Image-Based Phenotyping of the Mature Arabidopsis Shoot System

Marco Augustin[1], Yll Haxhimusa[1(✉)], Wolfgang Busch[2], and Walter G. Kropatsch[1]

[1] Pattern Recognition and Image Processing Group, Vienna University of Technology, Vienna, Austria
{marco,yll}@prip.tuwien.ac.at
[2] Gregor Mendel Institute of Molecular Plant Biology, Austrian Academy of Sciences, Vienna, Austria

Abstract. The image-based phenotyping of mature plants faces several challenges from the image acquisition to the determination of quantitative characteristics describing their appearance. In this work a framework to extract geometrical and topological traits of 2D images of mature *Arabidopsis thaliana* is proposed. The phenotyping pipeline recovers the realistic branching architecture of dried and flattened plants in two steps. In the first step, a tracing approach is used for the extraction of centerline segments of the plant. In the second step, a hierarchical reconstruction is done to group the segments according to continuity principles. This paper covers an overview of the relevant processing steps along the proposed pipeline and provides an insight into the image acquisition as well as into the most relevant results from the evaluation process.

Keywords: Image-based phenotyping · Geometrical/topological traits · Tracing · Hierarchical reconstruction · Network of curvilinear structures

1 Introduction

The functional analysis of genes became a popular and interesting challenge in life science in the last 30 years [31]. The correlation between genetic data and the likelihood for certain pathologies in humans or genes causing an increasing crop yield are only two important and relevant examples [9]. While sequencing the genome of model organisms can be solved effectively nowadays, the automatic extraction of quantitative characteristics describing the phenotype, so called *traits*, became the bottleneck for many large-scale functional genomic studies [11]. Modern image acquisition tools and the progress in computer vision offer new possibilities for high-throughput phenotyping studies. The development of image-based phenotyping pipelines can help to overcome the current drawbacks of large-scale genetic studies, which are the manipulation of thousands of samples and the subsequent manual determination of relevant traits [27]. *Arabidopsis thaliana* is a small flowering plant and a popular model organism. The determination of quantitative traits, e.g. the plant's root length, became

L. Agapito et al. (Eds.): ECCV 2014 Workshops, Part IV, LNCS 8928, pp. 231–246, 2015.
DOI: 10.1007/978-3-319-16220-1_17

an important challenge during large-scale genetic studies [15]. In this work 2D images of mature *Arabidopsis thaliana* are analyzed and traits concerning the *final* appearance of the plant are extracted. During the last stages of growing, the plant mostly consists out of stems and siliques (see sample image in Fig. 1). The analysis of the rosette and the roots of the plant are not of interest for this work. The plants used in this work grew in a natural environment where the environmental conditions were well known. Due to storage and shipping reasons the plants are dried and pressed before the images are taken. Thus some stems overlap and appear as possible branching points in a segmented image. Hence a correct reconstruction of the plant's "realistic" architecture has to be done before the relevant traits can be extracted. The complexity of the reconstruction increases with the number of critical points: branching-, crossing- and termination- points. Branching points are defined as points where one branch splits in multiple branches. Due to the projection into the 2D image space as well as the flattening of the plants, overlapping stems appear as branching points in the images without having a correlating physical connection in nature. A distinction between branching and overlapping regions has to be done and is crucial for the correct reconstruction of the branching topology. Termination points are defined as end points of the plant (e.g. siliques' tips) and start points are located nearby the rosette.

The study of plants using computer vision approaches is a rather *new* field compared to the study of other biological objects, e.g. blood vessels. The latest works focus on the extraction of quantitative traits describing the root system of different plants like *A. thaliana*, maize or rice [27]. While the early root

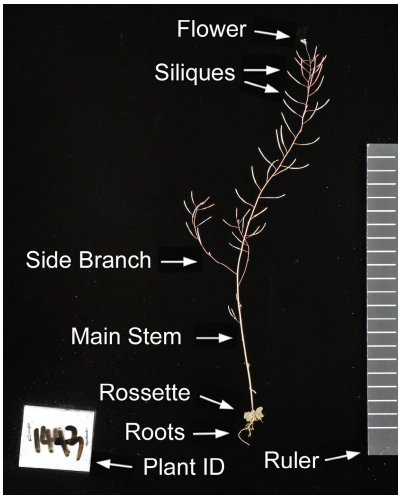


Fig. 1. Image of mature *A. thaliana*. The plant mainly consists out of stems/branches and siliques.

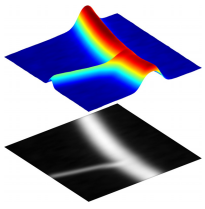


Fig. 2. Intensity profile of a branching region

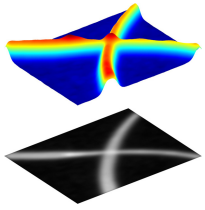


Fig. 3. Intensity profile of a crossing region

development (growth rate) is in focus of the works [4,14,23] other approaches investigate the ability of a root to respond to gravity changes in high-temporal analysis [28]. Systems, like in [1,32] analyze the root architecture only at a certain moment using a single image and traits like the root length or root angles are extracted. A number of high-throughput approaches exist to analyze details of plants' shoot systems (parts which are growing above ground). The authors in [2] developed a phenotyping pipeline for high-throughput studies which focuses on the analysis of the leaf growth behaviour. The interested reader is referred to [11] for an extended list of frameworks for *next-generation phenotyping*. It can be noticed that there is a lack of approaches for the analysis of the shoot system of *Arabidopsis*. Furthermore, most techniques focus on geometrical traits like curvature characteristics or lengths of certain plant parts. More complex traits describing the branching topology network of roots or stems, like branching patterns or branching orders are barely noted.

