

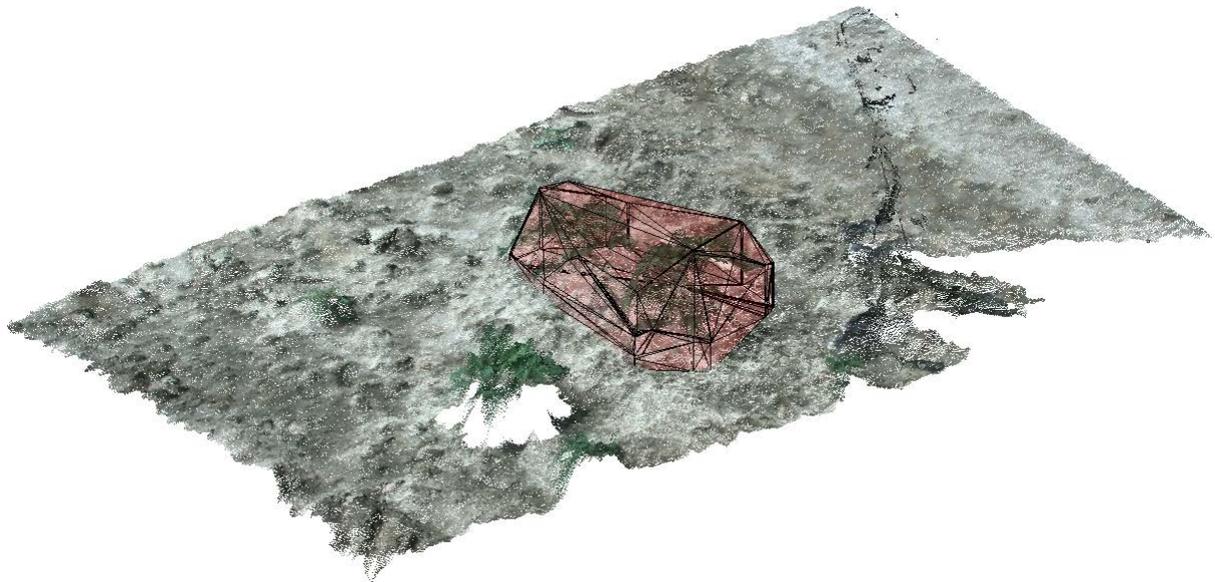
Above-ground potato plant dry matter estimation using stereo vision

Internal report for the Blightmanager project.

Anders Krogh Mortensen^{1,}, Jacob Glerup Gyldengren¹ and René Gislum¹*

¹Department of Agroecology, Aarhus University, Denmark

*Corresponding author: anmo@agro.au.dk



Contents

1	Introduction	3
2	Data acquisition	3
3	Potato plant segmentation	6
3.1	From images to point cloud	6
3.1.1	Orient point cloud according to ground plane.....	7
3.2	Subtract point clouds from before and after sampling plants.....	9
3.2.1	Align point clouds using iterative closest point	9
3.2.2	Resample point clouds to same spatial grid	10
3.2.3	Subtracting point clouds	10
3.3	Segment potato plant	11
3.4	Volume	12
4	Results	14
4.1	Effect of hyper-parameters.....	14
4.2	Dry matter estimation.....	15
5	References	18

1 Introduction

Aboveground crop biomass and the estimation of this during the growth and development of the potato crop is an important parameter for the general status of the crop. Crop biomass is also important if the aim is to optimise nitrogen application rate and/or optimise irrigation. It is difficult and time consuming to either estimate crop biomass by the human eye or take destructive plant samples in the field, process these samples and calculate the total biomass. Based on the importance of gathering information of potato crop biomass, one of the aims in the Blightmanager project is to use cameras to predict aboveground potato crop biomass.

This internal report for the Blightmanager project presents an algorithm for estimating the dry matter content of potato plants. The proposed algorithm uses a stereo vision camera to estimate the change in volume to predict the dry weight of sampled potato plants.

2 Data acquisition

Potato plant samples and corresponding images were collected on 11th and 16th August 2021 at Aarhus University, Flakkebjerg. The MultiSense S21 (Carnegie Robotics, Pittsburgh, PA, USA) stereovision camera was used to collect matching RGB and depth images (Figure 1). The depth images are derived by matching points in the MultiSense S21's built-in left and right RGB cameras. The MultiSense S21 calculated the depth image internally and with respect to its left camera. The two cameras had a resolution of 544×1024 pixels, a focal length of 6 mm and a baseline¹ of 28 cm. Exposure time and white balance were set automatically by the camera. The camera was mounted downward facing at approximately 1.7 m above the ground on the agricultural field robot Robotti (Agrointelli, Aarhus, Denmark). A laptop running the open source Robot Operating System (ROS, www.ros.org) was connected to the camera. The laptop was used to control and monitor the live images from camera and to store the collected images for offline processing. Images of the plant samples were acquired before and after collecting the individual plant samples (Figure 1). On the 11th August, the Robot remained stationary before, during and after collecting the plant samples. On the 16th August, the Robotti moved while collecting the images to provide images of the same plant sample from multiple view points.

The potato plants were sown in rows with a row distance of 0.75 m and 0.33 m between plants in in the same row. Ten plant samples were collected on both the 11th and 16th August. A given plant sample could contain multiple plants (Figure 1, a+c). The collected samples were weighted in the laboratory and then dried for 24 hours at 60 °C. The samples were weighed before and after the drying process. The samples from the 11th and 16th August ranged from 30 g to 157 g dry matter and from 30 g to 116 g dry matter, respectively (Figure 2).

The data collected on the 11th August were used for calibration, while the data collected on the 16th August were used for validation.

¹ The baseline is the distance between the left and right camera in a stereovision setup.

