Precise 3D Reconstruction of Plants from UAV Imagery Combining Bundle Adjustment and Template Matching

Elias Marks   Federico Magistri   Cyrill Stachniss

***Abstract*— Monitoring individual plants and computing precise 3D reconstructions is highly relevant for crop breeding. In the conventional breeding approach, humans measure phenotypic traits by hand, requiring substantial manual labor. This paper addresses precise 3D plant reconstructions in a crop field or breeding plot based on UAV imagery. We explicitly address the challenges resulting from the thin structures of leaves and naturally occurring self-occlusions. We combine photogrammetric bundle adjustment with a template-based matching approach and produce accurate 3D models that allow us to derive common, geometric traits used by breeders to phenotype plants. We provide a thorough experimental evaluation on commercially used sugar beet breeding plots to illustrate the capabilities of our method as well as its real world applicability.**

I. Introduction

Phenotyping refers to the task of measuring the observable characteristics or traits of an organism. It covers its physical form and structure, its developmental processes, and physiological properties. Plant phenotyping plays a vital role in crop production and especially plant breeding, where the appearance and performance of different plant varieties have to be monitored and assessed [10]. This monitoring process involves regularly measuring individual plants at several stages to support breeders and scientists in selecting plants that show desirable traits and using them for the following breeding generations. Traditionally, the in-field assessment is done by human inspection [11], which creates substantial manual work and limits the spatial and temporal throughput.

An attractive way to observe fields or breeding plots at a larger scale is the use of unmanned aerial vehicles (UAVs). A typical UAV can be equipped with a high-resolution camera and is able to cover multiple hectares of fields at high resolution (in our setup 6 ha/h at 1 mm/px ground sampling distance). The key advantage of UAVs is that acquiring data is fairly easy. Their downside, however, is that they provide only 2D image data, mainly in a top-down (nadir) view. Thus, performing 3D reconstruction of plants can be challenging, especially if required at high resolution. Furthermore, serious self-occlusions of the leaves pose great challenges to matching algorithms. The fact that leaves are thin structures additionally prevents the use of most volumetric environment representations commonly used in SLAM systems, such as voxel grids or TSDFs [13], [22], [27], [28], [31].

Fig. 1: Our UAV collecting images over commercial breeding plots of sugar beets (top row). By combining bundle adjustment (middle) and template matching, we can obtain precise 3D reconstructions (bottom right) of individual plants (bottom left). We also show a detailed view of a leaf reconstruction in the presence of occlusions.

This paper addresses the problem of building high-resolution 3D models of plants from UAV images. The models should be precise enough to compute common phenotypic traits such as individual leaves' length, width, or margin undulation. As a concrete application, we consider monitoring sugar beets, a value crop commonly grown in Germany and other European countries. We use data from real sugar beet breeding plots, grown and maintained by commercial breeding companies and the Federal Office of Plant Varieties in Germany. Fig. 1 shows our UAV during a flight mission and illustrates the results of the reconstruction approach presented in the paper.

The main contribution of this paper is a novel, template-based approach for the reconstruction of crops in agricultural fields and breeding. It combines photogrammetric bundle adjustment of canopy points with a technique that fits templates, which are well-suited to represent the thin structure of leaves, to the data. Our approach deforms these templates to match with the 3D points stemming from the

All authors are with the University of Bonn, Germany. Cyrill Stachniss is additionally with the Department of Engineering Science at the University of Oxford, UK.

This work has partially been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy, EXC-2070 – 390732324 – PhenoRob and by the Federal Ministry of Food and Agriculture (BMEL) based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support programme under funding no 28DK108B20 (RegisTer).

bundle adjustment using a gradient descent approach optimized to handle partial observations. Furthermore, we use least squares optimization to robustly pre-align the template by leveraging domain-specific constraints, such as the axis defined by plant center and leaf tip. We implemented and thoroughly evaluated our approach on real UAV images photographing sugar beet plants in agricultural fields and commercially used breeding plots. Our experiments suggest that our approach (i) is able to accurately reconstruct the leaf parts visible in the leaf point clouds, including fine details, (ii) reconstructs in a plausible way part of the leaves that not visibile due to occlusions, (iii) is well suited for reconstruction of partial plants in real field conditions, and (iv) and the obtained reconstructed meshes are correctly aligned to the corresponding leaf parts.

II. Related Work

In recent years, we have been witnessing an increase in agricultural robotics studies, with applications ranging from weed control [32] and ripeness estimation [12] to harvesting [15] and grasping [4]. However, phenotyping using mobile robots is still limited to basic traits such as average plant height and leaf area index. For instance, the works by Carlone et al. [5], Dong et al. [9], and Chebrolu et al. [6] estimate how the height of the plants changes over time by aligning point clouds of crop fields at different growth stages, using non-rigid registration techniques. Instead, in our work, we want to infer the 3D geometry of crops at a detailed leaf level and in the presence of occlusions by jointly deforming a template and estimating which leaf parts are missing in the observation. In this sense, our task can also be seen as a partial non-rigid registration problem. Non-rigid registration techniques can register scans with localized deformations in contrast to rigid registration techniques such as iterative closest point. One can divide these approaches into two categories. On one side, approaches that explicitly compute the data association between source and target point clouds [25], [33], with applications to monitor plant growth [7], [8], [17]. This class of problems has the drawback of an explicit data association step, which is an ill-posed problem in our setting. On the other side, approaches that cast the registration task as a probability density estimation problem [20], [21]. With these methods, however, it is not possible to consider prior knowledge about the global position of the template. Therefore, these approaches do not perform well on partially occluded scans.