The detailed analysis of curvilinear structures like roots or stems is limited in the field of image-based plant phenotyping. Computer vision approaches are successfully developed in other fields dealing with similar structures like blood vessels or biological neural networks [13,22]. Generally, the approach of analyzing such networks is executed as a two-step approach in which the object is segmented in a first step and labeled in a second step. It already exists a great amount of different approaches [12,20,24,33–35] for the automated segmentation or skeleton extraction, the amount of approaches for labeling and grouping of trees or sub-trees inside these networks is rather limited [5,17,18,21,30].

In this work we propose the use of a tracing algorithm with a semi-circular search window to extract centerline segments of the plant's shoot system. Geometrical and topological properties, which are determined during tracing, are further used to group and label these segments based on continuity principles. We show that geometrical traits can be extracted accurately, whereas, a detailed topological analysis is still limited by the morphological complexity of the network as well as by noise.

The paper is structured as follows. In Sec. 2 a short overview of the image acquisition setup is described. In Sec. 3 we introduce the methodology used to extract the traits of the plant. The paper is concluded with experiments (Sec. 4) and conclusion (Sec. 5).

2 Image Acquisition of the *Arabdiopsis* Shoot System

The images which are analyzed in this work were taken as part of a project entitled "*The molecular basis of local adaption in A. thaliana*" led by Benjamin Brachi (Bergelson Lab, University of Chicago, US). Originally the images were acquired for the use of a manual phenotyping study and the use of computer vision methods for a computer guided analysis was originally not planned. The images were acquired in a darkened room using a digital single lens reflex camera. Two additional flashlights were used to ensure equal light conditions during the whole image acquisition process. The plants were put on a black velvet board to

guarantee a high contrast between object and background. A wipeable plant ID sign and a ruler were added to the velvet for later analysis (see Fig. 1). While the sign can be used for an automated data handling during phenotyping, the ruler is used for an automated spatial calibration. The images are taken with 12.1 megapixels (4284 × 2844 pixels) and are stored as TIFF (tagged image file format) in RGB color mode with 8 bit color depth.

To simplify the task of analyzing the architecture of a plant a few assumptions are made based on the image acquisition setup and preliminary experiments. In this work, the plants' stems and branches are defined as piecewise linear and their medial axis is represented by connected line segments. The stems' and branches' crosswise intensity profile approximates a Gaussian profile and the gray level changes along the stems and branches are smooth. The region of the rosette is defined as the origin of all main stems and the plant forms a tree-like structure. The stems' and branches' diameter decreases coming closer to the plant's end points and does not change abruptly. Further, the siliques' diameter increases before it decreases until a termination point is reached. Figure 2 and 3 show 3D intensity profile plots of specific parts (branching and crossing region) of a plant to illustrate some of the characteristics defined above. The term *"architecture"* of a plant is used to describe the different components of a plant regarding to space and time. Describing the plant's architecture is done by topological as well as geometrical information. Topological information describes the physical connection between different parts (e.g. siliques, leaves, flowers, branches) of the plant [16]. Concerning the topological characteristics, this architecture can range from simple plants (low number of critical points) to complex plants (high number of critical points). Geometrical information is used for a detailed description of the plant's components like size, shape, orientation or the spatial location of the components [16]. The quantitative traits extracted in this work comprise topological and geometrical information and are denoted in the following either as geometrical or topological traits. While the number of siliques (on different stems) is considered as a topological trait, the individual silique's length is considered as a geometrical trait.

3 Analysis and Reconstruction of the Plant's Architecture

To analyze 2D images of mature *Arabidopsis* a pipeline based on several computer vision methods is proposed as shown in Fig. 4. This section is divided according to the steps of this pipeline.

3.1 Pre-Processing

Due to logistical reasons the plants dried and the RGB color information of different plant components is distributed uneven in one image and over the whole dataset. Hence the images are transformed into grayscale images using the R-channel of the RGB color space. The images contain three different regions-of-interest (ROI) which are the *plant ID sign*, the *ruler* and the *plant* itself. These

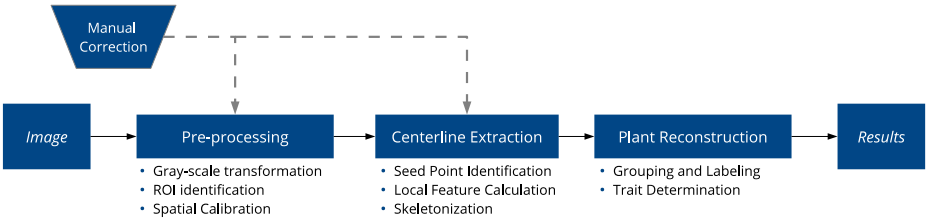


Fig. 4. Analysis and reconstruction framework

regions are extracted using image processing methods such as image pyramids, bit-plane slicing and mathematical morphology operators. These principles combined with constraints concerning the approximative location of the ROIs (e.g. the plant is always the most central object) are used to crop each image to the borders of the objects' bounding boxes. To make quantitative traits comparable to results of other studies the pixel units are transformed to *real-world metric units*, e.g. *mm*. The spatial calibration is achieved by use of the ruler. A routine based on gray-level thresholding and mathematical morphology operators is used to extract the conversion factor automatically in each image. A more detailed description regarding the sequence of steps during pre-processing is given in [3].

