

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Faculty Publications, Department of Statistics

Statistics, Department of

---

2020

## A High-Throughput Phenotyping Pipeline for Image Processing and Functional Growth Curve Analysis

Ronghao Wang





Follow this and additional works at: <https://digitalcommons.unl.edu/statisticsfacpub>

---

This Article is brought to you for free and open access by the Statistics, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications, Department of Statistics by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

## Research Article

# A High-Throughput Phenotyping Pipeline for Image Processing and Functional Growth Curve Analysis

Ronghao Wang <sup>1</sup>, Yumou Qiu <sup>2</sup>, Yuzhen Zhou,<sup>1</sup> Zhikai Liang <sup>3</sup>,  
and James C. Schnable <sup>4</sup>

<sup>1</sup>Department of Statistics, University of Nebraska-Lincoln, Lincoln 68503, USA

<sup>2</sup>Department of Statistics, Iowa State University, Ames 50011, USA

<sup>3</sup>Department of Plant and Microbial Biology, University of Minnesota, Saint Paul, MN 55108, USA

<sup>4</sup>Center for Plant Science Innovation, Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln 68503, USA

Correspondence should be addressed to Yumou Qiu; yumouqiu@iastate.edu

Received 13 November 2019; Accepted 16 May 2020; Published 14 July 2020

Copyright © 2020 Ronghao Wang et al. Exclusive Licensee Nanjing Agricultural University. Distributed under a Creative Commons Attribution License (CC BY 4.0).

High-throughput phenotyping system has become more and more popular in plant science research. The data analysis for such a system typically involves two steps: plant feature extraction through image processing and statistical analysis for the extracted features. The current approach is to perform those two steps on different platforms. We develop the package “implant” in R for both robust feature extraction and functional data analysis. For image processing, the “implant” package provides methods including thresholding, hidden Markov random field model, and morphological operations. For statistical analysis, this package can produce nonparametric curve fitting with its confidence region for plant growth. A functional ANOVA model to test for the treatment and genotype effects on the plant growth dynamics is also provided.

## 1. Introduction

High-throughput phenotyping is a newly emerging technique in the plant science research. Many automated systems have been constructed both in the greenhouse and field to study plant features [1–3]). One of the main innovations is to use automated cameras to take raw images for plants. Several types of high-resolution images, including RGB, infrared, fluorescence, and hyperspectral, are recorded for a large number of plants at designed time points. From the images, we are able to process and extract useful phenotypical features, such as plant height, width, and size. Compared to the traditional methods, the high-throughput system is able to provide the plant features of interest in a more efficient, accurate and nondestructive way.

In order to extract the plant traits, segmentation for parts of a plant or the whole plant is necessary. Thresholding is the simplest and the most commonly used method for image seg-

mentation [4], which separates the image into foreground and background classes by a cut-off value on the pixel intensities. Based on the thresholding methods, several platforms have been developed for the analysis of high-throughput plant phenotyping, including HTPheno [5], Image Harvest [6], and PlantCV [7]. Those software have admitted procedures in processing plant images to extract phenotypical features. However, these platforms solely focus on image processing. There is a lack of functionality on statistical modeling and inference for plant growth.

*K*-means clustering algorithm [8] is also well known for image segmentation, which assigns pixels into subgroups so that the within-group variation of pixel intensity is minimized. When the number of clusters is given, the *K*-means method is free of tuning parameter selection. The hidden Markov random field model (HMRF) [9] can be applied to refine the segmentation result from *K*-means clustering and thresholding. HMRF is a hierarchical model with a hidden

layer of Markov random field to model the class label of each pixel, which captures the spatial dependence of pixels to their neighborhood. As the thresholding and  $K$ -means methods ignore the spatial structure of an image, the HMRF model is able to provide a more accurate classification of pixels by incorporating their neighborhood class information.

Given an accurate segmentation, the measurements of phenotypical traits can be extracted from the images. These numerical measurements can be used for analyzing genotype and treatment effects on the plant growth over time. In a traditional growth curve analysis, the approach of a pointwise analysis of variance (ANOVA) is applied at each measurement time point. However, this approach analyzes each time point separately and thus cannot reflect the dynamics of plant growth. Parametric modeling for the growth curve is another popular tool. However, fitting the parametric models requires measurements of the plant traits over the whole growth stage which may not be available for some experiments, and the temporal dependence of the data is usually ignored in this approach. Functional ANOVA [10] is a recent nonparametric method for analyzing plant traits collected over time [11]. Instead of parametric regression, smoothing splines [10, 12] or local polynomial regression [13] is used to estimate the plant growth. This approach is nonparametric, fully data driven, and adaptive to temporal dependence of the data. Despite those advantages, the implementation of functional ANOVA for plant phenotyping data [11] is nontrivial. The current R package “fda” [14] for functional data analysis is complicated, and it is difficult to use for non-statisticians. There is no computation guidance of functional ANOVA on studying plant growth.

