

Field Robot for High-throughput and High-resolution 3D Plant Phenotyping

Felix Esser^{*,+,1}, Radu Alexandru Rosu^{*,2}, André Cornelißen¹, Lasse Klingbeil¹, Heiner Kuhlmann¹ and Sven Behnke²

¹ University of Bonn, Institute of Geodesy and Geoinformation,

² University of Bonn, Autonomous Intelligent Systems,

*Equal contribution, +Corresponding author, Email: esser@igg.uni-bonn.de



a) Ground robot with multimodal sensors



b) Laser-based reconstruction



c) Textured mesh reconstruction (color + geometry)

Fig. 1: a) UGV for high-resolution, georeferenced 3D in-field crop reconstruction with RTK, laser scanners and camera dome. b) Point cloud created with the laser scanning system. c) Textured mesh reconstructed from the multi-camera system.

Abstract—With the need to feed a growing world population, the efficiency of crop production is of paramount importance. To support breeding and field management, various characteristics of the plant phenotype need to be measured—a time-consuming process when performed manually. We present a robotic platform equipped with multiple laser and camera sensors for high-throughput, high-resolution in-field plant scanning. We create digital twins of the plants through 3D reconstruction. This allows the estimation of phenotypic traits such as leaf area, leaf angle, and plant height. We validate our system on a real field, where we reconstruct accurate point clouds and meshes of sugar beet, soybean, and maize.

I. INTRODUCTION

With today’s population growth and adverse climatic conditions, the need for crop monitoring and intervention has increased. However, manual field management is costly and laborious, which has encouraged the development of automated systems. In this article, we present such a system for high-throughput and high-resolution plant phenotyping. Phenotyping is the measurement and description of functional and structural plant traits such as fruit quality, leaf area, and biomass. These traits are critical for field management (fertilization, irrigation, or weeding) and breeding, where genotypes with desired traits are selected.

Existing phenotyping methods are diverse and range from visual observation of the crop to cutting down plants for chemical or structural analysis. In controlled environments such as greenhouses, phenotyping facilities have been established to automate the process. However, in the field, where the plant is in a natural environment subject to

weather conditions and competition from neighboring plants, the phenotyping process is still mostly performed manually. Although remote sensing platforms such as UAVs, airplanes, and satellites are increasingly being used for measurements from higher altitudes, the limited resolution of their sensors and plant self-occlusion can cause issues when phenotyping at the leaf and sub-leaf scale. The ground robot presented in this work enables automated high-throughput, high-resolution phenotyping in the field. It is equipped with multiple laser and camera sensors to capture accurate plant representations. The characteristics of the two system modalities are complementary: the 3D laser scanner can capture larger sections of the field in a short time, while the cameras focus more on individual plants and are suitable for reconstructing high-resolution color together with geometry. The acquisition system is precisely georeferenced using Real-Time Kinematic (RTK) GNSS positioning, allowing it to be combined with other relevant geospatial information, such as soil maps, and to revisit individual plants in the field. Fig. 1 shows the robot and exemplary 3D reconstructions from both sensing modalities. Both measurement systems provide a high geometric resolution, resolving plant organs such as individual leaves and even smaller structures. The laser-based reconstruction uses two laser-line triangulation sensors whose profile scans are georeferenced using a factor graph-based pose optimization technique. The camera-based system uses the neural implicit method of PermutoSDF [1] to reconstruct 3D geometry and texture from 2D images only. We validate the presented phenotyping system on the

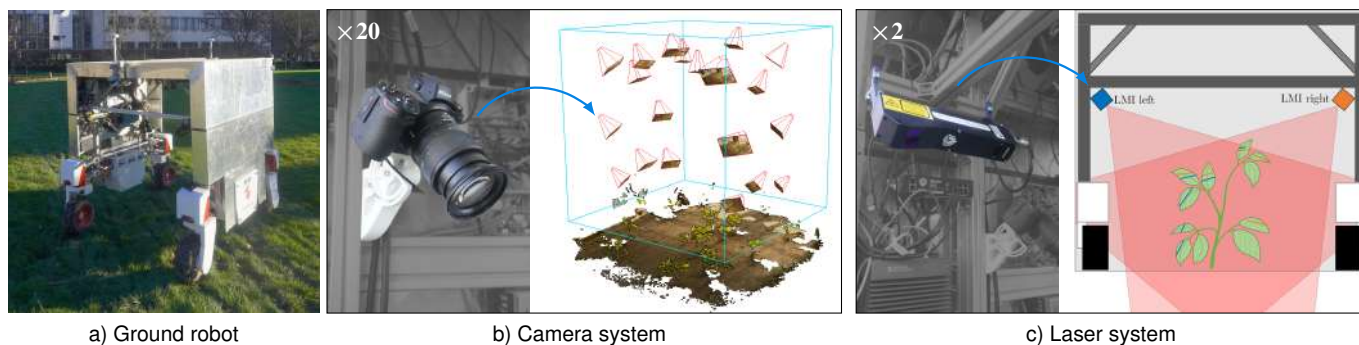


Fig. 2: **a)** UGV equipped with a multimodal sensor system for in-field plant phenotyping. **b)** 20 Nikon Z7 cameras are placed in a dome-like manner around the UGV’s center. **c)** Two profile laser scanners capture entire plots while driving at low speed.

PhenoRob experimental field. Accurate point clouds and textured meshes of sugar beet, soybean, and maize were created. In addition, we compare the laser and camera systems to each other given high-quality ground truth from a static indoor laser scanner. In summary, our contributions are:

- a ground robot equipped with high-resolution laser sensors, a camera dome, and a georeferencing system,
- methods to address outdoor lighting and localization challenges,
- demonstration of neural implicit representation reconstructions on a wide variety of plants imaged in the field,
- comparison between laser scanner and camera system against ground-truth 3D measurements, and
- experimental validation of the phenotyping system on a real field.

II. RELATED WORK

Existing field robots can be broadly categorized according to whether they are designed for intervention, such as weeding and seeding, or only for inspection and phenotyping. Although the distinction is not always clear, here we describe and compare some of the existing solutions.