3.2 Centerline Extraction

The medial axis of curvilinear structures is an efficient representation to analyze characteristics concerning the geometry as well as topology. In this work a tracing approach is used to extract the centerline of the plant iteratively. Different variation of tracing algorithms are already successfully used for blood vessel analysis [7,12,29] as well as biological neural network analysis [22,35]. Tracing is a direct exploratory centerline extraction approach where a structure (often curvilinear/tubular) is followed from a starting point to an end point. Tracing provides a sequence of extracted centerline points, which represent samples on the medial axis of an object. Local features like tracing direction or radius at these points are determined and used for travelling along the object.

Seed Point Identification. Tracing algorithms are initialized at certain (starting) positions, so called *seed points*. A set of seed points is chosen either manually or by an automated procedure. Automated seed point identification procedures focus either on the detection of ridges/edges along scan lines [8,12] or on detection of these points in a certain ROI's neighborhood [17]. The automated seed point identification method in this work, is based on the assumption that one seed point per main stem is sufficient for a complete centerline extraction. This assumption should hold under regular conditions: low gray-value discontinuities and little noise in the images. As all main stems originate from the rosette, the rosette is roughly segmented and possible seed point candidates (ridge points) in

its neighborhood are estimated before the seed points are validated. The interested reader is referred to [3] for a detailed description about the seed point identification procedure.

Iterative Tracing. The process of extracting the centerline of an object using tracing is executed iteratively until at least one of certain stopping criteria is fulfilled.

Initialization. Each iteration is initialized at a start point p^k. While at the very beginning of the tracing procedure the seed point is chosen from the set of points from the seed point identification routine, subsequent start points are for example neighbouring centerline points or branching points which are determined during tracing.

Estimation. The location of the next centerline point along the structure is estimated by a dynamic semi-circular search window (Fig. 5). This window guarantees a constant look-ahead distance in all directions and the next centerline point is expected to be located along its circumference. The gray-value profile $S_r^k\left(p^k\right)$ along the circumference of a semi-circle with radius r^k at p^k is defined as a sequence of equally spaced samples n [7]:

$$S_r^k\left(p^k\right) = \{c_i,\ i = 0, 1, \cdots, n-1\} \quad . \tag{1}$$

$$c_i = V\left(p_x^k + r^k \cdot \cos\left(i \cdot \delta\theta\right), p_y^k + r^k \cdot \sin\left(i \cdot \delta\theta\right)\right), \qquad \delta\theta = \frac{\pi}{n-1} \quad . \tag{2}$$

Since V (normalized intensity values of the image I) is defined in the discrete space the values of c_i are determined using the nearest pixel values. The gray-values along the circumference are smoothed using a Gaussian filter. The radius r^k of the semi-circle must be defined "big" enough to cover all different stem widths and must be "small" enough to not detect points along neighbouring structures. For this reason, the radius r^k is adapted dynamically to the current structure at each iteration of the algorithm by [7]:

$$r^k = \rho \cdot \left[\max\{R^k, R^{k+1}\}\right] \quad . \tag{3}$$

R^k and R^{k+1} are the radius of the stem at the current and the next centerline point position. The constant factor ρ should be defined bigger than 1 to guarantee the coverage of the whole width of a stem.

Identification. The stems in this work appear as bright structures on a dark background and thus the intensity profile across the stems appears as a Gaussian-like shape (Fig. 2 and 3). To identify relevant centerline candidate points q_i, the intensity profile along the circumference is searched for local maxima (as indicated in Fig. 6 with triangles). If two or more stems run very close to each other, it can occur that local maxima are identified which belong to a neighboring stem. This can normally be avoided by choosing a small value for r^k in (2). Additionally,

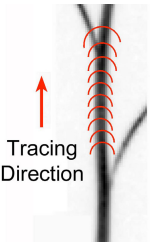


Fig. 5. A semi-circular template is propagating along a stem until a branching point is detected

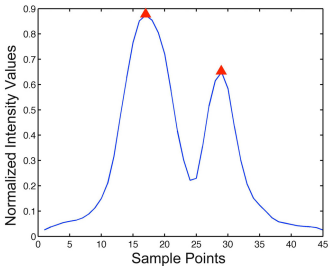


Fig. 6. The intensity profile along the semi-circular circumference of the search window is shown for the detection of the branching point. The triangles indicate the identified candidate points.

the intensity profile along the connection between p^k and the candidate points q_i is checked for values below 20% of the normalized intensity values [17]. If any point undercuts this threshold the candidate point is flagged as outlier-stem point. These points are stored in an additional list which is used for initialization at the very end of the tracing procedure. Figure 7 shows an overview of different points along the circumference of the search window.