Above-ground potato plant dry matter estimation using stereo vision

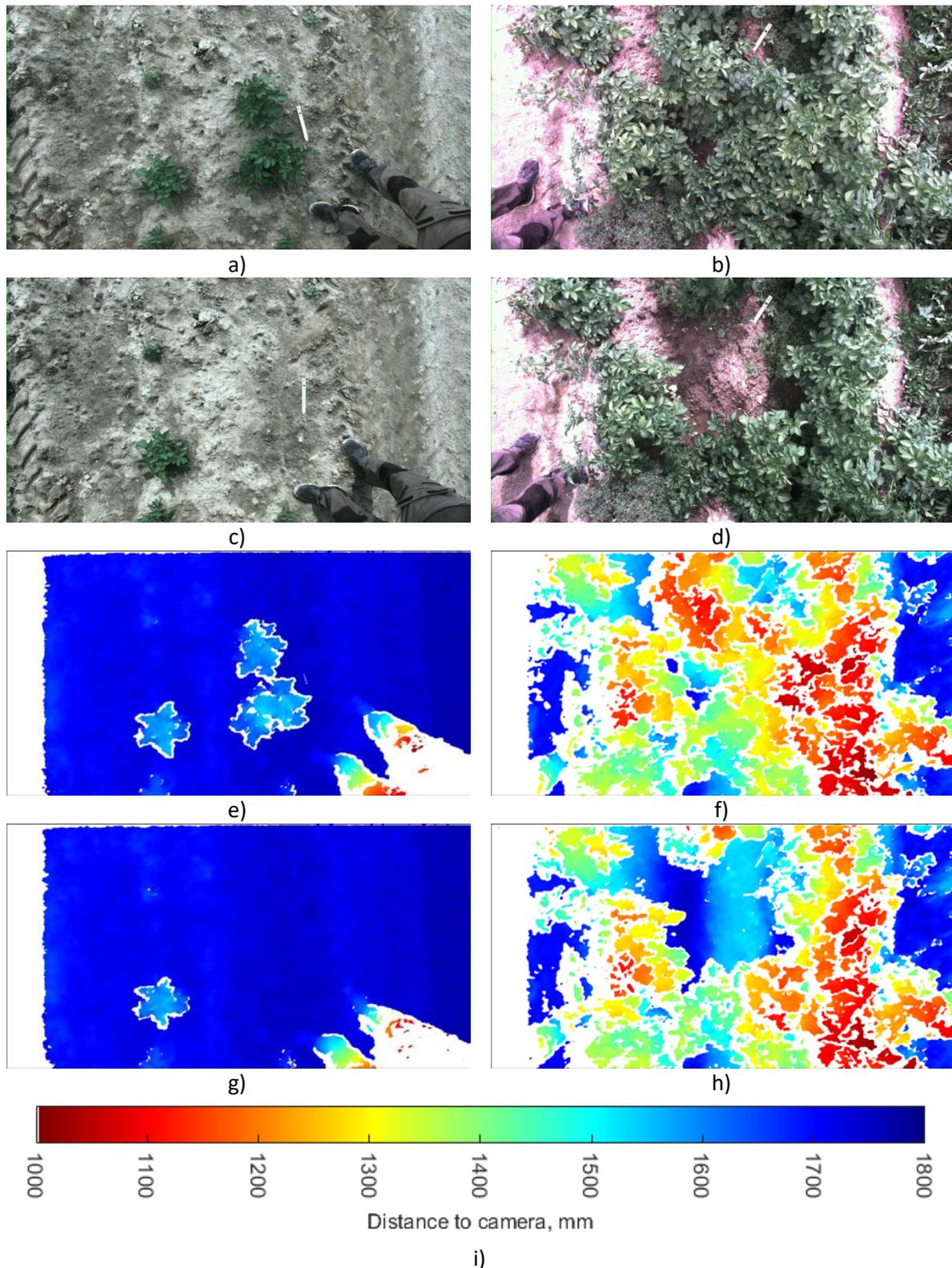


Figure 1. Examples of collected RGB images (a-d) and pseudo-coloured depth images (e-h) before (a-b and e-f) and after (c-d and g-h) collecting two different potato plant samples. i) shows the pseudo-colour to depth mapping used in the pseudo-coloured depth images. White corresponds to unknown depths.

Above-ground potato plant dry matter estimation using stereo vision

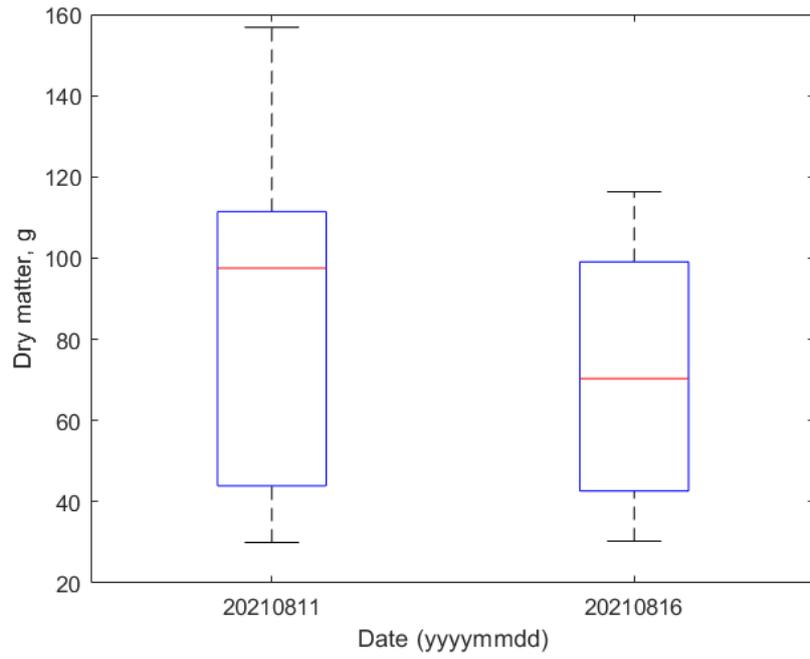


Figure 2. Dry matter distribution of the collected plant samples. The red lines indicate the median, the box edges are the 25th and 75th percentile and the end of the dashed lines show the extreme data points. “+” indicates outliers (none present).

3 Potato plant segmentation

To estimate the dry matter of the sampled potato plants from the collected RGB and depth images, the RGB and depth images are first converted to 3D point clouds. Then the point clouds before and after sampling are aligned and the changes between the two are determined. Assuming, that only the sampled plants have been removed between acquiring the “before” and “after” point clouds, the sampled plants can be identified by examining the changes between the two point clouds. The sampled potato plant is segmented in the “change” point cloud by through thresholding and clustering. Two volume estimates were explored. Each step is described in detail in the following sections.

3.1 From images to point cloud

Each pixel in the depth image provided by the stereovision camera corresponds to the distance from the camera to a point on the nearest object intersected by the pixel ray. The stereovision camera must be able to identify the same point in both the left and the right camera of the stereovision camera to calculate the distance to the point. Therefore, points, which are either not identifiable in both the left and right camera of the stereovision camera, lead to undefined depths in depth image (Figure 1, e-h).

The intensity of each pixel in the depth image are in metric units, while the spatial extend of the image are in pixel units. To have both metric units, the depth image was converted to a point cloud using the focal length and principle point of the camera. The conversion assumes a pinhole camera model without any distortion. As left and right images are rectified to create the depth images, this is a fair assumption. Let $p = [p_x, p_y, p_z]$ be a point in three dimensional space, and x_{px} and y_{px} be the image coordinates, where the ray from the optical centre to p intersects the image plane, then the coordinates of p are given by:

$$p_x = \frac{(x_{px} - \rho_x) \cdot D(x_{px}, y_{px})}{f_x} \quad \text{Eq. 1}$$

$$p_y = \frac{(y_{px} - \rho_y) \cdot D(x_{px}, y_{px})}{f_y} \quad \text{Eq. 2}$$

$$p_z = D(x_{px}, y_{px}) \quad \text{Eq. 3}$$

where $\rho = [\rho_x, \rho_y]$ is the principle point in the image plane, f_x and f_y are the focal length in pixels in the x- and y-directions of the image plane, respectively, and $D(x_{px}, y_{px})$ is the depth given by the depth image at coordinates $[x_{px}, y_{px}]$. We used the calibration values obtained by the manufacturer: $\rho = [252, 498]$ and $f_x = f_y = 556$ pixels.

As the depth image is derived from and fully aligned with the left RGB image from the stereovision camera, the points in the point cloud can without any assumptions be coloured from the RGB image.

Figure 3 show examples of the converted point clouds from the depth and RGB images in Figure 1. Note, that direction of the Z-axis is reversed, as the origin of the Z-axis is according to the image plane of the downward facing camera – and not the ground.