Despite the research efforts, few studies address the ubiquitous challenge of occlusions in agricultural environments. At an image level, Blok et al. [3] estimates the pixel-wise size of broccoli heads in the presence of occlusions using a deep neural network. Some works also integrate prior knowledge of plant structures into 3D measurements. Binney et al. [2] fit cylinders to point clouds of trees to recover missing data. Ando et al. [1] proposed to estimate 2D shapes leaves by capturing and extracting deformations. However, the most similar work to ours is the one by Sodhi et al. [24], which addresses the problem of mapping plant sub-units called

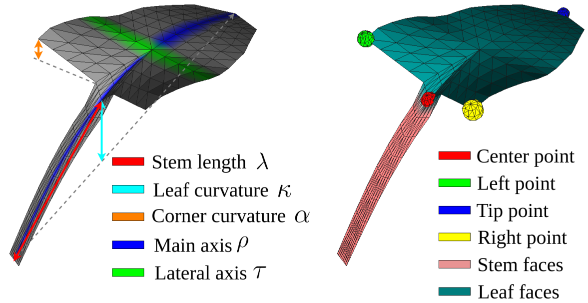

Fig. 2: We show the geometric parameters of our template λ, κ, α. The blue line represents the central leaf axis $\boldsymbol{\rho}$, while the green one is the lateral leaf axis $\boldsymbol{\tau}$ (left side). By encoding semantic information such as stem and leaf points as well as leaf keypoints (right side), we can compute fine-grained phenotypic traits.

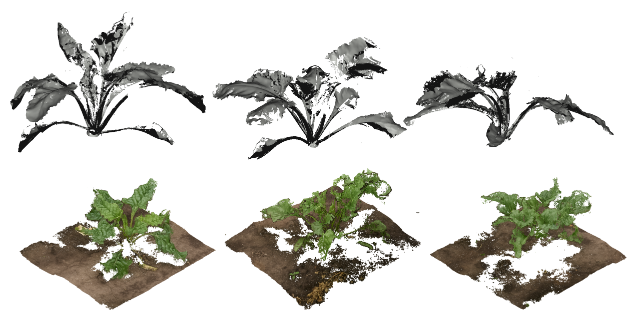

Fig. 3: A visual impression of the dataset collected in controlled environment (top) and the point cloud obtained from the bundle adjustment of real field images (bottom).

plant phytomers to their phenotype values. They sample 3D plant models from an underlying probability distribution, thus cast phenotyping as a search in the space of plant models. However, the results of their approach only model the basic leaf structure, omitting finer details such as the leaf blobbiness and margin undulation, which are prominent in certain crop varieties.

We exploited the idea of template matching in previous work [16] coupling differentiable rendering and registration techniques; however, our previous study was not able to solve the 3D reconstruction in occluded regions. We overcome this by introducing a parametric template model and explicitly estimating which parts of such template have no correspondence in the target observation.

III. Our Approach

Our leaf reconstruction approach works in 3D as this is needed for the computation of the plants' geometric features. However, UAVs commonly collect 2D image data in nadir view containing multiple plants. Hence, the first steps of our approach generate point clouds of the whole field and segment them into individual plants and leaves. We define a parametric leaf model M to reconstruct the leaves imaged by the (partial) scans $\mathcal{S} = \{\boldsymbol{s}_1, \dots, \boldsymbol{s}_N\}$ of points $\boldsymbol{s} \in \mathbb{R}^3$. We fit this parametric model to the points of a leaf scan to obtain a plausible estimate of the missing parts and to detect the different leaf parts. To further refine this, we then move the vertex positions by optimizing a reconstruction loss function.

A. From 2D Field Images to 3D Plant Point Clouds

We accomplish this by using bundle adjustment, which allows estimating the 3D pose of points in a scene that has been imaged from multiple viewpoints in a statistically optimal manner. See Triggs et al. [26] for an overview. To obtain dense and accurate point clouds, each part of the scene needs to be visible from different viewpoints. We, therefore, use a flight pattern to obtain images with an overlap of 75% in both height and width. Before starting the leaf reconstruction, we segment the point cloud generated by the bundle adjustment in individual plants and, for each of them, segment their leaves in separated instances. To achieve this, we use a recent work by Weyler [29], which extends prior work [30] towards semantic and instance segmentation of individual plant and its leaves on images. The obtained 2D masks can then be projected onto the 3D point clouds to segment individual leaves.

B. Leaf Model

We use a parametric leaf template M as a prior for the leaf reconstruction where the set of parameters $\Omega = (\lambda, \kappa, \alpha)$ describes the coarse deformation. In the following we refer to the (curved) line connecting the plant center to the leaf tip as the *central leaf axis* $\boldsymbol{\rho}$. We define the leaf shape as a triangular mesh composed of a set $\mathcal{V} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{209}\}$ with 209 vertices and a set $\mathcal{F} = \{f_1, \ldots, f_{352}\}$ containing 352 triangle faces lying in the xy-plane. We then model the stem length by spacing the stem vertices over the length λ. To model the leaf curvatures we define the vertical distance $d_{yi} = v_y - y_0$ of v_i from the leaf base, where v_y is the vertex y-component and y_0 is the y-component of the base. To obtain the principal leaf curvature, we increase the height of the vertices according to $\kappa \sin(d_{yni})$, where d_{yni} is d_{yi} normalized to $[0, \pi]$. Additionally, we define the distance $d_{xi} = dist(\rho, v_i)$ of v_i from the *central leaf axis*, where $dist$ is the point to line distance function.

