On the Evaluation of Methods for the Recovery of Plant Root Systems from X-ray Computed Tomography Images

Stefan Mairhofer[1], Craig Sturrock, Darren M. Wells, Malcolm J. Bennett, Sacha J. Mooney and Tony P. Pridmore

Centre for Plant Integrative Biology (S.M., C.S., D.M.W., M.J.B., S.J.M., T.P.) and School of Biosciences (C.S., D.M.W., M.J.B., S.J.M.), University of Nottingham, Nottingham LE12 5RD, United Kingdom; School of Computer Science(S.M., T.P.P.), University of Nottingham, Nottingham NG8 1BB, United Kingdom

1. Stefan.Mairhofer@nottingham.ac.uk

Abstract

X-ray micro computed tomography (µCT) allows non-destructive visualisation of plant root systems within their soil environment and thus offers an alternative to commonly used destructive methodologies for the examination of plant roots and their interaction with the surrounding soil. Various methods for the recovery of root system information from X-ray CT image data have been presented in the literature. Detailed, ideally quantitative, evaluation is essential, in order to determine the accuracy and limitations of the proposed methods, and to allow potential users to make informed choices between them. This, however, is a complicated task. Three-dimensional ground truth data is expensive to produce, and the complexity of X-ray CT data means that manually generated ground truth may not be definitive. Similarly, artificially generated data is not entirely representative of real samples. The aims of this work are to raise awareness of the evaluation problem and to propose experimental approaches that allow the performance of root extraction methods to be assessed, ultimately improving the techniques available. To illustrate the issues, tests are conducted using both artificially generated images and real data samples.

Keywords: Segmentation, Root Architecture, Root Image Analysis, Evaluation

Introduction

Plant roots can develop into a highly complex and diverse system, whose architecture is fundamental in providing stability to the plant while exploring the soil for nutrients and water. Because root systems are covered by soil, it is difficult to examine them non-invasively and, hence, they are often referred as the 'hidden half' of plants. Recent studies have shown that X-ray micro computed tomography (μCT) allows non-destructive observation of plant roots in soil without causing any harm to the plant or its surrounding environment (Tracy et al. 2010; Zappala et al. 2013). The recovery of root system architecture descriptions from X-ray CT data is a segmentation problem: the input density data must first be divided into voxels representing root, and those representing non-root, material. This process is often performed either manually or by using (semi-) automated extraction methods. Computational methods, exploiting results in image analysis and computer vision, have been presented in previous studies (Heeraman et al. 1997; Pierret et al. 1999; Lontoc-Roy et al. 2006; Kaestner et al. 2006; Perret et al. 2007; Mairhofer et al. 2012) and have demonstrated the possibility of recovering root material from the surrounding soil environment in X-ray μCT images. Systematic evaluation of these methods, however, proves to be very challenging.

Segmentation is a longstanding problem in computer vision, and many segmentation algorithms have both been proposed and evaluated. Evaluation of segmentation results takes one of two forms. In some cases results are assessed on their success in supporting some higher-level task. A segmentation algorithm intended for use in face recognition, for example, might be evaluated by measuring the proportion of faces that are correctly identified (Phillips et al. 2000). In most cases, however, segmentation is assessed by comparing the reported segments to the ideal solution - ground truth. Ground truth data is typically produced manually, and a number of standard data sets with associated ground truth have been created and made public for many different applications (Martin et al. 2001). In some fields, however, such data does not exist or is impossible to obtain (Bouixet al. 2007, Yan et al. 2010, Kohlbergeret al. 2012). The generation of ground truth is often complicated by the large number and ambiguity of voxels in the data, making manual selection very subjective, and so unreliable.

To assess their ability to recover roots from X-ray CT image data, Heeraman et al. (1997) destructively removed the plant sample from the sand after it had been scanned, and measured root length per unit volume. The same parameter was estimated from the CT image data and the two values compared. Results showed an overestimation of root length (76 cm/cm^3) compared to destructive sampling (44-60 cm/cm^3). An evaluation based on root length was also carried out by Perret et al. (2007) who, in contrast to Heeraman et al. (1997), reported an underestimation of average root length of about 10% compared to the destructively extracted sample. In addition to root length, volume, surface area and number of roots were determined. Root systems were also removed from the growth media by Lontoc-Roy et al. (2006), who took two-dimensional digital photographs which were skeletonised with WinRhizo (Arsenault et al. 1995) and their complexity determined using fractal dimension. The complexity value was then compared to the skeleton of the CT recovered root system and the Pearson's linear correlation coefficient calculated. Mairhofer et al. (2012) attempted to evaluate their approach by comparing it to previously presented methods. However, due to lack of availability of either an executable software tool or complete implementation details the authors had to generalise the method used for comparison, which led to a region growing approach with connectivity constraints as a reference method. Kaestner et al. (2006) evaluated their method on synthesised data composed of three-dimensional dichotomous branches. Gaussian noise was applied to obtain image data at various degradation levels to assess the performance of their method under different conditions. To evaluate the extraction of the root system, a visual comparison to a root-washed image was presented. Similar comparisons were also made in (Lontoc-Roy et al. 2006) and (Mairhofer et al. 2012). The above mentioned studies show the inconsistency found in the evaluation of root recovery methods from X-ray µCT image data, which is largely due to the lack of ground truth data and the difficulties in accurately measuring deviations from the original sample. It is important, however, to work towards a 'standard' evaluation that can be used to compare various methods, so that users can understand the limitations and strengths of each method as well as having the means to choose among available approaches.

In this work we suggest several evaluation procedures that aid in determining how well a method can recover the sought information and in identifying its limitations. In what follows we introduce the root segmentation method used throughout the paper, details on sample preparation and data acquisition, and various evaluation strategies, in which the performance of the root recovery method is assessed and the obtained results statistically analysed. The paper ends with discussion and conclusion.

Methods and Materials

Root Segmentation Method

The root recovery method considered here is RooTrak (Mairhofer et al. 2012, 2013), which to the authors' best knowledge, is the only method made publically available at the time of writing. The method uses a tracking based strategy to follow root cross-sections through a sequence of images, allowing objects to split as root branching occurs. This is accomplished by building on a level set framework (Sethian 1999) that is guided by greyscale intensity distributions to identify root object boundaries. While tracking root objects through the image stack, they are gradually assembled to reconstruct the complete root system architecture. For more details on the root segmentation method, the interested reader is referred to (Mairhofer et al. 2012, 2013).