Above-ground potato plant dry matter estimation using stereo vision

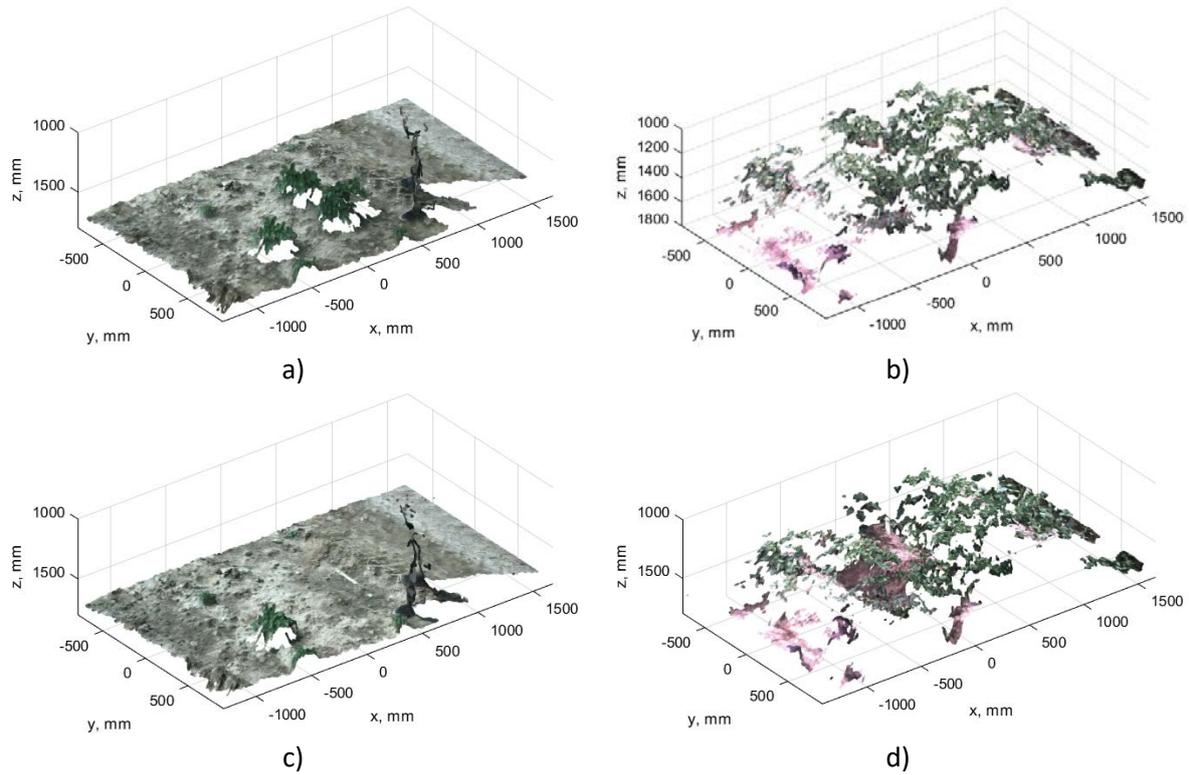


Figure 3. Examples of generated point clouds from before (a-b) and after (c-d) collecting potato plant samples from two different conditions. The point clouds correspond to the RGB and depth images in Figure 1.

3.1.1 Orient point cloud according to ground plane

As noted in the previous section, the z-axis of the generated point cloud is reversed (Figure 3). While not strictly necessary, it is convenient when working with and viewing the point clouds to have the z-axis pointing upwards and the origin at ground level.

To estimate the ground level, a plane was fitted to all soil points in a distance between 1600 to 1800 mm from the camera (Figure 4.a-d). The plane was fitted to the soil points using the M-estimator SAmple Consensus (MSAC) algorithm (Torr & Zisserman, 2000). A point was regarded as soil, if its excess red (ExR) value was larger than its excess green (ExG) value (Meyer & Neto, 2008). Let $p_{RGB} = [p_R, p_G, p_B]$ be the red, green and blue colour of the point p , then $ExG(p_{RGB})$ and $ExR(p_{RGB})$ are given by:

$$ExG(p_{RGB}) = \frac{2p_G - p_R - p_B}{p_R + p_G + p_B} \quad Eq. 4$$

$$ExR(p_{RGB}) = \frac{1.4p_R - p_G}{p_R + p_G + p_B} \quad Eq. 5$$

$$ExGR(p_{RGB}) = ExG(p_{RGB}) - ExR(p_{RGB}) \quad Eq. 6$$

The point p is then regarded as a soil point, if $ExGR(p_{RGB}) < 0$.

A rigid transformation was used to translate and rotate the point cloud. The translation of the z-coordinates corresponded to the distance from the ground plane to the origin was transformed using a rigid transformation with a translation corresponding to the distance from the ground plane

Above-ground potato plant dry matter estimation using stereo vision

The origin of the x- and y-axis in the untransformed point cloud and the sampled potato plants were located at approximately the centre of the point cloud. Therefore, the distance from the camera to the ground plane was evaluated with respect to the origin. The untransformed point cloud was translated according to the estimated distance to the ground plane and fixed rotations around the y-axis and z-axis of 180° was used to orient the point cloud correctly, such that it had its z-axis point upwards, a right handed coordinate system and the aligned with the original RGB and depth images (Figure 4).

