15th IAPR International Conference on Machine Vision Applications (MVA)
Nagoya University, Nagoya, Japan, May 8-12, 2017.

Phenotyping of Xylem Vessels for Drought Stress Analysis in Rice

Swati Bhugra*, Anupama Anupama**, Santanu Chaudhury*, Brejesh Lall* and Archana Chugh**

*Department of Electrical Engineering, Indian Institute of Technology, Delhi
**School of Biological Sciences, Indian Institute of Technology, Delhi
{*eez138301@ee, blz138031@bioschool, santanuc@ee, brejesh@ee, achugh@bioschool*}*.iitd.ac.in*

Abstract

Xylem vessels play a pivotal role in plant adaptation to drought stress. In this paper, we propose a novel framework that associates automatic segmentation of xylem vessels with its morphological features as a quantitative proxy to predict drought stress response (DSR). We develop an image processing pipeline that comprises of low level processing which enables high-throughput detection of xylem vessels. With no prior information about its size and location, the proposed detection methodology gives an accuracy of 98%. The labelled data for DSR are either not available or are subjectively developed, which is a low-throughput and error prone task. We resolve this problem by employing simplex volume maximisation (SiVM) algorithm. The convex representations obtained from SiVM for each xylem in microscopic images based on its shape factors are aggregated to get an automated scoring of the whole plant. Bhattacharya distance is then employed to obtain the divergence of these responses w.r.t. the control group. The proposed framework successfully captures the phenotypic difference between MTU-1010 (drought susceptible rice cultivar) and Sahbhagi Dhan (drought tolerant rice cultivar).

1 Introduction

Drought is one of the major abiotic stress factors that limits crop productivity [1]. In order to develop drought tolerant cultivars, researchers notably rely in part on collection of specific characteristics related to plant structure and function [1]. These measurable attributes are called phenotypic traits. Root anatomical traits, in particular, xylem vessels have an important effect on plant functions, including acquisition of nutrients and water from the soil and thus, is of major interest for understanding plant adaptation to drought stress [2]. Quantification of drought stress response (DSR) based on xylem vessels provide researchers a direct selection criteria of drought tolerant cultivars [2]. This requires high-throughput analysis of large volume of microscopic data. But, in recent studies [3, 4, 5], the use of manual or semi-automated approaches for example manually segmenting tissue regions and calculating approximate xylem diameter or areas by manually fitting polygons using existing tools ImageJ [6], RootScan [7] and Nikon NIS elements, creates a bottleneck. In addition, scoring the level of drought stress based on these image based features is a complex problem due to the differential response of cultivars after drought stress induction and its variance among the replicates of a cultivar. Thus, well defined labels for the corresponding digital traits are either not available or are subjectively created, which is again a low-throughput and often error prone task.

Machine learning approaches can be used to discover the underlying principles that are too complex to model directly, in the absence of a well defined mathematical representation of DSR [8]. But due to the aforementioned reasons that result in lack of phenotypic data, xylem vessels have not yet been utilized to model the behaviour of drought tolerant and susceptible cultivars, that will assist in breeding cultivars with higher water use efficiency [7].

To resolve these issues, we propose a novel automated framework that can be employed to characterize drought stress response of various cultivars based on xylem vessels. The main contributions of this paper are as follows:

(1) Existing softwares employed by authors [3, 4, 5] to extract the phenotypic traits offers manually controlled tools, which makes the detection and extraction of the traits laborious and time consuming. They also limit the types of traits needed to quantify subtle but statistically significant changes in the morphology of xylem vessels in response to drought. Thus, in the first step we develop an image processing pipeline for automatic segmentation of xylem vessels and quantify its morphology using shape factors [9].

(2) In the next step, we identify latent drought clusters and their corresponding convex representation based on shape factors of the detected xylem vessels using simplex volume maximisation (SiVM) [10]. The progressive stages of drought are then quantified by aggregation of the convex representation of each xylem vessel in the image.

(3) We provide a novel index to quantify relative drought response between different cultivars using Bhattacharya distance [11] as a measure of divergence of these drought stages w.r.t to the control group.