To respond to the needs of data analysis for the high-throughput phenotyping systems, we develop an R package “implant” that involves both image processing and functional data analysis for the extracted traits. The scope of this paper is to provide an easy-to-use pipeline to analyze high-throughput phenotyping data, from the raw images to statistical analysis. Compared to [11] that mainly focuses on introducing the methodology of nonparametric curve fitting, the proposed package provides a user friendly computation tool, which allows plant scientists to easily conduct functional data analysis for plant growth dynamics. Our package also provides the confidence regions for the time-varying regression coefficients, which is not thoroughly discussed in Xu et al. [11]. The flow chart in Figure 1 illustrates the main steps of this pipeline. In the first step, plant segmentation is done by double-criteria thresholding (DCT) or HMRF methods. Notice that if the image of an empty pot is available, DCT can be applied on the contrast image between the plant and the empty pot as demonstrated in Figures 1(a)–1(c).

Figure 2 In the second step, morphological erosion and dilation operations [15] and plant region identification can be applied to refine the segmentation. Then, plant traits are calculated based on the segmented image. In the last step, functional data analysis and statistical inference are conducted on the extracted traits. The pipeline is able to estimate both the main and interaction effects on the plant

growth, provide confidence regions for the effect curves, and deal with irregular observation time points. Those confidence regions can demonstrate the statistical significance for the treatment and genotype effects over time. A real data example is provided in Implementation and Results to illustrate the utility of our package.

## 2. Methods

In this section, we introduce the hidden Markov random field model for image segmentation and the functional ANOVA for growth curve analysis.

*2.1. Hidden Markov Random Field Model.* The hidden Markov random field (HMRF) model is a hierarchical model with an unobserved layer for the pixel class and an observed layer for the pixel intensity given its class. The hidden layer of the pixel class is modeled by a Markov random field, where the probability of a pixel from the plant category depends on the classes of its neighborhood pixels. As the plant pixel is more likely to be surrounded by plants, this transition probability matrix models the spatial dependence of the pixel classes. We assume that the pixel intensity follows a normal distribution where its mean and variance are determined by the class of this pixel. And, the joint distribution of the unobservable classes for all the pixels follows the Gibbs distribution, according to the Hammersley-Clifford Theorem [16]. The aim of this method is that for each pixel, given the observed pixel intensity, we predict its class label by maximizing the probability that the pixel is classified into this class.

We use the relative green intensity as our response variable. In order to fit the HMRF model, we apply the expectation maximization (EM) algorithm [17, 18]. By using the segmentation results obtained by  $K$ -means clustering as the initial class label for the EM algorithm, we iteratively find the maximal likelihood estimators for the mean and variance of the relative green intensities for the plant class and the background class. Then, the class label of each pixel is predicted as the class with higher posterior probability of the pixel belonging to given the observed intensities.

*2.2. Functional ANOVA.* Phenotypical traits extracted from the segmented images can be used to build functional data models and draw statistical inference for the plant growth dynamics [10, 11]. We consider functional data analysis for one type of the extracted traits, denoted by  $y_i(t)$  as the measurement of the  $i$ th plant at time  $t$ . Since the high-throughput measurements for each plant are relatively dense over time and the plant growth curve is smooth, we directly use the extracted trait as the response instead of its smoothed values. Suppose the trait is potentially affected by  $q$  factors. With the number of levels of the  $j$ th factor denoted by  $\ell_j$ , we define  $\mathbf{X}_{ij} = (x_{ij2}, \dots, x_{ij\ell_j})^T$  as the categorical indicators of the  $j$ th factor of the  $i$ th plant. Specifically,  $x_{ijk}$  is set to one if the  $j$ th factor of the  $i$ th plant has level “ $k$ ”; otherwise,  $x_{ijk} = 0$ . With the Kronecker product of matrices denoted by  $\otimes$ , a functional multiway

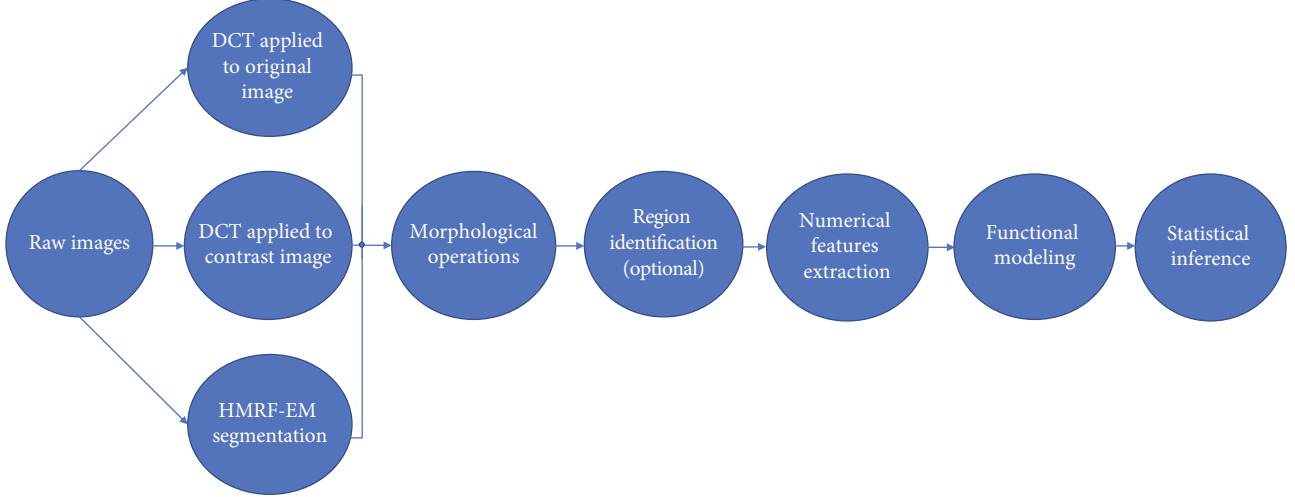


FIGURE 1: Flow chart of the proposed “implant” pipeline. In the first step of segmentation, multiple methods could be jointly applied and the common plant area is considered to be the final segmentation.