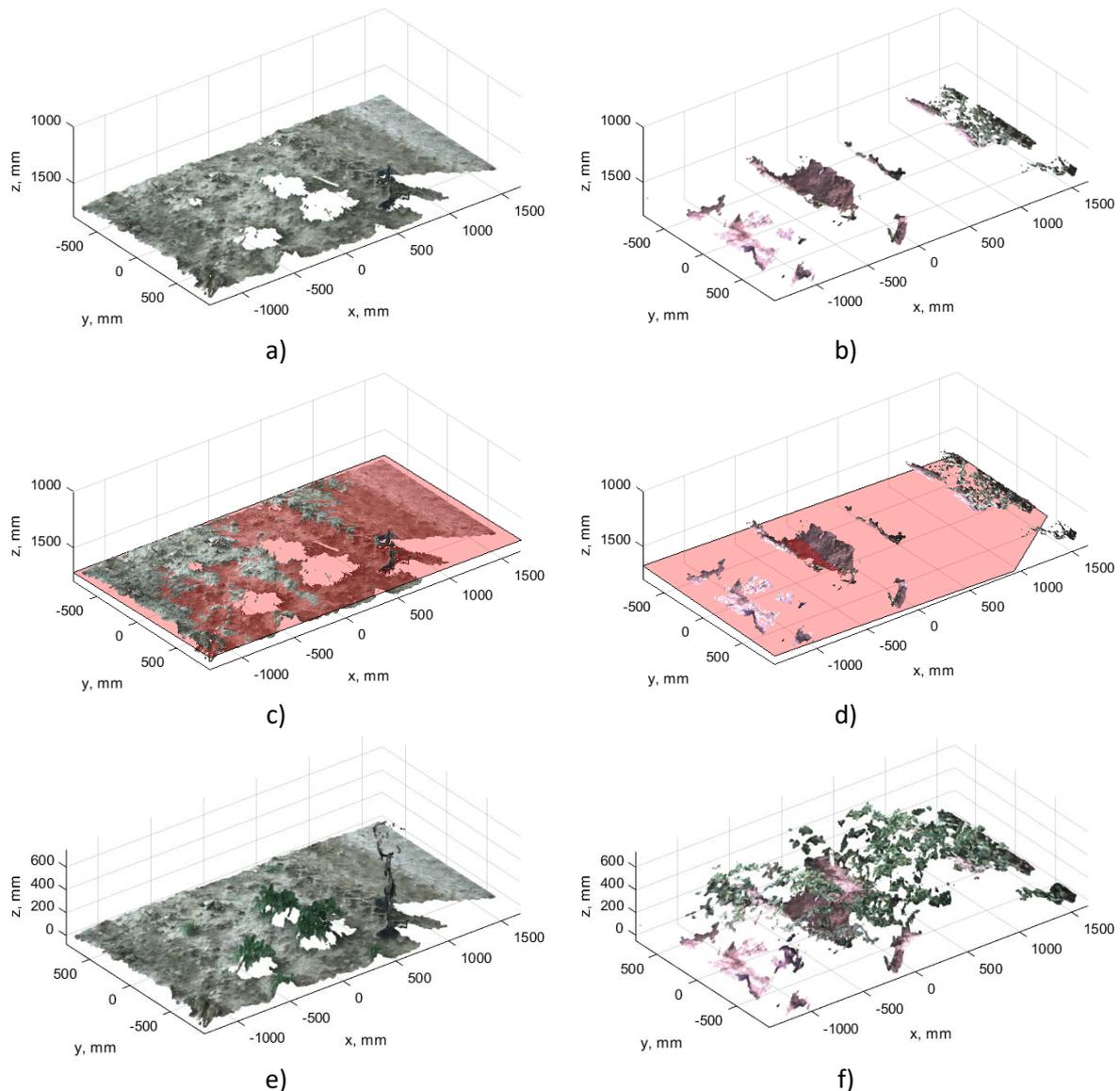


Figure 4. Examples of identified soil points (a-b), ground plane fitted to soil points (c-d), and rotated and translated point cloud (e-f). The examples corresponds to the point clouds seen in Figure 3.a (a, c, and e) and Figure 3.d (b, d, and f).

3.2 Subtract point clouds from before and after sampling plants

To subtract the point cloud after plant sampling from the point cloud before plant sampling, the two point clouds need to be aligned and resampled to the same spatial grid.

3.2.1 Align point clouds using iterative closest point

The iterative closest point (ICP) algorithm (Besl & McKay, 1992) was used to align the point clouds before and after sampling a potato plant. The ICP algorithm iteratively estimates the translation and rotation of a “moving” point cloud to align it with a “fixed” point cloud. In each iteration, the ICP algorithm matches the points in the “moving” point cloud to the points in the “fixed” point cloud, uses the matched points to estimate the translation and rotation, and translates and rotates the “moving” point cloud before rematching the points again in the next iteration. The ICP algorithm terminates, when the translation and rotation in a given iteration is sufficiently small or the maximum number of iterations has been reached.

In the present case, the point cloud collected after plant sampling was the “fixed” point cloud, while the point cloud from before sampling the potato plant was the “moving” point cloud. To reduce the processing time and reduce false matching between points while aligning the point clouds with ICP, the centre 1.5 m of the point clouds were ignored and each point cloud were downsampled using an averaging grid with 27 mm grid cells (Figure 5.a-b).

Examples of the aligned point clouds superimposed on each other are shown in Figure 5.c-d. Note that the soil/ground below the sampled plants are now present compared to Figure 3.a-b.

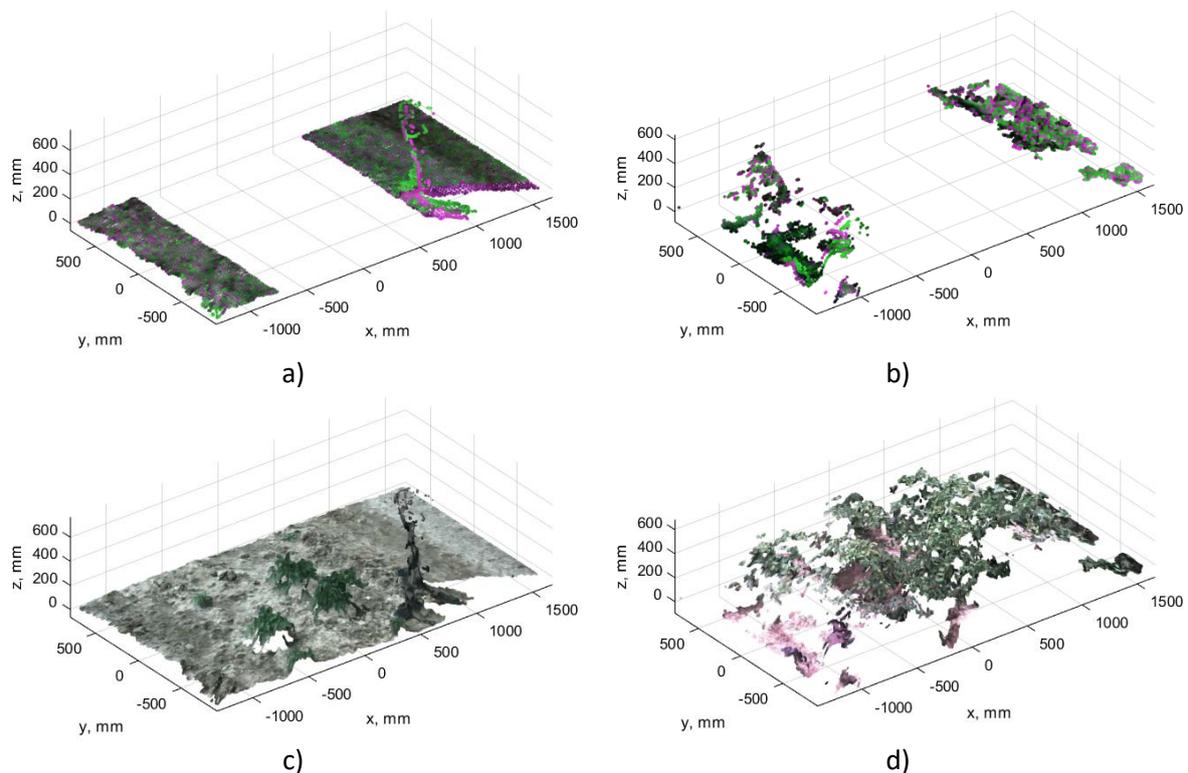


Figure 5. Examples of aligned point clouds. a) and b) show the downsampled point clouds without the centre used to determine the translation and rotation of the “before” point cloud. Green, magenta, and black points are the “before”, “after” and aligned points. c) and d) show the aligned “before” and “after” point clouds superimposed on each other. Note the soil below the sampled plants.