Validation. At the end of an iteration each candidate point is validated by determining its local features. The accuracy of the features depends on the tracing direction as well as on small intensity variations across the stems. To gain a higher accuracy regarding the location, the first estimate q_0 (position of the local maxima) is refined by taking into account the stem's edge points. The refinement involves the following steps (see Fig. 8 for illustration) [7,17,29]:

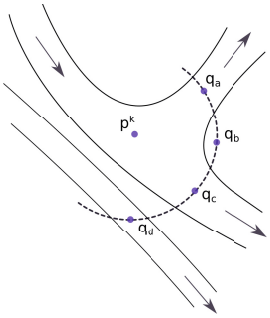


Fig. 7. Different types of candidate points can be identified: Stem points (q_a, q_c), outlier points (q_d) and non-stem points (q_b)

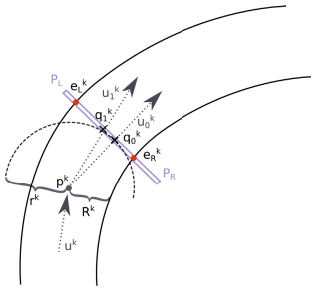


Fig. 8. The geometric determination and extraction of the local features at a centerline point

1. Calculate tracing direction for q_0^k

$$\overrightarrow{u_0^k} = \frac{p^k - q_0^k}{\|p^k - q_0^k\|} \quad . \tag{4}$$

2. Detect edge points along two linear intensity profiles P_{L} and P_{R} perpendicular to the direction $\overrightarrow{u_0^k}$ by finding the gradient's maxima. The width of the intensity profiles is chosen to be the same value as r^k.
3. The location of the final stem point q_1^k is calculated as the medial point between the edge points. The radius R^{k+1} is determined.
4. Update tracing direction

$$\overrightarrow{u_1^k} = \frac{p^k - q_1^k}{\|p^k - q_1^k\|} \quad . \tag{5}$$

The final position of the validated points can be influenced by a parameter α (set to 0.9 in this work) which regulates the step size according to:

$$p^{k+1} = p^k + \alpha \overrightarrow{u_1^k} \quad . \tag{6}$$

A centerline point, denoted as *STEL (STem-ELement)*, is defined as a set of geometrical and topological properties in this work:

- p^k: Location of current $STEL_k$ in pixel units
- $\overrightarrow{u}^k$: Tracing direction (unit vector) at p^k
- R^k: Radius of the stem at position p^k
- e_{L}^k, e_{R}^k: Edge point positions
- s^k: Normalized intensity level at p^k
- γ^k: Percent dynamic range from perpendicular intensity profile at p^k
- ID: A unique ID for each STEL
- $Parent$: ID of the previous STEL
- $Type$: Regular (one child), branching/crossing (two or more children), root (no parent/seed point), termination (no child), outlier (no parent)

Stopping Criteria. The skeleton of the plant structure is extracted by executing the previously described steps iteratively until certain criteria are fulfilled:

- Any of the pixels of the search window is outside the image range.
- No *valid* candidate points are identified.
- More than one valid candidate point is identified (branching/crossing).
- Connection between the current point p^k and one of the current candidate points q_i^k intersects the actual skeleton.
- Percent-dynamic-range falls below a threshold (e.g. 10 %).

To prevent tracing of already traced stem parts a centerline image is created and updated during the tracing procedure. The centerline image is a binary image with the size of the plant's image I (initialized with pixel values equal zero). During tracing the pixel values of the current stem-segments are set to one. Two STELs are connected using the Bresenham line drawing algorithm [10]. To validate for example a seed point, its neighborhood (e.g. 5×5) in the centerline image is checked for non-zero elements.

3.3 Plant Reconstruction

The tracing procedure results in a set of *unconnected* centerline stem segments (set of STELs). Due to the fact that the branching pattern of the plants do not follow an a-priori known principle and the *local* knowledge at branching and crossing points is not sufficient, the plant's realistic architecture cannot be reconstructed during tracing. A hierarchical reconstruction is proposed in this work to rebuilt the realistic architecture of the plants (see Fig. 9 for illustration).

Filtering. The structure of centerline segments is represented as an undirected graph $G = (V, E)$ where the segments are the nodes V and possible connections between two segments are represented as edges E. Segments shorter than 5 pixel are removed.

Clustering. Assuming that each centerline segment V_i can be connected to any other centerline segment V_j we obtain a complete graph. We know that a segment at the very top of the image is unlikely to be connected to a segment at the very bottom. Thus we are looking for a representation (i.e. a layer on top of the G) with a reduced set of possible connections. At this level each centerline segment is represented by an edge (S) and a set of possible connections between adjacent centerline segments as a (cluster) node (CN). Since we have topological information (STEL relations) that we found during tracing, we can reduce the set of possible connections. The tracing procedure can not distinguish between