The rest of the paper is organised as follows: In section 2 the experiment protocol for data acquisition is elucidated, section 3 explains the proposed framework, in section 4 the results are discussed and section 5 concludes the paper.

2 Dataset

Two series of drought experiments were conducted on rice pots during the kharif season of 2014 and

2015 at Indian Agricultural Research Institute, New-Delhi, India. In 2014, each genotype (Sahbhaghidhan, IR64 and MTU-1010) was divided into three groups (two replicates each) with differing irrigation intensities i.e. well-watered, reduced watered and unwatered. In 2015, each genotype (Pusa-44, NL-44 and CR262) was divided into three replicates of reduced watered and one of well-watered plants. These plants were harvested by gently excavating soil to obtain main root systems on each observation day for a period of 5 days. The tip of the primary roots as described by Bashar et. al [12] was used to obtain root sections. The images of these sections containing the stele region were taken with a scanning electron microscope (model EVO 50, Zeiss, UK).

3 Methodology

The input to the proposed framework is time series microscopic images of stele cross-sections. It mainly comprises of four steps, explained in the following subsections.

3.1 Pre-processing step

The first step involves the isolation of stele region from the background. This can be achieved by thresholding. But the threshold may differ for each image which requires empirical estimation. Thus, Otsu's method [13] was employed for automatic threshold selection and to obtain the corresponding binary image. This algorithm groups pixels into two classes, foreground (i.e. the stele and any debris) and background, by minimizing the intra-class variances of these two classes. Prior to thresholding, Contrast Limited Adaptive Histogram Equalisation (CLAHE) [14] was employed to enhance the brightness difference between stele and the background. The binary image obtained from Otsu's thresholding comprises of the stele region, scale bar and isolated noise components. To obtain the stele, connected components labelling [13] was utilized. Since in the investigated binary image, stele comprises of majority of the foreground pixels, the label corresponding to maximum number of pixels was extracted.

3.2 Detection of xylem vessels

The main part of the xylem detection method uses a border following algorithm introduced by Suzuki [15]. Border between a connected component of pixels that correspond to foreground (pixel value = 1), and a connected component of pixels that belong to hole (pixel value = 0) is described by the chain codes. Then topological structures of the image are extracted and these structures are described by contours. For each contour detected using this algorithm, the corresponding minimum enclosed circle (radius and centre) was also collected. Since the binary image representing the stele contains both xylem and meta-xylem vessels as contours, both are segmented from this algorithm. Xylem vessels were then identified within the stele based on their size, since they are generally larger than the meta-xylem. The ranked list of the radius of minimum enclosed circles was created for each image and the contours above the maximum step difference are labelled as xylem vessels.

3.3 Feature Extraction

While Kadam et al. [3] and Burton et al. [4] extracted the mean xylem vessel area, Haworth et al. [5] extracted the frequency distribution of diameter of xylem vessels as digital traits. These extracted traits are dependent on the magnification of the microscopic images. To eliminate the magnification as an input parameter to the framework, shape factors were employed to characterise the morphology of detected xylem vessels. The shape factors considered in this work are the following:

(1) Aspect Ratio (A_R) of the contour is a function of the largest diameter (d_{max}) and the smallest diameter (d_{min}) orthogonal to it: $A_R = d_{min}/d_{max}$.

(2) Circularity (C_R) is defined in terms of the perimeter P and the area A of the contour: $C_R = 4\pi A/P^2$.

(3) Solidity (S) is the measurement of the overall concavity of the contour. It is a function of contour area (A) and convex hull area (A_c): $S = A/A_c$.

(4) Roundness (R) is defined as a function of contour area (A) and the largest diameter (d_{max}): $R = 4A/\pi d_{max}^2$.

These shape based features contains complementary information about the morphology of xylem vessels. Thus, they were fused at the feature level to quantify drought stress response, as explained in the next subsection.

3.4 Unsupervised labelling to identify different stages of drought stress response

Well-defined labels to denote different drought stress responses (DSR) are not available or obtained from the experts which introduces subjectivity. Hence, the objective of the unsupervised method was to group similar phenotypic traits to categorize DSR and obtain interpretable representation of the patterns in the phenotypic data. Clustering approach was employed for the unsupervised response analysis.

