40th Anniversary of the IEEE International Conference on Robotics and Automation (ICRA@40),
Rotterdam, Netherlands, September 2024.

Automated Leaf-Level Inspection of Crops in Agricultural Fields by Combining Aerial and Ground Robot Systems

Felix Esser[1,*], Elias Marks[1], Federico Magistri[1], Jan Weyler[1], Simon Bultmann[2],
Tobias Zaenker[3], Alireza Ahmadi[4], Michael Schreiber[2], Heiner Kuhlmann[1], Chris McCool[4], Marija Popović[1],
Cyrill Stachniss[1], Sven Behnke[2], Maren Bennewitz[3], Lasse Klingbeil[1]

I. Introduction

Robotic systems play a major role for realizing the vision of sustainable crop production [1]. While Unmanned Aerial Vehicles (UAVs) are increasingly used to monitor the health status of agricultural fields using sensors like RGB cameras, multi-spectral cameras, and LiDAR, it is often still necessary to literally walk into the field to do close-up inspections of individual plants or even leaves for the detection of diseases or nutrient deficiencies in early stages of plant growth.

The main contribution of this work is demonstrating the integration of aerial and ground robotic systems to automate plant inspection processes, thereby enhancing efficiency in field monitoring tasks. A UAV identifies areas or plants of interest from higher altitudes whose coordinates are transferred to an Unmanned Ground Vehicle (UGV). The UGV then autonomously navigates to the specified location and a mounted robotic arm with five cameras captures close-range images with automatically optimized camera poses. The result is a high-resolution 3D reconstruction suitable for further plant analysis. This integration of UAV and UGV allows us to 'zoom in' on any coordinate in the field with a few centimeters of accuracy.

II. Aerial Inspection of Crop Fields

An attractive way to observe crop fields or breeding plots at a larger scale is the use of UAVs. Fig. 1 left shows our UAV that is equipped with a high-resolution Phase One iXM 100 MP camera with an 80 mm lens. The UAV can cover 6 ha/h at 1 mm/px ground sampling distance [2].

The UAV is used to obtain accurate, geo-referenced orthophotos. To this end, we use a flight pattern to capture images with an overlap of 75% in both height and width. Each of these images is tagged with a centimeter-accurate RTK-GNSS 3D position. These positions are used as initial guess in the structure from motion procedure [3], which outputs a high-resolution orthophoto.

From the orthophoto, we compute a semantic mask differentiating between crops and weeds using the approach proposed by Weyler et al. [4] as it generalizes well to unseen crop fields. Its semantic perception network consists of one encoder and two decoders based on the ERFNet [5] architecture. The encoder produces highly descriptive features from the input pixels and the two decoders predict the semantics and the instances, respectively. This method can be trained on partial annotations, where only a subset of pixels is annotated, exploiting a set of loss functions based on common operators used in physics to analyze the behavior of vector fields, i.e., divergence and curl.

Fig. 1. Left: UAV used to create high-resolution semantic orthophotos in the field. Right: UGV and robotic camera arm setup for crop inspection.

To provide a navigation map for the UGV, we detect crop plots and rows. To segment the plots, we perform unsupervised clustering with HDBSCAN [6] on the crop pixels. We find the crop rows by performing a Hough transform on the crop pixels, see Fig. 2 left. The rows are then used to create a traversability map for the UGV by defining occupied areas around the crop row lines and intersecting the resulting polygons with the crop plot polygons coming from the plant sowing machine.

We extract coordinates of interest from the semantic orthophoto depending on the downstream tasks. For example, we can automatically extract coordinates corresponding to high weed pressure for precise weeding or corresponding to slow-growing regions to identify deficits in nutrition.

III. Close-Up Inspection From the Ground

Fig. 1 right shows our UGV that is based on the Thorvald II [7] platform. For detailed close-up plant inspections, we integrate a robotic arm with five high-resolution cameras. The UGV's dimensions are about two meters in height,

[1] University of Bonn, Institute of Geodesy and Geoinformation
[2] University of Bonn, Autonomous Intelligent Systems
[3] University of Bonn, Humanoid Robots Lab
[4] University of Bonn, Agricultural Robotics & Engineering
[*] Corresponding Author, E-Mail: esser@igg.uni-bonn.de

length, and width. Its U-shaped design allows plants to be observed with the camera arm even if crops are large. For autonomous navigation in the field, the UGV has four electric wheels powered by a lithium battery that can be controlled using a ROS (Robotic Operating System) interface. For more details about the UGV platform, we refer to Esser et al. [8]. The robot arm is a 7-DOF UFactory xArm7, mounted upside-down on aluminum profiles in the center of the UGVs U-shape. It can be moved side-ward on a linear axis between the active and the home position to avoid plant-destroying while the UGV is driving in the field. The arm moves a sensor setup consisting of five 20 MP Basler ace 2 Pro cameras [9].

