Topological Image Analysis and (Normalised) Representations for Plant Phenotyping

Ines Janusch, Walter G. Kropatsch
Pattern Recognition and Image Processing Group, Institute of Computer Graphics and Algorithms Vienna University of Technology Vienna, Austria
Email: ines@prip.tuwien.ac.at, krw@prip.tuwien.ac.at

Wolfgang Busch
Gregor Mendel Institute of Molecular Plant Biology Austrian Academy of Sciences Vienna, Austria
Email: wolfgang.busch@gmi.oeaw.ac.at

Abstract—This paper discusses the use of topological image analysis to derive characteristics needed in plant phenotyping. Due to certain features of root systems (deformation over time, overlaps of branches in a 2D image of the root system) a topological analysis is needed to correctly derive these characteristics. The advantages of such a topological analysis are highlighted in this paper and root phenotyping is presented as a new application for computational topology. Characteristics used in plant phenotyping that can be derived from root images using methods of topological image analysis are further presented. A Reeb graph based representation of root images is shown as an example for such a topological analysis. Based on a graph representation a new, normalised representation of root images is introduced.

I. INTRODUCTION

An organism’s phenotype is defined as the set of its observable characteristics. In comparison, the genotype of an organism is defined its genetic configuration. The phenotype of an organism is based on the interaction of the genotype with the environment. The genome (the entire DNA sequence) of an increasing number of species is completely characterised [11]. However, their phenotype (the full set of phenotypes of an individual species) can never be completely characterised as phenotypes can vary strongly [7]. How genotypes translate into phenotypes is a major question in Biology. Efficient extraction of characteristics for a large number of individuals is a key aspect in this research. For phenotyping of plants their root structure can be analysed. The characteristics derived describe the branching pattern; length and distance measurements are obtained as well. Zheng et al. present in [14] a method to reconstruct the 3D volume of a root system based on several 2D images. For analysis of a single 2D image a medial axis based approach is for example presented by Leitner et al. in [10].

Branched structures such as root systems, pictured as 2D images, can be well represented using topological graphs (based on a medial axis or for example Reeb graphs). The characteristics used in root phenotyping are captured and represented by these topological graphs. Within the scope of this paper a representation of the root system based on Reeb graphs (similar to the representation introduced in [8]) is presented. Branching points and endpoints of the roots are represented by the nodes in the Reeb graph, the edges of the graph represent the branches of the root system. Using a geodesic distance as Morse function the length of individual roots can be directly measured based on the Morse function value of the according nodes in the graph. To allow for other measurements of length or angles geometric attributes as for example the position of the branching points can be stored with the nodes. This representation of topological and geometric characteristics in a Reeb graph is called augmented Reeb graph according to Tung et al. [12]. The Reeb graphs presented in this paper consider the positions of the nodes in the graph and are therefore augmented Reeb graphs even though the graphs are simply called Reeb graphs throughout the paper.

Such a graph representation of a root system allows for analysis of the root system’s characteristics as well as for comparisons of root systems based on graph comparison. Furthermore, the topological graphs can serve as a basis for new representations. For a comparison of roots of the same plant on different days of the growth cycle or of different plants on the same day of growth a normalised representation is needed. Additionally, such a normalised representation provides a tool for the combination of different graphs.

The rest of the paper is structured as follows: Section II provides information on plant phenotyping, while section III presents a Reeb graph based representation of root characteristics and section IV discusses advantages of such a topological image analysis for the application of plant phenotyping. A new root representation - a normalised representation based on topological graphs - is introduced in section V. Section VI summarises the results and concludes the paper.

II. PLANT PHENOTYPING

As the phenotype of a plant is formed by the observable characteristics of a plant, it is linked to plant performance and plant productivity. These are especially important aspects in food supply. Plant phenotyping aims at analysing how the genotype of a plant and the environmental conditions interact and therefore promote a certain phenotype (see Figure 1). In plant phenotyping, to understand phenotype to genotype relations, model organism are often analysed first. The plant Arabidopsis thaliana is for example used as a model organism because it is of small size, has a fast life-cycle, it can be grown in transparent media and its genome is fully sequenced [6]. The analysis done in plant phenotyping can be based on the
These critical points form the nodes in the Reeb graphs.