By adding $\alpha \sin(d_{xni})$ to the z-component of each vertex, we obtain a deformation symmetric to $\boldsymbol{\rho}$, that represents the typical corner arching of sugar beet leaves. Here d_{xni} is d_{xi} normalized to $[0, \frac{\pi}{2}]$ to obtain maximum arching at the leaf corners. See Fig. 2 for a visual explanation of the template and its parameters.

C. Alignment and Optimization of the Template

The first step in the fitting process is the alignment of the template with the scan. We align the template by translating the stem base onto the plant center and then scale and rotate it to align the tip of the template with the leaf tip in the scan. We can predict the plant center with high accuracy by using the work of Weyler et al. [30]. We assume that the leaf tip is always present in the scan, as the characteristic growth of sugar beet plants leads to no self-occlusions that cover the tips. We then search for the scan point that lies furthest away from the plant center and define it as the leaf tip.

This leads us to the pose of the leaf up to a rotation of an angle θ around the central leaf axis $\boldsymbol{\rho}$. We find this parameter θ along with the deformation parameters of the template M by optimization, by minimizing the distances of the scan points to the closest vertex of the template. To solve this, we limit the parameters to predefined ranges and apply the Levenberg-Marquardt algorithm [19].

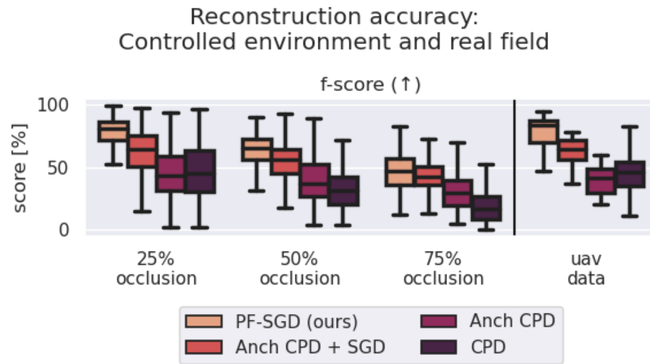

Fig. 4: F-score for both datasets, our approach yields better reconstruction accuracy, handling high percentage of missing data.

D. Partiality Filtered Stochastic Gradient Descent

The previous step delivers a rough alignment of the template to the scan, allowing us to infer which parts of the leaf are likely to be observed and which likely are not. To get a reconstruction that resembles the actual leaf more closely, we use stochastic gradient descent (SGD) to optimize a combined loss consisting of four parts:

- **Chamfer distance** between template vertices $\mathcal{V}$ and scan vertices $\mathcal{S}$

$$\mathcal{L}_{\text{cd}}(\mathcal{V}, \mathcal{S}) = \frac{\bar{d}(\mathcal{V}, \mathcal{S})}{2} + \frac{\bar{d}(\mathcal{S}, \mathcal{V})}{2}, \tag{1}$$

$$\bar{d}(\mathcal{V}, \mathcal{S}) = \frac{1}{|\mathcal{V}|} \sum_{\boldsymbol{v} \in \mathcal{V}} \min_{\boldsymbol{s} \in \mathcal{S}} \|\boldsymbol{v} - \boldsymbol{s}\|_2^2. \tag{2}$$

- **Vertex normal consistency** computes the normal consistency for each pair of neighboring faces of the template mesh:

$$\mathcal{L}_{nc} = \sum_{f_i \in \mathcal{F}} \sum_{f_j \in \mathcal{N}(f_i)} 1 - \boldsymbol{n}_i^\top \boldsymbol{n}_j, \tag{3}$$

where $\mathcal{F}$ is the set of faces in the template, and $\mathcal{N}(f)$ defines the neighborhood of adjacent faces of a given face f. $\boldsymbol{n}_i$ and $\boldsymbol{n}_j$ are the normals associated to f_i and f_j, where we assume that $||\boldsymbol{n}_i||_2 = 1$ and $||\boldsymbol{n}_j||_2 = 1$.

- **Laplacian smoothing** computes the Laplacian smoothing objective for the template mesh:

$$\mathcal{L}_s = \sum_{\boldsymbol{v} \in \mathcal{V}} \frac{1}{|\mathcal{N}(\boldsymbol{v})|} \sum_{\boldsymbol{u} \in \mathcal{N}(\boldsymbol{v})} \boldsymbol{u} - \boldsymbol{v}, \tag{4}$$

where $\mathcal{N}(\boldsymbol{v})$ defines the direct neighborhood of vertex $\boldsymbol{v}$, given the triangle mesh, i.e., all vertices that are connected to $\boldsymbol{v}$ with an edge.