A. Path Planning and Navigation in the Field

Autonomous navigation of the UGV in agricultural fields needs 1) precise online vehicle position and orientation estimation, 2) path planning based on a traversability map of the field to avoid destroying plants, and 3) controlling the UGVs electric motors to execute the planned trajectory.

For precise pose (position and orientation) estimation, the heading and 3D position data of a dual antenna RTK-GNSS receiver are combined with inertial measurements of an IMU using an extended Kalman Filter (EKF). The resulting pose is published to ROS navigation at 10 Hz.

For path planning to the waypoints of interest coming from the UAV orthophoto, we use a simple A* algorithm similar to [10], and the traversability map of the field that was created based on the semantics in the UAV orthophotos and the crop polygons of the sowing machine. The path planning outputs the shortest path from the current robot position and the next point of interest by minimizing the costs on the map.

To execute the path, the current robot EKF pose and the angle and distance to the next point of the planned path are used to calculate linear and angular velocities for the four motor controllers. After reaching a waypoint of interest, the UGV stops and sends a message to the five-camera robotic arm to closely inspect the plants at that position.

B. Viewpose Planning and Image-Based Reconstruction

After the UGV stops at the coordinate of interest, the arm moves from its home position to the active position in the center of the UGV using a linear axis. The arm is then moved to a set of pre-defined discovery viewposes, looking in the general direction of the plant of interest.

At each discovery pose, we identify the leaves to approach to take detailed close-up images using color segmentation and online multi-view stereo processing. The cameras' intrinsic and extrinsic calibration is obtained offline using Kalibr [11] and they are software-triggered to provide synchronized frame-sets. We employ a simple stereo block matching [12] between each camera pair of the central camera and one of the outside cameras, project the obtained depth images to 3D, and transform the resulting point clouds to the central camera coordinates. We fuse the point clouds, mask them with the green color obtained from the central camera image, and finally post-process them using voxel and outlier filters. We then segment planes that represent

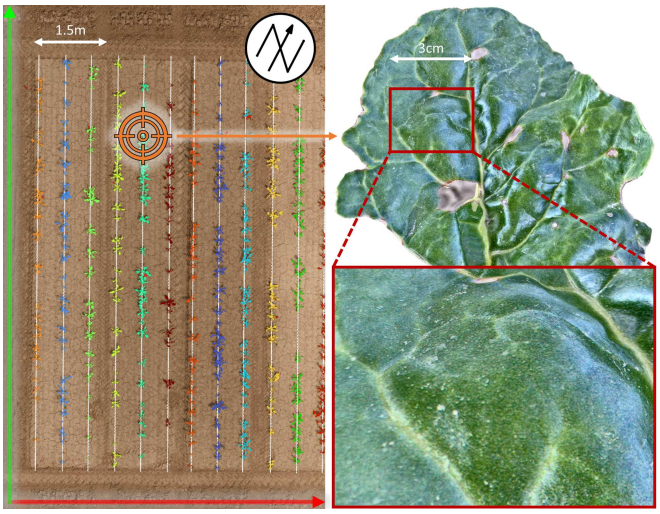

Fig. 2. Left: Georeferenced orthophoto with extracted crop rows and one example for an individual sugar beet plant of interest. Right: Colored high-resolution 3D reconstruction of a single sugar beet leaf created with the multi-camera robot arm of the UGV.

the leaves in the post-processed point cloud and calculate the camera pose to look at the largest perceived leaf along the plane's normal direction from the optimal focus distance. The arm then approaches the calculated pose.

This process can be iterated several times to further improve the capture pose. The obtained images are saved for the offline reconstruction of the high-resolution leaf model. To ensure that the arm moves safely within the UGV without collisions, we integrate our setup with MoveIt [13]. Our robot model includes a collision box to keep the robot arm and the cameras at a safe distance from both the ground and the robot's inner enclosure. To compute the joint configuration to reach a desired view pose, we utilize BioIK [14].

