Fruit Tracking Over Time Using High-Precision Point Clouds

Alessandro Riccardi Shane Kelly Elias Marks Federico Magistri
Tiziano Guadagnino Jens Behley Maren Bennewitz Cyrill Stachniss

***Abstract*— Monitoring the traits of plants and fruits is a fundamental task in horticulture. With accurate measurements, farmers can predict the yield of their crops and use this information for making informed management decisions, and breeders can use it for variety selection. Agricultural robotic applications promise to automate this monitoring task. In this paper, we address the problem of monitoring fruit growth and investigate the matching of fruits recorded in commercial greenhouses at different growth stages based on data recorded from terrestrial laser scanners. This is challenging as fruits appear highly similar, change over time, and are subject to severe occlusions. We first propose a fruit descriptor, which captures the topology of the fruit surroundings to facilitate the matching between different points in time. We capture and describe the relationship between a fruit and its neighbors such that our descriptors are less affected by the growth over time. Furthermore, we define a matching cost function and use an optimal assignment algorithm to match the fruit observations taken in different weeks. The experiments show that our descriptor achieves a high spatio-temporal matching accuracy, which is superior to the commonly used geometric point cloud descriptors.**

I. Introduction

To satisfy food demand for an ever-growing population, we have to grow more crops more consciously in the following years than ever. Robotics and automation have become increasingly popular in supporting farmers in monitoring and management actions. Autonomous robots have the prospect of revolutionizing agricultural production systems [12], [40] and can also boost phenotyping to assess plant performances when breeding new varieties. Phenotyping is the task of measuring plant properties and assessing the performance of plant varieties [13], [41] to better select cultivars for the following seasons. Today, phenotyping involves humans manually measuring individual plants [15], which limits spatial and temporal phenotyping throughput. Instead, robots, when equipped with high-resolution sensors, provide an attractive way to monitor plants at a large scale continuously, having, at the same time, the ability to inspect plants and fruits closely.

A. Ricardi, E. Marks, F. Magistri, T. Guadagnino, J. Behley, M. Bennewitz, and C. Stachniss are with the University of Bonn, DE. C. Stachniss is additionally with the Department of Engineering Science at the University of Oxford, UK, and with the Lamarr Institute for Machine Learning and Artificial Intelligence, DE. S. Kelly is with the ETH Zurich, CH.

This work has partially been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy, EXC-2070 – 390732324 – PhenoRob, by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under STA 1051/5-1 within the FOR 5351 (AID4Crops), 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).

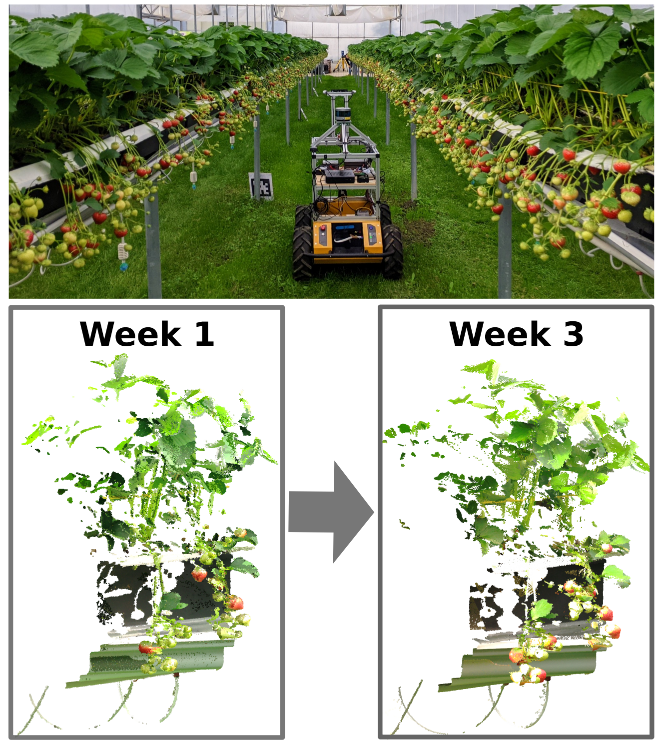

Fig. 1: Top, our UGV system is navigating inside the greenhouse to collect data. Bottom, the same section of the greenhouse one week apart, where we can appreciate the different sizes and the different ripeness of the fruits. Such changes in the scene's appearance make it difficult to establish correspondence between fruits.