Experimental and treatment conditions

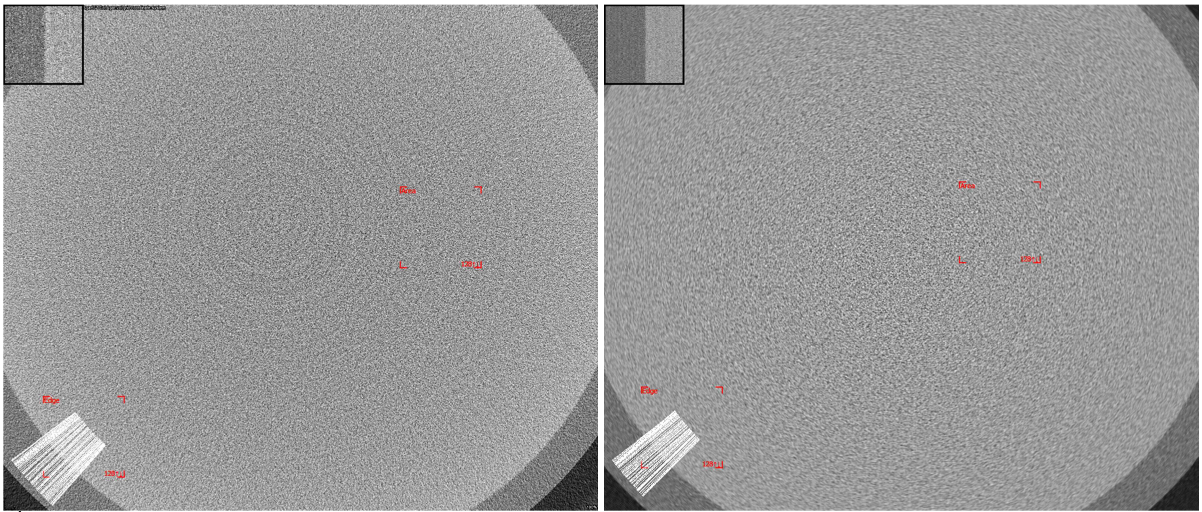
In this work, samples of maize Jubilee F1 (*Zea mays L. convar. saccharata var. Rugosa*), winter wheat Cordiale (*Triticumaestivum L.*) and tomato (*Solanumlycopersicum L.*) were used. All plants were grown in a Newport series loamy sand (brown soil) and a Worcester series clay loam soil (argillicpelosol) from the University of Nottingham farm at Bunny, Nottinghamshire, UK (52.52°N, 1.07°W), which were air-dried and sieved to <2mm. The plants grew for 10 days after germination in environmental controlled growth rooms with a 16/8 hours light cycle at a temperature of 23/18 degree Celsius. The seeds were germinated in Petri dishes on wet filter papers, covered with an aluminium foil to shield them from sunlight, and planted after two days in plastic columns of 30mm diameter. The water status of the samples at the point of imaging was approximately at field capacity. The imaging device used in this experiment was a Nanotom (Phoenix X-ray / GE Measurement & Control Systems) X-ray µCT scanner. The samples of maize were scanned at 120KeV/120µA, taking 1,200 projections at an exposure time of 750ms with 4(1 skip) signal averaging and placed 160mm away from the X-ray source, acquiring a volume data with a voxel size of 48.48µm. The samples of wheat and tomato were scanned at 120KeV/250µA, taking 1,200 projections at an exposure time of 750ms with 3(1 skip) signal averaging and placed 200mm away from the X-ray source, acquiring a volume data with a voxel size of 25.00µm. The scan time was 75 minutes for the maize and 67 minutes for the wheat and tomato plants.

Evaluation Strategies

When evaluating an automatic segmentation method, the most obvious approach is to test it on data that it was designed for. However, this is also the data for which the least information is available, which makes the evaluation very difficult. This will be discussed later, but to begin with we start with data for which everything is known. Such data can be artificially generated and is described in more detail in the next subsection. From artificially generated data we will then move gradually to more realistic data, proposing various strategies for the evaluation of methods for root system recovery from X-ray µCT image data.

Ground truth by artificially generated test data

When generating artificial data, we first need to decide on an object that will serve as our test sample. For the present we propose to keep the object simple. A cone is often considered in root measurement (Iyer-Pascuzzi et al. 2010) and modelling (Lynch et al. 1997) tools as the most simplistic shape from which more complex root structures are formed, and therefore forms a basic element. It is also an object that can be easily generated. If placed vertically, in the centre of the image stack, and viewed from a cross-sectional perspective, it appears as a circle that gradually shrinks in radius and therefore allows methods to be evaluated on their ability to extract fine and decreasing structures.



a) b)

Figure 1: (a) Cross-sectional X-ray µCT image of a column filled with agar solution and (b) an artificially generated image. Areas selected for the calculation of the NPS and MTF are highlighted by red squares. The extracted and aligned signals used in the calculation of the MFT are shown in the top-left corner of each image. Note that the original image has slight ring artefacts. At present we do not simulate any CT scanning artefacts in the artificially generated data.

It is important that artificially generated data displays characteristics similar to those found in real CT. This was achieved by taking projections of the test object using the Radon transform (Radon 1917; Kak and Slaney 1988), which results in shadow images of a rotating object as obtained with CT. These images were degraded by adding noise and blur. The generation of X-ray photons, their interaction with matter and the detection of their intensity, are described by a Poisson process (Herman 2009) and therefore the projections were superimposed with Poisson noise. A Gaussian kernel was applied to blur the images in order to simulate the inaccuracy of the detector panel when measuring the signal. Using the filtered back-projection reconstruction (Kak and Slaney 1988; Toft 1996), the projections were then converted back into a single volume. In order to compare the noise characteristics found in real CT with those in the artificially generated data, we used as a reference a cross-section from a column that was filled with an agar solution and imaged with an X-ray µCT scanner (figure 1). The modulation transfer function (MTF) and noise power spectrum (NPS) of the scanned specimen and a comparable artificially generated image were estimated and compared (figure 2).

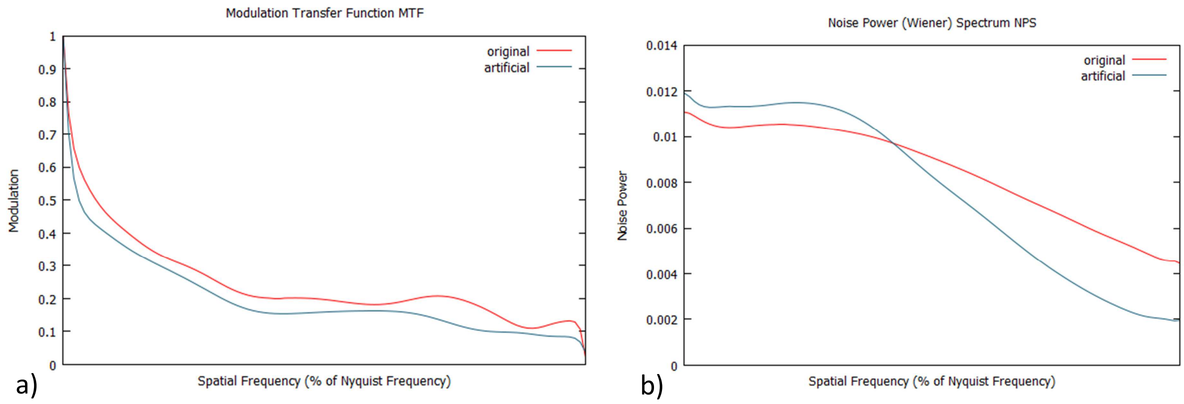


Figure 2: (a) MTF and (b) NPS from the original and the artificially generated image as shown in figure 1.