3.2.2 Resample point clouds to same spatial grid

To resample the point clouds to the same spatial grid, the coordinates of the point clouds were rounded to a fixed grid size. Let $p = [p_x, p_y, p_z]$ be a point in the point cloud, then the resampled point $\tilde{p} = [\tilde{p}_x, \tilde{p}_y, \tilde{p}_z]$ is given by:

$$\tilde{p}_x = \left\lfloor \frac{p_x - (p_{x,0} - \frac{1}{2}\delta)}{\delta} \right\rfloor \delta + (p_{x,0} - \frac{1}{2}\delta) \quad \text{Eq. 7}$$

$$\tilde{p}_y = \left\lfloor \frac{p_y - (p_{y,0} - \frac{1}{2}\delta)}{\delta} \right\rfloor \delta + (p_{y,0} - \frac{1}{2}\delta) \quad \text{Eq. 8}$$

$$\tilde{p}_z = \left\lfloor \frac{p_z - (p_{z,0} - \frac{1}{2}\delta)}{\delta} \right\rfloor \delta + (p_{z,0} - \frac{1}{2}\delta) \quad \text{Eq. 9}$$

where $p_{x,0}$, $p_{y,0}$ and $p_{z,0}$ are the lower spatial limits of the spatial grid, δ is the step size of the grid, and $\lfloor \cdot \rfloor$ rounds the number to the nearest integer. A step size δ of 3 mm was used. The colour of the resampled point is the same as the original point. While resampling, resampled points outside the spatial limits of either point cloud are discarded.

3.2.3 Subtracting point clouds

The points in the “before” point cloud were linked to the points in the “after” point cloud by comparing their xy-coordinates. If the distance in xy-coordinates between the points was less than 3 mm, the “after” points were considered a link candidate of the “before” point. The link candidate with the highest z-coordinate was linked to the “before” point, and the remaining link candidates were discarded. If there were no points within 3 mm, the “before” point was discarded.

After linking a point in the “before” and “after” point clouds, the height was determined change z-coordinate (Δz) by subtracting the z-coordinate of the “after” point from the z-coordinate of the “before” point. Furthermore, the change in ExGR ($\Delta ExGR$) was determined by subtracting the ExGR value of the “after” point from the ExGR value of the “before” point.

After determining the changes in height ExGR, a “change” point cloud was created by using the xy-coordinates of the points in the linked points in the “before” point cloud, the change in height as z-coordinate and the change in ExGR as intensity (Figure 6).

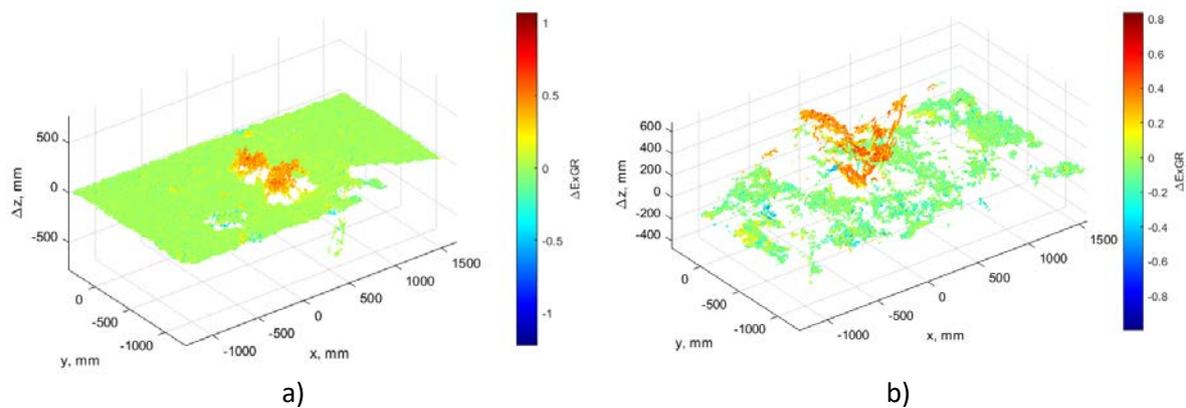


Figure 6. Examples of difference in height (Δz) and ExGR ($\Delta ExGR$) between pairs of “before” and “after” point clouds. The two examples (a) and (b) corresponds to the pairs of point clouds in Figure 3.

3.3 Segment potato plant

As an initial step to segment the potato plant in the “change” point cloud, all points with a change in ExGR ($\Delta ExGR$) less than $\Delta ExGR_{th}$ were removed (Figure 7a-b). A numerical small $\Delta ExGR$ corresponds to an area, which are either soil before and after sampling the plant or green plant material before and after plant sampling. A large positive $\Delta ExGR$ corresponds to green plant material removed and revealing soil below. This is mainly the sample potato plant. Small areas not associated with the sampled plant may also experience a positive $\Delta ExGR$ (Figure 6). This is due to small changes in canopy making the soil below visible, e.g. due to wind. Similarly, small changes in the canopy can also leads to a large negative $\Delta ExGR$, when soil, which was visible in the “before” point cloud, is covered by green plant material in the “after” point cloud.