For autonomous phenotyping, Bonirob [2] is a large four-wheeled robot with both electric and internal combustion engines designed for long-term scanning of sugar beet plots. It is equipped with a 4-channel multispectral JAI camera and an RGB-D Kinect v2 camera. However, the sensor does not have the accuracy and resolution required for highly detailed plant reconstruction. Later, Bonirob was also extended to support weeding tasks. TerraSentia¹ is another robot designed for phenotyping but on a smaller scale. The robot is small enough to fit under the canopy and navigate between crop rows. It captures camera images to the front, sides, and top, as well as horizontal lidar distance measurements, which are analyzed offline to estimate multiple plant traits.

¹EarthSense: TerraSentia <https://www.earthsense.co>

In contrast, our system is designed to navigate between rows and is flexible and high-resolution enough to scan plants from less than a centimeter to almost a meter high.

Several works exist on robots specialized in field intervention, such as Robotti2², Oz4403³, Dino4⁴, and BonnBot-I [3], all focusing on weed or pest control. Although they typically include laser scanners and cameras, the sensors are used to detect weeds or infected plants and not to reconstruct the full structure of the crop. They are also limited to observing smaller and younger plants, which is when weed control is most important. In contrast to these robots, we focus on high-resolution and accurate above-canopy 3D crop reconstructions under field conditions, with a quality that allows modeling and phenotyping at the plant and plant-organ level.

III. MATERIALS AND METHODS

A. Field Robot

Our field robot extends the Thorvald II platform [4], a lightweight modular base designed for agricultural purposes. Thorvald II features an electric 4-wheel drive that can be adjusted to different track widths and robot sizes. The onboard computer coordinates the motor wheels and can be remotely controlled using a Bluetooth-enabled Xbox controller, implemented in an ROS environment. This interface provides the potential for autonomous driving capabilities and enables easy communication and integration of the various sensor systems. We extend the default Thorvald base by an aluminum enclosure for the sensors of width and length 1.5 m × 1.5 m as seen in Fig. 2a. Together with the wheels, which are attached to a suspension module, the robot has a height of about 2 m. We add two additional computers to the platform for the laser and cameras. Both computers are equipped with multiple USB and Ethernet ports for easy connection to all the sensors and a large SSD hard

²Agrointelli: Robotti <https://agrointelli.com>

³Naio Technologies: Oz <https://www.naio-technologies.com/en/oz>

⁴Naio Technologies: Dino <https://www.naio-technologies.com/en/dino>

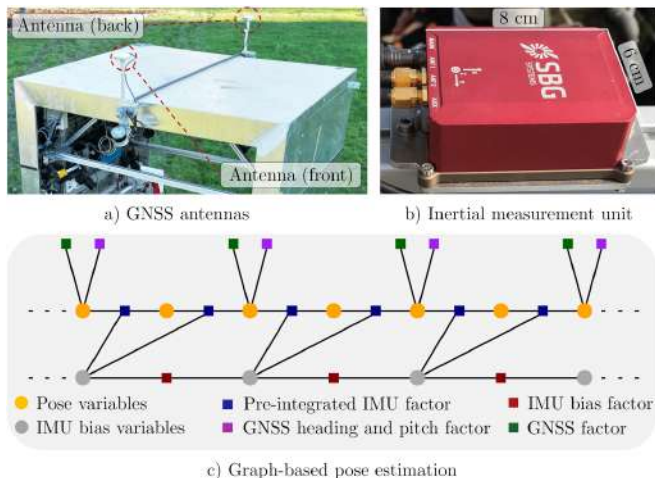


Fig. 3: UGV georeferencing. a) Antennas for global positioning with centimeter accuracy using RTK GNSS. b) Inertial navigation system with integrated GNSS receiver. c) Factor-graph-based pose estimation using GNSS and IMU data.

drive for storing the captured data. They also have ROS interfaces for connecting to and triggering the cameras and laser. Processing of the data for 3D reconstruction is done offline, as it usually takes longer than the capturing session. A foldable table holds a laptop while working in the field and motorized curtains on the front and rear openings reduce excessive sunlight inside the robot while capturing plant data. The robot's 48 V lithium battery is transformed to the power needed by our sensor and compute systems.

B. Georeferencing System

For robot navigation and spatio-temporal mapping, we need a global position accuracy on the order of a single plant, i.e. a few centimeters. To achieve this, we equip the platform with an Inertial Navigation System from SBG Systems, which includes a dual-antenna multi-frequency GNSS receiver and an inertial measurement unit (IMU) as shown in Fig. 3.

1) *RTK GNSS Hardware*: Real-time kinematic (RTK) GNSS positioning with centimeter accuracy requires a 4G internet connection streaming the necessary data from the reference station network provider SAPOS NRW. Additionally, the receiver computes the GNSS baseline between the front and rear antennas, which provides an absolute estimate of the UGV's heading and pitch. We use both the position and angle observations within the pose estimation as presented in section III-B.2. The MEMS-based industrial-grade IMU shown in Fig. 3b is installed under the robot's roof in the back. It measures in 3D accelerations and angular velocities with a sampling rate of 100 Hz. The GNSS and IMU data are streamed to an industrial-grade computer that we also use for sensor data synchronization with the laser scanning data. We use the PPS (pulse per second) signal from the GNSS receiver to assign GPS time stamps to the laser scanner measurements and link those to the estimated pose data with high precision.

2) *Pose Estimation*: To compute the poses of the robot while driving in agricultural fields, a factor-graph-based approach is used [5]. In the factor graph shown in Fig. 3c, the parameters of the robot state (position and orientation) are represented as variable nodes. The factor nodes represent the non-linear functions of the sensor measurements and parameters of the variable nodes. They consist of the pre-integrated IMU factor, the GNSS factor, and the GNSS heading-and-pitch factor. The IMU factor takes into account the sensor measurements of acceleration and angular velocity by integrating them, thereby imposing constraints on the robot's change of position and orientation. The GNSS factor depends on the position measurements of the rear antenna and the heading-and-pitch factor incorporates the global orientation information of the GNSS receiver. Further variable nodes represent IMU bias parameters to compensate for its systematic errors. From the linearization of the factor nodes, a normal equation system can be established to define a quadratic cost function for global pose estimation. The iSAM (incremental smoothing and mapping) algorithm [6] is used to minimize this cost function, as implemented in the GTSAM⁵ library. This yields smoothed robot poses to georeference the sensor data of the robot's camera and laser system for crop reconstruction.