For experimental purposes, we artificially generate a set of image stacks of the cone with differing degradation levels, the aim being to assess the efficiency of the extraction as image quality falls. One could generate variations in image quality that reflect the data obtained from using different imaging parameters (exposure time, X-ray power, etc), but due to the wide range of CT scanners available and the different conditions under which samples can be prepared, it would be difficult to generalise. Instead, we alter contrast (2, 4, 8, 16, 32, 64) and noise level (0, 2, 4, 6, 8, 16, 32, 64) to cover a wide range of possible scenarios, resulting in a total of 42 different image stacks. Contrast values represent the difference in greyscale intensities between back- and fore-ground. The noise is modelled by a Poisson distribution with a parameter equal to the signal intensity in the projections and multiplied by a factor of noise level intensity. The Gaussian blur in all generated samples is kept constant at σ=1 for a radius of 3σ. The radius of the here generated cone is 23 pixels in the first image slice, reducing to 1 pixel in the final image. The image stack is 512x512x512 voxels in size.

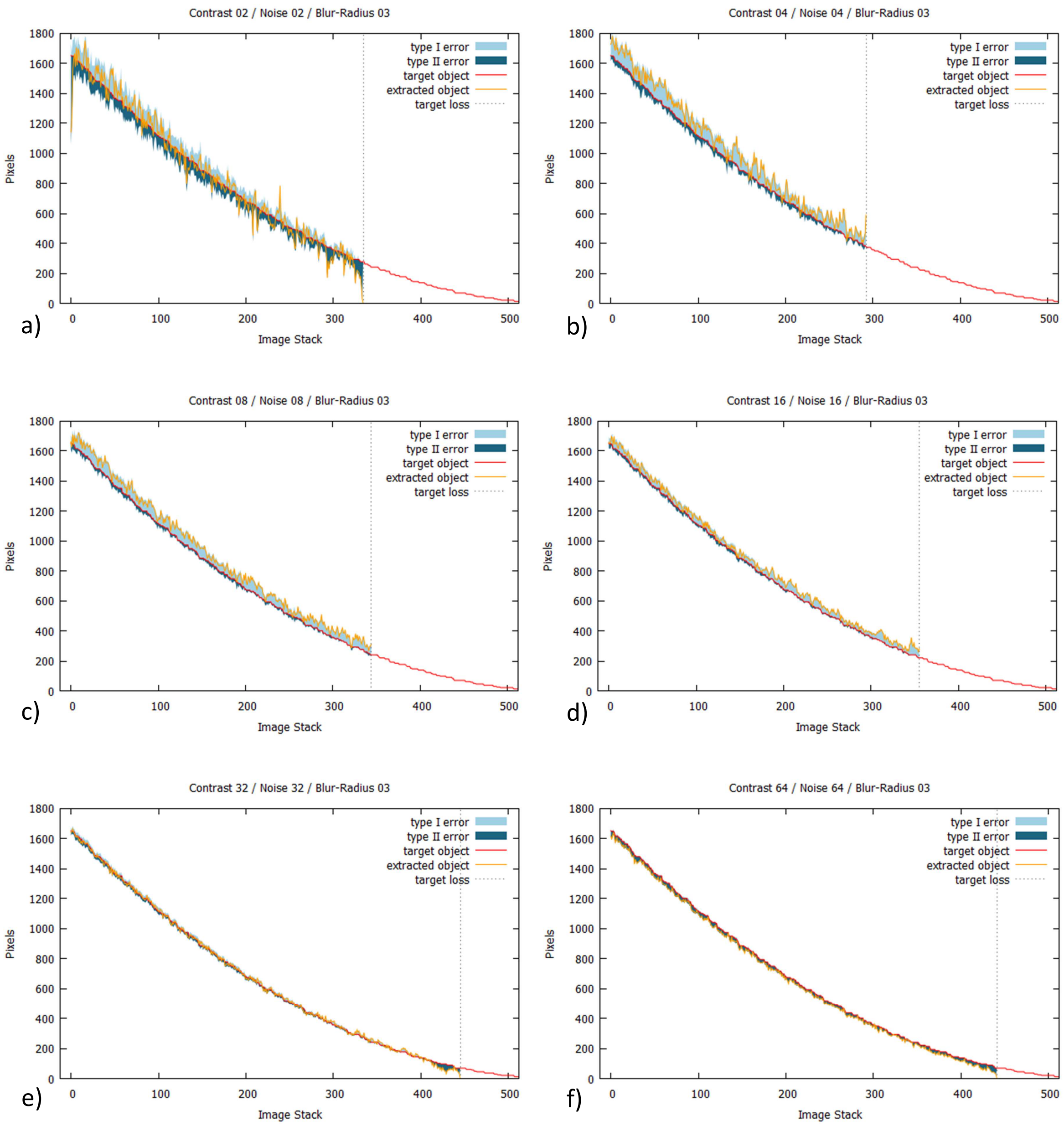


Figure 3: Subset of extraction accuracy plots with type I-II errors shown for contrast and noise levels of (a) 2/2, (b) 4/4, (c) 8/8, (d) 16/16, (e) 32/32, (f) 64/64

The method being evaluated was initialised on the first image, in the centre of the object. Being a tracking-based approach, in the course of the extraction the target object (root section) may be lost, either because the level set function shrinks until there is no inside area left or because it grows past the object boundary, including both the target and the background. To deal with the latter case we have to define a threshold value which, if exceeded, will be considered to signal loss of the target. In this experiment we set the maximum acceptable area to 125% of the actual target object.

Knowing the exact geometry of the artificially generated data, allows quantification of false positives (type I error - the number of voxels extracted as root material but belonging to the background) and false negatives (type II error - the number of voxels belonging to roots not extracted) and so determination of the accuracy of the method. Figure 3 shows the plot of type I and type II errors obtained for selected image stacks with different contrast and noise levels.

While in this experiment a cone shaped object was used as a test sample, we would ideally want to move from simple shapes to more complicated structures that resemble the shape of complete root systems. This could be achieved with root simulation tools, such as SimRoot (Lynch et al. 1997), but at the time of writing this data was not available to the authors. A soil simulation tool could add the complexity to the background and as such making the artificial test data complete and, if made publically available, could serve as reference data to the community for future development of root recovery methods.