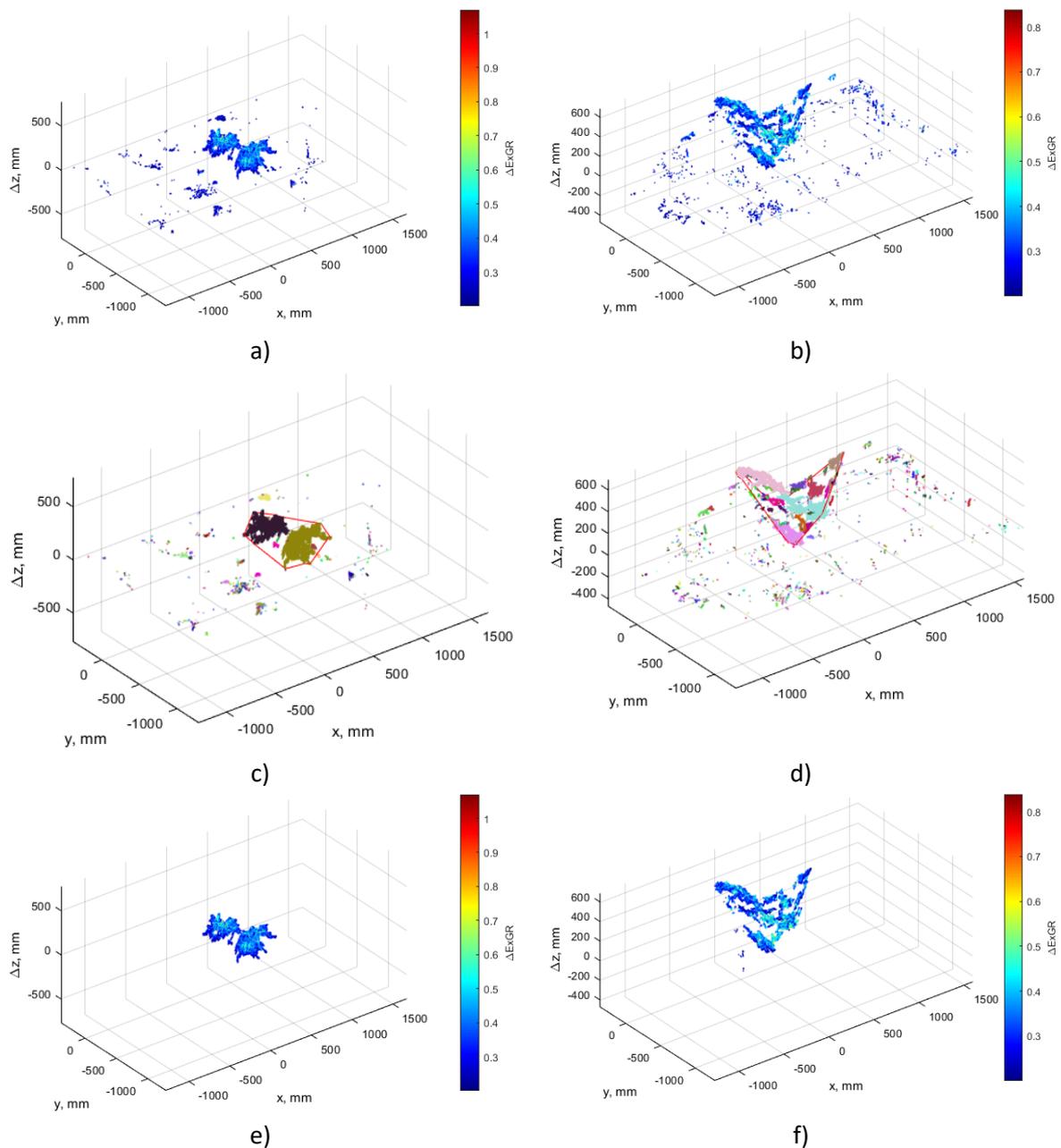


Figure 7. a) and b) show the thresholded “change” point clouds ($\Delta ExGR > 0.2$) from Figure 6. c) and d) show the clustering of the remaining points after thresholding. Each colour represents a different cluster. The red lines are the convex hull of the large clusters (#points < 717). e) and f) are the final segmented “change” point cloud after small clusters within the convex hull of the large clusters have been added.

After thresholding the “change” point cloud, the points are clustered based on their Euclidean distance. As the points were resampled to the same spatial grid with a step size of 3 mm, the points within $\sqrt{3 \cdot \delta^2} = \sqrt{3 \cdot 3^2} = \sqrt{27} \cong 5.2$ mm from each other are assigned to the same cluster. After thresholding, the largest clusters are formed from points associated with the sampled plant (Figure 7c-d). Small clusters exist are formed from both points associated with the sampled plant and from other minor changes in between the “before” and “after” point clouds, which has led to a large $\Delta ExGR$. To separate the large clusters of the sampled plant from the small clusters, a threshold on the cluster size were employed. All clusters larger than C_{th} was considered large clusters and assumed to be part of the sampled plant. To include the small clusters associated with the plant, all small clusters with a centre of mass inside the convex hull of the large clusters were assumed to be part of the sampled plant as well (Figure 7c-d).

Examples of the final segmented potato plant in the “change” point cloud are shown in Figure 7e-f.

3.4 Volume

Two different approached to estimating the volume change were explored: 1) convex hull (Barber, Dobkin, & Huhdanpaa, 1996) and 2) sum-of-columns. Furthermore, two segmentations were considered: One using only the large clusters to form the segmented point cloud and one using both the large and the small clusters to form the segmented point cloud.

The convex hull of a set of points is the extreme points, which encases all the other points (Figure 8c-d). The convex hull volume is then the volume spanned by these points. A virtual point with zero height was added for each point in the point cloud to ensure that the height was captured by the convex hull. If no virtual point were added, the change in height below the lowest point in the segmented point cloud would no contribute to the estimated volume.

As the point clouds were resampled into a regular spatial grid, each point in the segmented point cloud effectively represent a column with a volume of $v_{col} = \delta^2 \cdot \Delta z$. The sum-of-columns volume is then the sum of the volume of each point in the segmented point cloud (Figure 8e-f).

Above-ground potato plant dry matter estimation using stereo vision

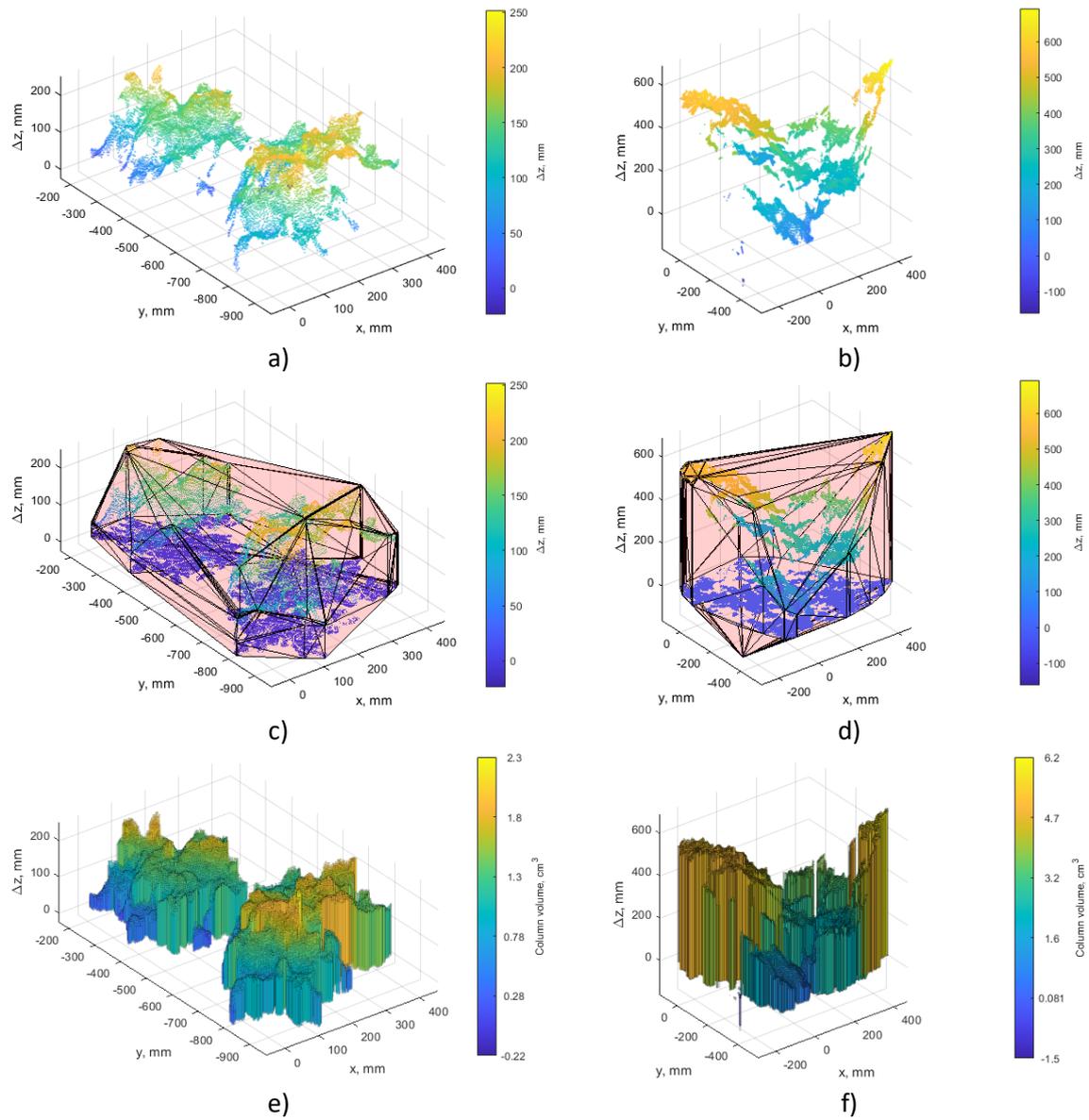


Figure 8. a-b) Segmented point clouds, where the colour corresponds to the change in height. c-d) Convex hull (red shaded area) of the segmented point clouds (a-b). e-f) column representation of the segmented point cloud (a-b) used to calculate the sum-of-columns volume.