The goal of clustering is to find a set of k basis vectors expressed as $\mathbf{W} \in R^{d\times k}$ from the d-dimensional data-set with n samples represented as $\mathbf{V} \in R^{d\times n}$ with rank $r \leq \min(d, n)$. The coefficient matrix $\mathbf{H} \in R^{k\times n}$ is defined as the belongingness of data points to each centroid. This can be cast as a matrix factorisation task, and factor matrices $\mathbf{W} \in R^{d\times k}$ and $\mathbf{H} \in R^{k\times n}$ are determined using following minimisation problem:

$$\min_{\mathbf{W},\mathbf{H}} \|\mathbf{V} - \mathbf{WH}\|, \tag{1}$$

where $\|\cdot\|$ denotes matrix norm and $k \leq r$.

Common approaches to achieve this factorisation include Principal component analysis (PCA) [16], k-means [16], Non-negative Matrix Factorisation (NMF) [16], fuzzy k-means [16] and recently developed SiVM. Although these methods minimise the same criteria, they differ in terms of constraints and produce different matrix factors. PCA constraints $\mathbf{W}$ with uncorrelated basis vectors, k-means computes $\mathbf{W}$ as centroids of the clusters and $\mathbf{H}$ comprises of unary column vectors and NMF assumes $\mathbf{V}$,$\mathbf{W}$ and $\mathbf{H}$ to be non-negative

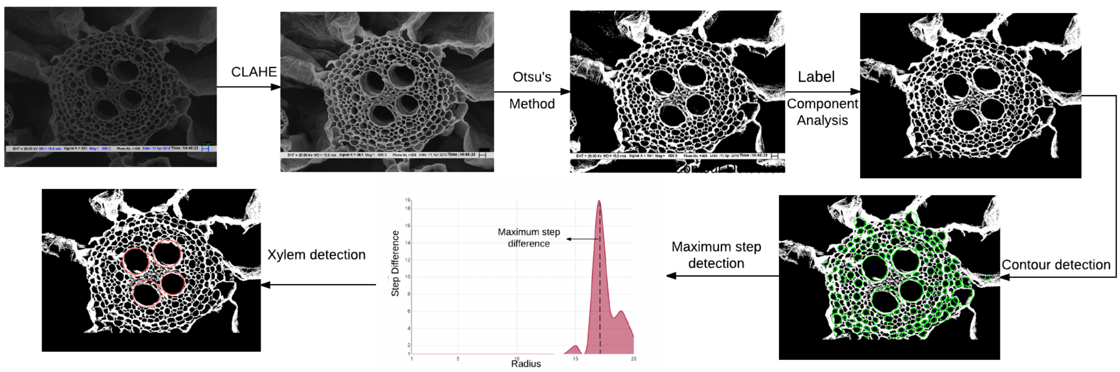


Figure 1. Automatic detection of Xylem Vessels

matrices. On the other hand, SiVM calculates the basis vectors by fitting a simplex with the maximum volume to the data. SiVM selects the basis vectors from the data matrix $\mathbf{V}$. Thus, the matrix $\mathbf{W}$ is defined as $\mathbf{W} = \mathbf{VG}$ where, $\mathbf{G} \in R^{n \times k}$ and is restricted to unary column vectors and $\mathbf{H}$ is restricted to convexity. We compared these commonly used clustering algorithms [8]: k-means, NMF, PCA, fuzzy k-means and SiVM on the data matrix $\mathbf{V}$, where every column $C_i = [A_R, R, C_R, S]^T$ and $i = 1, 2, \cdots, n$ to obtain a descriptive representation of drought responses in terms of the basis vectors. These are evaluated based on their interpretability and separability for different stages of drought.