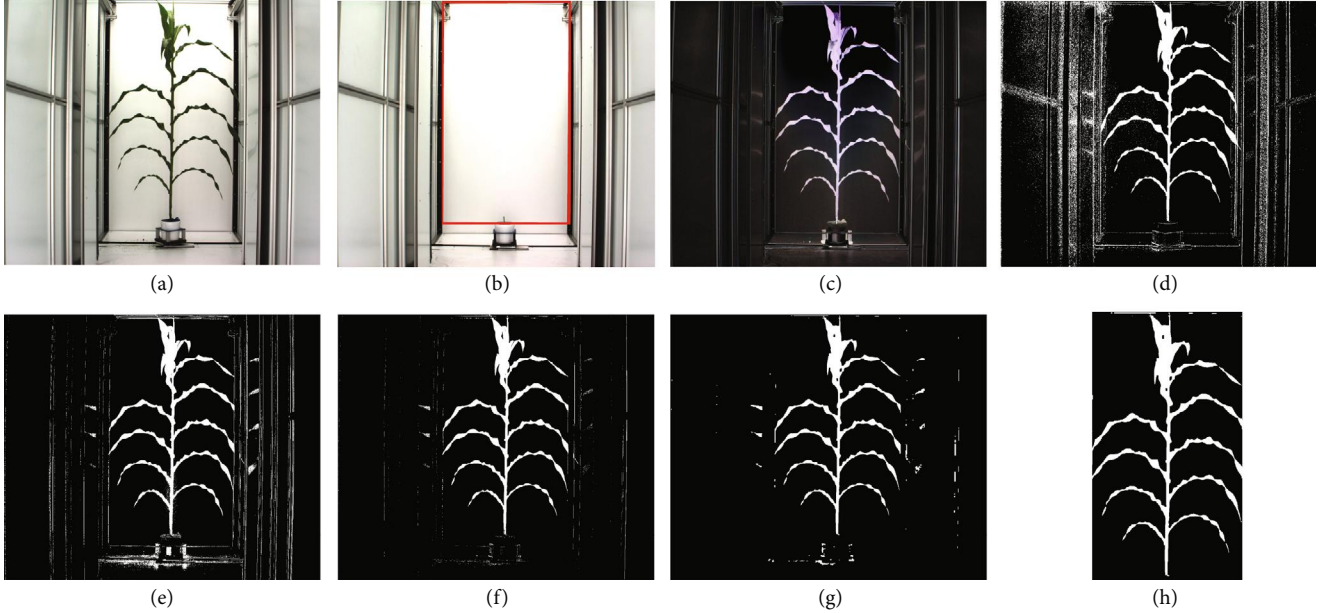


FIGURE 2: (a) Original plant image. (b) Original empty pot image; the red square is the identified region of interest by the functions “ColorB” and “ColorG.” (c) Contrast of (a) and (b). (d) Segmented image of (a) using DCT. (e) Segmented image of (c) using DCT. (f) Intersection of (d) and (e). (g) Dilated-eroded-eroded-dilated image of (f). (h) Final segmented image by identifying the region of interest.

ANOVA model with interactions can be written in the following form:

$$\begin{aligned}
 y_i(t) = & \mu(t) + \mathbf{X}_{i1}^T \mathbf{a}_1(t) + \mathbf{X}_{i2}^T \mathbf{a}_2(t) + \dots + \mathbf{X}_{iq}^T \mathbf{a}_q(t) \\
 & + (\mathbf{X}_{i1}^T \otimes \mathbf{X}_{i2}^T) \mathbf{a}_{1,2}(t) + \dots + (\mathbf{X}_{iq-1}^T \otimes \mathbf{X}_{iq}^T) \mathbf{a}_{q-1,q}(t) \\
 & + \varepsilon_i(t),
 \end{aligned} \tag{1}$$

where  $\mathbf{a}_j(t) = (a_{j2}(t), \dots, a_{j\ell_j}(t))^T$  values are the treatment effect functions of the  $j$ th factor with dimension  $\ell_j - 1$ ,

$\mathbf{a}_{j_1, j_2}(t)$  values are the pairwise interaction effect functions between factors  $j_1$  and  $j_2$  with dimension  $(\ell_{j_1} - 1)(\ell_{j_2} - 1)$ , and  $\varepsilon_i(t)$  values are temporal dependent random processes with zero means. We have implemented this multifactor model ( $q \geq 2$ ) in our package such that researchers can specify the main and interaction effects as needed. A real data example without interaction is illustrated in the next section, and another example with interaction is provided in the user guide.

