diameter (PD1). Finally, the linear regression model is used to fit the relationship between predicted value and actual value of PH1 and PD1 to get predict PD and PD, where AR^2 of model is 94.4% and 91.8%, respectively. When average PH is 13cm and average PD is 23cm, RMSE is as low as 0.816cm and 0.980cm, respectively.

Table 1. Optimal model of PH and PD.

Parameters	model	AR^2	RR	RMSE
PH	0.974PH1+0.778	4.59	7.6	16.83
PD	0.0536PD1+2.1697	4.47	9.33	18.49

The second step is predict LA, FW, and DW of different types of lettuce. Due to large difference of lettuce shape among different species and growing period(GP), we build the model of fresh weight and dry weight separately. The solution of each parameter is divided into five parts, and they are all varieties at seedling stage, Aphylion, Lugano, Satine and Salanova at mature stage. According to data of different stages of lettuce, we use PH1, GP and RPA to fit the prediction model. In terms of LA, we find that all varieties at seedling stage and Salanova at mature stage have a significant linear relationship with GP, RPA, PH and the product of two, respectively. Lugano and Satine have a significant linear relationship with the power function of GP, PH and RPA. While Aphylion has a significant linear relationship with GP and RPA. The average AR^2 of LA is 91.8% and the average RMSE is 192.58, accounting for 14.92% of actual average LA, as is shown in Table 2. In terms of FW, we find that there is a significant relationship between the power function of all varieties and LA at seedling stage, while Lugano at mature stage has an obvious linear relationship with GP, RPA and their product. FW of Aphylion, Salanova and Satine at mature stage is significantly related to exponential function of GP and power function of LA. The average AR^2 of FW is 92.08% and the average RMSE is 11.63, accounting for 16.75% of actual average FW, as is shown in Table 3. In terms of DW, we find that all varieties at seedling stage and mature stage have a significant relationship with exponential function of GP and power function of LA. The average AR^2 of DW is 93.42% and the average RMSE is 0.50, accounting for 15.44% of actual DW, as is shown in Table 4.

Table 2. Optimal model of LA.

GP	variety	model	AR^2	RR	RMSE
seedling	All	-16.54GP+0.000294LA+3.329PH+0.00317GP∗LA+32.99	0.876	11.36%	11.69
mature	Aphylion	395.83GP+0.0154RPA-449.93	0.965	13.03%	217.2
mature	Lugano	$86.68GP^{1.1}+0.000053RPA^{1.47}+1.66PH^{2.32}-45.02$	0.906	18.42%	217.2
mature	Salanova	398.3GP-0.00437RPA-93.03PH+0.00159RPA∗PH +451.64	0.911	17.86%	328.59
mature	Satine	$111.28GP^{0.979}+0.000315RPA^{1.3281}+0.946PH^{1.902}-30.547$	0.932	13.95%	148.85

Table 3. Optimal model of FW.

GP	variety	model	AR^2	RR	RMSE
seedling	All	$0.0143RPA^{1.19}-0.0444$	0.809	18.45%	0.659
mature	Aphylion	$3.47e^{0.62GP}+0.0153RPA^{1.15}-1.003$	0.942	19.83%	18.02
mature	Lugano	0.254GP-0.00526RPA+0.0179GP∗RPA+23.7	0.958	15.33%	16.48
mature	Salanova	$0.018e^{1.56GP}+0.0076RPA^{1.2}+5.88$	0.975	11.03%	8.86
mature	Satine	$5.5e^{0.5GP}+0.0042RPA^{1.32}-0.102$	0.920	19.10%	14.11

Table 4. Optimal model of DW.

GP	variety	model	AR^2	RR	RMSE
seedling	All	$0.0133e^{1.37GP}+3.02E\text{-}05RPA^{1.84}-0.0251$	0.884	18.34%	0.0495
mature	Aphylion	$0.452e^{0.52GP}+0.0026RPA^{0.96}-0.0954$	0.954	15.59%	0.7138
mature	Lugano	$0.686e^{0.44GP}+0.00183RPA^{0.98}-0.179$	0.947	14.52%	0.662
mature	Salanova	$0.307e^{0.5GP}+0.00169RPA^{1.002}-0.183$	0.968	11.14%	0.4578
mature	Satine	$0.448e^{0.45GP}+0.00052RPA^{1.16}-0.01$	0.918	17.60%	0.6048

3. RESULT AND DISCUSSION

According to optimal model we have selected, verification result based on test data is shown in Table 5. It can be clearly seen that MAPE is below 20% basically. The correlation analysis of actual value and predicted value is shown in Figure.2−6, NMSE of PH, PD, LA, FW and DW are 0.00773, 0.00763, 0.01688, 0.02370 and 0.02123, respectively, whose sum is the value of TE. More importantly, when our model proposed in this paper is used to predict five phenotypic parameters of test data, final TE can reach 0.07717. While the first place's TE was 0.08128 in the 3rd Greenhouse Independent Challenge based on this dataset, it is obvious that we have exceeded the first place in the above competition, which further proves that our model has good applicability.