<u>Root segments</u>

Artificial data is valuable, but currently limited, and does not fully represent the complexity of root objects found in soil. To keep a certain degree of control over the test data while moving to more realistic data sets, we use root segments of real plants. In this experiment we consider the roots of a winter wheat Cordiale (*Triticumaestivum L.*) that were cut into segments of approximately 10mm length. Their length was measured with a digital high-precision calliper that has an accuracy of approximately 10µm. The root segments were viewed under a dissecting microscope to determine their diameter, measured by line counting on a haemocytometer slide, from which the surface area and volume were estimated. For this calculation it was assumed that the roots were perfectly cylindrical. This might not always be the case but, as previously mentioned, root measurement (Iyer-Pascuzzi et al. 2010) and modelling (Lynch et al. 1997) tools, however, often make the assumption that root systems are composed of multiple conical frustums, and hence that their cross-section is always circular with a given radius. For a short segment, that radius can be assumed to be constant, resulting in a cylinder. After measurement, the root segments were buried in soil. Samples were watered from beneath with tap water, before being scanned using X-ray µCT. After the scan, all root segments were recovered from the soil and examined for possible deformation or breakage. Alongside the root segments, a plastic wire was also buried in soil, providing a non-root reference object. To perform this experiment we used 12 randomly selected root segments. To prevent the root segments from drying out, they were kept covered with a thin layer of water between being measured under the microscope and buried in the soil.

While having some information about the root object, and so the test data, the exact performance of the root recovery method cannot be determined. It is impossible to manually assign each voxel correctly as root or non-root material, since this is not always apparent from the image data and users will therefore select them by intuition, making the extracted reference object highly subjective and therefore unsuitable for a precise evaluation. However, parameters such as length, surface area and root volume can be measured relatively accurately beforehand and therefore are known at the time of evaluation. Data recovered using root extraction methods can therefore be assessed by comparing these parameters. This allows the determination, to a certain degree, of the method's ability to quantify root traits.

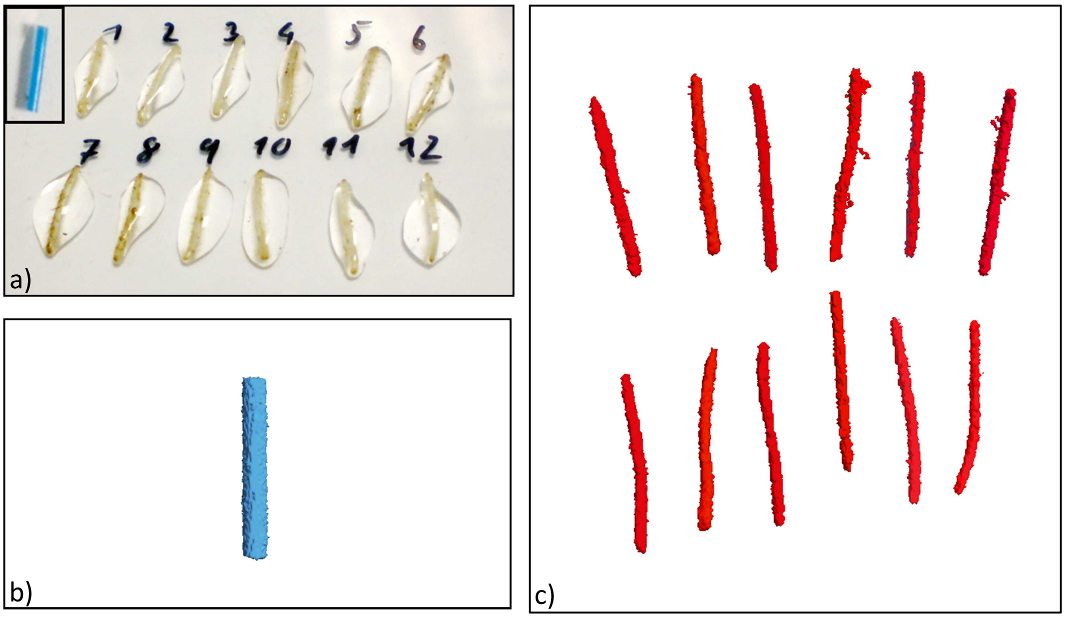


Figure 4: (a) Root segments and plastic wire, (b) plastic wire from X-ray CT data and (c) extracted segments and

Plant root systems washed free from soil

Although imaging and comparing data obtained from root segments comes close to the characteristics we encounter in real samples, the structural complexity of these objects is by no means comparable to complete root system architectures. Root recovery methods need therefore to be applied to CT image data of real and complete root system samples. While this allows determination of whether or not the method is able to extract the root system it becomes almost impossible to evaluate the obtained result. As mentioned previously, comparison to manually extracted root systems is not reliable, as they cannot be considered a suitable ground truth for evaluation. For that, manually extracted data is too inaccurate and subjective. As an alternative it is possible to destructively remove root systems by washing them free from soil. Using a flatbed scanner, a two-dimensional image of the root system can be acquired. In doing so, the three-dimensional structure of the root system is lost, but soil removal allows image analysis tools developed for two-dimensional root system images to be used.

To perform this experiment, various plant species were grown in different soil textural types and prepared for examination. After X-ray CT analysis the plants were root-washed free from soil and analysed with WinRhizo (version 2002c). WinRhizo is a popular and widely used tool in plant root studies and a de facto standard (Bouma et al. 2000; Himmelbauer et al. 2004), which is why it was used here. The roots were placed on a water tray and scanned with a flatbed scanner at 400dpi. The images were analysed using WinRhizo's automatic thresholding for normal roots, ignoring speckles that had an area less than $2mm^2$. The image data used in this experiment is the data acquired from the root systems of maize Jubilee F1 (*Zea mays L. convar. saccharata var. Rugosa*), winter wheat Cordiale (*Triticumaestivum L.*) and tomato (*Solanumlycopersicum L.*). For each plant species, eight samples were prepared of which half were grown in loamy sand and the other half in clay loam. From the total of 24 samples, 12 were scanned and used to evaluate the extraction method, two of each plant species in both soil textural types. Figure 6 shows some of the two-dimensional scanned images in comparison to the recovered and rendered three-dimensional root systems.

WinRhizo allows measurement and estimation of a range of two- and three-dimensional root parameters from two-dimensional images. These can be used and compared to assess the root recovery from CT images. As for the root segments, this method only allows determination of confidence in the quantification of these parameters and does not allow determination of the exact accuracy of the segmentation, since they are not representing ground truth but instead are measured secondary variables.