We approximate all of the coefficient functions with rank  $K$  B-spline expansion. For example,  $a_{jk}(t) = \sum_{v=1}^K \beta_{jk,v} B_{d,v}(t)$ , where  $\{B_{d,v}(t)\}_{v=1}^K$  values are order  $d$  B-spline basis functions,

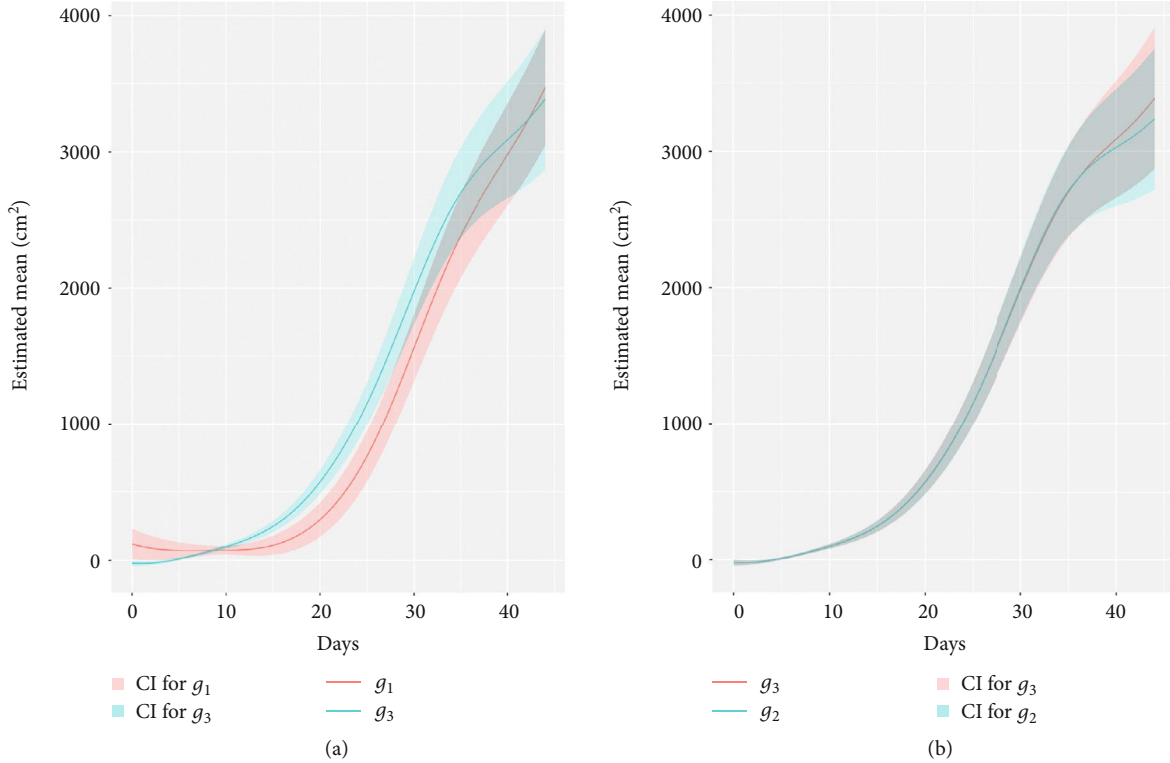


FIGURE 3: (a) 95% confidence regions for the average plant size of genotypes 1 and 3 over the three blocks. (b) 95% confidence regions for the average plant size of genotypes 2 and 3 over the three blocks.

and  $\{\beta_{jk,v}\}_{v=1}^K$  values are coefficients of the basis functions. The rank  $K$  of the B-spline basis functions is equal to the order (degree + 1) of the spline plus the number of interior knots. To avoid overfitting, we choose the rank that is less than half of the number of observation time points  $m$ . Since the growth curves of plants are relatively smooth as shown in Figure 3, we use splines with degree 3 and  $m/2 - 4$  equally spaced interior knots. Then, the estimator  $\hat{\beta}$  for the spline coefficients can be found explicitly via the penalized smoothing splines. We penalize the  $L_2$  norm of the second derivatives of the spline expansion functions and choose a common smoothing parameter  $\lambda$  by the generalized crossvalidation (GCV). We develop the function “fanova” in our package to implement the estimation procedure. It turns out that  $\hat{\beta}$  is a linear estimator of the response  $y_i(t)$ . Given the sample covariance of the response  $y_i(t)$ , it is straightforward to estimate the covariance of  $\hat{\beta}$  and hence the covariance of  $\hat{a}_{jk}$  based on the spline expansion. Confidence regions of the treatment effects can be constructed accordingly, which can be obtained by the functions “CI\_contrast” and “CI” in the package.

### 3. Implementation and Results

In this section, we illustrate the implementation of our “implant” package by a maize experiment conducted at the University of Nebraska-Lincoln (UNL) Greenhouse Innovation Center. The package, documentation, and user guide are available online at <https://github.com/rwang14/implant>.

Detailed descriptions for the methods are presented in Methods.

**3.1. Experiment.** The experiment involved 420 maize plants with 140 different genotypes. There were three replicates for each genotype. The pots in the greenhouse were divided into three blocks based on the layout of the belt conveyor system. We conducted the randomized complete block design (RCBD) such that each of the 140 genotypes was randomly located within a block and the three replicates of the same genotype were assigned to different blocks. Each plant was imaged about every two or three days from May to July, and the imaging time points were irregular due to the large number of plants in the experiment.