4 Results

4.1 Effect of hyper-parameters

To study the effect of the two hyper-parameters $\Delta ExGR_{th}$ and C_{th} on the estimated dry matter, a set of linear models are fitted to the training data and evaluated using leave-one-out cross validation (Figure 9). The average root mean square error (RMSE) of the cross validation is used as metric for the performance of a given set of hyper-parameters. Five different values of $\Delta ExGR_{th}$ were evaluated. For each $\Delta ExGR_{th}$ a set of logarithmic spaced values of C_{th} ranging from 1 to 1585 were evaluated. Both sum-of-columns (Figure 9a) and convex hull (Figure 9b) volumes as well as using only the large clusters and including the small clusters were explored.

The models using the sum-of-columns volume estimation generally showed a much lower average RMSE and appeared more stable in terms of C_{th} . This is likely due to the convex hull being very sensitive to outliers, which can greatly increase the estimated volume (Figure 8d), while the same outlier will only increase the volume by the equivalent to its column in the case of the sum-of-columns method (Figure 8f). Moreover, the convex hull will also tend to overestimate the volume of point clouds, which have a concave shape, i.e., low point in the centre and high points near the edge (Figure 8c-d).

For the models using the sum-of-columns volume including the small clusters generally improved the models, particularly when increasing C_{th} (Figure 9a). In some cases, not including the small clusters resulted in a lower average RMSE, but in those cases, the average RMSE of the model without small clusters is very similar to the average RMSE of the model with small clusters.

Overall, the models with $\Delta ExGR_{th} = 0.1$ and $C_{th} = 282$ produces the lowest average RMSE. While these parameters provide the lowest RMSE for relating the volume to the dry matter, they do not necessarily provide the most accurate segmentation of the plants.

Above-ground potato plant dry matter estimation using stereo vision

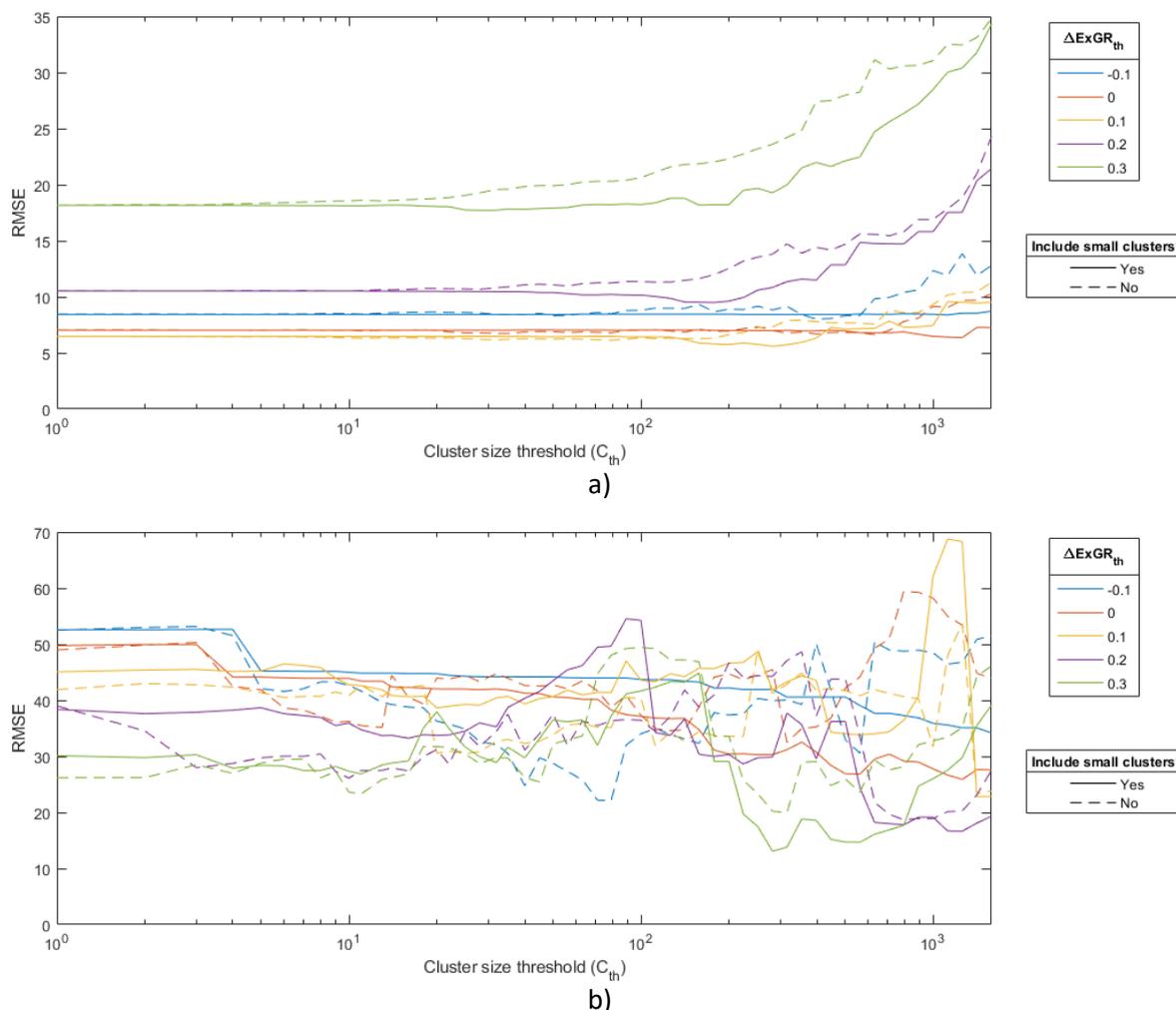


Figure 9. The effect of $\Delta ExGR_{th}$ threshold, cluster size thresholds (x-axis), inclusion of small clusters and volume calculation method on the average RMSE in leave-one-out cross validation on the training set. a) Volume calculated using sum-of-columns. b) Volume calculated using convex hull. Note that while the y-axis of both the top and bottom graphs start at 0, the y-axis on the bottom graph span twice the range of the y-axis on the top graph.

4.2 Dry matter estimation

The dry matter was estimated using the values for the hyper-parameters ($\Delta ExGR_{th} = 0.1$ and $C_{th} = 282$) from the previous section and using the sum-of-columns method for calculating the volume. Both large and small clusters were included in the calculation of the volume. A linear model was fitted on the training set and evaluated on the test set. A single outlier from the training set was excluded from the model fitting as it underestimated the volume across all hyper-parameters tested in the previous section. Unfortunately, this was also the sample with the highest dry matter.

The fitted model showed a strong correlation ($R^2 = 0.98$) on the training set (Figure 10), but a weaker correlation to the test set ($R^2 = 0.62$). The RMSE of the training and test sets were 4.4 g and 18.5 g, respectively. The test set indicates a slight trend in the model to underestimate the dry matter in the larger plants (90 – 120 g), however, the number of samples are low and including more samples may easily change the slope of the model. Including the outlier from the training set may have created a model better correlated with the test set, however, as all decisions regarding the model should be taken independently of the test set, it was not included.