For the root images the plants can be grown in a transparent media: for young plants this can be a nutrient containing agar gel surface in vertically oriented plastic petri dishes. For larger root systems plants can be grown in a volume of a transparent gellan gum system. Another option is to grow the plants in soil and take them out of the soil for measuring and imaging. Moreover, transparent substrates with qualities comparable to soil were introduced.

To study the development of root systems younger plants are imaged on successive days of their growth cycle. For each plant a series of 2D images is taken over time.

### III. REEB GRAPH REPRESENTATION

The presented method computes Reeb graph representations for segmented root images. Therefore an image segmentation is needed as a pre-processing step. The root region is defined as the foreground region. On the segmented image critical points are computed according to a Morse function. Reeb graphs describe the topological structure of a shape (e.g. 2D or 3D content) as the connectivity of its level sets. For the root images the topological structure is given as the branches of the roots that are topologically persistent considering the connectivity and branching angles.

Critical points of a Morse function represent a change in topology, e.g. a change in the connectivity. For the roots this is the case at branching points and endpoints of roots. These critical points form the nodes in the Reeb graphs.

The Reeb graph approach presented in this paper implements the geodesic distance as Morse function. The geodesic distance is defined as the shortest distance in a curved space (manifold) or a restricted area measured between two points of this area or space. It is measured from a source pixel, which for this representation is selected to be located in the very top of the root. As we assume the shoots to be in the top half of the image and the roots in the lower half, the source pixel is set to the center of the topmost pixel line in the foreground region of the segmented root images.

According to Morse theory, Reeb graphs are defined in the continuous domain as follows:

A smooth, real-valued function \( f : M \to \mathbb{R} \) is called a Morse function if it satisfies the following conditions for a manifold \( M \) with or without boundary:

- **M1**: all critical points of \( f \) are non-degenerate and lie inside \( M \),
- **M2**: all critical points of \( f \) restricted to the boundary of \( M \) are non-degenerate,
- **M3**: for all pairs of distinct critical points \( p \) and \( q \), \( f(p) \neq f(q) \) must hold [3].

However, as the presented approach computes Reeb graphs for the discrete pixels of an image, a definition for discrete Reeb graphs is needed. A similar approach to derive a Reeb graph according to a geodesic distance from a source pixel as Morse function was presented for discrete 3D data by Werghi et al. in [13]. A definition for discrete Reeb graphs is given there as well:

- Two point sets are connected if there exists a pair of points (one point of each point set) with a distance between these two points below a fixed threshold.
- If all non-empty subsets of a point set, as well as its complements, are connected, such a point set is called connective.
- A group of points that have the same Morse function value and that form a connective point set, is called a level-set curve [13].

The nodes in a discrete Reeb graph represent level-set curves, the edges connect two adjacent level-set curves, therefore the underlying point sets are connected [13].

In 2D the nodes in the resulting Reeb graphs are of type minimum, maximum and saddle. We will further distinguish the saddle nodes to be a saddle node of type split (increase in the number of connected components) and saddle node of type merge (decrease in the number of connected components) [3].

Based on these definitions critical points according to the geodesic distance capture endpoints of roots as local maxima of the geodesic distance. Saddle nodes in such Reeb graphs either represent branching points in a root (saddle node type split) or overlaps of root branches (saddle node type merge).

These saddle nodes are found as locations where parts of the foreground region with the same geodesic distance to the source pixel are split in two connected components or are merged from two into one connected component. The so found critical points then serve as nodes in the Reeb graphs and are connected by edges according to the foreground (the

![Diagram of plant phenotyping](image-url)
Fig. 2: Reeb graph representation of the root region in a segmented image.

Due to the segmentation needed as a pre-processing step, segmentation artefacts as frayed borders of the foreground region may appear in the images. These artefacts introduce additional critical points and spurious branches in the Reeb graphs. A simple graph pruning approach based on the length of branches (as for example described by Attali et al. in [1]) can be used to reduce the number of spurious branches, as spurious branches due to segmentation artefacts typically introduce short edges compared to edges representing true branches. However, this may not be generally true, false positives and false negatives are possible: spurious branches may be accepted as true branches during the graph pruning process, while true branches may be discarded, as they resemble spurious branches. For the result in Figure 2 graph pruning was already applied.