#### 3.2. Image Processing

**3.2.1. Image Segmentation Using DCT.** Figure 2 shows the general process of the double-criteria thresholding (DCT) for one of the plant images from the experiment. Here, Figures 2(a) and 2(b) are the RGB maize image and the empty pot image, respectively. Figure 2(d) is obtained by the function:

$$\begin{aligned} \text{imageB} &= \text{imageBinary}(\text{original\_image}, \text{weight} \\ &= c(-1, 2, -1), \text{threshold1} \\ &= 30/255, \text{threshold2} = 0.02), \end{aligned} \quad (2)$$

where “threshold1” is applied to the sum of the RGB intensities, and “threshold2” is applied on the green contrast



intensity by the specified weight in the function [19]. The two thresholds are to delete the black pixels and to segment the plant green pixels, respectively. We choose a small level 0.02 for the second threshold to retain most part of the plant. The background noises in Figure 2(d) can be much reduced by applying the DCT procedure on the contrast image in Figure 2(c) resulting from the difference between Figures 2(a) and 2(b). The image in Figure 2(e) is obtained by setting “threshold1 = 0.7,” “threshold2 = 0,” and “weight = c(1, -2, 1)” in the “imageBinary” function for Figure 2(c). Then, we take the intersection between Figures 2(d) and 2(e) to obtain Figure 2(f). As we observed from Figure 2(f), most of the background noises are eliminated and the plant body is segmented well. Those thresholding parameters work consistently well over the whole experiments for the UNL greenhouse. The double-criteria procedure makes the results less sensitive to the threshold levels than the procedure using only one criterion. Though for a different system, those parameters need to be properly tuned for good segmentation results. In the case of no empty pot images, we should set a more restrictive threshold for the green contrast intensities.

Morphological operations can be applied on the thresholding results to further reduce the segmentation errors (see Figure 2(g)), which can be performed by the “dilation” and “erosion” functions in our packages as follows:

```
imageBD = dilation(imageB, mask = matrix(1, 5, 5)),
imageBDE = erosion(imageBD, mask = matrix(1, 5, 5)).
imageBDEE = erosion(imageBDE, mask = matrix(1, 3, 3)),
imageBDEED = dilation(imageBDEE, mask = matrix(1, 3, 3)),
```

(3)

where “mask” is a structuring matrix specifying the neighborhood structure of a pixel [15]. The dilation operator is applied to a binary image to enlarge the boundaries of the segmented object and fill in the holes within the object. Erosion is the opposite operator to dilation, which erodes away the boundaries of the segmented object. We call dilation followed by erosion as a morphological closing operation, and erosion followed by dilation as a morphological opening operation.

Region of interest can be automatically identified by some specific characteristics on the background of an imaging system. For the UNL greenhouse, we can identify the inner black bars and the border top of the pot to obtain the region of interest for the plants; see the red rectangle in Figure 2(b). Notice that this identification strategy is for the images from the UNL greenhouse system only. Different systems need different strategies for locating the region of interest. It is worth mentioning that although identifying the region of interest can help us easily remove most of the background noises, parts of the plants might be lost as well (see Figure 2(h) as the chopped image by the identified region in Figure 2(b)).

**3.2.2. Image Segmentation Using HMRF.** The segmentation method by the HMRF model is also provided in the package.

Compared to the former thresholding procedure, the HMRF model is data driven and free of tuning parameter selection. Figure 4(c) shows the segmentation result by the HMRF model with initial class assignment by the  $K$ -means clustering algorithm on relative green intensity  $G/(R + G + B)$  with  $K = 2$ . From Figure 4(c), we see that the HMRF model provides a quite good segmentation for the plant with few classification errors. Comparing to the  $K$ -means result in Figure 4(b), the HMRF is able to fill in the missing plant pixels by using their neighborhood class information. Comparing the thresholding result in Figure 2(f), the result from the HMRF approach eliminates most of the background noises. This method is implemented by the function “HMRF” in the package as

$$\text{HMRF}(X, Y, \dots) \text{image}_{\text{matrix}}, \quad (4)$$

where  $X$  is a matrix of initial class labels (for example, results from  $K$ -means clustering), and  $Y$  is a matrix of relative green intensities. Description on other arguments of this function can be found in the help documentation of the “implant” package. Morphological operations can be applied on the segmented result from HMRF model, see Figure 4(d) which shows a better segmentation than the morphological operations on the thresholding result in Figure 2(g). The HMRF method can generally get a good segmentation result without identifying the region of interest, which broadens the scope of its application. Moreover, it can be used to refine the segmentation results obtained by other methods.

## 4. Plant Feature Extraction

Based on the segmented images, we can extract the phenotypical features of plants. Given the information of the pixel size in millimeters, we can obtain plant height, width, and size using the functions.

```
extract_pheno(processed_image, Xsize = 1, Ysize = 1, ...),
```

(5)

where “processed\_image” is the segmented image of a plant as in Figures 2(h) and 4(d), and “Xsize” and “Ysize” are the actual horizontal and vertical lengths of one pixel.