4 Results

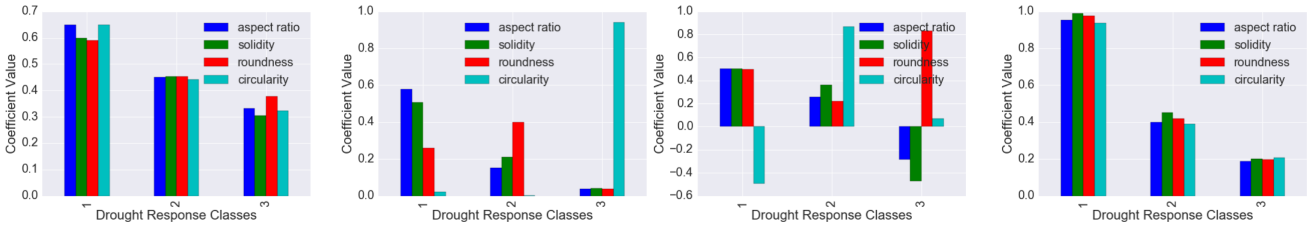
In this section, we present both quantitative and qualitative results obtained by the proposed framework. The result of automatic detection of xylem vessels in microscopic images is shown in Figure 1. Microscopic images from both the experiments were employed to obtain the accuracy of proposed methodology, which was calculated to be 98%. Since xylem vessels are much larger in size than the meta-xylems, they were not misidentified. This remains valid even for deformed xylem vessels under drought stress, as the radius of the minimum enclosing circle of these deformed vessels was used in the determination of maximum step difference. But it was observed that for the images with sheared cross-sections, vessels were not clearly defined, and the portions of the cross-sections appeared compressed. This resulted in false positives as some of the contours in the outer border were detected as xylem.

From the detected xylem vessels, different shape factors were extracted and fused to obtain compact representation of the vessel's morphology. The first two components of this feature vector is shown in supplementary Figure 1 for Sahbhaghi Dhan under control and third day of drought. To obtain the objective scoring of the drought responses ($\mathbf{V}$), PCA, NMF, k-means and SiVM methods were compared. Since, it was observed that the results of both k-means and fuzzy k-means were similar, we included only k-means results. The drought classes/basis vectors were fixed to $k = 3$, since the cultivars were subjected to three types of drought treatments. Figure 2 shows the basis vectors $\mathbf{W}$ computed from the aforementioned methods. The basis vectors obtained using PCA lacked physical meaning since its components have negative values that do not correspond to actual drought responses (shape factors lies between 0 and 1). NMF gave characteristics parts rather than the archetypal behaviour. It was observed that the factorisation obtained using k-means and SiVM resulted in interpretable basis vectors. From Figure 2(c) and (d) the first basis vector corresponds to control group (as its components corresponds to perfect circular xylems), second is labelled as mild stress response and third vector as severe stress response (as the deviation from circular shape is the maximum). But more distinctive basis vectors characterising drought stages were obtained using SiVM as compared to k-means, since it provides data representation as a convex combination of extreme data points. Also, the belongingness matrix $\mathbf{H}$ in k-means comprises of unary column vectors. Thus, each response is associated with only one class. Since drought is not a discrete process, the convex representation of drought response obtained using SiVM provides a better representation of this continuous change. Thus, SiVM was further employed to quantify DSR. The convex representation for each xylem vessel present in the image was obtained. These local states of drought stress response were then aggregated to obtain a histogram of class frequencies H_d, which characterises the response of the whole plant. Supplementary Figure 2 contrasts H_d of Sahbhagi Dhan under control and second day of drought.

To study the divergence of the response of rice cultivar under drought condition based on H_d, control group was used as a reference. This divergence was computed using Bhattacharya distance [11] for each observed day. We then tested the hypothesis that for the same drought treatment, various cultivars respond differently based on this divergence measure. Linear regression model was employed to quantify this difference. The functional form of the linear regression model between the day of drought stress (x) and the divergence (y) is given by:

$$y = \beta_0 + \beta_1 \cdot x + \beta_{2i} \cdot i + \beta_{3i} \cdot x \cdot i \qquad (2)$$

where i is the genotype coded as 0/1. Student's t-test was used to detect statistically significant differences in the variation of drought response. From the model (supplementary Figure 3), it was observed that the slope showed statistical differences between the two