Another method to detect and discard spurious branches is based on knowledge about the dataset and the topology of the plants in the dataset. For the root analysis the imaged Arabidopsis thaliana plants were grown up to 20 days at 10°C Celcius. For plants of this age and grown at this temperature there is usually a primary root and lateral roots that appear as branches to the left and the right out of the primary root. Therefore branches (lateral roots) do not branch again at this age under these conditions. All short branches that may appear as lateral roots of lateral roots can therefore be discarded as artefacts. Evaluation results of this representation on a dataset of Arabidopsis thaliana plants on day 7 and day 10 of their growth cycle are presented in [9]. An example of a Reeb graph representation that was built using this knowledge about the root structure is shown in Figure 3. For this representation a simple graph pruning based on length was additionally used to discard spurious branches that appear as small lateral roots of the primary root. However, due to this graph pruning two small true branches were wrongly discarded during graph pruning.

Fig. 3: Reeb graph representation using knowledge about root structure to avoid spurious branches due to segmentation artefacts. Two small true branches were wrongly discarded during graph pruning.

IV. Topological Characteristics in Plant Phenotyping

When growing, roots may transform non-rigidly: they may for example bend around an obstacle. Moreover, when plants that were grown in soil are imaged, they need to be taken out of the soil. Roots may thereby be bended or rearranged. The root’s shape and therefore the root’s geometric properties are changed by these actions. However, the connectedness and the branching structure of the root are not affected by these actions - the roots topological properties are invariant to these deformations of the root. Topological properties therefore provide a stable representation of root characteristics that can be used in plant phenotyping.

Advantages of topological properties over geometric properties regarding key aspects of root phenotyping are discussed in more detail in the following subsections:

A. Deformation of Roots

Roots posses the ability to transform non-rigidly. When running into an obstacle roots may bend and grow around that obstacle. Although the usual primary direction of growth of a root is the direction of gravity, some environmental conditions or chemical perturbation of plant growth cause roots to grow agravitropically (i.e. not in the direction of the vector of gravity).
Fig. 4: Agravitropically growing *Arabidopsis thaliana* root.

gravity. Figure 4 shows an example for such a root structure. While geometric properties (as for example the positions of endpoints of roots) change with a bend of the root, its topological properties (for example its connectedness) stays the same. Topological characteristics are not altered by such non-rigid transformations.

B. Detection of Branches

Based on geometric properties branches can be detected by an analysis of the root contour. Methods for corner detection are a possible option. However, a sharp bend in a root may also be detected as a branching point by such an analysis. Using topological properties, such a bend will not be mistaken for a branching point. Branches appearing in the root structure change the topology of the root. These changes in topology can for example be detected as critical points according to a Morse function (for example the geodesic distance) or based on a medial axis skeleton. For the medial axis a branching point is found as a pixel in the skeleton with three neighbouring pixels (while regular skeleton pixels have two neighbours and endpoints of roots have only one neighbour). According to the geodesic distance as Morse function, a branching point is found as the location for which a part of the foreground region of one geodesic distance is split in two components, while for smaller distances, there is only one connected component. This critical point is represented by a saddle node of type split in a Reeb graph.

A first attempt to represent the branched structure of roots using Reeb graphs based on the height function is given in [8]. The results given in [8] show that all major branches (primary root and long lateral roots) were correctly detected. However, for 23.5% of the images in the dataset presented in this paper smaller branches were wrongly discarded as they resembled spurious branches due to segmentation artefacts.

For the root shown in Figure 3 two of the lateral roots were discarded as spurious branches in the graph representation as well. Six lateral roots are visible in the graph representation, while the root image shows in total eight lateral roots.

C. Length Measurements

The length of individual roots is used as a characteristic in plant phenotyping. Additional length of roots due to bends in the roots should be taken into account for length measurements. The distance between the start point of a root (for example a branching point for a lateral root) and the endpoint of the root measured as Euclidean distance may not provide an accurate measurement for deformations (for example deflections as shown in Figure 4) in the root. The geodesic distance however measures the distance as the shortest path from a source pixel to any other pixel of the root inside the root region. Additional length due to curvature of the root is taken into account by this measurement.

As the geodesic distance is used as Morse function in the approach presented in this paper, the length according to the geodesic distance is implicitly given by this Morse function. The length of lateral roots can be easily computed as the difference in the geodesic distances of the branching point and the endpoint of the root.