We obtain an accurate, dense leaf point cloud from the close-up images captured on the field using Colmap multi-view stereo [15] and obtain a mesh using Poisson surface reconstruction [16]. The mesh is cropped using the leaf's segmentation mask. The high-resolution color image from the central camera is then added to the mesh as texture, where the texture coordinates are obtained by projecting the vertex coordinates onto the camera image plane. The final, highly detailed leaf model is depicted in Fig. 2 right.

IV. Outlook

In this work, we demonstrated how a set of close-range cameras can autonomously observe a single plant in an agricultural field, selected in a map generated by a UAV.

With the same set of methods, it is also possible to monitor specific plants over longer periods, e.g. to understand diseases or nutrient deficiencies on the leaf-level or to perform selective weeding at single-plant scale.

ACKNOWLEDGMENT

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

REFERENCES

[1] H. Storm, S. J. Seidel, L. Klingbeil, F. Ewert, H. Vereecken, W. Amelung, S. Behnke, M. Bennewitz, J. Börner, T. Döring, J. Gall, A.-K. Mahlein, C. McCool, U. Rascher, S. Wrobel, A. Schnepf, C. Stachniss, and H. Kuhlmann, "Research priorities to leverage smart digital technologies for sustainable crop production," *European Journal of Agronomy*, vol. 156, p. 127178, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1161030124000996

[2] E. Marks, F. Magistri, and C. Stachniss, "Precise 3D reconstruction of plants from UAV imagery combining bundle adjustment and template matching," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2022.

[3] Agisoft LLC, "Agisoft metashape professional," https://www.agisoft.com/, accessed: April 2024.

[4] J. Weyler, T. Läbe, J. Behley, and C. Stachniss, "Panoptic segmentation with partial annotations for agricultural robots," *IEEE Robotics and Automation Letters*, vol. 9, no. 2, pp. 1660–1667, 2024.

[5] E. Romera, J. M. Álvarez, L. M. Bergasa, and R. Arroyo, "ERFNet: Efficient residual factorized ConvNet for real-time semantic segmentation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 1, pp. 263–272, 2018.

[6] L. McInnes, J. Healy, and S. Astels, "hdbscan: Hierarchical density based clustering," *Journal of Open Source Software*, vol. 2, no. 11, p. 205, 2017. [Online]. Available: https://doi.org/10.21105/joss.00205

[7] L. Grimstad and P. J. From, "The Thorvald II agricultural robotic system," *Robotics*, vol. 6, no. 4, p. 24, 2017.

[8] F. Esser, R. A. Rosu, A. Cornelißen, L. Klingbeil, H. Kuhlmann, and S. Behnke, "Field robot for high-throughput and high-resolution 3D plant phenotyping: Towards efficient and sustainable crop production," *IEEE Robotics and Automation Magazine*, vol. 30, no. 4, pp. 20–29, 2023.

[9] Basler, "Basler a2a4504-18umpro specifications," https://docs.baslerweb.com/a2a4504-18umpro, 2024, accessed: 2024-06-25.

[10] F. Duchoň, A. Babinec, M. Kajan, P. Beňo, M. Florek, T. Fico, and L. Jurišica, "Path planning with modified A Star algorithm for a mobile robot," *Procedia Engineering*, vol. 96, pp. 59–69, 2014.

[11] J. Rehder, J. Nikolic, T. Schneider, T. Hinzmann, and R. Siegwart, "Extending Kalibr: Calibrating the extrinsics of multiple IMUs and of individual axes," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 4304–4311.

[12] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, vol. 25, no. 11, pp. 120–125, 2000.

[13] S. Chitta, I. Sucan, and S. Cousins, "Moveit![ros topics]," *IEEE Robotics & Automation Magazine*, vol. 19, no. 1, pp. 18–19, 2012.

[14] P. Ruppel, N. Hendrich, S. Starke, and J. Zhang, "Cost functions to specify full-body motion and multi-goal manipulation tasks," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 3152–3159.

[15] J. L. Schönberger, E. Zheng, M. Pollefeys, and J.-M. Frahm, "Pixelwise view selection for unstructured multi-view stereo," in *European Conference on Computer Vision (ECCV)*, 2016, pp. 501–518.

[16] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson surface reconstruction," in *Fourth Eurographics Symposium on Geometry Processing (SGP)*, ser. ACM International Conference Proceeding Series, vol. 256. Eurographics Association, 2006, pp. 61–70.