Comparison to other extraction techniques: X-ray scans of complete root architectures

Comparing a segmentation method to other methods developed to accomplish the same task is a commonly applied approach for evaluation in computer vision (Heimann et al. 2009). This is easier in fields where several methods have already established themselves as efficient solutions, and assessment can determine how the method under evaluation relates to a reference method (or methods). When considering the extraction of root systems from X-ray CT data, however, this proves to be very difficult, as the discipline is still in early development and not many methods have been presented in the literature. Alternative extraction techniques have been proposed, but none of them, to the authors' best knowledge, have been made publically available. This leaves re-implementing those methods as the only option. Unfortunately, many of the papers describing these techniques were written for a biological, rather than computer science, audience and emphasise the use, rather than technical detail behind, the proposed methods. As a result, insufficient information is provided to allow them to be accurately and confidently re-implemented. Assumptions would have to be made that may lead to an implementation that differs from the authors' and would do them an injustice. This approach is not, therefore attempted here. In the spirit of defining evaluation strategies, however, it deserves to be mentioned and should in the future become a norm in the evaluation process.

Results

In this section we will pick up on the proposed evaluation strategies and present the results obtained when RooTrak is applied to each of the presented test data samples. The data is statistically analysed in order to provide information on how well the method performs in each experiment.

Statistical Analysis

Artificially generated test data

Artificially generated data provides ground truth and therefore determination of the number of type I (false positive) and type II (false negative) errors, which are shown in figure 3, where a subset of plots are chosen for the samples with corresponding contrast (2-64) and noise level (2-64). Ground truth data also allows calculation of Fisher's scores for segmentation accuracy, which are given in table 1.

Noise/Contrast	c = 2	c = 4	c = 8	c = 16	c = 32	c = 64
n = 0	0.9213	0.9228	0.9316	0.9531	0.9702	0.9885
n = 2	0.8391	0.9324	0.9346	0.9492	0.9483	0.9618
n = 4	0.0198	0.7421	0.9291	0.9443	0.9412	0.9620
n = 8	0.0002	0.0092	0.8793	0.9143	0.9399	0.9521
n = 16	0	0.0053	0.0058	0.8989	0.9337	0.9554
n = 32	0	0.0038	0.0046	0.0064	0.9227	0.9493
n = 64	0	0	0.0027	0.0049	0.0117	0.9248

Table 1: Fisher's score from artificially generated test data for varying contrast and noise level - (n) noise, (c) contrast

Root segments

The root segments and plastic wire were extracted from the X-ray µCT images after the scan using the method under evaluation. The rendered data is shown in figure 4. To this data we applied the measurement tool integrated into RooTrak, to determine surface area and volume, and compared them to measured and calculated parameters of the root segments obtained under the microscope. The mean error in surface area and volume are 12.9% and 10.4% respectively. Figure 5 shows the measured results in comparison. The Pearson's product moment correlation coefficient between the measured and extracted data are r_{area}=0.8683 and r_{volume}=0.8909 for the measurements of surface area and volume respectively, with a p-value of 0.0002 and 0.0001 based on Fisher's Z transform. The paired Wilcoxon signed rank test gives a p-value of 0.0009 and 0.0068 for the two datasets.

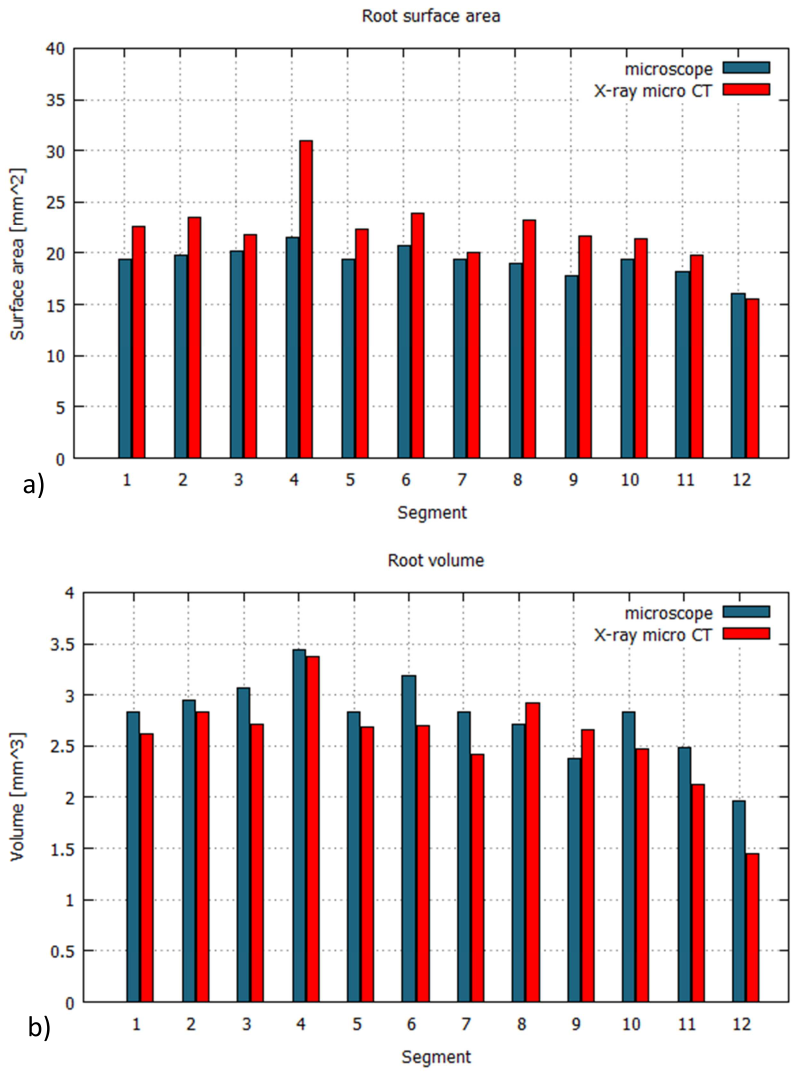


Figure 5: Measured root segments viewed under microscope compared to the objects extracted from the image data: (a) Surface area and (b) Volume

Plant root systems washed free from soil

The root systems extracted from the X-ray µCT image data were compared to the root-washed images quantified using the WinRhizo analysis tool. Both set of images for plants of each species are shown in figure 6, with the thresholded and skeletonised architecture of the root system highlighted in colour. Figure 7 shows the plotted surface area and volume for each sample. The average error in surface area and volume are 19.6% and 39.1% for maize, 54.1% and 56.5% for wheat and 9.0% and 14.0% for tomato. The Pearson's product moment correlation coefficient between WinRhizo and the data extracted from X-ray µCT images are r_{area}=0.5670 and r_{volume}=0.5924 with a p-value of 0.433 and 0.4076 for maize, r_{area}=0.9831 and r_{volume}=0.9802 with a p-value of 0.0168 and 0.0197 for wheat, and r_{area}=0.9847 and r_{volume}=0.9209 with a p-value of 0.0152 and 0.079 for tomato.