An approach using Reeb graphs based on the geodesic distance as a Morse function was presented in [9]. Moreover, root lengths measured using the Reeb graph approach are compared to manually obtained ground truth measurements. The results presented in [9] show in general longer roots for length measurements based on the Reeb graph approach compared to the ground truth measurements. The ground truth lengths of individual roots are well approximated by the Reeb graph measurements, the mean deviation from the ground truth is 0.8% at maximum for the dataset evaluated.

D. Overlaps of Branches

A common problem that may arise for root images are overlaps of branches in the images. The roots grow as 3D shape
but are imaged as a projection to a 2D image. Branches of the roots may overlap in the root image due to this projection. Using geometric properties an overlap of branches is detected the same way a branching point is. But these two, overlaps and branching points, cannot be distinguished without further processing: the direction of growth may for example be taken into account. If several images over time exist for one plant, these images can be used to distinguish an overlap from a branching point to represent the root structure correctly. Using a topological analysis, overlaps and branching points share properties as well: in a topological graph both are represented by saddle nodes. Using a skeleton representation of a root an overlap is detected just the same way a branching point is. As it is the case for geometric properties, a branching point and an overlap cannot be immediately distinguished based on the derived skeleton, without further processing of the skeleton.

The major advantage that a Reeb graph representation provides for root images is that branching points and overlaps can be distinguished immediately. When computing critical points according to the geodesic distance as a Morse function, overlaps are detected as locations for which a part of the foreground region of one geodesic distance forms one connected component, while for smaller distances there are two connected components. This critical point is represented by a saddle node in the Reeb graph, as is a branching point. However, in a Reeb graph saddle nodes can be further distinguished to be split or merge saddle nodes. While split saddle nodes represent branching points, merge saddle nodes represent overlaps in the root structure. These overlaps can therefore be distinguished from branching points simply by the node type. Figure 5 shows an example of a root image with overlapping branches and the according Reeb graph representation. The saddle node of type merge, that indicates and represents the overlap, is highlighted in red.

E. Persistence in Time Series of Images

As the plants are imaged on successive days of the growth cycle a stack of 2D images over time is available for the plants. Figure 6 shows an example - the same plant was imaged every fourth day, the images show day 12, day 16 and day 20 of the growth cycle. Figure 7 shows the according Reeb graphs built for the segmented images of the plant images shown in Figure 6.

Based on such a time series of images the persistence of topological properties can be exploited. When growing, plants form new branches but do not (naturally) lose branches, therefore the knowledge about the topology of the plant on an older day can be used to verify the topology of a plant on an earlier day. In this way noise and segmentation artefacts can be detected.

A small branch on an earlier day can for example be verified as a true branch or discarded as spurious branch based on the fact whether this branch exists in an image of an older day as well.

As mentioned before, overlaps can be easily detected in a Reeb graph based on the saddle node of type merge, that is introduced for such an overlap. However, to correctly represent the root structure an overlap not only needs to be detected but resolved as well. Based on an image at a point in time before the overlap appeared, the general direction of growth of the
two branches is known. Assuming the direction of growth to be consistent for the small time frame of two images (one before the overlap, one showing an overlap), the overlap can be resolved. The merge node is doubled and is in each case connected as a node of degree 2 to the respectively succeeding node in the two branches. Nodes of degree 2 are regular nodes and do not represent any topological changes, as critical points do. The saddle node of type merge is changed to a regular node when resolving an overlap. It can therefore be smoothed out of the graph after clearing the overlap.

V. NORMALISED REPRESENTATION

In this section a normalised representation for root structures is presented. This normalised representation is based on a graph representation (as for example a Reeb graph) of a root image. For the normalised root representation the roots are organised as branches (lateral roots) of a main root (primary root). The main root is determined as the longest branch in the root structure. The main root and all side branches are drawn as straight lines of their actual length. Branches occur perpendicular to the left or the right of the main root. This normalised representation therefore captures the overall branching pattern which is the number of branches, the side to which a branch forms (to the left or the right of a main root), distances at which branches appear are included, as are the lengths of the branches. This representation does not represent branching angles, although the angles can be computed for the original graph and stored as node attributes in the normalised representation. Moreover, curvature and bends of roots are not represented.