C. Laser Scanning System

Typical laser scanners, often used in robotic mapping systems, do not provide the measurement precision necessary for detailed 3D reconstruction of single plants and plant organs. We therefore decided on a 3D sensing technology, which is often employed for 3D inspection tasks in factory assembly lines. Here, triangulation-based laser line scanners allow for capturing sub-millimeter accurate 3D measurements of the objects as they move through the scanner's active area. In our system, we turn this concept around by moving the scanner with the robot to scan the plants.

1) *Laser-line Triangulation Sensors*: The kinematic laser scanning system shown in Fig. 2c consists of two laser triangulation scanners (LMI Gocator 2490) attached to robot side panels at a height of 1.2 m and a lateral offset of 1.4 m, measuring vertically towards the ground between the wheels. We use two scanners to reduce occlusions due to the crops themselves. The scanners are tilted at an angle of about 50° with respect to the side panels to capture lower plant structures. The sensors use a red laser (600 nm) diode to project a laser line into the plot (see Fig. 2c right), which is reflected to the sensor's 2D CCD array. Based on the known geometry between the center of the projector and the lens focusing on the CCD, the distance for each point along the line is triangulated, yielding 2D points in the sensor coordinate system. For more information on the methodology of laser triangulation sensors, we refer to [7]. The measurement distance ranges from 390 to 2000 mm with a repeatability of 12 μm. Each laser line consists of

⁵GTSAM: Factor graphs for Sensor Fusion in Robotics <https://gtsam.org>

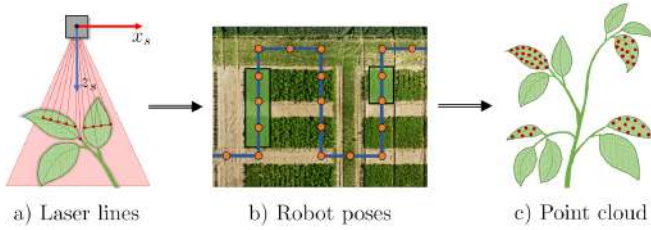


Fig. 4: Kinematic laser scanning pipeline for 3D point cloud creation. a) 2D laser lines given in the sensor frame. b) Robot poses known from factor graph optimization. c) Generated 3D point cloud. See eq. 1.

1920 points, resulting in point-to-point distances along the laser line of about 0.5 mm at a depth of 1 m. Important sensor settings are the exposure time Δt and the scan rate r . The exposure time needs to be adjusted to obtain proper reflections from the crop surfaces without over-exposure. The scan rate, together with the driving speed, determines the line distances along the driving direction. An exposure time of 1200 μ s and a scan rate of 200 Hz show reliable scans of crop structures in terms of spatial resolution and reflectivity. At a robot speed of 10 cm/s and a scan rate of 200 Hz the line spacing is 0.5 mm.

2) *Kinematic Laser Scanning*: For 3D crop point cloud creation, we georeference the 2D laser line measurements from the scanners using the globally optimized robot poses, estimated by the georeferencing system. This kinematic laser scanning pipeline is shown in Fig. 4 with the corresponding equation (1). The line distance measurements of the laser triangulation scanner are given in the sensor coordinate frame (s), only populating the XZ-plane (Fig. 4a). In the first step, the measurements are transformed into the body frame (b) of the robot, which coincides with the IMU sensor frame, by applying the system calibration parameters. These parameters describe the orientation \mathbf{R}_s^b and translation $[\Delta x, \Delta y, \Delta z]$ between the scanners and the body frame. They have been estimated using a plane-based calibration approach, as described in [8].

$$\begin{bmatrix} x_g \\ y_g \\ z_g \end{bmatrix} = \begin{bmatrix} p_x \\ p_y \\ p_z \end{bmatrix}_g + \mathbf{R}_b^g \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} + \mathbf{R}_s^b \begin{bmatrix} x_s \\ 0 \\ z_s \end{bmatrix} \quad (1)$$

The robot poses (Fig. 4b), describe the position $[p_x, p_y, p_z]$ and orientation \mathbf{R}_b^g of the robot with respect to the global coordinate system g . Since the poses are needed at the exact times of the laser scan measurements, we accurately tag the time of the laser scans in the same time frame as the poses and then performed a cubic interpolation of the poses with respect to the laser time. The application of this direct georeferencing procedure to the left and the right scanner data leads to the creation of a globally consistent point cloud, as seen in Fig. 4c. For more details on the georeferencing of 2D laser scanner measurements, we refer to [9].

D. Color Camera System

To capture individual plants in the center of plots from all sides, the UGV has been equipped with a dome of high-resolution color cameras.

1) *Camera Hardware*: We use 20 Nikon Z7 cameras arranged around the center of the UGV, as shown in Fig. 2b. We use the variable focal length Nikkor Z 24-70mm lens and capture images at a resolution of 8256 \times 5504 pixels. The zoom of each camera is manually adjusted depending on the growth stage of the plant to maximize the effective resolution on the plant, which is in the range of 10-30 pixels/mm. The cameras are attached to the robot frame with adjustable mounts for fast on-field modifications. We distribute the cameras on a hemisphere of approximately 1.5 m radius. We also experimented with placing the cameras in stereo pairs with short baselines but found that a roughly uniform distribution works better for plant reconstruction as it minimizes possible occlusions or under-constrained areas. Diffuse lighting is provided by multiple LED panels mounted in the robot's enclosure.