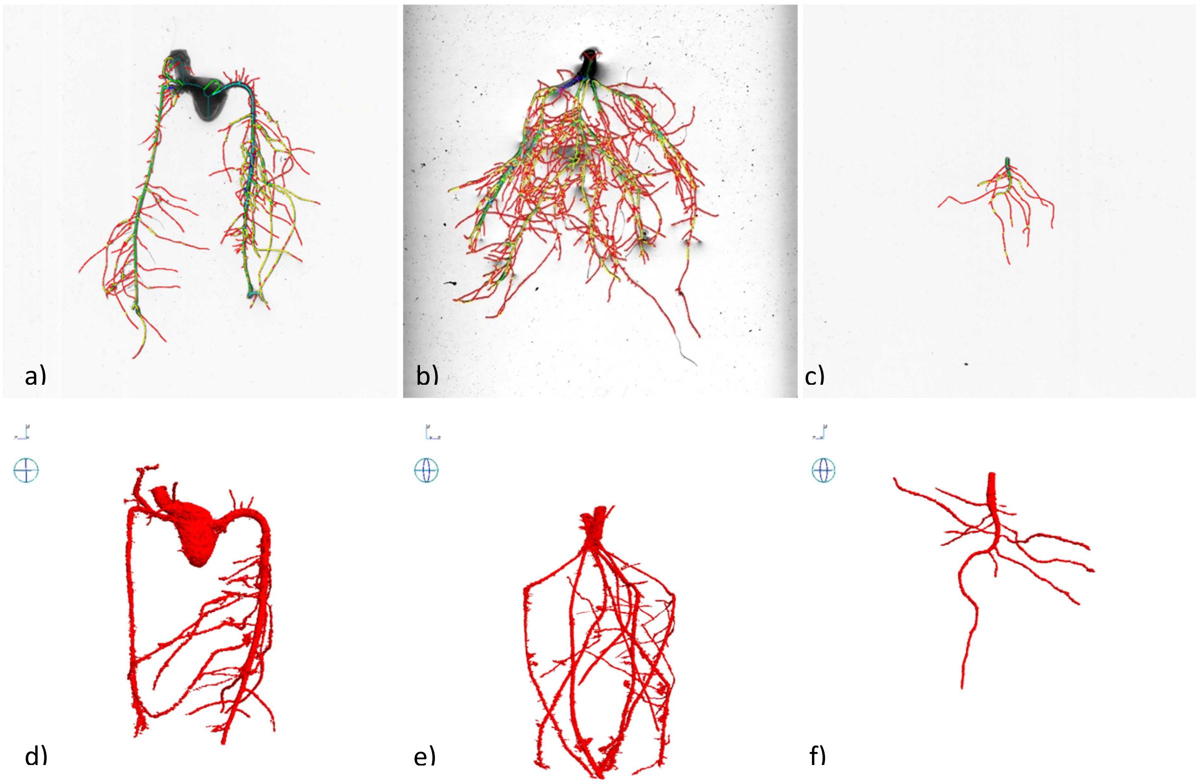


Figure 6: Plant root systems of (a,d) maize grown in clay loam, (b,e) wheat grown in loamy sand and (c,f) tomato grown in loamy sand, showing (a,b,c) WinRhizo, (d,e,f) X-ray CT extracted root systems

Discussion

A number of experimental approaches have been presented for the evaluation of methods for the recovery of plant root systems from X-ray µCT image data. These range from analysing artificially generated data, to measuring root segments buried in soil, to the extraction of complete root systems of real plants. To provide a point of reference we focused on RooTrak, a visual tracking-based extraction method that is publically available via SourceForge.

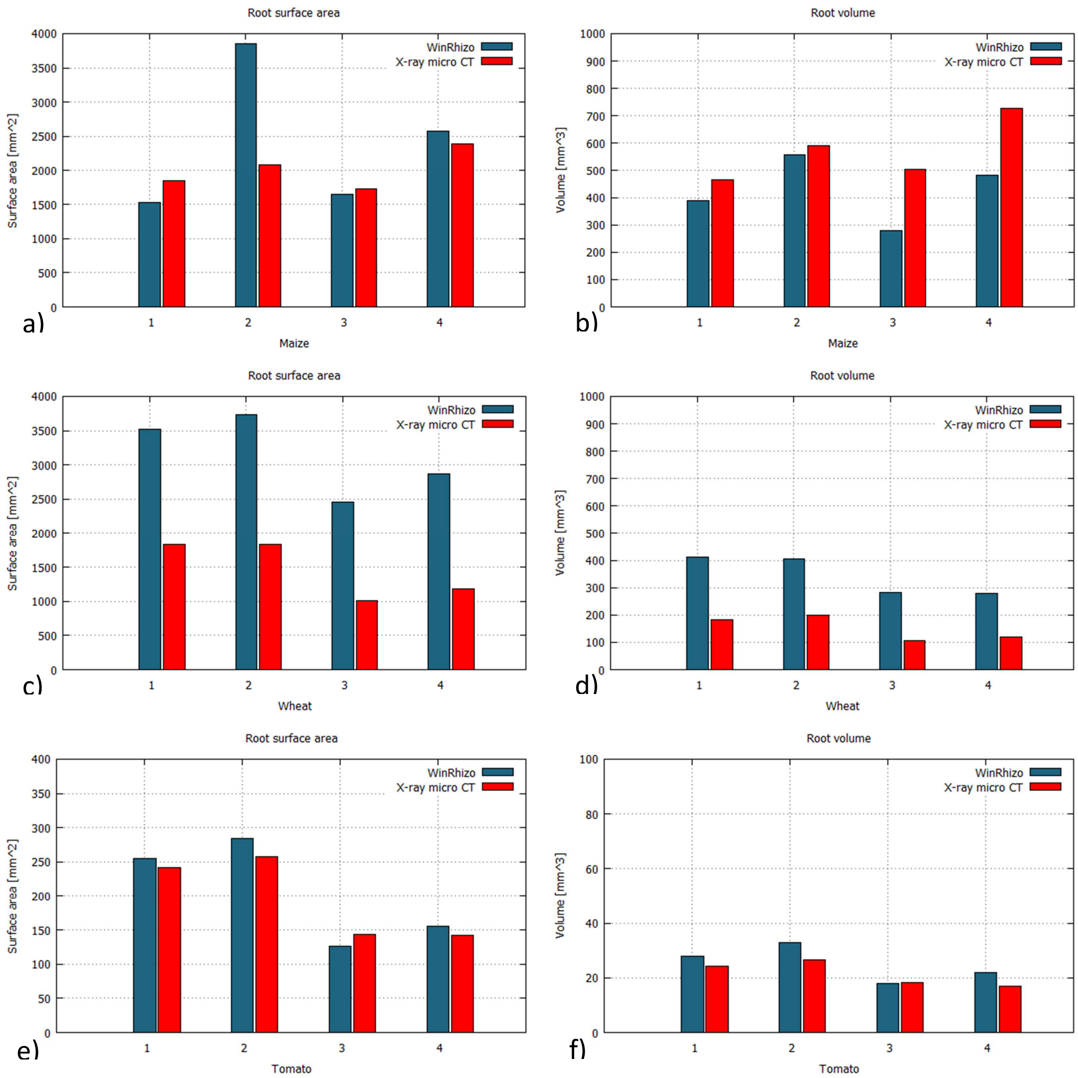


Figure 7: (a,c,e) Surface area and (b,d,f) volume between WinRhizo and X-ray µCT extracted data using RooTrak for (a,b) maize, (c,d) wheat and (e,f) tomato