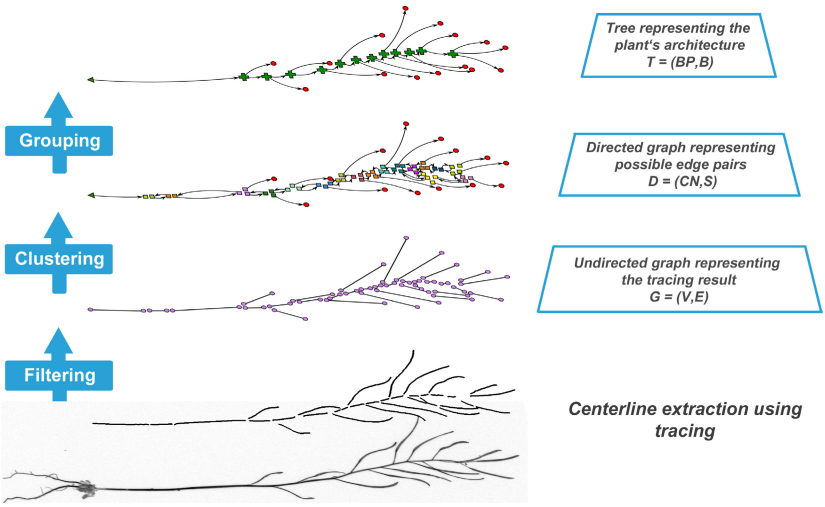


Fig. 9. Hierarchical reconstruction of the plants' architecture. The base layer consists of the centerline segments from the tracing procedure. The top layer represents the plant with use of branches and branching points. The root node is shown by a triangle, end nodes are represented by circles and branching points are represented by "+"-signs.

crossing and branching points. Thus intersection and unrelated stems can occur. Therefore we defined heuristic rules and an Euclidean distance to merge nearby critical points.

Grouping. Determining the costs for grouping possible edge pairs is based on continuity principles in this work. A cost term $c(S_i, S_j)$ for traveling from one segment S_i to another segment S_j is proposed as

$$c(S_1, S_2) = \theta(S_1, S_2) + d_1 \cdot \exp^{tor(S_1, S_2)} \quad . \tag{7}$$

Where $\theta(S_1, S_2)$ is defined as the edge direction similarity between two segments S_1, S_2 with directions $\overrightarrow{u}_1$, $\overrightarrow{u}_2$ [30]:

$$\theta(S_1, S_2) = acos\left(\frac{\overrightarrow{u}_1 \bullet \overrightarrow{u}_2}{|\overrightarrow{u}_1| \cdot |\overrightarrow{u}_2|}\right) \quad . \tag{8}$$

The "•" operator is denoting the dot product. Experiments during the empirical modeling of the cost function showed that the edge direction similarity is a good feature to group segments if the lengths of involved edge candidates are similar. To include information regarding the curvature of the resulting path B_{12} into the cost function an additional term in (7) was added. This curvature term is modeled using the tortuosity, which is calculated as the ratio between the path length l_{12} and the Euclidean distance between the endpoints d_{12} [30]:

$$tor(S_1, S_2) = \frac{l_{12}}{d_{12}} \quad . \tag{9}$$

The Euclidean distance d_1 between segments' end points is used to weight the curvature term of the cost function. The grouping process follows a trace-back principle which means that the tree is constructed from the exterior regions to the interior regions. Further, the cluster nodes are visited from less complex to more complex nodes. The number of possible edge-pairs in a cluster is defined as the grade of complexity. A tree is iteratively build by minimizing the cost function in (7).

3.4 Geometrical and Topological Traits

After constructing the realistic architecture of the plant, the quantitative traits for each plant are determined. *Topological traits* are used to quantify the "network topology" of the plant. Different branch types are identified in the reconstructed tree and their occurrence is determined:

- *Main Stem (MS):* Stems which originate from the rosette area.
- *Silique:* Terminating branch containing the seeds and having a characteristic width variation.
- *Leaf:* Terminating branch which is not classified as a silique.
- *Side Branch (SB):* Branch originating from a main stem/side branch.

The *geometrical traits* are calculated for each branch B:

- *Path Length* l_{B}*:* The length of the branch centerline calculated by the number of odd N_{o} and even N_{e} Freeman (8-directional) chain codes along the centerline:
$$l_{\mathrm{B}} = N_{\mathrm{e}} + N_{\mathrm{o}} \cdot \sqrt{2} \ . \tag{10}$$
- *Euclidean Length* d_{B}*:* The Euclidean distance between the branches' end points [21].

Silique Detection. The classification into siliques and *leaves* is based on the assumption that the majority of the exterior parts of a mature plant are siliques. Furthermore, siliques generally tend to have a higher variation in width and the lengths of siliques on one plant is more evenly distributed than the length of other leave parts. For this reason, the path length, the average width and the standard deviation of the width of the exterior branches are used to differentiate between siliques and other leaves. If the number of leaves on a plant is greater or equal 20, the robust *Minimum Covariance Determinant* estimator is used (outlier detection based on all three features) [25]. Otherwise the *Median Absolute Deviation* is used based on the branches' path lengths [26].

4 Experiments and Evaluation

The evaluation in this work was done with a set of 106 images (see Sec. 2). The traits extracted with the proposed framework (*PPFW*) are compared with traits from a semi-automatic analysis using *Fiji Simple Neurite Tracer* [19] (*GT Fiji*).