2) *Camera Triggering*: All cameras are connected to an onboard PC via USB. We interface with the cameras using the libgphoto2 library⁶ which allows to trigger all cameras synchronously. Various parameters such as aperture, ISO, and shutter speed can be controlled. This interface is made available to ROS. Before triggering the cameras, we send a signal to focus all cameras on a central patch of the image where we expect the plant of interest to be. After focusing, we send a capture signal to all cameras in parallel such that the images captured are as time-synchronized as possible. In our experiments, we observed that the maximum delay between camera captures is below 33 ms which is precise enough for our use case, as the UGV is stationary during image capture. We store both JPEG and RAW formats to allow for flexible post-processing.

3) *Camera Auto-exposure and Calibration*: During in-field capture, we want to ensure that all cameras have the same ISO, aperture, and shutter speed to obtain consistent exposures. However, it is impossible to use fixed parameters because the lighting conditions in the field are constantly changing due to the weather and passing clouds. To address this issue, we developed a method for multi-camera auto-exposure. At a frequency of 5 Hz, all cameras capture a low-resolution 800 \times 600 image which is transferred to the onboard PC. Using all 20 images, we compute a gray-scale histogram of eight bins. We are interested in reducing the number of pixels that saturate to either too dark values (gathered in the first bin of the histogram) or too bright values (last bin), thus ensuring that the image is correctly exposed with most of the values falling within the measurable range of the camera. For this, we start with initial values for ISO, aperture, and shutter speed and also define reasonable minimum and maximum values for them. We follow a simple but robust algorithm to modify the camera parameters:

⁶H. U. Niedermann and H. Figueire: GPhoto2 <http://www.gphoto.org>

- 1) While the first and last bin of the histogram are not sufficiently equal, we increase or decrease the ISO of all cameras by 100 such that their values are brought closer together.
- 2) If we reach the acceptable ISO limit, we switch to modifying the aperture of all cameras by adjusting the F-stop.
- 3) If we reach the acceptable F-stop limit, we are left with modifying the shutter speed in increments of 5 ms until the image becomes correctly exposed.

We choose this priority list based on the expected effect of each parameter on the final image. Increasing the ISO can cause noise grain in the images. However, this can be alleviated since we are fusing information from multiple cameras. Therefore changing the ISO will have the highest priority since it's the easiest effect to correct. Increasing the aperture of the camera can result in a shallow depth of field and a blurred background. However, since the plant in the center of the image is expected to stay mostly sharp, the final reconstruction is usually unaffected by the depth of field. Lastly, increasing the shutter speed can cause severe motion-blurring artifacts which are difficult to correct. Therefore, we place this parameter last in our list of priorities and increase it only if the other two parameters fail to achieve a correct exposure. We initialize all cameras with an ISO of 400, an f/14 aperture, and 10 ms shutter speed and let our algorithm dynamically modify them based on the weather conditions. To obtain the camera poses, it is not sufficient to rely on offline calibration methods as the robot is non-rigid and deforms while driving, causing the relative camera poses to change. To address this issue, we estimate the camera poses using Colmap [10] after each multi-view image capture. We initialize the camera poses with a reasonable prior for the location and orientation and perform bundle adjustment to estimate both intrinsic and extrinsic parameters.

4) *Image Capture GUI*: During robot operation in the field, it is often necessary to inspect the captured images and to manually trigger the cameras. For this, we use the Foxglove GUI⁷. Foxglove enables us to remotely connect to the onboard PC to visualize selected ROS topics. We show in real time the captured low-resolution 800×600 images from all cameras. Foxglove can be used either as a standalone app or as a web application which extends its usage to environments that don't have a ROS installation.

5) *3D Plant Reconstruction using Neural Implicit Surfaces*: For the 3D reconstruction of plants given multiple 2D RGB images from different viewpoints, we take inspiration from recent neural implicit models [1], [11]–[13]. This family of models represents a 3D scene with a neural network that encodes both color and geometry. While the color is modeled as a function of position and viewing direction, geometry is modeled as either a signed-distance function (SDF) or an occupancy field. The neural networks are trained using volumetric rendering so that the renderings match the captured images as closely as possible. We rely on neural

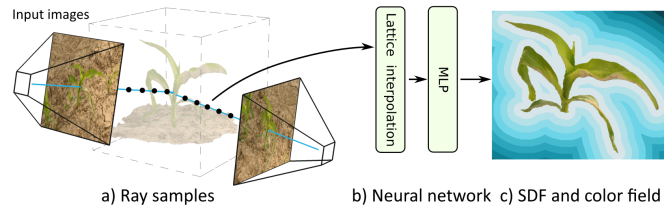


Fig. 5: 3D plant reconstruction from multiple 2D RGB images through volumetric rendering. We follow the approach of PermutoSDF [1] and shoot rays through every pixel of the 20 cameras (a). 3D point samples on the ray are passed through a neural network (b) to output both geometries as a signed-distance function (SDF) and a color field (c). The network is trained so that the resulting rendered images match the given captured images.

implicit models for reconstructing plants due to their ability to recover very fine detail and complete geometry even in occluded areas. We choose the recent PermutoSDF [1] model which optimizes an SDF and a color field using local features embedded in a permutohedral lattice. Fig. 5 gives an overview of the PermutoSDF pipeline. We refer to Rosu and Behnke [1] for more details. While PermutoSDF obtains accurate models, visualizing them requires performing volumetric rendering or sphere tracing. This is both expensive and requires specialized software solutions. To address this issue, we extract a textured mesh from the trained PermutoSDF model that can be visualized in any 3D package. For our visualizations, we use EasyPBR [14]. To obtain the mesh, we first run Marching Cubes [15] on the reconstructed signed-distance field. Secondly, we decimate the mesh to 300 k faces using QSlim [16] to reduce triangle density without sacrificing geometric quality. We rely on texture mapping to obtain high-resolution color. For this, the mesh is UV-unwrapped using the automatic unwrapping tool from Blender⁸ which establishes a UV-parameterization. We use a 4000×4000 RGB texture on the mesh and set each texel to the average color of all cameras that have an unobstructed view of the texel:

$$\mathbf{x}_i = \sum_{j=1}^{20} v_i^j \cdot \mathbf{I}_j(P_j(\mathbf{u}_i)) \quad (2)$$

where \mathbf{x}_i is the value of i -th texel, P projects a texel position \mathbf{u}_i into the j -th camera coordinate system, \mathbf{I}_j represents the j -th camera image, and v_i^j is a binary visibility indicator that is 1 if texel i is visible in camera j and 0 otherwise. Reconstructing plants using PermutoSDF is done offline on a computer with an RTX 3090 GPU and takes approximately 30 minutes per plant, followed by 10 minutes for mesh extraction and texture creation. We note that PermutoSDF can also run on less powerful GPUs by slightly downsampling the images.