Figure 8 shows an examples of the normalised representation for the three roots in Figure 6. For day 12 there are no side branches therefore the plant is only represented by one straight line representing the main root. For day 16 side branches appear and they are longer and better visible for day 20. As can be seen in the original images in Figure 6 the plants of day 16 and day 20 are very similar as are their normalised representations. The growth of the side branches is immediately visible. The difference in the distance of the branching points on the main root is due to the discretisation during the segmentation.

Figure 9 shows a possibility to efficiently compare the root structure of a plant on different days of the growth cycle. This overlay of the normalised representations especially highlights the development of the root through the growth period. The individual representations shown in Figure 8 are drawn as an overlay. The representations of the three days are therefore matched at their starting points (at the origin). Day 12 is represented by red nodes, day 16 by blue nodes and day 20 by green nodes; black nodes belong to the representations of more than one day. There is an artefact clearly visible in this representation: the first lateral root to the right is closer to the starting point for day 20 than for day 16. As negative growth does not exist for plants, this does not correctly represent the root development but represents an artefact, due to the segmentation and the selection of the start point of the root. Apart from the starting points of the roots, the normalised representations can just as well be aligned at the branching points, in case all representations share a common branching point.
As the roots may not be imaged every day but only every representations. The branching points may in this case provide a more stable landmark to align the different
images of several days. The branching points in this case may hence provide a more stable landmark to align the different representations.

As the roots may not be imaged every day but only every nth day, the presented normalised representation finds further application in a simulation of the root structure on days that were not originally imaged. An interpolation between the representation of two originally imaged days allows to build a root representation of a day in between. This can help to more reliably compare images of different datasets for which the plants were not imaged on the same days of their growth cycle. Furthermore, the growth of a root (e.g. its length) can therefore be associated with a point in time. Root growth can be mapped to time and vice versa. Using this interpolation between several images, respectively their graph representations, stages of development of the root can be projected to the oldest recording of the root and the development of the root can be visualised on this image.

The normalised representation presented in this paper is well suited for small branched structures as it is the case for root images in plant phenotyping up to day 20 in the growth cycle. For larger structures, with a higher number of branches, the representation may for repeated branching cause overlaps of the side branches due to the orthogonal branching representation. However, keeping the true branching angles, instead of the artificial orthogonal branching angles used for this representation, may not eliminate this problem completely: branches, that overlap in the images, still overlap in the representation and a solution of overlapping branches through this new arrangement of the branches is no longer given. A trade-off between clear representation of the branching structure, correct presentation of distances and length and prevention of overlaps needs to be found depending on the represented dataset and the size of the branched structures within.

VI. CONCLUSION

The representations presented in this paper employ topological image analysis to obtain properties used within the application of root phenotyping. Root structures are represented by Reeb graphs, the nodes in the graphs represent branching points and endpoints of roots, while edges represent the actual branches. Characteristics needed in plant phenotyping as the branching structure, number of branches, lengths of branches and distances between them are captured by this Reeb graph representation. Compared to geometric properties the advantages of topological properties lie in their invariance to deformations of the root (as for example bends when growing around an obstacle) and the ability to overcome the common problem of overlapping branches in root images. As reasoned in Section IV geometric properties are not suitable for capturing and representing characteristics used in plant phenotyping. Due to the ability of roots to transform non-rigidly, plant phenotyping poses a suitable application for methods of topological image analysis.

Additionally, a new normalised representation of root structures is presented in this paper. This representation focuses on the branching pattern of the roots and captures lengths and distances, while branching angles are neglected. Thus, the normalised representation allows for a simplified comparison of a plant’s development on different days of growth or of various plants on the same day of growth. Especially the option to draw several different representations aligned by one node as an overlay highlights the differences in the plant’s development.

For future work the temporal correlation of the root images needs to be taken into account. The roots are imaged on several days of their growth cycle. A stack of 2D images over time is provided for each plant. The persistence of the topological properties over time provides additional knowledge to detect noise and artefacts. To distinguish a small true branch from a spurious branch due to artefacts may for example not be possible based on one image. A later image of the same root may help with this decision. The information and knowledge about a root gained on the images of later days in the growth cycle can be used for decisions on earlier images.

REFERENCES