- **Relative edge length** compares the length of each vertex normalized by the sum of vertex lengths with the

TABLE I: Precision, Recall, F-score, and Chamfer distance in controlled environment

	Occlusion Percentage 25%	Occlusion Percentage 50%	Occlusion Percentage 75%	Occlusion Percentage 25%	Occlusion Percentage 50%	Occlusion Percentage 75%	Occlusion Percentage 25%	Occlusion Percentage 50%	Occlusion Percentage 75%	Occlusion Percentage 25%	Occlusion Percentage 50%	Occlusion Percentage 75%
Approach	**precision** [%] ↑ - avg (std)			**recall** [%] ↑ - avg (std)			**f-score** [%] ↑ - avg (std)			D_c [%] ↓ - avg (std)		
CPD [21]	74.51 (23.08)	**72.07 (22.07)**	**65.33 (26.56)**	36.25 (20.47)	22.61 (14.34)	12.12 (9.82)	46.66 (21.28)	32.72 (16.65)	19.09 (12.73)	7.00 (2.82)	11.04 (3.69)	18.33 (6.47)
Anch CPD	57.96 (20.85)	55.06 (19.59)	45.87 (18.18)	38.97 (20.91)	33.49 (17.81)	25.76 (15.44)	45.02 (20.15)	40.01 (17.62)	31.55 (15.61)	8.58 (5.38)	11.16 (6.43)	15.40 (9.13)
Anch CPD + SGD	66.01 (19.43)	63.99 (19.53)	59.88 (19.26)	60.02 (17.20)	50.24 (15.01)	35.77 (12.60)	61.95 (16.95)	55.15 (15.31)	43.36 (12.88)	6.17 (3.46)	8.54 (4.51)	12.58 (6.64)
PF-SGD (ours)	**77.44 (16.21)**	66.86 (16.43)	50.71 (16.12)	**77.12 (15.15)**	**61.89 (14.16)**	**44.54 (15.25)**	**76.75 (14.41)**	**63.53 (13.44)**	**46.65 (14.21)**	**4.21 (3.01)**	**6.56 (4.27)**	**10.97 (7.25)**

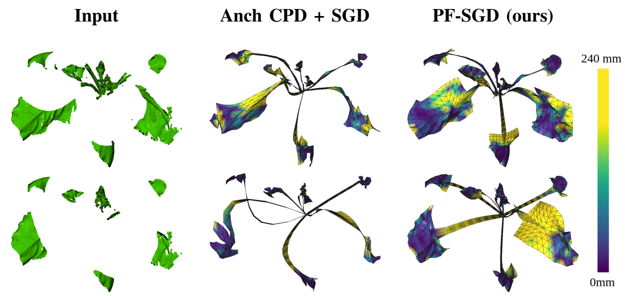

Fig. 5: Controlled environment. Qualitative and quantitative reconstruction results at different occlusion levels: 50% top, 75% bottom. The lighter the color, the bigger the distance between reconstructed mesh and scan.

same value before the optimization began to preserve the leaves morphology:

$$\mathcal{L}_{\text{e}} = \frac{\sum_{e \in \mathcal{E}} ||\boldsymbol{u} - \boldsymbol{v}||_2}{\frac{1}{|\mathcal{E}|}\sum_{e \in \mathcal{E}} ||\boldsymbol{u} - \boldsymbol{v}||_2}, \tag{5}$$

where $\mathcal{E}$ is the set of edges e in the template. Each edge is defined by two connected vertices $\boldsymbol{u}$ and $\boldsymbol{v}$.

Our loss function for stochastic gradient descent is, then, defined as the weighted sum of the previously defined terms:

$$\mathcal{L} = w_{\text{cd}}\mathcal{L}_{\text{cd}} + w_{\text{nc}}\mathcal{L}_{\text{nc}} + w_{\text{s}}\mathcal{L}_{\text{s}} + w_{\text{e}}\mathcal{L}_{\text{e}}. \tag{6}$$

During the optimization, the non-observed vertices tend to collapse onto the partial scan. Therefore, we filter them out dynamically during the process. To accomplish this, we divide the set of vertices $\mathcal{V}$ into a subsets $\mathcal{V}_{\text{obs}}$ of observed vertices and a subset $\mathcal{V}_{\text{occl}}$ of occluded and therefore unobserved vertices (see Fig. 7), such that $\mathcal{V}_{\text{obs}} \cup \mathcal{V}_{\text{occl}} = \mathcal{V}$ and $\mathcal{V}_{\text{obs}} \cap \mathcal{V}_{\text{occl}} = \emptyset$. We define the threshold $\delta = \min ||\boldsymbol{v} - \boldsymbol{s}||_2$, where $\boldsymbol{s} \in \mathcal{S}$, representing the maximum distance at which we consider a vertex to be observed. We initialize δ to an initial value δ_0 and then decay δ linearly over the SGD iterations to a final value δ_e. At each iteration, we redefine the sets $\mathcal{V}_{\text{obs}} := \{\boldsymbol{v} \in \mathcal{V} \,|\, \min_{\boldsymbol{s} \in \mathcal{S}} ||\boldsymbol{v} - \boldsymbol{s}|| < \delta\}$ and $\mathcal{V}_{\text{occl}} := \mathcal{V} - \mathcal{V}_{\text{obs}}$. The loss function is computed on the vertices contained in $\mathcal{V}_{\text{obs}}$ and their position is optimized. The vertices contained in $\mathcal{V}_{\text{occl}}$ instead are reset to their initial position and excluded from the optimization process.