⁷Foxglove Technologies Inc.: Foxglove <https://foxglove.dev>

⁸Blender Organization: Blender <https://www.blender.org>

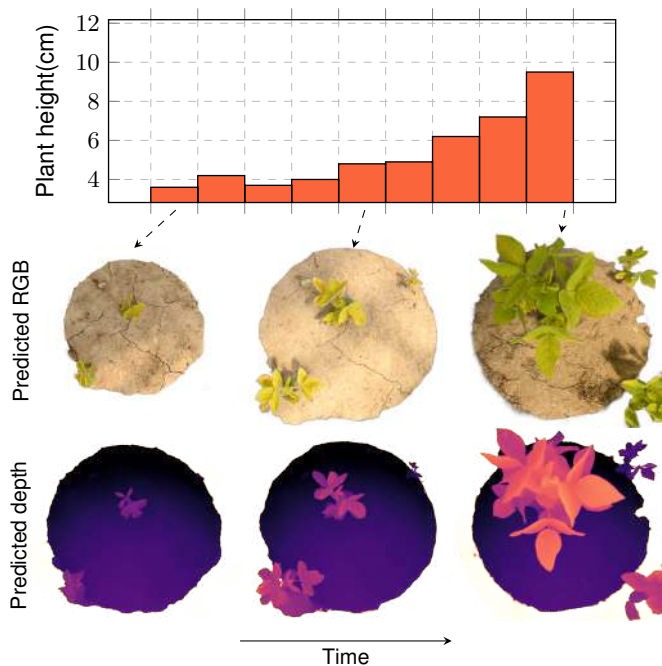


Fig. 6: Image-based 3D reconstruction over multiple days. The same plant is captured over the course of nine days and textured meshes are reconstructed for each individual day. The computed plant height is shown over time and the predicted RGB and depth from the PermutoSDF network are shown for three days.

IV. DATASETS AND EXPERIMENTS

We performed several experiments to validate the proposed plant phenotyping methods. They include field experiments with both laser-based and camera-based captures of plots and individual plants, respectively. We extract phenotypic traits from the 3D reconstructions and compare the precision and completeness of the 3D reconstructions from the cameras and the laser scanners.

A. In-field Image Capture and Laser Scanning

Field experiments were performed on the PhenoRob Central Experiment field at Campus Klein-Altendorf in Germany which features various crops and species that allow for a wide variety of phenotypic traits to be captured. We focused on measuring sugar beet, maize, and soybean and acquired laser and image data with our robot platform for a total of 14 days spanning from May 13 until August 2nd, 2022. This covers the bulk of the vegetation period, from the initial sprouting of the plants until the plants became too large to fit inside the robot. With the multi-camera system, we captured images from approximately 30 plants from each species. This resulted in a dataset of ≈ 2.3 TB consisting of 24 046 images. With the laser system, we captured a large dataset on August 2nd, 2022 including scans of multiple crops rows of sugar beet, soybean, and wheat leading to ≈ 100 GB high-resolution point cloud data. In terms of the time taken to measure and reconstruct 3D using the laser system, we need about 30s to collect data from one plot

(1.5 m \times 3 m) and a few minutes to post-process and create the point cloud on a standard office computer at a driving speed of 10 cm / s and a scanner rate of 200 Hz. For the camera system, we stop on top of each plot to trigger the cameras which takes about 2s followed by a copy to disk which is done in the background. We then reconstruct the scene using the 20 images on a computer with an RTX 3090 which takes about 30 min. In addition to the images and laser scan data, we also recorded the position and orientation of the robot's georeferencing system which enables us to embed our reconstructions within other datasets that were captured in the same field, as shown in Fig. 7 and Fig. 8.

B. 3D Reconstructions of Individual Plants and Plots

Fig. 7 and Fig. 8 show camera-based and laser-based 3D reconstructions of plants in the field, respectively. Both types of reconstructions are overlaid with UAV images, highlighting the highly accurate georeferencing. The camera-based reconstructions provide detailed textured meshes of the measured individual plants. The output of the laser scanning system is a high-resolution 3D point cloud of entire plots. In the soybean and sugar beet plots, single leaves are visible. Within the wheat plot, even the ears and the stems are recognizable. Fig. 6 shows image-based reconstructions of the same soybean plant over multiple days and its estimated height. Analyzing plant growth is needed for growth models and allows to identify potential plant stressors that can stunt growth. Therefore, 4D (3D space + time) reconstruction is of great interest to breeders and plant scientists. Structural analysis and tracking of plant organs such as [17] registers the plant temporally and provides insight into how the structure of the plant changes as it grows.