Above-ground potato plant dry matter estimation using stereo vision

In the hyper-parameter search in the previous section, the models fitted with a low cluster size C_{th} performed almost as good as the corresponding models fitted a higher C_{th} (Figure 9a). Likewise, the models fitted with $\Delta ExGR_{th} = -0.1$ performed almost as well as the models fitted with the optimal $\Delta ExGR_{th} = 0.1$ (Figure 9a). A low $\Delta ExGR_{th}$ corresponds to including almost all the points in from the “change” point cloud in segmentation, while a very low C_{th} of 1 corresponds to including almost all points in the large clusters. Combined this almost equals skipping the segmentation step. This naturally leads to the question, if the segmentation step can be skipped altogether to fit a model on a sum-of-columns volume estimate of the “change” point cloud rather than the segmented point cloud. The model fitted directly to the “change” point cloud showed a similar, but slightly weaker correlation (Figure 11) to both the training ($R^2 = 0.97$) and test set ($R^2 = 0.58$) compared to the model fitted to the segmented point cloud (Figure 10). The RMSE of the training and test sets were 6.3 g and 19.5 g, respectively. The test set shows a greater trend underestimate the volume and dry matter across all samples. The outlier from the training set appears to be more aligned with the test set (Figure 11b).

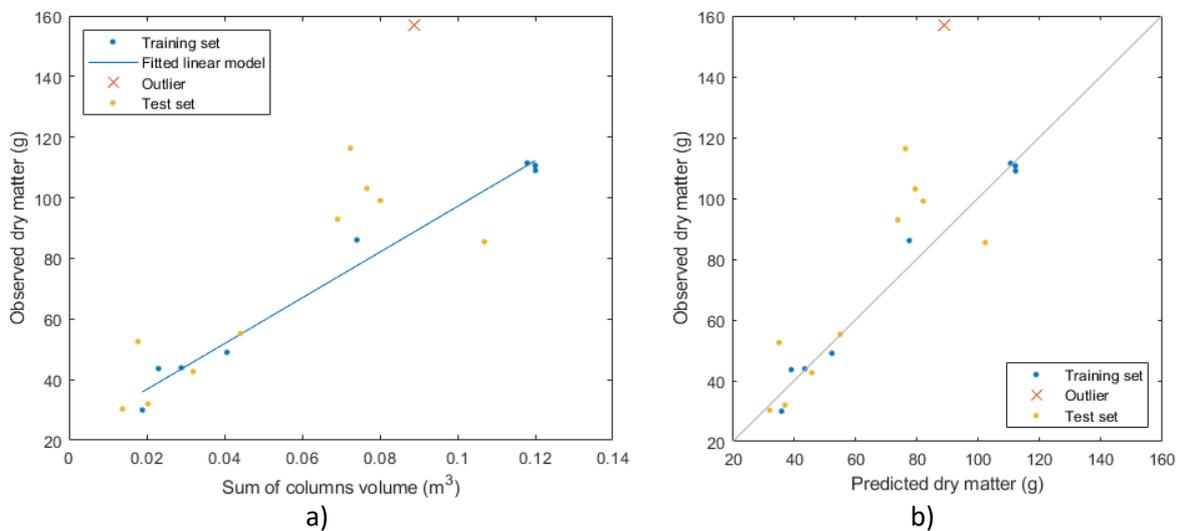


Figure 10. Correlation plots for the model using the hyper-parameters: $\Delta ExGR_{th} = 0.1$ and $C_{th} = 282$. a) Correlation plot of the estimated volume and the observed dry matter. b) Correlation plot between predicted dry matter and the observed dry matter.

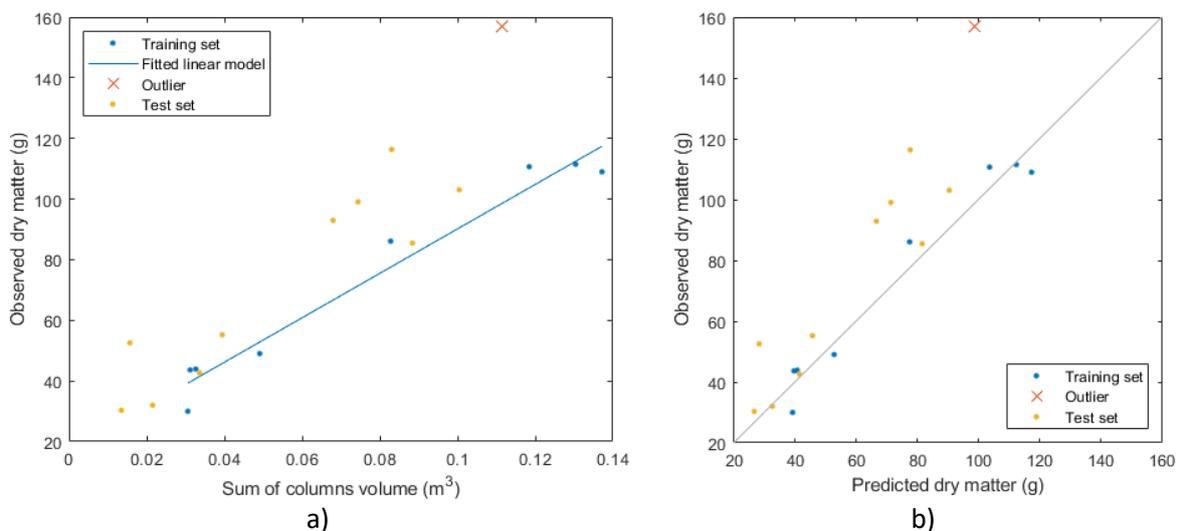


Figure 11. Correlation plots for the model without segmenting the potato plant. a) Correlation plot of the estimated volume and the observed dry matter. b) Correlation plot between predicted dry matter and the observed dry matter.

5 Conclusion

This report has presented a novel method for estimating the dry matter change between two point clouds generated from coloured 3D point clouds generated from depth and RGB images from a stereovision camera. While the results are promising, they are preliminary as the number of samples are very low. The results showed that the segmentation step could potentially be omitted for a slightly weaker correlation, but for a simpler algorithm with less hyper-parameters.

While the proposed method has focused on estimating the dry matter, it could be used for estimating new growth. This be done by comparing the “before” point cloud to a reference point cloud of the bare soil acquired the plant emerges. Weekly scans of the potato plants would give weekly estimates of the dry matter – and changes in the weekly dry matter would provide estimates of new growth. A deceleration in new growth could indicate plant stress, e.g. from fungal infection.

Future work should focus on acquiring more samples – ideally throughout the growing season, across different years and locations to capture as much variability in the conditions as possible. Future work may also focus on estimating the below-ground yield based on the above-ground plant material. This work may need support from a growth model combining the expertise of engineers (point cloud segmentation) and agronomists (growth model).

6 References

- Barber, C. B., Dobkin, D. P., & Huhdanpaa, H. T. (1996). The Quickhull Algorithm for Convex Hulls. *ACM Transactions on Mathematical Software*, 22(4), 469-483.
- Besl, P. J., & McKay, N. D. (1992). A Method for Registration of 3-D Shapes. *IEEE Transactions on Pattern Analysis and Machine Learning* (pp. 239-256). Los Alamitos, CA: IEEE Computer Society.
- Meyer, G. E., & Neto, J. C. (2008). Verification of color vegetation indices for automated crop imaging applications. *Computers and electronics in agriculture*, 282-293.
- Torr, P. H., & Zisserman, A. (2000). MLESAC: A New Robust Estimator with Application to Estimating Image Geometry. *Computer Vision and Image Understanding*.