## 5. Plant Growth Dynamics Analysis

Based on the extracted traits, we can study the treatment and genotype effects on plant growth. Let  $t_{ij}$  be the  $j$ th observation time of the  $i$ th plant, and let  $y_i(t_{ij})$  denote a specific trait of the  $i$ th plant measured at time  $t_{ij}$ . Note that there are 140 different genotypes with one replicate in each of the 3 blocks in this study. Let  $\{G_{ik}\}_{k=2}^{140}$  be the genotype indicators for the  $i$ th plant, where  $G_{ik} = 1$  if the  $i$ th plant has the  $k$ th genotype, and  $G_{ik} = 0$  otherwise. Similarly, let  $\{P_{ik}\}_2^3$  be the block indicators. The first genotype

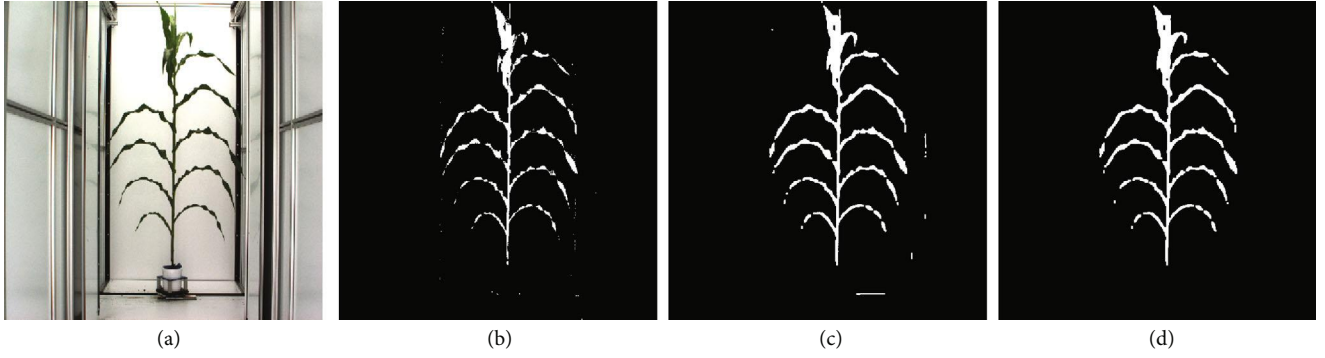


FIGURE 4: (a) Original image. (b) Initial classification using  $K$ -means ( $K = 2$ ) on relative green intensity  $G/(R + G + B)$ . (c) Segmentation result using HMRF. (d) Applying morphological closing and opening to (c).

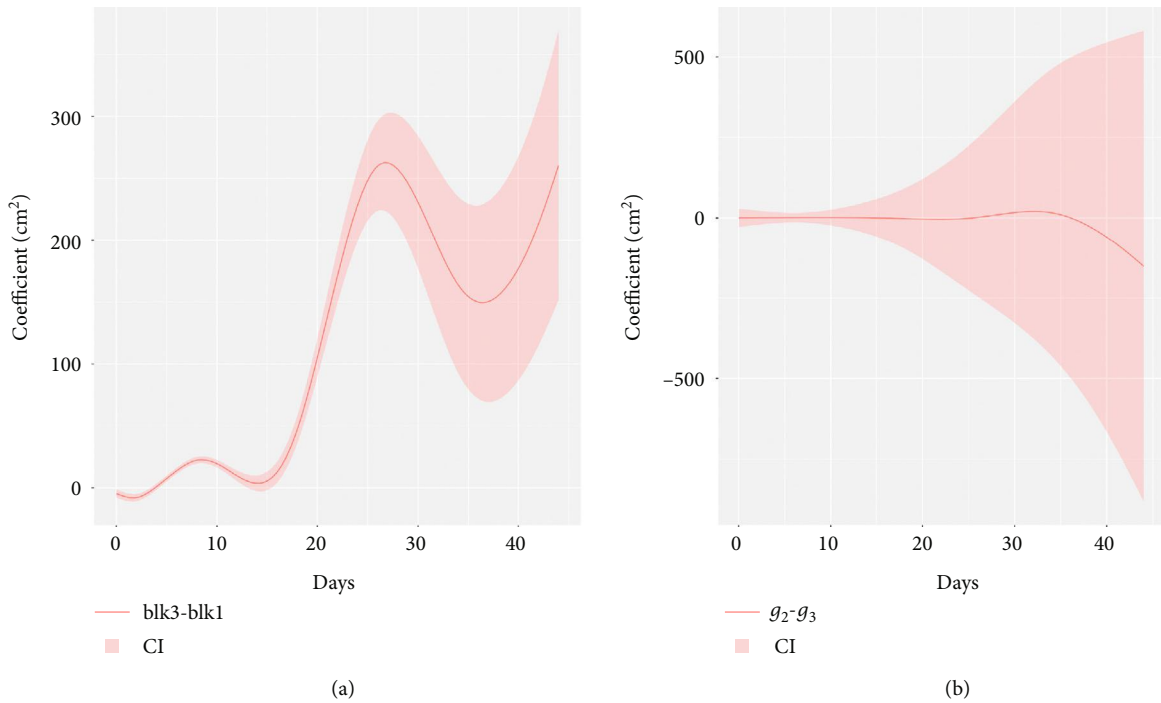


FIGURE 5: (a) 95% confidence region for the block effect between block 3 and block 1. (b) 95% confidence region for the genotype effect between genotype 2 and genotype 3.

and the first block are treated as the baseline. The functional ANOVA model is

$$y_i(t_{ij}) = \mu(t_{ij}) + \sum_{k=2}^{140} G_{ik}g_k(t_{ij}) + \sum_{k=2}^3 P_{ik}p_k(t_{ij}) + \varepsilon_i(t_{ij}), \quad (6)$$

where  $\mu(t)$ ,  $g_k(t)$ , and  $p_k(t)$  are the intercept, genotype effect, and block effect functions, respectively. The regression error  $\varepsilon_i(t)$  is modeled by a temporal dependent random process with mean zero (see Methods for more details).