One challenge is the temporal tracking of fruits that allows for monitoring fruit growth and traits such as shape and color over time. However, recognizing the same fruits in different recordings is challenging due to the similarity of most fruits and the changes in the scene, and the fruits themselves deriving from growth and recording conditions. Fig. 1 shows the same portion of a strawberry greenhouse as an example. Within a week, plants can grow substantially, new fruits develop, existing fruits change shape and color, and some fruits may get harvested. Such changes must be considered when tracking individual fruits over time.

This problem somewhat resembles detecting loop closures in SLAM systems [36] in changing environments and shares similarities with visual place recognition [39]. However, solutions to those problems typically rely on distinctive man-made structures, such as buildings, poles, etc.

Therefore, this paper's main contribution is a practical approach that allows for fruit matching across different times,

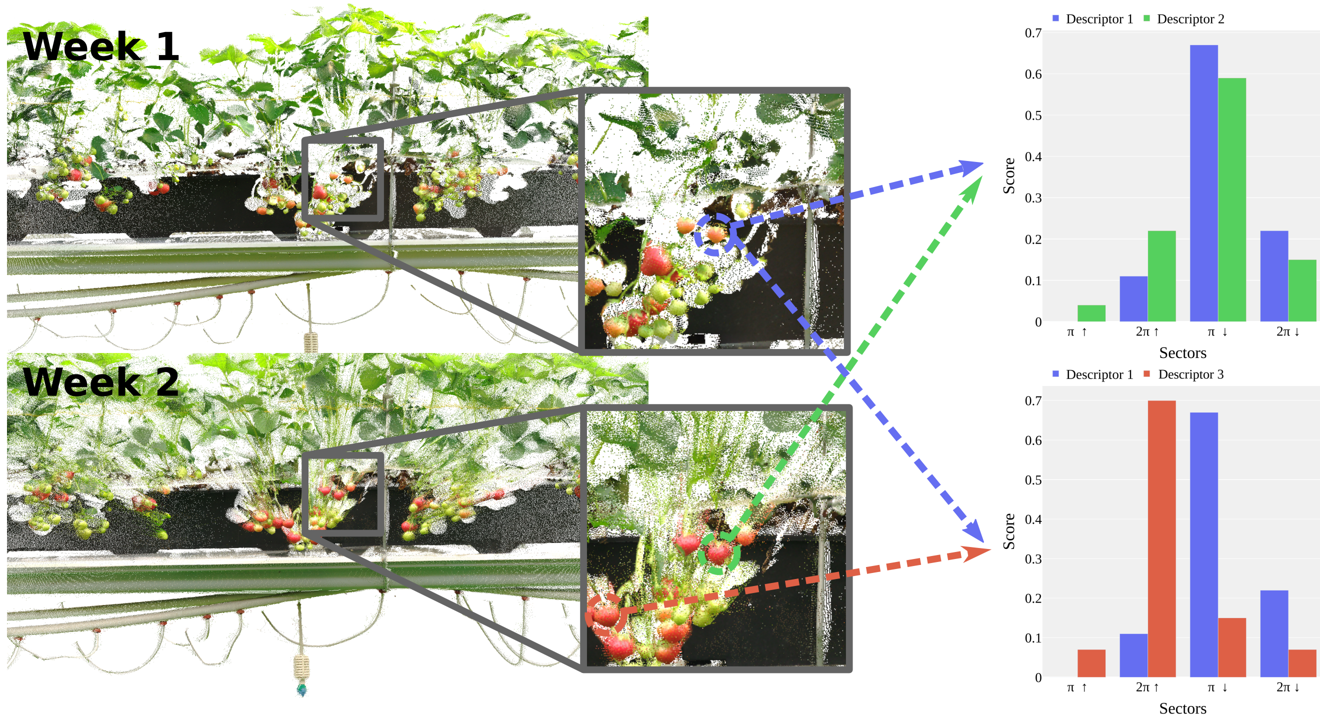

Fig. 2: We define a descriptor that captures the relationship between each fruit and its neighbors. This allows us to match individual fruits over time, even if the scene looks different because of harvested or newly developed fruits, but also due to changes in color and radius. We show here the descriptors for fruits in different weeks. We observe a high similarity of the corresponding descriptors for the same fruit (blue and green). For different fruits (blue and red), we see a significant difference between the corresponding descriptors.

