Work in Progress, NOT YET FINISHED

Francel Lamprecht, Christine Robinson, Regina Wehler

Winter Semester 2018/19

Contents

1 Introduction

Plant phenotyping is a research area in the plant sciences that focuses on the quantitative measurement of both functional and structural properties of plants. During the last decade plant phenotyping evolved to non-destructive, image-analysis-based phenotyping. This computer driven approach allows the characterization of plants in a high-throughput manner (Walter et al., 2015).

The focus of this plant phenotyping research project was on the processing, segmentation and classification of *A. thaliana*, to gain valuable insight on how the image-analysis-based phenotyping in the life sciences works. *A. thaliana* is often used as a model plant, as the plant is small in size, with a rapid life cycle of about six weeks (Koornneef and Meinke, 2010).

The project was divided into two main parts: a) Image Analysis with Fiji in Java b) Explorative Data Analysis with Python in Jupyter Notebook.

2 Methodology

Each image was processed in a consecutive series of algorithms and tools which are implemented as plugins in the Fiji distribution (Schindelin et al., 2012) of the open-source software ImageJ (Rueden et al., 2017). The image analysis approach consists of three main steps (figure 1). First, the input RGB input image is segmented into two classes, plant and background by a trainable Weka segmentation classifier. Second, the binary class output is used to identify single leaf objects by watershed segmentation. The resulting binary mask is utilized to analyze properties of the leaf objects in the single red, green and blue channel, respectively.

2.1 Segmentation

The plant is separated from the background with the help of a trainable Weka segmentation classifier (Arganda-Carreras et al., 2017). Weka provides a GUI to train machine learning algorithms to produce pixel-based segmentations. The user can add traces to classes and train the classifier with those. Afterwards, traces/regions of interest can be adjusted and the classifier can be re-trained to improve classification. Six representative pictures are selected from the 2017 data set for training (plant029, plant145 and plant159 from A1; plant032, plant034 and plant037 from A2). Pictures are chosen to cover the whole range of plant green shades and the range of background characteristics of the given data sets. The Weka Experimenter was used to assess the performance of different machine learning algorithms. Based on these results (table 1), FastRandomForest was used with default parameters as a classifier. By applying the trained classifier to the RGB input images, for each input image a binary classification image is obtained (figure 1-b).

2.2 Objects Recognition

A plant consists of leaves that are attached to each other. To be able to analyze leaves individually, they need to be separated. Here, the watershed algorithm was used to separate touching objects. Its implementation in Fiji is based on the algorithm described by Kunt (1990). The algorithm first calculates an Euclidean distance map and determines center points as points which are, from a topological view, the ultimate eroded points, As the algorithm's name indicates, this topological map is "flooded" with water and at each collision of two "watersheds", a line separating two objects is drawn. Before watershed
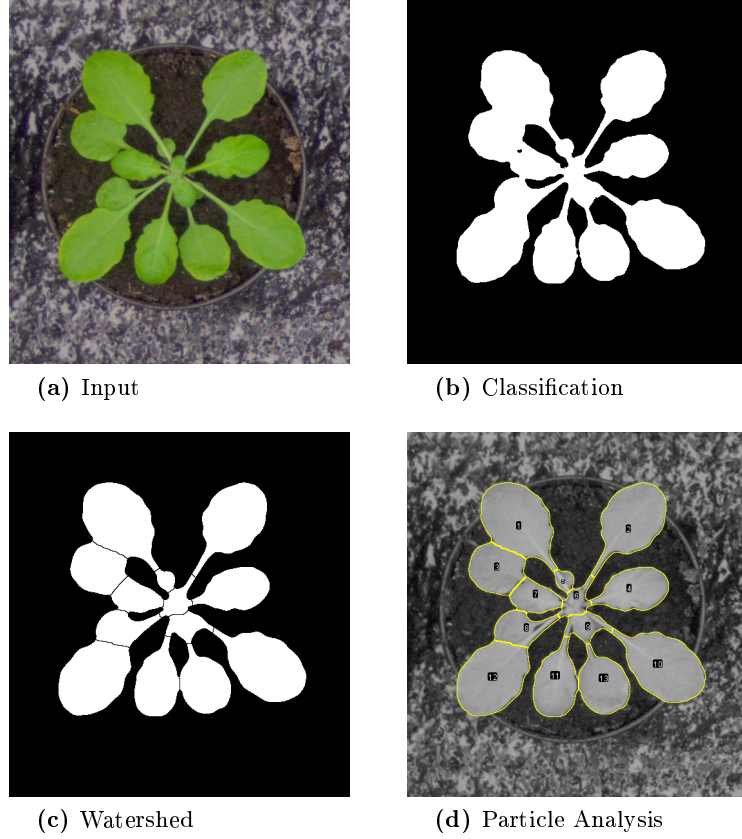
(a) Input

(b) Classification

(c) Watershed

(d) Particle Analysis

Figure 1 – The analysis of plant leaves is conducted in three steps. The RGB input image (a) is classified into plant and background by a trainable Weka segmentation classifier (b). Single leaf objects are identified by watershed segmentation (c). The binary mask is used to analyze properties in the split red, green and blue channel (d).