C. Precision and Completeness

To analyze the reconstruction quality of the two approaches in terms of precision and completeness, we generate 3D reference point clouds of two distinct plants using a high-precision scanning system in a lab. We capture and scan the same plants using cameras and laser sensors of the robot system in an outdoor environment, mimicking the in-field situation, and compare the reconstruction results with the lab reference. Fig. 10b shows the lab scanner. It consists of a measuring arm and a laser triangulation sensor commonly used in industrial applications. The sensor poses are calculated from joint encoder measurements by forward kinematics and used to reference the laser measurements to create a high-resolution 3D reference point cloud with a point accuracy below 1 mm. This measurement system was used to create the popular Pheno4D dataset [18]. For this experiment, we chose two semi-rigid plants to minimize plant movements while scanning and between the scans enabling a proper comparison of the reconstructions. Fig. 10c shows the chosen Hen-and-chick (*Echeveria Pelusida*) and the Lamb's tail (*Chiastophyllum oppositifolium*) plants. We obtain 3D textured meshes from the captured images and 3D point clouds from the laser measurements. We call the standard deviation of the distances between the reconstructed mesh

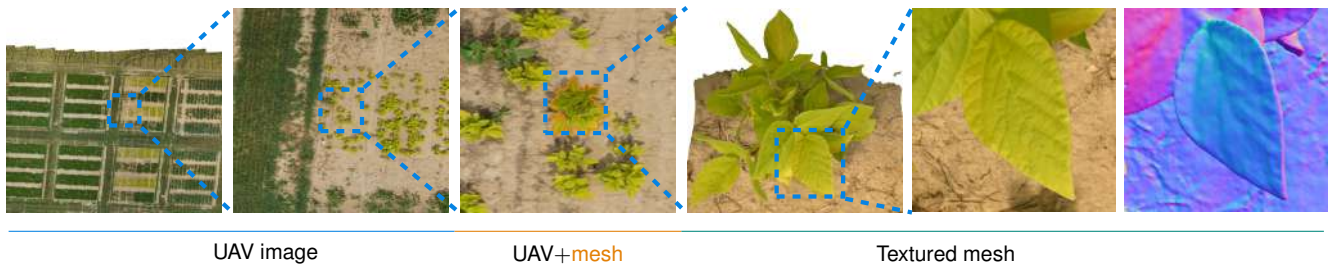


Fig. 7: Georeferencing of the 3D textured meshes reconstructed from multi-camera images. Our field robot globally localizes with RTK while capturing images. This allows us to embed our highly detailed single-plant reconstructions into an ortho mosaic map captured by a UAV.

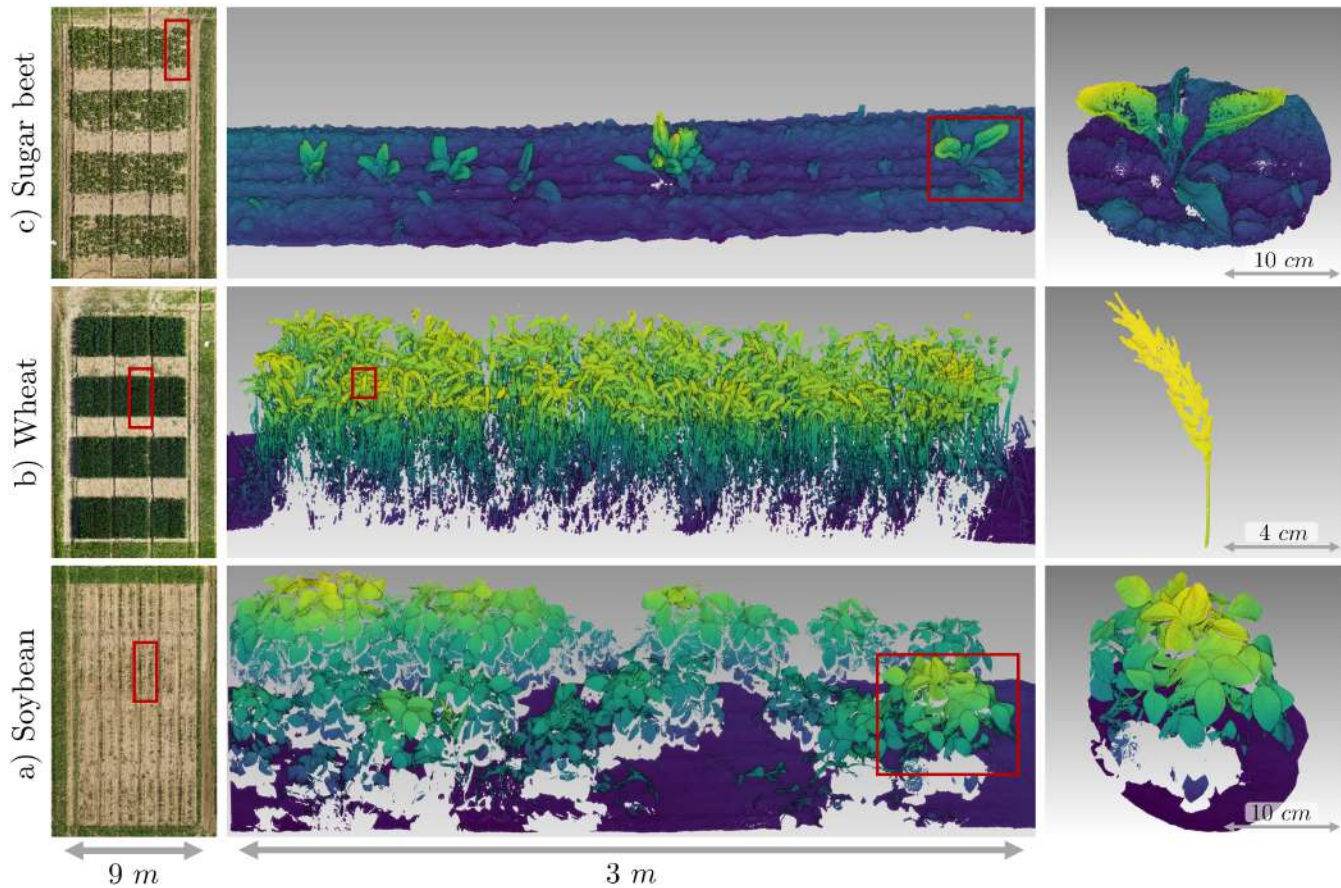


Fig. 8: Point clouds created with the laser scanning system of the robot while driving in the field. a)-c) Soybean, wheat, and sugar beet in the UAV image of the field (left), on plot level (middle), and single plant level (right).