4.1 Measurement of the Plant Size

The accuracy of the final traits is evaluated using only images where the grouping process resulted in one connected object. This was the case in 89 out of 106 images. In the remaining approximately 15% the grouping process yield to multiple separated (sub)trees instead of one tree per main stem. This is caused by wrong local decisions in certain cluster nodes and an error propagation during the systematic grouping of edge candidates. Wrong local decisions were frequently observed in big cluster nodes with multiple smaller segments (multiple overlapping regions). These limitations originating from the grouping process should be reconsidered in future works. The lengths of certain branch types (MS, SB and Siliques) are averaged per plant and compared to GT Fiji. The evaluation concerning the stems' lengths is summarized in Tab. 1. The average relative error concerning the MS length between the two measurements is $(3.64 \ \pm 3.19)\,\%$. The average relative error concerning the siliques length is $(6.73 \ \pm 4.77)\,\%$. Scatter plots showing the individual measurements as well as the linear regression line (solid black) and the identity function (dashed line) for both measurements are shown in Fig. 10 and 11. The evaluation concerning the side branches shows a much higher deviation between GT Fiji and PPFW (see Tab. 1). This can be

Table 1. Comparison regarding the branches' lengths (path length l_B in pixels (Px)) and number of branches (N_B)

Branch Type	Ground Truth Fiji		Proposed Framework	
	l_B [Px]	N_B	l_B [Px]	N_B
MS	1983 (±731)	99	2030 (±756)	97
SB	667 (±441)	143	253 (±367)	540
Siliques	179 (±24)	2953	168 (±21)	2680

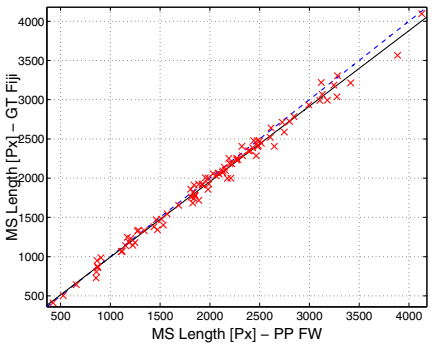


Fig. 10. Scatter plot comparing the GT Fjiji and the PPFW MS lengths (Pearson's r = 0.99 ($p < 0.01$))

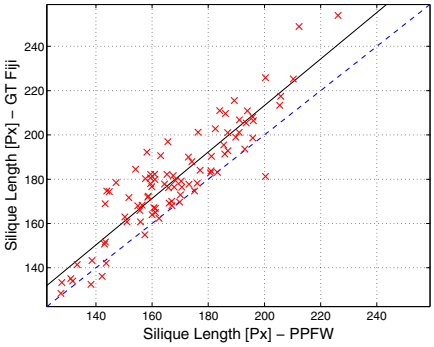


Fig. 11. Scatter plot comparing the GT Fjiji and the PPFW Siliques lengths (Pearson's r = 0.90 ($p < 0.01$))

explained by extensive overlappings in the exterior parts of the plant. Thereby a crossing is wrongly resolved as a new (mostly very short) side branch and terminating branches instead of only terminating branches. During the trait extraction this results in an overestimation of number of side-branches and an underestimation of number of siliques (compare deviation of number of siliques and SB in Tab. 1). Furthermore, this is also an explanation of the underestimation of the siliques' lengths which is noticeable in the scatter plot (Fig. 11) and the underestimation of the averaged side-branches lengths in Tab. 1.

4.2 Number of Siliques

The identification of the siliques is one of the major interests for biologists. The relative error is calculated by the difference between the number of siliques of GT Fiji and the number of siliques determined by the proposed framework. The mean relative error between GT Fiji and PPFW is $(11.66 \pm 9.09)\,\%$. To compare the deviation between GT Fiji and the PPFW more in detail a Bland-Altman plot [6] is shown in Fig. 12. While the mean number of siliques per plant is shown on the x axis, the absolute differences between the number of siliques determined for both measuring methods are shown on the y-axis. It can be noticed that more samples are spread below the zero line and the deviation increases by a rising number of siliques. An example of a labeled plant image

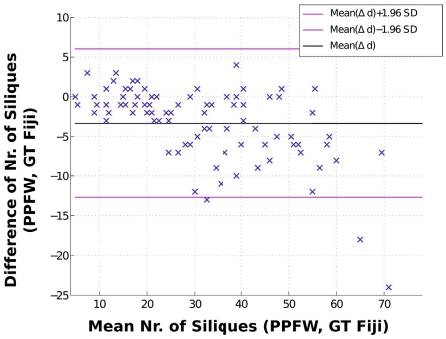
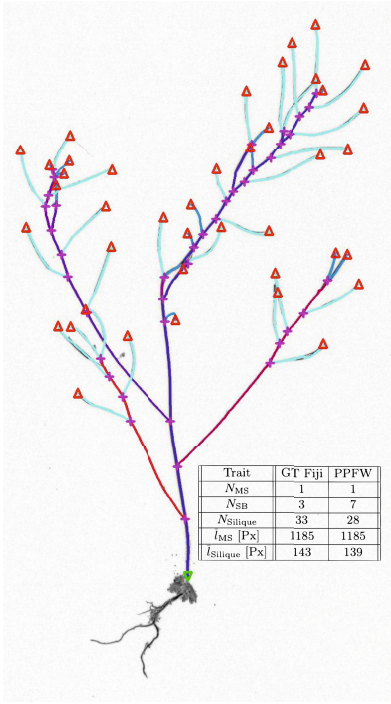


Fig. 12. Bland-Altman plot for number of siliques evaluation. An underestimation of number of siliques and can be observed. This underestimation increases with the number of siliques on a plant. The underestimation yields mainly from an error-prone classification of siliques in regions where many siliques tend to overlap with each other. This occurs in complex plants at the very upper parts.