The regression coefficient curves  $\mu(t)$ ,  $g_k(t)$ , and  $p_k(t)$  can be estimated by the function “fanova” in our package,

$$\text{fit} = \text{fanova}(Y.\text{na.mat}, X, \text{formula}, \dots), \quad (7)$$

where  $X$  specifies levels of genotype and block factors,  $Y.\text{na.mat}$  is the matrix of the extracted traits, and formula specifies the model, namely, Equation (6) in this example. Then, we can construct the confidence regions for the significance of the treatment and genotype effects by the functions:

$$\begin{aligned} & \text{CI\_contrast}(\text{fit}, j_1, j_2, \dots), \\ & \text{CI}(\text{fit}, L, \dots), \end{aligned} \quad (8)$$

where fit is the fitted model from the output of *fanova*. The first function provides the confidence regions for the comparison of two treatments/genotypes, where  $j_1$  and  $j_2$  specify the columns of the design matrix corresponding to the treatments/genotypes of interest. The second one offers the confidence regions for general linear combination of coefficients,

where  $L$  is a contrast vector under model (8). This includes estimating the average growth curve of a particular genotype over all the blocks.

As an example, we consider plant size in this study. The confidence region of the block effect function  $p_3(t)$  infers whether there is a significant difference in plant size between block 1 and block 3, given the same plant genotype. By choosing  $j_1 = 142$  (the last two columns in the design matrix correspond to blocks 2 and 3, respectively) and  $j_2 = 1$  (intercept) in the function “CI\_contrast,” Figure 5(a) shows the 95% confidence region of  $p_3(t)$  from day 1 to day 44, which shows a significant positive effect of block 3 compared to block 1 especially in the later dates. Similarly, we can construct the confidence regions for  $g_2(t) - g_3(t)$  by setting  $j_1 = 2$  and  $j_2 = 3$ , which test for the significance of the difference between the 2<sup>nd</sup> and the 3<sup>rd</sup> genotypes. Figure 5(b) shows the 95% confidence region for the genotype effect, which is not significant over the whole time course.

Under model (6), when the contrast vector  $L$  in the function CI takes  $(1, 0, \dots, 0, 1/3, 1/3)$  and  $(1, 0, 1, 0, \dots, 0, 1/3, 1/3)$ , respectively, we obtain the estimated growth curves for genotypes 1 and 3 averaging over the three blocks with their 95% confidence regions, see Figure 3(a). We see that genotype 3 significantly grew faster than genotype 1 from day 10 to day 30. Similarly, Figure 3(b) shows the growth curves for genotypes 2 and 3 with their 95% confidence regions, which overlap with each other and demonstrate no significant difference between those two genotypes. This coincides with the result from Figure 5(b).

## 6. Discussion

In this paper, we developed a comprehensive pipeline for analyzing high-throughput plant phenotyping data that includes RGB image preprocessing, plant feature extraction, and functional data modeling and inference for growth curve dynamics. In recent literature, there has been an increasing trend of using deep convolutional neural networks (CNN) to extract plant traits from images, see Lu et al. [20] and Miao et al. [21] for counting tassels and leaves of maize plants, respectively; Pound et al. [22] for identifying wheat images containing spikes and spikelets as well as counting their numbers; and Aich et al. [23] for estimating emergence and biomass of wheat plants. Compared to the traditional approaches used in the proposed pipeline (image segmentation by thresholding + feature calculation + statistical modeling), the deep learning methods are able to work under an unconstrained field environment, where the thresholding method may not give a reliable and robust separation of the plants from backgrounds. However, such deep neural nets typically have a vast number of parameters. A sufficiently large training sample and intensive computation are required to train those models. Preparing the training data is both time and labor consuming. As a comparison, thresholding methods are computationally efficient and easy to implement without a training set, and they can give accurate segmentation for plants under homogeneous backgrounds, as in a greenhouse environment. More importantly, the statistical analysis is able to model and study the biological mechanism

on the plant growth, which is an advantage over the ‘black box’ methods.

In a recent work, Adams et al. [24] proposed a neural network model trained on over half million pixels from maize images for plant segmentation. Although, it is more time consuming than the thresholding method in computation, this method is able to achieve robust plant segmentation under noisy backgrounds. We will include such neural network methods in the next version of the “implant” package to improve the current functionality for field images.

Beside the RGB images, the UNL greenhouse also takes the hyperspectral images for every plant. Compared to RGB images which only have three channels, the hyperspectral images record the pixel intensities at every 5 nm over the whole spectrum, which contain more information than RGB images. The hyperspectral images can be used to separate plant organs and predict chemical concentration within a plant [19]. In future works, we will extend the HMRF model and functional ANOVA to hyperspectral images for studying traits from different plant organs.