(a) k-means (b) NMF (c) PCA (d) SiVM

Figure 2. Drought response classes obtained using k-means, NMF, PCA and SiVM

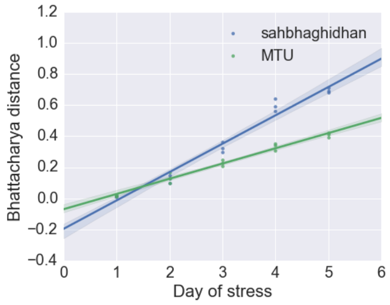


Figure 3. Linear regression models for genotypes MTU-1010 and Sahbhagi Dhan

genotypes, thus providing an index to quantify relative drought response between different cultivars. In Figure 3 the divergence of Sahbhaghi Dhan (drought tolerant cultivar) is more as compared to that of MTU-1010 (drought susceptible cultivar). This demonstrates that morphological changes in the xylem vessels as a response to drought is more in Sahbhaghi Dhan than MTU-1010. These results show the potential of our framework to characterise the response of unknown cultivars using the response of known cultivars and selection of the drought tolerant cultivars based on it.

5 Conclusion

In this article, we presented a novel framework that provides an automated description of drought stress using SiVM based on fusion of shape based features of xylem vessels. The proposed method also captured the phenotypic difference between a drought susceptible rice cultivar (MTU-1010) and a drought tolerant rice cultivar (Sahbhagi Dhan). This result demonstrates the potential of this approach to quantify the relative drought responses between different cultivars, which can further be used for desired trait selection. The findings contribute to the ongoing studies to predict drought stress based on phenotypic traits. We plan to conduct similar experiments to quantify other phenotypic traits of different rice genotypes and quantify the relation between them. This will enable high throughput characterisation of different drought mechanisms of rice.

References

[1] Furbank, Robert T and Tester, Mark: "Phenomics–technologies to relieve the phenotyping bottleneck", *Trends in plant science*,vol. 16, no. 12, pp. 635–644, 2011.

[2] Comas et al.: "Root traits contributing to plant productivity under drought", *Ecophysiology of root systems-environment interaction*, pp. 18, 2014.

[3] Kadam et al.: "Does morphological and anatomical plasticity during the vegetative stage make wheat more tolerant of water-deficit stress than rice?", *Plant physiology*, pp. 114, 2015.

[4] Burton et al.: "Phenotypic diversity of root anatomical and architectural traits in species", *Crop Science*, vol. 53, no. 3, pp. 1042-1055, 2013.

[5] Haworth et al.: "Xylem morphology determines the drought response of two Arundo donax ecotypes from contrasting habitats",*GCB Bioenergy*, 2015.

[6] Abràmoff et al.: "Image processing with ImageJ", *Biophotonics international*, vol. 11, no. 7, pp. 36–42, 2004.

[7] Burton et al.: "RootScan: software for high-throughput analysis of root anatomical traits", *Plant and Soil*, vol. 357, no. 1-2, pp. 189-203, 2012.

[8] Singh et al.:"Machine learning for high-throughput stress phenotyping in plants", *Trends in plant science*, vol.21, no. 2, pp. 110–124, 2016.

[9] Yonekawa et al.: "Identification of idealized leaf types using simple dimensionless shape factors by image analysis", *Transactions of the ASAE*, vol. 39, no. 4, pp. 1525–1533, 1996.

[10] Thurau et al.: "Yes we can: simplex volume maximization for descriptive web-scale matrix factorization", *ACM*, pp. 1785–1788, 2010.

[11] Comaniciu et al.: "Kernel-based object tracking", *IEEE Transactions on pattern analysis and machine intelligence*, vol. 25, no. 5, pp. 564–577, 2003.

[12] Bashar MK: "Xylem vessel variability at three positions of rice root (Oryza sativa L.)", *Bangladesh Rice Journal (Bangladesh)*, 1990.

[13] Gonzalez et al.: "Digital image processing", 2002, Prentice hall Upper Saddle River, NJ.

[14] Zuiderveld, Karel: "Contrast limited adaptive histogram equalization", *Graphics gems IV*, pp. 474–485, 1994.

[15] Suzuki et al.: "Topological structural analysis of digitized binary images by border following", *Computer Vision, Graphics, and Image Processing*, vol. 30, pp. 32–46, 1985.

[16] Friedman et al.: "The elements of statistical learning", vol. 1, 2001, Springer series in statistics Springer, Berlin.