Trait	GT Fiji	PPFW
N_{MS}	1	1
N_{SB}	3	7
N_{Silique}	33	28
l_{MS} [Px]	1185	1185
l_{Silique} [Px]	143	139

Fig. 13. Exemplary output of PPFW. The table summarizes the plant's traits determined by GT Fiji and PPFW. (N_{B} Number of branches type B; l_{B} Avg. length of branches type B.).

as well as its corresponding table of traits is shown in Fig. 13. For illustration reasons the plant image is inverted and the (overlaid) centerline is thickened. The different branches are marked in different colors while the siliques and leafs are shown in turquoise respectively blue. The green triangle denotes the start of the MS, red triangles denote the termination points and the purple "+" signs denote branching points. The table of traits contains the number of different branches (MS, SB, Siliques) as well as the lengths of the MS and the siliques.

4.3 Performance

Besides the accuracy of the measurements the performance of image-based phenotype systems is crucial when it comes to a laboratory use. Thus the performance of the framework was evaluated concerning the need of manual corrections (clicks). A manual correction of preliminary results is provided for the steps of the automated seed point identification and the tracing algorithm. While during the seed point identification 7 clicks were needed to overcome false positives and

false negatives, the tracing procedure required 112 additional clicks to complete the centerline extraction process (due to missed branchings or crossings). In total 119 clicks were necessary while processing all 106 images. The user decides if a mistakes has been made and if an interaction with the framework is needed.

5 Conclusion

In this work a semi-automatic framework is presented to extract geometrical and topological traits from 2D images of mature *Arabidopsis* plants. We showed that the use of a tracing approach to gain geometrical and topological information already during the segmentation of curvilinear objects can be valuable when topological characteristics are in focus of analysis. The accuracy and the grade of automation during the plant reconstruction depends on the quality of the pictures and on the morphological complexity of the plant structure. Unsupervised trait extraction using the proposed framework is reserved to plants with a limited morphological complexity and images with a uniformly high contrast.

The biggest source of error could be identified with overlapping branches and stems originating from the image acquisition. If bushy parts of the plants could be untangled before taking the images, the quality of the trait extraction would be improved. We have encountered problems finding reliable topological properties if multiple overlaps occur close to each other. In this case our tracing method produces a lot of small segments and the *"local"* systematic classification between branching and crossing regions gets error-prone. The problem lays in the way of how we set the rules for clustering and grouping. Especially when grouping small segments in large clusters (i.e. more than four candidate edges), there are many plausible way to connect them. Another drawback of is the side-effect of the error propagation as a wrong, local decision in a cluster node can lead to a propagation of this error to other nodes. An alternative to overcome this drawback would be the use of a global optimization approach, such as finding a minimum spanning tree. This could raise the quality of the results and should be part of future investigations.

From the phenotypic point of view an accurate topological reconstruction would rise the number of extractable traits in future works. Such traits can be the bifurcation ratio or the internodal distance between types of branches. Furthermore, with use of reliable topological and geometrical traits a context between these properties could be established. Image-based phenotyping of mature plants can be a valuable approach for the understanding of the correlation between the genotype, the phenotype and the environmental conditions of a plant. Combined with approaches to phenotype plants in earlier stages of growing, these approaches would open the way to (high-throughput) longitudinal studies covering the whole life-cycle of a plant.

Acknowledgments. We want to thank Svante Holm (Mid Sweden University, SE) and Alison Anastasio (University of Chicago, US) for planting and harvesting the plants, Man Yu and Andrew Davis for taking the photos and Benjamin Brachi

(Bergelson Lab, University of Chicago, US) for his valuable inputs and support along the stages of development.

References

1. Armengaud, P., Zambaux, K., Hills, A., Sulpice, R., Pattison, R.J., Blatt, M.R., Amtmann, A.: EZ-Rhizo: integrated software for the fast and accurate measurement of root system architecture. The Plant Journal **57**(5), 945–956 (2009)
2. Arvidsson, S., Perez-Rodriguez, P., Mueller-Roeber, B.: A growth phenotyping pipeline for Arabidopsis thaliana integrating image analysis and rosette area modeling for robust quantification of genotype effects. New Phytologist **191**(3), 895–907 (2011)
3. Augustin, M.: Extraction of quantitative traits from 2D images of mature arabidopsis plants. Vienna University of Technology, Master's Thesis (2014)
4. Basu, P., Pal, A., Lynch, J.P., Brown, K.M.: A novel image-analysis technique for kinematic study of growth and curvature. Plant Physiology **145**(2), 305–316 (2007)
5. Benmansour, F., Fua, P., Turetken, E.: Automated reconstruction of tree structures using path classifiers and mixed integer programming. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 566–573 (2012)
6. Bland, J.M., Altman, D.G.: Measuring agreement in method comparison studies. Statistical Methods in Laboratory Medicine **8**(2), 135–160 (1999)
7. Boroujeni, F.Z., Rahmat, O., Wirza, R., Mustapha, N., Affendey, L.S., Maskon, O.: Coronary artery center-line extraction using second order local features. Computational and Mathematical Methods in Medicine (2012)
8. Boroujeni, F.Z., Wirza, R., Rahmat, O., Mustapha, N., Affendey, L.S., Maskon, O.: Automatic selection of initial points for exploratory vessel tracing in fluoroscopic images. Defence Science Journal **61**, 443–451 (2011)
9. Brachi, B., Morris, G., Borevitz, J.: Genome-wide association studies in plants: the missing heritability is in the field. Genome Biology **12**(10), 232–240 (2011)
10. Bresenham, J.E.: Algorithm for computer control of a digital plotter. IBM Systems Journal **4**(1), 25–30 (1965)
11. Cobb, J.N., DeClerck, G., Greenberg, A., Clark, R., McCouch, S.: Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotypephenotype relationships and its relevance to crop improvement. Theoretical and Applied Genetics **126**(4), 867–887 (2013)
12. Delibasis, K.K., Kechriniotis, A.I., Tsonos, C., Assimakis, N.: Automatic model-based tracing algorithm for vessel segmentation and diameter estimation. Computer Methods and Programs in Biomedicine **100**(2), 108–122 (2010)
13. Fraz, M.M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A.R., Owen, C.G., Barman, S.A.: Blood vessel segmentation methodologies in retinal images - a survey. Computer Methods and Programs in Biomedicine **108**(1), 407–433 (2012)
14. French, A.P., Ubeda-Tomas, S., Holman, T., Bennett, M., Pridmore, T.: High-throughput quantification of root growth using a novel image-analysis tool. Plant Physiology **150**(4), 1784–1795 (2009)
15. Furbank, R.T., Tester, M.: Phenomics technologies to relieve the phenotyping bottleneck. Trends in Plant Science **16**(12), 635–644 (2011)
16. Godin, C., Costes, E., Sinoquet, H.: A method for describing plant architecture which integrates topology and geometry. Annals of Botany **84**(3), 343–357 (1999)