Artificially generated test data allows precise determination of missed and incorrectly classified voxels, giving unambiguous measures of segmentation performance. A further clear advantage is the amount of control provided over the test object's shape and the image noise characteristics. This supports evaluation of methods in particular scenarios and under well-defined conditions, some of which can be difficult to create when growing real plants and using a real scanner. However, tests should not be limited to artificial data, as it is less representative of real plants grown in soil. Although certain characteristics can be simulated that resemble real CT data, the object and its background are highly simplified. They do not reflect the full complexity and heterogeneity that is found in real image data.

The artificial data experiments conducted here show, as expected, that low contrast and high noise level makes it impossible to accurately identify the target object. The accuracy however increases if either the contrast increases or the noise level drops. Best performance is obtained with high contrast and no added noise.

Using root segments buried in soil gives a more realistic test environment without having to sacrifice too much control over the test object. Segments can be selected so that root branches are included, for instance. These can then be placed in the soil column at known locations and in any orientation. Image noise does not need to be simulated and can be varied by using different scan settings. While more realistic, this approach limits the noise characteristics tested to the model of scanner being used. Roots can be manually measured, though this is easier if short and straight segments are used. Branching structures are more difficult to measure accurately. Measuring root segments manually gives a very good approximation of the actual size, but because roots are easily deformable they might respond to the pressure that they experience while being buried in soil, whereas no force is applied when they are exposed to the air. In addition, roots are able to store a large amount of water but at the same time dry out very quickly if kept in the air for too long. While the root segments were viewed under the microscope, they had to be removed from the water. It would have made the measurement difficult if the roots were kept in water due to root hairs becoming visible and blurring the edges of the root in the image. It therefore remains unknown whether the water content in the root segments truly remained constant while segments were viewed under the microscope and imaged with X-ray µCT. Water content and soil pressure might have caused variations in the physical shape of the root during the experiment and led to small errors in measurement. The acceptability of the measurement error under the different circumstances in which the analysis was performed is debateable. An overestimation of surface area from the X-ray µCT recovered data can be observed in comparison to the data obtained by microscope, while root volume is slightly underestimated. This could be explained by the assumption we made that the manually measured root segments were perfectly cylindrical with a smooth surface, whereas in X-ray µCT we find a more natural, rough and uneven surface, which lead to higher values. However, the results obtained show a strong correlation between the manual measurements and those extracted from the CT scans, which strongly suggests that the measurements obtained are representative of the samples.

To evaluate methods on their performance when recovering the structure of complex root systems in a natural environment, real plants must be used. Extracted root system descriptions can be compared to excavated plants, by measuring characteristics such as volume or surface area. Because of the root system's complexity, these are difficult to determine manually, but two-dimensional analysis tools, such as WinRhizo, can be used. In the experiment performed here, the best results were obtained from the tomato samples, which yielded the smallest error and a high correlation. The root system of the tomato is the least complex among all the samples, with only a few lateral roots. Relatively good data was also obtained from the maize samples, although the mean error is slightly increased and the results are not as well correlated as the traits recovered from the other samples. Maize 2 showed a large error in surface area, which is very likely due to the high number of fine lateral roots that were not present in the other maize plants, and were missed during the extraction process. It should be noted that the maize plants have been scanned at a lower resolution than the tomato and wheat samples. Roots can only be extracted from the CT data if they are large enough to be visible in the cross-sectional images that are processed. The main root architecture, however, has been fully recovered, as shown in figure 6. Surface area and volume measurements recovered from the wheat samples are largely underestimated. The extraction of the many fine later roots proved to be problematic, though the primary roots have been successfully recovered. This is clearly shown in figure 6. However, it should be noted that the exact surface area and volume of the root systems are unknown. Two-dimensional analysis tools often estimate root system characteristics from other parameters and therefore fail to completely describe the structural complexity. Measures, such as volume or surface area, are valuable in the comparison between root systems, but do not provide information on the amount of root material that has been missed or incorrectly classified. The field of view from the CT scans can be limited and in our experiment did not reach the far bottom of the sample. When removing the roots from the soil, it becomes unclear what portion of the roots has grown below the visible depth and therefore gives rise to an unknown error when compared to the flatbed scanned images.

None of the above tests is clearly better than the others, since all have their advantages and limitations, but they do complement each other very well and together provide a picture of the ability of the segmentation method. Where possible, proposed extraction methods should also be compared against alternative solutions. The comparative approach relies upon real scanner data, which is characterised by many factors. Data obtained with one model of scanner may not be achieved with another, limiting the generality of the evaluation. Different segmentation methods also have different strengths and weaknesses, and may be better suited to some image acquisition methods than others.

Conclusion

Evaluation of root extraction methods from X-ray CT data is a complicated task, since reliable and accurate ground truth data of real samples is not available and artificially generated data is less representative of real roots. Ground truth data is difficult to achieve, as root architectures cannot be predicted before scanning, and manual creation of ground truth from complex three-dimensional imagery is challenging. Though segmentation is performed to support the extraction of plant root traits, root phenotyping is in its infancy and the acquisition of ground truth trait descriptions is also problematic. On the other hand, artificial data can be easily generated and allows an exact evaluation of the extraction methods, but does not necessarily share the same image complexity and characteristics. Together, however, they can reduce bias and increase the reliability of evaluation, as well as covering a wide range of different scenarios. The experiments discussed in this work form a set of possible tests that can be easily performed and reproduced, but is not limited by additional examinations. Interesting, though not accessible to everyone, is the use of 3D printers that would allow the creation of complex structures obtained from mesh data of previously extracted root systems. Through registration of the mesh and the recovered root system data of the printed object, it would be possible to determine the number of correct and incorrectly classified voxels. The material used for printing such an artificial root structure should ideally have a density value similar to root material, but be robust enough to survive being buried and excavated. 3D printed objects are showing potential applications in tomographic imaging. O'Callaghan et al. (2014) have recently presented details of 3D printable objects used in the calibration of pre-clinical, high resolution MRI systems. Another useful tool in evaluating extraction methods for plant root systems would be a combination of mathematical root and soil modelling tools that could be used to generate more realistic artificial image data. Both approaches would narrow the gap between the artificially generated data discussed here and the use of real plant samples in soil, providing better ground truth data for accurate evaluation while at the same time preserving their natural characteristics.