Table 1 – Comparison of classifier performance by Weka Experimenter. The default algorithm parameters were kept.

Algorithm	Percent correct	Precision	Recall	F score	Matthews correlation	AUC
FastRandomForest	99.93	1.00	1.00	1.00	1.00	1.00
SMO	92.52	0.89	0.93	0.91	0.85	0.93
k-nearest Neighbors	95.38	0.95	0.93	0.94	0.90	0.95
RandomSubSpace	99.85	1.00	1.00	1.00	1.00	1.00
Bagging	99.77	1.00	1.00	1.00	1.00	1.00
DesicisonTable	99.05	0.99	0.9	0.99	0.98	1.00

was applied, outliers were removed in both classes of the binary image by using the Fiji remove outliers function separately on each class. This function uses a median filter, for which the pixel radius was set to 6 and the threshold was set to 50. The output of the watershed algorithm is the binary image with added watershed lines (figure 1-c).

2.3 Object Analysis

After an image was subjected to the aforementioned steps, the resulting watershed image was then analyzed using Fiji's Analyze Particles tool. The measurements that were selected to be performed during the particle analysis are:

- Area - Measures the area of selection in square pixels. This was used to obtain the object's size.
- Mean gray value - Average gray value of the selection.
- Centroid - Provides the center point of the selection as the average X and Y coordinates for all pixels in the selection or image.
- Bounding rectangle - This is the smallest possible rectangle that encloses the selection. It gives the X and Y coordinates in respect of the upper left corner of the rectangle, as well as the height and width, from which the width-to-height-ratio will be calculated.
- Shape descriptors - These descriptors are used to attain the roundness of each object.
- Feret's diameter - Can be used to find the longest distance between any two points on the selection boundary.
- Integrated density- Used to indicate the sum of the brightness, as it sums the values of the pixels in the image or selection

Particles with a pixel size less than 70 were excluded from the "Analyze Particles" command. The value of 70 was established by trial and error, by inspecting the lines drawn by the watershed algorithm to separate objects.

Fiji automatically generates a Results Table, with the selected measurements. These results can be accessed by using the `getResultsTable()` method in the source code, to select the appropriate results for further use. By using this method, a csv file with all the results deemed necessary was created.

The generated Results Table results ("columns") that were included in the csv file can be seen in Table 2.

Table 2 – Measurement options and the corresponding value in the csv file.

Measurement option	Corresponding value in the csv file
Display Labels	objectID and imageID
Area	size
Centroid	pos_x and pos_y
Shape descriptors	roundness
Integrated density	brightness_sum
Mean gray value	brightness_average
Bounding rectangle	width_to_height_ratio

As brightness values calculated from binary watershed images is insufficient for plants, Fiji's "Analyze Particles" command was also applied to the original RGB images. The Particle Analyzer however, can only work with one channel (ie. grayscale) images. Working with the original RGB images required a split in the color channels. The analysis was then done on the chosen channel. The brightness (both the sum and the average) of each of the

three color channels (red, green and blue) were individually calculated. These individual channel results were then added together to calculate the sum of the brightness and average brightness for each object. To ensure computational efficiency not all the measurements were selected again, as this would have resulted in the duplication of results, eg. the size of the objects will remain constant, it is independent of the type of input image. The measurements that were necessary to be applied to the original RGB images, are the mean gray value and the integrated density. Results regarding the brightness of each color channel were attained and added to the csv file.

An additional requirement of the project was to include the width to height ratio. As this is not a built-in measurement option in Fiji, the width and the height that were included in the Results Table, as part of the "Bounding rectangle" measurement were extracted and the ratio was calculated, by dividing the leaf width by the leaf height. The calculated width to height ratio was added to the csv file to be available for further use.

The Feret Ratio was additionally calculated from the minimum caliper diameter (MinFeret) and the maximum caliper (MaxFeret) values. These values are calculated as part of the Feret's diameter measurements. The ratio was calculated, by dividing MinFeret by MaxFeret, as another approach to get a width to height ratio overview for each leaf.