E. Reconstruction of the Unobserved Part

The approach presented in the previous section only refines the estimate of observed template vertices $\mathcal{V}_{\text{obs}}$, while it does not update the poses of the unobserved vertices $\mathcal{V}_{\text{occl}}$. To propagate the refinement from $\mathcal{V}_{\text{obs}}$ to the neighboring vertices in $\mathcal{V}_{\text{occl}}$ and obtain a smoother reconstructed leaf, we make use of the idea proposed by Sorkine et al. [25]. We propagate the deformations estimated for visible vertices $\mathcal{V}_{\text{obs}}$ to the non-observed ones $\mathcal{V}_{\text{occl}}$. Each $\boldsymbol{v} \in \mathcal{V}_{\text{occl}}$ undergoes a similar transformation to its neighbors by solving:

$$\boldsymbol{T}^* = \min_{\boldsymbol{T}} \sum_{\boldsymbol{v}_i \in \mathcal{V}_{occl}} \sum_{\boldsymbol{v}_j \in \mathcal{N}(\boldsymbol{v}_i)} (\boldsymbol{v}'_i - \boldsymbol{v}'_j) - \boldsymbol{T}(\boldsymbol{v}_i - \boldsymbol{v}_j), \tag{7}$$

where $\boldsymbol{v}'_i$ and $\boldsymbol{v}'_j$ are the vertices after applying the transformation $\boldsymbol{T} \in \mathbb{R}^{4\times4}$ and $\mathcal{N}(\cdot)$ is the neighboring function.

IV. Experimental Evaluation

We present our experiments to show the capabilities of our method, from now on called PF-SGD (Partially Filtered Stochastic Gradient Descent optimization). We use the same parameters for all experiments, namely $w_{\text{cd}} = 1.0$, $w_{\text{nc}} = 0.2$, $w_{\text{s}} = 0.1$ and $w_{\text{e}} = 0.5$. To ensure convergence for all leaves we set the number of optimization iterations to 1000. The presented results support our key claims, which are: (i) our approach is able to closely reconstruct the leaf parts visible in the leaf scan, including finer details, (ii) parts of the leaves that are missing due to occlusions are reconstructed plausibly, (iii) our reconstruction process can complete partial scans obtained on the field, (iv) and the obtained reconstructed meshes are correctly aligned to the corresponding leaf parts

A. Datasets

We collected two different datasets of sugar beet plants, a manually created high precision point cloud dataset and a dataset of UAV images of breeding plots. The first one contains nine plants with 90 individual leaves in sum. We acquired it with a high-resolution laser scanner (Perceptron V5) attached to a measuring arm (Romer Infinite) to obtain ground truth data with sub-millimeter accuracy and without any occlusion (due to numerous viewpoints during manual recording). We refer to Schunck et al. [23] for more details on the scanning system. This dataset allows us to evaluate the reconstruction accuracy in a controlled environment. The second dataset consists of 13 different plants resulting in 138 individual leaves. We collected this dataset with the Phaseone iXM-100 camera attached to a UAV. As reference data, we flew three missions on the same field with a camera angle of 45, 90, and 135 degrees from the ground plane. We use the canonical nadir view of 90 degrees as input of our reconstruction pipeline and the other missions to obtain more accurate reference data. We process the resulting 100-megapixel images into 3D point clouds using bundle adjustment and then segment them into individual plants and their leaves. We show few exemplary plants in Fig. 3. All the leave point clouds used in this evaluation are labeled manually to ensure ground truth information.

TABLE II: Precision, recall, f-score, and Chamfer distance computed on data from an agricoltural field

Approach	**precision** [%] ↑ avg (std)	**recall** [%] ↑ avg (std)	**f-score** [%] ↑ avg (std)	D_c [%] ↓ avg (std)
CPD [21]	71.05 (17.86)	33.12 (15.63)	43.59 (17.07)	7.48 (2.57)
Anch CPD	52.83 (8.69)	33.80 (13.35)	39.91 (11.56)	11.10 (4.92)
Anch CPD + SGD	69.39 (9.84)	56.77 (15.24)	61.41 (11.81)	7.54 (2.90)
PF-SGD (ours)	**81.65 (14.72)**	**73.69 (16.66)**	**77.25 (15.60)**	**4.74 (2.93)**

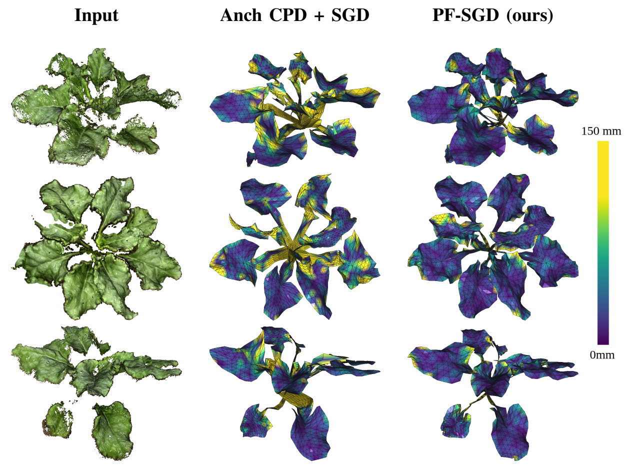

Fig. 6: Real field. Qualitative and quantitative reconstruction. The lighter the color, the bigger the distance between reconstructed mesh and scan.