17. Huang, Y., Zhang, J., Huang, Y.: An automated computational framework for retinal vascular network labeling and branching order analysis. Microvascular Research **84**(2), 169–177 (2012)
18. Lin, K.S., Tsai, C.L., Tsai, C.H., Sofka, M., Chen, S.J., Lin, W.Y.: Retinal vascular tree reconstruction with anatomical realism. IEEE Transactions on Biomedical Engineering **59**(12), 3337–3347 (2012)
19. Longair, M.H., Baker, D.A., Armstrong, J.D.: Simple neurite tracer: Open source software for reconstruction, visualization and analysis of neuronal processes. Bioinformatics (2011)
20. Martínez-Pérez, MElena, Hughes, Alun D., Stanton, Alice V., Thom, Simon A., Bharath, Anil A., Parker, Kim H.: Retinal blood vessel segmentation by means of scale-space analysis and region growing. In: Taylor, Chris, Colchester, Alain (eds.) MICCAI 1999. LNCS, vol. 1679, pp. 90–97. Springer, Heidelberg (1999)
21. Martinez-Perez, M.E., Hughes, A.D., Stanton, A.V., Thom, S.A., Chapman, N., Bharath, A.A., Parker, K.H.: Retinal vascular tree morphology: a semi-automatic quantification. IEEE Trans. Biomed. Engineering **49**(8), 912–917 (2002)
22. Meijering, E.: Neuron tracing in perspective. Cytometry Part A **77A**(7), 693–704 (2010)
23. Naeem, A., French, A.P., Wells, D.M., Pridmore, T.: High-throughput feature counting and measurement of roots. Bioinformatics **27**(9), 1337–1338 (2011)
24. Nguyen, U.T.V., Bhuiyan, A., Park, L.A.F., Ramamohanarao, K.: An effective retinal blood vessel segmentation method using multi-scale line detection. Pattern Recognition **46**(3), 703–715 (2013)
25. Rousseeuw, P.J., Driessen, K.V.: A fast algorithm for the minimum covariance determinant estimator. Technometrics **41**(3), 212–223 (1999)
26. Rousseeuw, P.J., Leroy, A.M.: Robust regression and outlier detection. John Wiley & Sons Inc., New York (1987)
27. Sozzani, R., Benfey, P.: High-throughput phenotyping of multicellular organisms: finding the link between genotype and phenotype. Genome Biology **12**(3), 219–225 (2011)
28. Subramanian, R., Spalding, E.P., Ferrier, N.J.: A high throughput robot system for machine vision based plant phenotype studies. Machine Vision and Applications **24**(3), 619–636 (2013)
29. Sun, Y.: Automated identification of vessel contours in coronary arteriograms by an adaptive tracking algorithm. IEEE Trans. Med. Imaging **8**(1), 78–88 (1989)
30. Tueretken, E., Gonzalez, G., Blum, C., Fua, P.: Automated reconstruction of dendritic and axonal trees by global optimization with geometric priors. Neuroinformatics **9**(2–3), 279–302 (2011)
31. Weigel, D.: Natural variation in arabidopsis: From molecular genetics to ecological genomics. Plant Physiology **158**(1), 2–22 (2012)
32. WinRHIZO: Winrhizo pro 2004a software: Root analysis. Regent Instruments Inc., Quebec, Canada (2004)
33. Yin, Y., Adel, M., Bourennane, S.: Retinal vessel segmentation using a probabilistic tracking method. Pattern Recognition **45**(4), 1235–1244 (2012)
34. Zana, F., Klein, J.C.: Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation. Trans. Img. Proc. **10**(7), 1010–1019 (2001)
35. Zhang, Y., Zhou, X., Degterev, A., Lipinski, M., Adjeroh, D., Yuan, J., Wong, S.: A novel tracing algorithm for high throughput imaging screening of neuron-based assays. Journal of Neuroscience Methods **160**(1), 149–162 (2007)