2.4 Explorative Data Analysis

Data analysis was conducted in Jupyter Notebooks using several Python packages. Data were handled with the python package pandas. For visual statistical analysis, the seaborn package was used. Numerical statistical analysis was mainly performed using the stats module from the python package scipy. For general plotting of graphs, the pyplot module from the python package matplotlib was used.

2.5 Data

For this study, datasets of the Leaf Segmentation and Counting Challenges were used (Minervini et al., 2016). The imaging data were manually annotated to provide ground truth information (Scharr et al., 2014). The datasets contain images of greenhouse images with *A. thaliana* and *N. tabacum*. *A. thaliana* is a rosette plant, with the individual leaves being on separate, distinguishable stems. This characteristic of *A. thaliana* makes leaf separation from each other less complex. Given that we were only exploring viable methods in this research area, *A. thaliana* was our chosen plant, as it gave us the opportunity to explore with different algorithms in an easier way than what the *N. tabacum* dataset would have allowed us to do. Since the datasets for 2014, 2015 and 2017 are nearly the same, only differing in the amount of provided meta data, datasets from a single year were used for analysis.

3 Results

3.1 Overview of the 2014 datasets

3.2 Analyses of correlations

- Correlation of leaf size and roundness - Correlation of leaf size and color - Correlation of leaf number and total plant area
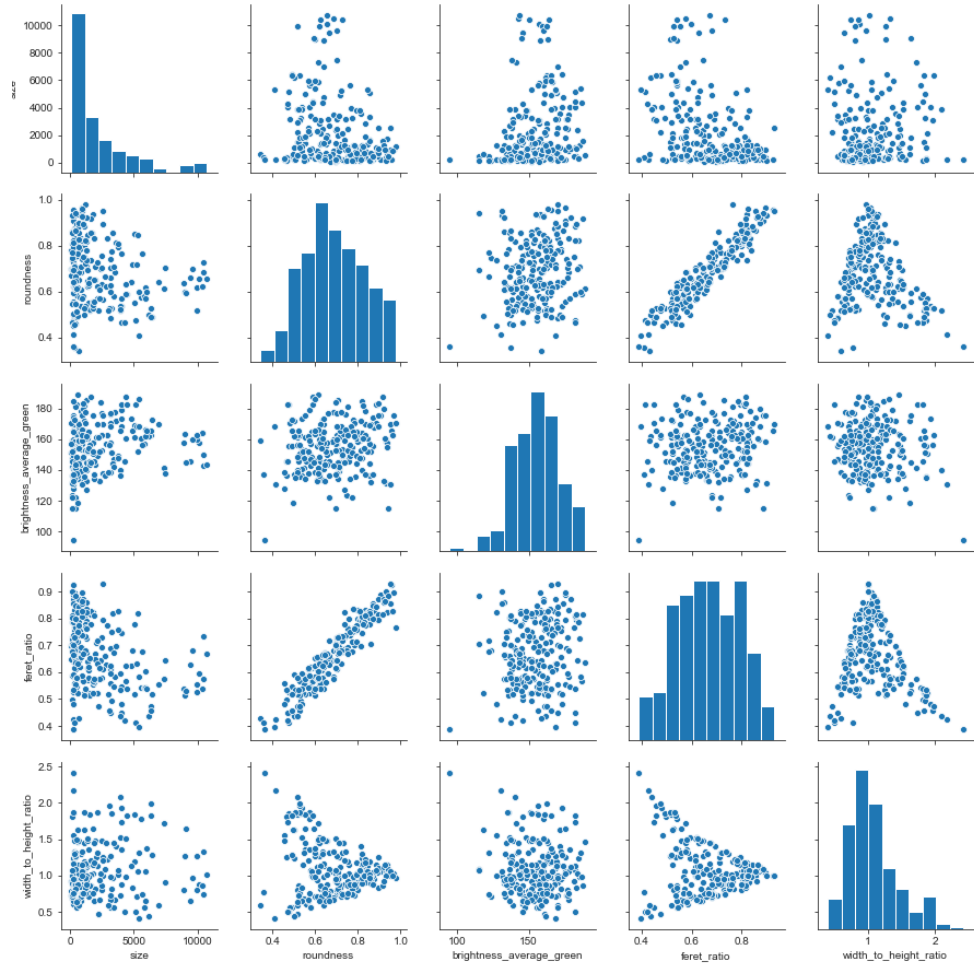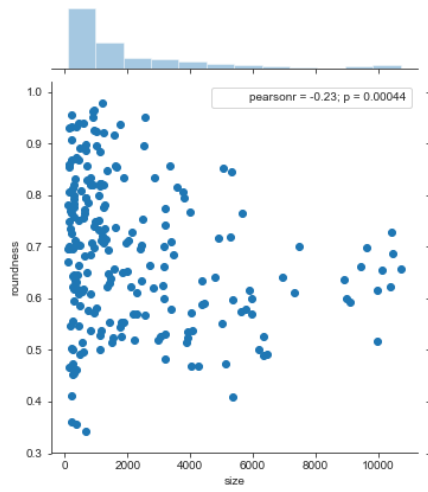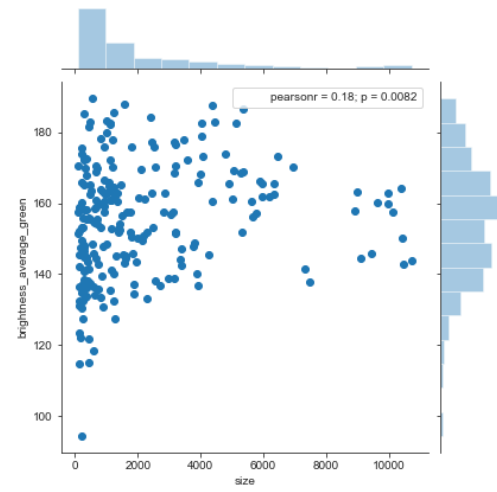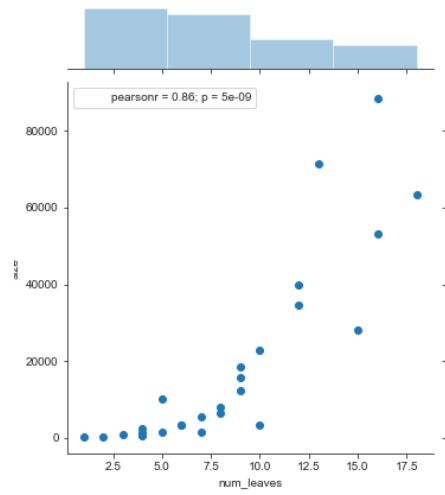
Figure 2 − Overview of correlations in the 2014 datasets shown in a pairplot matrix.
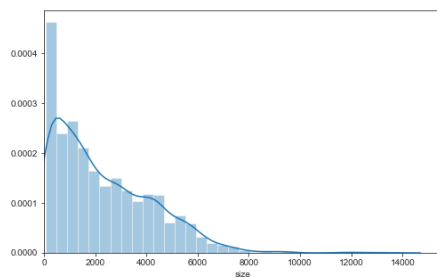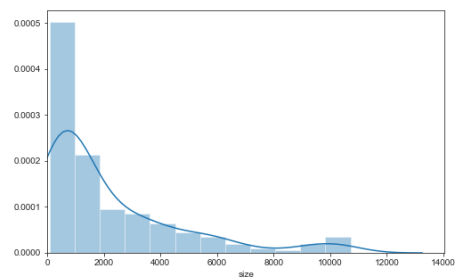
(a) Size vs. Roundness