B. *Metrics*

To measure the accuracy of our approach, we use different metrics: Chamfer distance, f-score, precision, and recall. To use the Chamfer distance defined in Eq. (1) as a metric, we define $\mathcal{G}$ and $\mathcal{R}$ as the ground truth point cloud and the point cloud obtained by densely sampling the reconstructed mesh. Then, the Chamfer distance $D_{\mathrm{c}} = \mathcal{L}_{\mathrm{cd}}(\mathcal{R}, \mathcal{G})$.
To illustrate the performance in a clearer way we define an f-score metric as proposed by Knapitsch et al. [14]. We first define precision p, and recall r, given a threshold δ:

$$\begin{aligned} p(\delta) &= \frac{100}{|\mathcal{R}|} \sum_{\boldsymbol{r} \in \mathcal{R}} \left[\!\left[\min_{\boldsymbol{g} \in \mathcal{G}} ||\boldsymbol{r} - \boldsymbol{g}|| < \delta \right]\!\right], \\ r(\delta) &= \frac{100}{|\mathcal{G}|} \sum_{\boldsymbol{g} \in \mathcal{G}} \left[\!\left[\min_{\boldsymbol{r} \in \mathcal{R}} ||\boldsymbol{g} - \boldsymbol{r}|| < \delta \right]\!\right], \end{aligned} \tag{8}$$

where $\boldsymbol{g}$ and $\boldsymbol{r}$ are points from $\mathcal{G}$ and $\mathcal{R}$ and the operator $[\![\cdot]\!]$ is the Iverson bracket, i.e., if the condition within the brackets is satisfied, it evaluates to 1, otherwise to 0. Intuitively, such metrics compute the percentage of points in one set whose distance to the closest point in the other set is smaller than a fixed threshold. The f-score is simply the harmonic mean of precision and recall $f(\delta) = \frac{2 \cdot p(\delta) \cdot r(\delta)}{p(\delta) + r(\delta)}$. For all the experiments, we set the threshold $\delta = 5\,\mathrm{mm}$.

C. *Comparison to the State of the Art*

For all the experiments, we compare our approach to the coherent point drift algorithm (CPD) [21], the state-of-the-art non-rigid registration algorithm because of its registration performance and scalability to large point sets [18]. The algorithm works quite well for the reconstruction of completely observed leaves, it fails for the ones presenting occlusions. We tried many parameter configurations but the mesh resulting from CPD always collapsed on the visible part of the leaves. As the algorithm's ability to correctly deform our template onto the partial leaf point clouds is not satisfactory and to ensure a fair comparison, as we used leaf tip and plant center as prior knowledge, we made two extensions to the approach.

We considered the squared distance of each template vertex v to the closest point in the scan $d = \min_{\boldsymbol{s} \in \mathcal{S}} ||\ \boldsymbol{v} - \boldsymbol{s}\ ||^2$, normalized to a maximum distance threshold and clipped to obtain values in $[0, 1]$ as the prior probability of each vertex of being observed. We then multiply these prior probabilities onto the association probability matrix of CPD, P. For details on this, please refer to the original publication. To use our prior knowledge about the position of the plant center and the leaf tip that we obtained as explained in III-C, we set the association probability of the corresponding entries in P to 1 and exclude those vertices from the optimization process. We refer to this as anchored CPD ("Anch CPD"). Furthermore, as the coherent point drift algorithm is not reconstructing finer details of the leaves accurately, we refine the obtained result by stochastic gradient optimization, using our loss Eq. (6). To ensure a fair comparison, the losses and their weights are the same as the ones explained in Sec. III-D. We refer to these results as "Anch CPD + SGD".

D. *Reconstruction Evaluation in Controlled Environment*

We design the first experiment to support the claims that our approach is (i) able to closely reconstruct visible leaf parts in the point clouds, including fine details and (ii) reconstruct in a plausible way parts of the leaves that are not visible.

To obtain a quantitative evaluation with a precise reference, in this experiment we use the dataset obtained with the laser scanner. As the point clouds in this dataset cover the entire leaves completely, we can measure the performance precisely. We first define three levels of occlusion, namely 25%, 50%, and 75%, of the points in each leaf. Then, we synthetically generate ten randomly occluded leaves for each of the occlusion levels for each leaf in our dataset. This enables us to test our approach on increasingly complex scenarios while having a precise ground truth model for the evaluation. In total, we test our method on 2 700 leaf variations. In particular in Fig. 4, we show the estimated distributions for the f-score. In detail, we present the median (bar inside the box), the lower and upper quartile (box extent), and the minimum and maximum values of the distribution. We also show qualitative examples in Fig. 5. An overview of the other metrics can be seen in Tab. I. From the evaluation follows that our approach outperforms the baselines in most of the metrics. The precision of the CPD approach is better for low occlusions, but, as can be seen in Fig. 5, the results of this class of methods tend to fit the entire template onto the partial observation. This leads to high precision, counterbalanced by a low recall, resulting in worse overall reconstructions, as can be seen from the f-

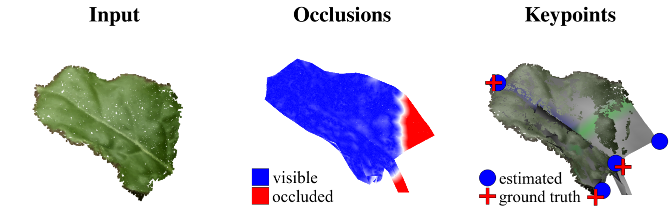

Fig. 7: Given a leaf (left), we show the estimated occlusions (mid) and the semantic components of the deformed template (right).

score. Notably, our approach can achieve a lower score for the chamfer distance across all occlusion levels by a large margin, see Tab. I.