4. CONCLUSION

In this research, we propose a series of a comprehensive and integrated approach to predict different parameters of four varieties of lettuces. By building different prediction models, AR^2 of predicted value and actual value have reached more than 90%. Test results indicate that prediction models we have constructed in this study are reliable and feasible. Although it is specific to different varieties, the early treatment methods can be better transferred to the research of other crops, and the later parameter prediction models can be better analogized to other similar crops. According to the five phenotypic parameters obtained by this method, producers can adjust the growing environment of crops in time, so as to obtain higher yield.

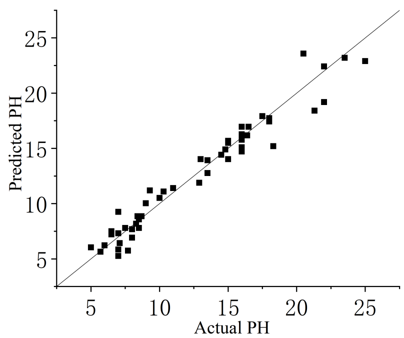

Figure 2. Correlation analysis of PH.

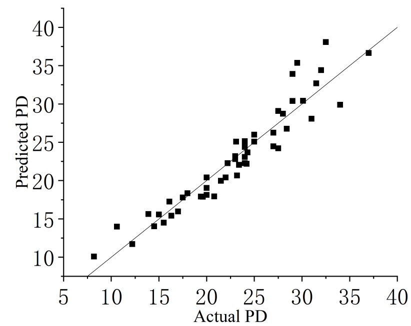

Figure 3. Correlation analysis of PD.

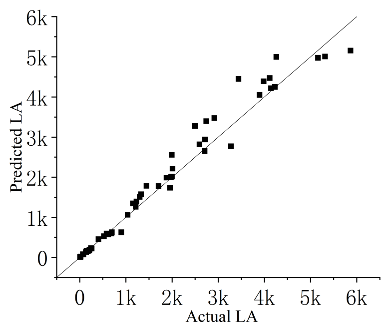

Figure 4. Correlation analysis of LA.

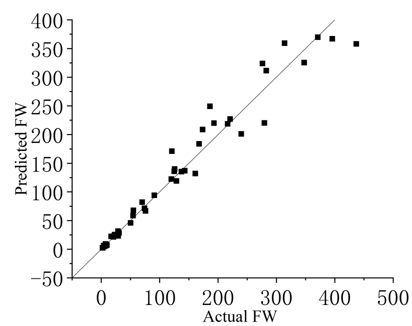

Figure 5. Correlation analysis of FW.

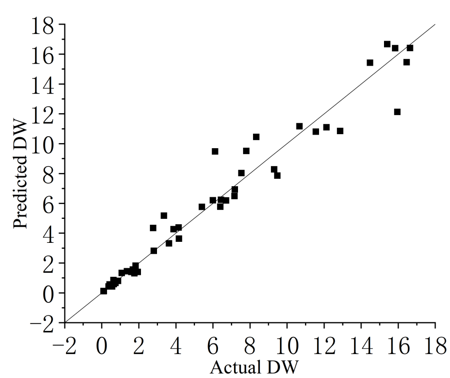

Figure 6. Correlation analysis of FW.

Table 5. Test results of different models.

parameters	GP	variety	R^2	MAPE	RMSE	RR	NMSE
PH	all	all	0.948	7.92%	1.209	9.04%	0.007735
PD	all	all	0.888	7.20%	2.084	7.45%	0.00763
LA	seedling	all	0.918	8.43%	17.656	11.26%	0.01688
	mature	Aphylion	0.897	13.86%	545.828	18.92%	
		Lugano	0.895	10.07%	312.733	12.75%	
		Salanova	0.963	11.51%	269.897	11.93%	
		Satine	0.847	10.33%	269.140	10.38%	
FW	seedling	all	0.697	15.48%	1.271	23.22%	0.02370
	mature	Aphylion	0.876	15.96%	51.234	24.90%	
		Lugano	0.936	13.17%	26.398	16.26%	
		Salanova	0.951	9.85%	20.303	12.11%	
		Satine	0.861	10.65%	20.502	11.25%	
FW	seedling	all	0.846	15.69%	0.097	20.50%	0.02123
	mature	Aphylion	0.896	23.30%	1.968	20.83%	
		Lugano	0.954	10.67%	1.490	19.64%	
		Salanova	0.950	13.18%	0.968	12.35%	
		Satine	0.910	9.70%	0.719	8.53%	

REFERENCES

[1] Rahaman, M., et al., Advanced phenotyping and phenotype data analysis for the study of plant growth and development. Frontiers in plant science, 2015. 6: p. 619.

[2] Watt, M., et al., Phenotyping: new windows into the plant for breeders. Annual review of plant biology, 2020. 71: p. 689-712.

[3] Ronneberger, O., P. Fischer and T. Brox. U-net: Convolutional networks for biomedical image segmentation. in International Conference on Medical image computing and computer-assisted intervention. 2015: Springer.