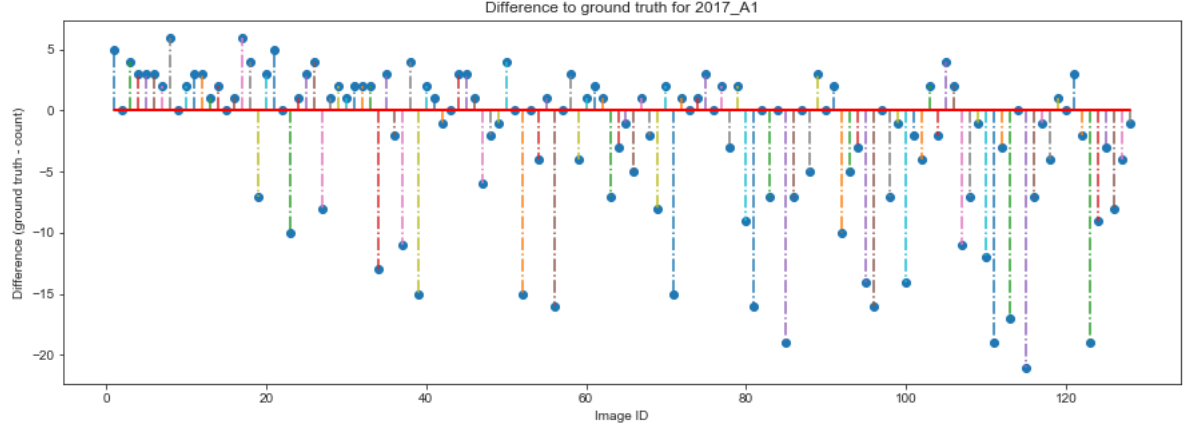**(b)** Size vs. Color

(c) Number of Leaves vs. Total Area

Figure 3 – XXX
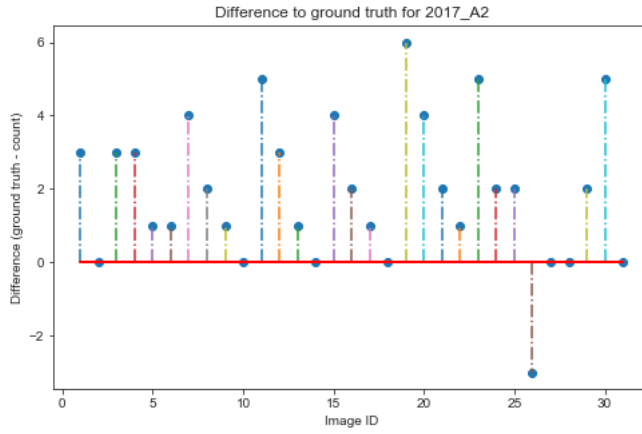

(a) A1 Dataset

(b) A2 Dataset

Figure 4 – XXX

(a) A1 Dataset

(b) A2 Dataset

Figure 5 – XXX

3.3 Comparison of A1 and A2 datasets

3.4 Comparison of leaf count to ground truth

4 Discussion

4.1 Challenges in the Identification of Single Leaves

- quality of classification depends on the selection of images for training and on the selection of traces

- need to find a balance between number of labeled pixels and classification accuracy because with increasing number of labeled pixels the reading of the classifier later requires more time

- watershed is working best for circular objects, that's why it sometimes fails on leaves with long stems

- detection and removal of outliers is difficult for very small plants -> with our pipeline we have lost very small plants (pixel number?)

4.2 Object Analysis

Object analysis to be included in the generated Results Table are clearly formulated in the ImageJ documentation (Ferreira and Rasband, 2012). A decision that was needed to be made, was whether to take an object's position from the the top left appearing pixel or from the centroid. In this project we used the centroid to indicate the object's position, as by using the top left appearing pixel position has no advantage over the centroid. By using the centroid positions, we ensured that the position coordinates are more robust to possible rotations. If the top left appearing pixel was used, problems could arrive if the leaves were to be rotated.

In the Fiji Results Table, circularity and roundness are two similar measurements. Circularity in ImageJ is calculated in such a way that it excludes local irregularities, therefore the roundness measurement which gives a value directly relative to the aspect ratio, was included in the csv file.

After calculating both the width to height ratio and the Feret ratio for each leaf, it was found that although the two ratio's did render corresponding values in some instances, there were other instances where the ratio values differed significantly. Manual calculations were done (measuring the length and width of a leaf on-screen with a ruler) and it was found that in the cases where there was a difference between the ratio's, the Feret ratio was more accurate. This is because the bounding rectangle is not turned by Fiji according to the leaf's position, instead the bounding rectangle remains parallel to the image border.

5 References

Arganda-Carreras, I., Kaynig, V., Rueden, C., Eliceiri, K. W., Schindelin, J., Cardona, A., and Sebastian Seung, H. (2017). Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification. *Bioinformatics (Oxford, England)*, 33(15):2424–2426.

Ferreira, T. and Rasband, W. (2012). *ImageJ User Guide: IJ 1.42r*.

Koornneef, M. and Meinke, D. (2010). The development of Arabidopsis as a model plant. *The Plant journal : for cell and molecular biology*, 61(6):909–921.

Kunt, M., editor (1990). *Visual Communications and Image Processing '90: Fifth in a Series*, SPIE Proceedings. SPIE.

Minervini, M., Fischbach, A., Scharr, H., and Tsaftaris, S. A. (2016). Finely-grained annotated datasets for image-based plant phenotyping. *Pattern Recognition Letters*, 81:80–89.

Rueden, C. T., Schindelin, J., Hiner, M. C., DeZonia, B. E., Walter, A. E., Arena, E. T., and Eliceiri, K. W. (2017). ImageJ2: ImageJ for the next generation of scientific image data. *BMC bioinformatics*, 18(1):529.

Scharr, H., Minervini, M., Fischbach, A., and Tsaftaris, S. A. (2014). Annotated Image Datasets of Rosette Plants.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., and Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature methods*, 9(7):676–682.

Walter, A., Liebisch, F., and Hund, A. (2015). Plant phenotyping: from bean weighing to image analysis. *Plant methods*, 11:14.