E. Reconstruction Evaluation in Real Field Conditions

In our next experiment, we support the claim that our method is well suited for the reconstruction of partial plants in real field conditions, where wind and real-world lighting create noise and artifacts. In this experiment, we perform the reconstruction on the point clouds obtained by processing only the images obtained from a drone mission with the optical axis perpendicular to the ground plane. This resembles the process that one would typically follow in a real application, as it considerably reduces the effort of acquiring and processing data. As obtaining complete ground truth scans is infeasible, we bundle adjust the full 3 flight missions with different camera angles instead. We use them as a reference model of the field, as they are more precise and cover the lower parts of the canopy more accurately. We show the f-score analysis in Fig. 4 and an overview of all the metrics in Tab. II. This experiment shows that our work is better suited at reconstructing leaves from real scans than the other approaches, as expected given the results of the previous experiment. This can also be seen from the qualitative examples in Fig. 6.

F. Semantic Accuracy Evaluation

In our last experiment we show that our approach is not only able to reconstruct the leaf surface accurately, but the reconstructed mesh is also correctly aligned with the main leaf parts. This backs up our last claim that our reconstructions correctly align with the corresponding leaf parts. As correctly detecting the leaf parts such as the tip and the lateral extremities is important in phenotyping applications, the metric here is the distance between the manually annotated keypoints, described in Fig. 2, and their corresponding vertices in the template mesh. We report such metrics for both datasets in Tab. III. Our approach shows better performances for each keypoint except the leaf tip. Such behavior can be attributed to our heavy reliance on the least squares pre-alignment.

In summary, our evaluation suggests that our method provides competitive registration accuracy in both geometry and semantics compared to the baselines. Thus, we supported all our claims with this experimental evaluation.

TABLE III: Semantic accuracy after template deformation

(a) Semantic fitting in controlled environment

	center	left	tip	right
Approach	distance [mm] ↓ - avg (std)			
CPD [21]	55.9 (22.1)	55.5 (28.3)	23.5 (11.0)	50.1 (21.9)
Anch CPD	45.2 (23.8)	59.9 (29.5)	7.9 (6.0)	51.7 (22.5)
Anch CPD + SGD	43.4 (22.6)	59.6 (30.3)	9.6 (5.7)	45.9 (23.5)
PF-SGD (ours)	**27.6 (15.5)**	**27.3 (17.6)**	**6.7 (4.7)**	**24.7 (16.5)**

(b) Semantic fitting in field conditions

	center	left	tip	right
Approach	distance [mm] ↓ - avg (std)			
CPD [21]	36.5 (10.0)	29.5 (13.1)	19.4 (9.1)	31.9 (14.6)
Anch CPD	26.5 (9.8)	34.2 (20.5)	**15.3 (7.8)**	36.1 (23.4)
Anch CPD + SGD	24.6 (12.0)	28.0 (18.4)	15.7 (8.5)	31.6 (23.9)
PF-SGD (ours)	**11.9 (7.3)**	**24.9 (14.5)**	18.1 (8.5)	**21.1 (12.7)**

V. Conclusion

In this paper, we presented a novel approach to precisely reconstruct sugar beet plants in field conditions. Our approach operates on point clouds of plants obtained by UAV imagery recorded in nadir view. Our method exploits existing photogrammetric bundle adjustment and a deep neural network to segment the field in plants and separate its individual leaves. Once we have segmented the leaves, we use SGD to deform the vertices of a leaf template mesh, while jointly estimating which mesh vertices have no correspondence in the point cloud due to occlusions. This allows us to successfully obtain 3D models of whole plants under real field conditions starting from 2D images. Furthermore, by encoding additional semantic information on the leaf template, our approach can be used to derive important phenotypic traits at a sub-plant level. We implemented and evaluated our approach on different datasets, provided comparisons to other techniques, and supported all claims made in this paper. The experiments suggest that, compared to other registration methods, our approach provides more reliable geometric and semantic reconstruction results.

Our approach can be extended along different dimensions. An interesting research direction would be the evaluation of plant health by leveraging the leaf shapes and the presence of damaged parts. Currently, it is not possible to differentiate between occluded and damaged leaves, but this could be achieved by analyzing inter-leaf effects and color information. Furthermore, to allow for a more robust reconstruction, reliable prior detection of the leaf keypoints to support the reconstruction process would be of great help. This would most probably improve the reconstruction accuracy and the robustness against noise in the measurements. Another interesting future direction is to apply this method to different plant species to reconstruct their leaf and fruit shape. An adaptation of the template and the prealignment strategy would probably suffice to achieve good results.

Acknowledgments

We thank R. Marcuzzi, L. Nunes, M. Sodano, and L. Wiesmann for supporting us with making annotations.

References

[1] R. Ando, Y. Ozasa, and W. Guo. Robust surface reconstruction of plant leaves from 3d point clouds. *Plant Phenomics*, 2021.

[2] J. Binney and G. Sukhatme. 3d tree reconstruction from laser range data. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, pages 1321–1326.

[3] P.M. Blok, E.J. van Henten, F.K. van Evert, and G. Kootstra. Image-based size estimation of broccoli heads under varying degrees of occlusion. *Biosystems Engineering*, 208:213–233, 2021.

[4] J. Brown and S. Sukkarieh. Design and evaluation of a modular robotic plum harvesting system utilizing soft components. *Journal of Field Robotics (JFR)*, 38(2):289–306, 2021.

[5] L. Carlone, J. Dong, S. Fenu, G. Rains, and F. Dellaert. Towards 4d crop analysis in precision agriculture: Estimating plant height and crown radius over time via expectation-maximization. In *ICRA Workshop on Robotics in Agriculture*, 2015.

[6] N. Chebrolu, T. Läbe, and C. Stachniss. Robust long-term registration of uav images of crop fields for precision agriculture. *IEEE Robotics and Automation Letters*, 3(4):3097–3104, 2018.