or point cloud and the reference point cloud 'precision' and provide it as a measure of the reconstruction quality. In the absence of georeferencing information about the reference cloud, we first register the reconstruction results with the reference using the ICP (iterative closest point) algorithm. We then calculate the distances using the multi-scale model-to-model (M3C2) algorithm [19]. M3C2 uses the surface normal vectors of the reference point cloud to build point correspondences at averaged core points and computes signed distances between the point clouds. For detailed information on the M3C2 method, see [19]. For plant 2, we additionally extract leaves and perform an ICP fine registration for each

of the five leaves individually. The standard deviation of the M3C2 distances is used as a value for the reconstruction precision of the robot's two 3D capturing systems. To evaluate the completeness of the reconstruction, we compare the leaf area of the reconstruction with the ground truth. To obtain the ground truth leaf area, we use the reference point cloud and perform surface reconstruction using the ball-pivoting algorithm [20] which generates a 3D triangle mesh. The areas of all triangles are summed up to estimate the total leaf area. Since the output of the robot's laser scanning system is also a point cloud, we repeat this pipeline for these scans as well. We compute the differences in leaf area for the

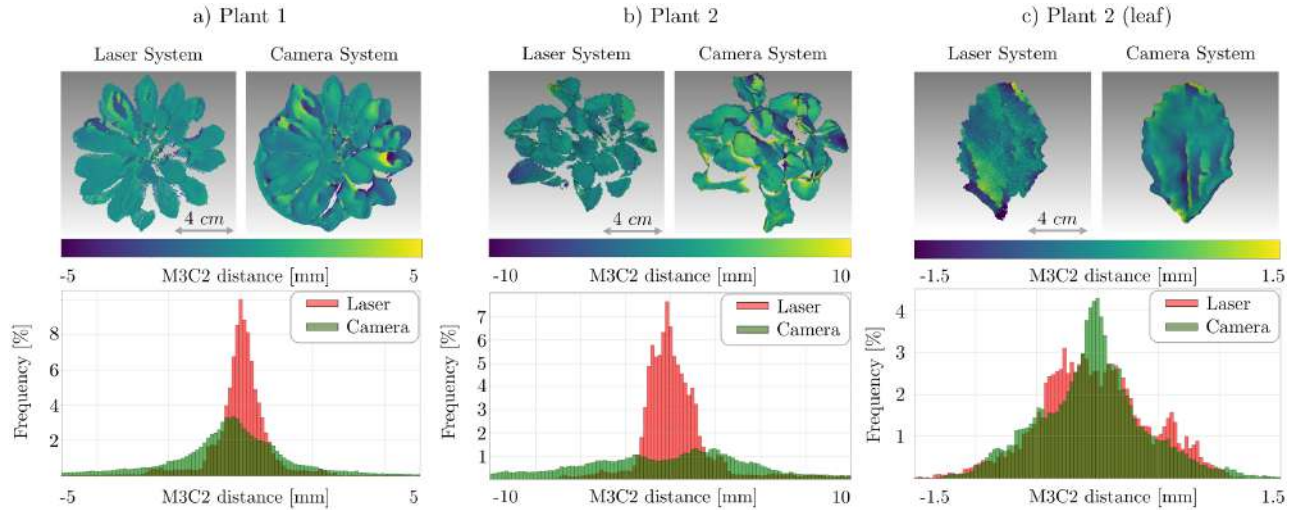


Fig. 9: M3C2 comparison between the reference measuring arm scan and the 3D reconstructions of the camera and laser system for the two plants of Fig. 10a. a), b) M3C2 distances and their histograms of Plant 1 and Plant 2 for laser and camera system. c) M3C2 distances and histograms for an example leaf of plant 2.

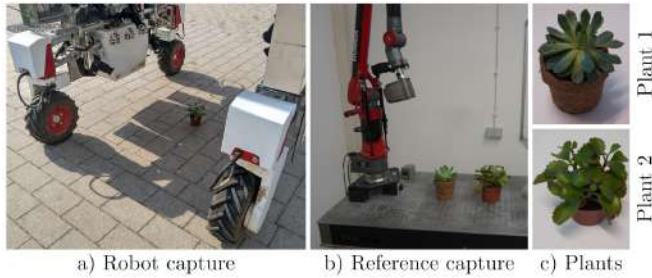


Fig. 10: Measurements for precision and completeness analysis. a) Plants 1 and 2 are captured with the robot’s camera and laser system. b) Reference capture using a precise scanning system in the lab. c) Plant 1: Hen-and-chick (*Echeveria Pelusida*), Plant 2: Lamb’s tail (*Chiastophyllum oppositifolium*).

laser and camera systems relative to the ground truth. Fig. 9 shows the distribution of the M3C2 distances of the laser and the camera-based reconstruction for the two plants and one example leaf of the Lamb’s tail plant. We observe that the reconstruction from the camera system exhibits a long tail distribution compared to the laser which is more peaked around zero and therefore more precise. The precision of the single leaf reconstruction on the right-hand side of figure Fig. 9 is in general better due to the separate leaf-level fine registration before the comparison, since the sugar beet leaf comparison includes an additional fine ICP registration the deviations from the reference are smaller. Tab. I reports the reconstruction precision and the leaf area differences to the reference scans for five example leaves of plant 2. The precision of the laser system ranges from 0.17 mm (Leaf 5) up to 0.48 mm (Leaf 1) with a mean of 0.27 mm. In comparison, the precision of the image-based reconstruction ranges from 0.24 mm (Leaf 3) to 0.55 mm (Leaf 1) and a mean of 0.43 mm. The comparison w.r.t. leaf areas are shown

TABLE I: Mean M3C2 distances and leaf areas compared to the reference scans for the Lamb’s tail plant. The means of the leaf area comparisons in the last row are average absolute percent differences.