## Data Availability

Data will be made available on request.

## Conflicts of Interest

The authors declare that they do not have any commercial or associative interest that represents a conflict of interest in connection with the work.

## Authors’ Contributions

R.W., Y.Q., and Y.Z. developed the package and conducted the analysis. R.W., Y.Q., Y.Z., and J.C.S. wrote the manuscript. Z.L. and J.C.S. conceived the experiment.

## Acknowledgments

The authors wish to acknowledge Yuhang Xu, Jinyu Li, and Zheng Xu for their helpful suggestions and feedback.

## References

- [1] A. Bucksch, J. Burrige, L. M. York et al., “Image-based high-throughput field phenotyping of crop roots,” *Plant Physiology*, vol. 166, no. 2, pp. 470–486, 2014.
- [2] N. Fahlgren, M. A. Gehan, and I. Baxter, “Lights, camera, action: high-throughput plant phenotyping is ready for a close-up,” *Current Opinion in Plant Biology*, vol. 24, pp. 93–99, 2015.
- [3] A. Hairmansis, B. Berger, M. Tester, and S. J. Roy, “Image-based phenotyping for non-destructive screening of different salinity tolerance traits in rice,” *Rice*, vol. 7, no. 1, p. 16, 2014.
- [4] E. R. Davies, *Computer and Machine Vision: Theory, Algorithms, Practicalities*, Academic Press, 2012.
- [5] A. Hartmann, T. Czauderna, R. Hoffmann, N. Stein, and F. Schreiber, “Htpheno: an image analysis pipeline for high-throughput plant phenotyping,” *BMC Bioinformatics*, vol. 12, no. 1, p. 148, 2011.



- [6] A. C. Knecht, M. T. Campbell, A. Caprez, D. R. Swanson, and H. Walia, "Image harvest: an open-source platform for high-throughput plant image processing and analysis," *Journal of Experimental Botany*, vol. 67, no. 11, pp. 3587–3599, 2016.
- [7] M. A. Gehan, N. Fahlgren, A. Abbasi et al., "Plantcv v2: image analysis software for high-throughput plant phenotyping," *PeerJ*, vol. 5, article e4088, 2017.
- [8] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis, Volume 5*, Prentice hall, Upper Saddle River, NJ, 2002.
- [9] G. Celeux, F. Forbes, and N. Peyrard, "Em procedures using mean field-like approximations for markov model-based image segmentation," *Pattern Recognition*, vol. 36, no. 1, pp. 131–144, 2003.
- [10] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*, Springer Series in Statistics, Springer, 2005.
- [11] Y. Xu, Y. Qiu, and J. C. Schnable, "Functional modeling of plant growth dynamics," *The Plant Phenome Journal*, vol. 1, no. 1, pp. 1–10, 2018.
- [12] C. Gu, *Smoothing Spline ANOVA Models*, vol. 297 of Springer Science & Business Media, Springer, 2013.
- [13] D. N. Da Silva and J. D. Opsomer, "Local polynomial regression," *Survey Methodology*, vol. 165, 2009.
- [14] J. O. Ramsay, H. Wickham, S. Graves, and G. Hooker, *fda: functional data analysis. r package version 2.2.6*, 2010.
- [15] M. Goyal, "Morphological image processing," *International Journal of Computer Science and Telecommunications*, vol. 2, p. 161, 2011.
- [16] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 36, no. 2, pp. 192–225, 1974.
- [17] Q. Wang, *Hmrf-em-image: implementation of the hidden markov random field model and its expectation261 maximization algorithm*, 2012, <http://arxiv.org/abs/1207.3510>.
- [18] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain mr images through a hidden Markov random field model and the expectation-maximization algorithm," *IEEE Transactions on Medical Imaging*, vol. 20, no. 1, pp. 45–57, 2001.
- [19] Y. Ge, G. Bai, V. Stoerger, and J. C. Schnable, "Temporal dynamics of maize plant growth, water use, and leaf water content using automated high throughput rgb and hyperspectral imaging," *Computers and Electronics in Agriculture*, vol. 127, pp. 625–632, 2016.
- [20] H. Lu, Z. Cao, Y. Xiao, B. Zhuang, and C. Shen, "Tasselnet: counting maize tassels in the wild via local counts regression network," *Plant Methods*, vol. 13, no. 1, p. 79, 2017.
- [21] C. Miao, T. P. Hoban, A. Pages et al., *Simulated Plant Images Improve Maize Leaf Counting Accuracy*, bioRxiv, 2019, 706994.
- [22] M. P. Pound, J. A. Atkinson, D. M. Wells, T. P. Pridmore, and A. P. French, "Deep learning for multi-task plant phenotyping," *017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2017, pp. 2055–2063, Venice, Italy, October 2017.
- [23] S. Aich, A. Josuttis, I. Ovsyannikov et al., "Deepwheat: Estimating phenotypic traits from crop images with deep learning," *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2018, pp. 323–332, Lake Tahoe, NV, USA, March 2018.
- [24] J. Adams, Y. Qiu, Y. Xu, and J. Schnable, "Plant segmentation by supervised machine learning methods," *The Plant Phenome Journal*, vol. 3, 2019.