[7] N. Chebrolu, T. Läbe, and C. Stachniss. Spatio-temporal non-rigid registration of 3d point clouds of plants. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2020.

[8] N. Chebrolu, F. Magistri, T. Läbe, and C. Stachniss. Registration of Spatio-Temporal Point Clouds of Plants for Phenotyping. *PLOS ONE*, 16(2), 2021.

[9] J. Dong, J. Burnham, B. Boots, G. Rains, and F. Dellaert. 4D Crop Monitoring: Spatio-Temporal Reconstruction for Agriculture. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2017.

[10] F. Fiorani and U. Schurr. Future scenarios for plant phenotyping. *Annual review of plant biology*, 64:267–291, 2013.

[11] R.T. Furbank and M. Tester. Phenomics – technologies to relieve the phenotyping bottleneck. *Trends in Plant Science*, 16(12):635 – 644, 2011.

[12] M. Halstead, C. McCool, S. Denman, T. Perez, and C. Fookes. Fruit quantity and ripeness estimation using a robotic vision system. *IEEE Robotics and Automation Letters (RA-L)*, 3(4):2995–3002, 2018.

[13] A. Hornung, K. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard. OctoMap: An Efficient Probabilistic 3D Mapping Framework Based on Octrees. *Autonomous Robots*, 34:189–206, 2013.

[14] A. Knapitsch, J. Park, Q.Y. Zhou, and V. Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Trans. on Graphics (TOG)*, 36(4), 2017.

[15] C. Lehnert, C. McCool, Y. Sa, and T. Perez. Performance improvements of a sweet pepper harvesting robot in protected cropping environments. *Journal of Field Robotics (JFR)*, 37(7):1197–1223, 2020.

[16] F. Magistri, N. Chebrolu, J. Behley, and C. Stachniss. Towards In-Field Phenotyping Exploiting Differentiable Rendering with Self-Consistency Loss. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2021.

[17] F. Magistri, N. Chebrolu, and C. Stachniss. Segmentation-Based 4D Registration of Plants Point Clouds for Phenotyping. In *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2020.

[18] B. Maiseli, Y. Gu, and H. Gao. Recent developments and trends in point set registration methods. *Journal of Visual Communication and Image Representation*, 46:95–106, 03 2017.

[19] D.W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.

[20] Z. Min, J. Pan, A. Zhang, and M.Q.H. Meng. Robust non-rigid point set registration algorithm considering anisotropic uncertainties based on coherent point drift. In *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, pages 7903–7910, 2019.

[21] A. Myronenko and X. Song. Point set registration: Coherent point drift. *IEEE TPAMI*, 32(12):2262–2275, 2010.

[22] R.A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A.J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon. KinectFusion: Real-Time Dense Surface Mapping and Tracking. In *Proc. of the Intl. Symposium on Mixed and Augmented Reality (ISMAR)*, pages 127–136, 2011.

[23] D. Schunck, F. Magistri, R. 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*, 16(8):1–18, 2021.

[24] P. Sodhi, H. Sun, B. Póczos, and D. Wettergreen. Robust plant phenotyping via model-based optimization. In *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, pages 7689–7696, 2018.

[25] O. Sorkine and M. Alexa. As-rigid-as-possible surface modeling. In *Proc. of the Symposium on Geometry Processing*, volume 4, pages 109–116, 2007.

[26] B. Triggs, P.F. McLauchlan, R.I. Hartley, and A.W. Fitzgibbon. Bundle adjustment – a modern synthesis. In *Proc. of the Intl. Workshop on Vision Algorithms: Theory and Practice*, page 298–372, 1999.

[27] E. Vespa, N. Nikolov, M. Grimm, L. Nardi, P. Kelly, and S. Leutenegger. Efficient octree-based volumetric slam supporting signed-distance and occupancy mapping. *IEEE Robotics and Automation Letters (RA-L)*, 2018.

[28] I. Vizzo, T. Guadagnino, J. Behley, and C. Stachniss. Vdbfusion: Flexible and efficient tsdf integration of range sensor data. *Sensors*, 22(3), 2022.

[29] J. Weyler, , F. Magistri, P. Seitz, J. Behley, and C. Stachniss. In-Field Phenotyping Based on Crop Leaf and Plant Instance Segmentation. In *Proc. of the IEEE Winter Conf. on Applications of Computer Vision (WACV)*, 2022.

[30] J. Weyler, A. Milioto, T. Falck, J. Behley, and C. Stachniss. Joint Plant Instance Detection and Leaf Count Estimation for In-Field Plant Phenotyping. *IEEE Robotics and Automation Letters (RA-L)*, 6:3599–3606, 2021.

[31] T. Whelan, S. Leutenegger, R.S. Moreno, B. Glocker, and A. Davison. ElasticFusion: Dense SLAM Without A Pose Graph. In *Proc. of Robotics: Science and Systems (RSS)*, 2015.

[32] X. Wu, S. Aravecchia, P. Lottes, C. Stachniss, and C. Pradalier. Robotic weed control using automated weed and crop classification. *Journal of Field Robotics*, 37(2):322–340, 2020.

[33] Q. Zheng, A. Sharf, A. Tagliasacchi, B. Chen, H. Zhang, A. Sheffer, and D. Cohen-Or. Consensus skeleton for non-rigid space-time registration. In *Computer Graphics Forum*, volume 29, pages 635–644, 2010.