Leaf	σ_{M3C2} [mm]		Differences of Leaf Area		
	Laser	Camera	Laser [%]	Camera [%]	Ref. [cm^2]
1	0.48	0.55	-3.3%	4.1%	23.78
2	0.18	0.33	-6.2%	7.8%	8.65
3	0.32	0.35	1.73%	-3.6%	9.38
4	0.22	0.67	-6.3%	-1.2%	13.18
5	0.17	0.24	-14.9%	-2.4%	8.54
mean	0.27	0.43	6.5%	1.9%	

in the right part of Tab. I. The leaf areas estimated from the laser scans deviate from the reference from -14.9% (Leaf 5) to 1.73% (Leaf 3) with an overall absolute mean difference of the leaves of 6.5%. The camera system shows leaf area differences between -3.6% (Leaf 3) and 7.8% (Leaf 2) and an absolute average difference to the reference of 1.9%. We observe that the laser system tends to underestimate the leaf area more than the camera system. This is mostly because the point cloud from the laser scanning system can contain holes caused by self-occlusion by the plant itself. In contrast, the camera system uses more viewing angles for the reconstruction, resulting in more complete plant structures. The advantage of the camera system is that it provides a textured mesh with a high-resolution color texture, whereas the laser provides only 3D information and laser intensity. The results show that the laser reconstructions are more accurate but less complete than the camera system. However, the camera system struggles to recover detail in reflective areas such as raindrops on the plant while the laser has problems with bright sunlight conditions and wind-induced plant movement during the scans. The two systems therefore complement each other and have different advantages and

limitations.

V. SUMMARY AND OUTLOOK

We presented a ground robot for in-field phenotyping. We developed an integrated system equipped with laser scanners, cameras, and accurate pose estimation capable of reconstructing 3D models of crops in the field with high throughput, resolution, and precision. Using the reconstructed plants over multiple days of field acquisition, we show that it is possible to obtain georeferenced 3D models that offer enough resolution to reconstruct single plants and plant organs to derive phenotypic traits such as leaf area, leaf angles, and plant height. We believe that our system can help automate various phenotyping tasks that currently require significant manual work in the field. In the future, we plan to extend our methods by recovering structural plant models, performing automated point cloud segmentation for crop organ detection, and registering plants over time for a better understanding of the growth processes.

ACKNOWLEDGEMENT

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC 2070—390732324.

REFERENCES

- [1] R. A. Rosu and S. Behnke, "PermutoSDF: Fast multi-view reconstruction with implicit surfaces using permutohedral lattices," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.
- [2] N. Chebrolu, P. Lottes, A. Schaefer, W. Winterhalter, W. Burgard, and C. Stachniss, "Agricultural robot dataset for plant classification, localization and mapping on sugar beet fields," *The International Journal of Robotics Research*, vol. 36, no. 10, pp. 1045–1052, 2017.
- [3] A. Ahmadi, M. Halstead, and C. McCool, "Bonnbot-i: A precise weed management and crop monitoring platform," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 9202–9209.
- [4] L. Grimstad and P. J. From, "The Thorvald II agricultural robotic system," *Robotics*, vol. 6, no. 4, p. 24, 2017.
- [5] F. Dellaert, "Factor graphs and GTSAM: A hands-on introduction," Georgia Institute of Technology, Tech. Rep., 2012.
- [6] V. Indelman, S. Williams, M. Kaess, and F. Dellaert, "Information fusion in navigation systems via factor graph based incremental smoothing," *Robotics and Autonomous Systems*, vol. 61, no. 8, pp. 721–738, 2013.
- [7] M. A.-B. Ebrahim, "3D laser scanners' techniques overview," *International Journal of Scientific Research (IJSC)*, vol. 4, no. 10, pp. 323–331, 2015.
- [8] E. Heinz, C. Holst, H. Kuhlmann, and L. Klingbeil, "Design and evaluation of a permanently installed plane-based calibration field for mobile laser scanning systems," *Remote Sensing*, vol. 12, no. 3, p. 555, 2020.
- [9] F. Esser, L. Klingbeil, L. Zabawa, and H. Kuhlmann, "Quality analysis of a high-precision kinematic laser scanning system for the use of spatio-temporal plant and organ-level phenotyping in the field," *Remote Sensing*, vol. 15(4), 2023.
- [10] J. L. Schönberger and J.-M. Frahm, "Structure-from-motion revisited," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [11] P. Wang, L. Liu, Y. Liu, C. Theobalt, T. Komura, and W. Wang, "NeuS: Learning neural implicit surfaces by volume rendering for multi-view reconstruction," in *Advances in Neural Information Processing Systems 34 (NeurIPS)*, 2021, pp. 27 171–27 183.
- [12] L. Yariv, J. Gu, Y. Kasten, and Y. Lipman, "Volume rendering of neural implicit surfaces," in *Advances in Neural Information Processing Systems 34 (NeurIPS)*, 2021, pp. 4805–4815.
- [13] M. Oechsle, S. Peng, and A. Geiger, "UNISURF: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction," in *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 5589–5599.
- [14] R. A. Rosu and S. Behnke, "EasyPBR: A lightweight physically-based renderer," in *16th International Conference on Computer Graphics Theory and Applications (GRAPP)*, 2021.
- [15] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm," in *14th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, vol. 21, no. 4, 1987, pp. 163–169.
- [16] M. Garland and P. S. Heckbert, "Surface simplification using quadric error metrics," in *24th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, 1997, pp. 209–216.
- [17] N. Chebrolu, F. Magistri, T. Läbe, and C. Stachniss, "Registration of spatio-temporal point clouds of plants for phenotyping," *PLoS ONE*, vol. 16, no. 2, p. e0247243, 2021.
- [18] D. Schunck, F. Magistri, R. A. Rosu, A. Cornelißen, N. Chebrolu, S. Paulus, J. Léon, S. Behnke, C. Stachniss, H. Kuhlmann, and L. Klingbeil, "Pheno4D: A spatio-temporal dataset of maize and tomato plant point clouds for phenotyping and advanced plant analysis," *PLoS ONE*, vol. 16, no. 8, p. e0256340, 2021.
- [19] D. Lague, N. Brodu, and J. Leroux, "Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (NZ)," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 82, pp. 10–26, 2013.
- [20] F. Bernardini, J. Mittleman, H. Rushmeier, C. Silva, and G. Taubin, "The ball-pivoting algorithm for surface reconstruction," *IEEE Transactions on Visualization and Computer Graphics*, vol. 5, no. 4, pp. 349–359, 1999.