Acknowledgment

This work was supported by Biotechnology and Biological Sciences Research Council and Engineering and Physical Sciences Research Council Centre for Integrative Systems Biology program funding to the Centre for Plant Integrative Biology; European Research Council Advanced Grant funding to M.J.B, T.P.P, S.J.M, C.J.S, S.M; BBSRC Professorial Fellowship funding to M.J.B.; Belgian Science Policy Office (grant IAP7/29) funding to M.J.B.; Royal Society Wolfson Research Merit Award to M.J.B.; EU FP7-KBBE-2011-5 project EURoot: Enhancing resource Uptake from roots under Stress in Cereal Crops to T.P.P. and M.J.B.

References

J.-L. Arsenault, S. Pouleur, C. Messier, R. Guay. WinRHIZO™, a root-measuring system with a unique overlap correction method. *HortScience*, **30**:906, 1995.

Tjeerd J. Bouma, Kai L. Nielsen, and Bas Koutstaal. Sample preparation and scanning protocol for computerised analysis of root length and diameter. *Plant and Soil*, **218**(1-2):185-196, 2000. ISSN 0032-079X.

S. Bouix, M. Martin-Fernandez, L. Ungar, M. Nakamura, M. S, Koo., R. W. McCarley, and M. E, Shenton. On evaluating brain tissue classifiers without a ground truth. *Neuroimage*, **36**(4):1207-1224, 2007.

Deo A. Heeraman, Jan W. Hopmans, and Volker Clausnitzer. Three dimensional imaging of plant roots in situ with x-ray computed tomography. *Plant and Soil*, **189**:167-179, 1997. ISSN 0032-079X.

T. Heimann, B. Van Ginneken, M. A. Styner, Y. Arzhaeva, V. Aurich, C. Bauer, A. Beck et al. Comparison and evaluation of methods for liver segmentation from CT datasets. *Medical Imaging, IEEE Transactions on,***28**(8):1251-1265, 2009.

Gabor T. Herman. Fundamentals of Computerized Tomography: Image Reconstruction from Projections. Springer Publishing Company, Incorporated, 2nd edition, 2009. ISBN 185233617X, 9781852336172.

Margarita L. Himmelbauer, Willibald Loiskandl, and Ferdinand Kastanek. Estimating length, average diameter and surface area of roots using two different image analyses systems. *Plant and Soil*, **260**(1-2):111-120, 2004. ISSN 0032-079X.

Anjali S. Iyer-Pascuzzi, Olga Symonova, Yuriy Mileyko, Yueling Hao, Heather Belcher, John Harer, Joshua S. Weitz, and Philip N. Benfey. Imaging and analysis platform for automatic phenotyping and trait ranking of plant root systems. *Plant Physiology*, **152**(3):1148-1157, 2010.

Anders Kaestner, Martin Schneebeli, and Frank Graf. Visualizing three dimensional root networks using computed tomography. *Geoderma*, **136**(12):459-469, 2006. ISSN 0016-7061.

Avinash C. Kak and Malcolm Slaney. Principles of Computerized Tomographic Imaging. IEEE Press, 1988.

T. Kohlberger, V. Singh, C. Alvino, C. Bahlmann, and L. Grady. Evaluating segmentation error without ground truth. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012, 528-536, 2012.

Melinda Lontoc-Roy, Pierre Dutilleul, Shiv O. Prasher, Liwen Han, Thomas Brouillet, and Donald L. Smith. Advances in the acquisition and analysis of ct scan data to isolate a crop root system from the soil medium and quantify root system complexity in 3-d space. *Geoderma*, **137**(12):231-241, 2006. ISSN 0016-7061.

Jonathan P. Lynch, Kai L. Nielsen, Robert D. Davis, and Andrei G. Jablokow. Simroot: Modelling and visualization of root systems. *Plant and Soil*, **188**(1):139-151, 1997. ISSN 0032-079X.

Stefan Mairhofer, Susan Zappala, Saoirse R. Tracy, Craig Sturrock, Malcolm J. Bennett, Sacha J. Mooney, and Tony P. Pridmore. Rootrak: Automated recovery of three-dimensional plant root architecture in soil from x-ray micro computed tomography images using visual tracking. *Plant Physiology*, **158**(2):561-569, 2012.

Stefan Mairhofer, Susan Zappala, Saoirse R. Tracy, Craig Sturrock, Malcolm J. Bennett, Sacha J. Mooney, and Tony P. Pridmore. Recovering complete plant root system architectures from soil via x-ray μ-computed tomography. *Plant Methods*, **9**(1):1-8, 2013. ISSN 1746-4811.

D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. *In Computer Vision, 2001.ICCV 2001.Proceedings. Eighth IEEE International Conference on*, **2**: 416-423, 2001.

J. O'Callaghan, J. Wells, S. Richardson, H. Holmes, Y. Yu, S. Walker-Samuel, B. Siow, and M. F. Lythgoe. Is Your System Calibrated? MRI Gradient System Calibration for Pre-Clinical, High-Resolution Imaging. *PloS one*, **9**(5), 2014.

J. S. Perret, M. E. Al-Belushi, and M. Deadman. Non-destructive visualization and quantification of roots using computed tomography. *Soil Biology and Biochemistry*, **39**(2):391-399, 2007. ISSN 0038-0717.

P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET evaluation methodology for face-recognition algorithms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **22**(10):1090-1104, 2000.

Alain Pierret, Yvan Capowiez, Christopher J. Moran, and Andre Kretzschmar. X-ray computed tomography to quantify tree rooting spatial distributions. *Geoderma*, **90**(3-4):307-326, 1999. ISSN 0016-7061.

James A. Sethian. Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science ... on Applied and Computational Mathematics. Cambridge University Press, 2nd edition, 1999. ISBN 0521645573.

Peter Toft. The Radon Transform - Theory and Implementation. PhD thesis, Department of Mathematical Modelling, Technical University of Denmark,1996.

Saoirse R. Tracy, Jeremy A. Roberts, Colin R. Black, Ann McNeill, Rob Davidson, and Sacha J. Mooney. The X-factor: visualizing undisturbed root architecture in soils using X-ray computed tomography. *Journal of experimental botany,* **61**(2):311-313, 2010.

Y. Yan, R. Rosales, G. Fung, M. W. Schmidt, G. H. Valadez, L. Bogoni, L. Moy, and J. G. Dy. Modeling annotator expertise: Learning when everybody knows a bit of something. *In International conference on artificial intelligence and statistics*, 932-939, 2010.

Susan Zappala, Jonathan R. Helliwell, Saoirse R. Tracy, Stefan Mairhofer, Craig J. Sturrock, Tony Pridmore, Malcolm Bennett, and Sacha J. Mooney. Effects of x-ray dose on rhizosphere studies using x-ray computed tomography. *PLoS ONE*, **8**(6):1-8, June 2013.