exploiting a novel 3D fruit descriptor to robustify matching results and thus enabling the monitoring of fruit traits such as color and radius over time. Our approach exploits dense high-precision point clouds recorded with a terrestrial laser scanner (TLS) together with fruit detections represented as bounding boxes. For matching fruit detections at different points in time, we utilize a relational histogram descriptor that leverages the fact that fruits maintain a similar constellation over time while individual fruits change substantially. With this, we can match fruits using the Hungarian method by defining a cost function that combines positional distance, descriptor distance, and radius variation. Experiments on strawberries in a greenhouse show that the proposed fruit descriptor combined with Hungarian matching is more effective than matching with descriptors based solely on the appearance of individual fruits. As shown in the evaluation, using the proposed pipeline enables us to identify fruits and model them over time robustly.

II. Related Work

Image-based phenotyping for the automatic monitoring of plants is gaining increasingly importance in various agricultural environments. Thanks to the advances in deep learning in the agricultural context [31],such techniques have been applied to yield estimation in vineyards [20], [27], to ripeness estimation and fruit counting in greenhouse [18], [33], but also to phenotyping [6], [42], [43] and disease spotting [17] in crop fields. While image-based phenotyping has received significant attention, few works propose methods to extract phenotypic traits from 3D data. Lehnert et al. [22] use a camera array to compute the next best view to maximize fruit coverage. In follow-up work, Zaenker et al. [44] combine local and global viewpoint planning for improving fruit coverage. Gibbs et al. [16] propose an active vision approach to compute high-quality 3D surface reconstruction of plants. Underwood et al. [38] measure canopy volume from 3D models of almond orchards and estimate flowers and fruits density. Binney et al. [5] fit cylinders to point clouds of trees to recover missing data. Sodhi et al. [34] address the problem of mapping plant sub-units called plant phytomers to their phenotype value. In our prior works, we estimate the shape of fruits [25] and plants [23], [26] in the presence of occlusions without considering the temporal growth of plants and fruits. One of the reasons behind this gap between 2D and 3D analysis can be found in sensor limitations. For example, traditional sensors equipped on robots like 3D LiDARs and RGB-D cameras often provide a 3D spatial resolution too coarse for agricultural scenes. A possible solution is using terrestrial laser scanners to obtain dense, colored, and precise point clouds. Sun et al. [37] exploit such a sensor for stalk and node detection in cotton fields showcasing TLS potential. The drawbacks of TLS are the long data acquisition time required and the need for manually placing the TLS in the environment. While reducing the acquisition time reduces the scan precision, placing the TLS on a robotic platform aallows the data to be collected automatically [29].

Compared to the 3D case, fewer works address the problem of finding plant and fruit correspondences over time, i.e., in 4D. Chebrolu et al. [8] exploit a skeletal structure to compute correspondences between the same plant over a period of two weeks. In follow-up works [9], [24], the same authors propose an improved skeletal representation by adding semantics information to the skeleton's nodes and being able to estimate leaf growth parameters. The main differences with our work are two-fold: On one hand, they use data collected in controlled lab environments where each plant was spatially separated from the other plants and manually scanned thoroughly. We, instead, use more realistic data from commercial greenhouses. On the other hand, the skeletal structure, a clever way to represent a plant, cannot be easily extended to our fruit-matching task. For computing data association over time, we rely on the Hungarian method [21], which has been used for various applications, including robot exploration [35].

Carlone et al. [7] propose an expectation-maximization framework to register point clouds of a crop field at different points in time. In this way, they can monitor plant growth at a plot level. Dong et al. [10] solve a similar problem using a factor graph. The main difference to our work is the resolution of the spatial registration since we are interested in monitoring individual fruits rather than plots consisting of several plants.

III. Our Approach

Our work aims to accurately match fruits over time to track the development of individual fruits. More formally, given a pair of aligned point clouds $\mathcal{P}^t$ and $\mathcal{P}^{t+1}$, taken within the same environment but collected at different times. Given a set of fruits $\mathcal{F}^t = \{f_0^t, \dots, f_N^t\}$ and $\mathcal{F}^{t+1} = \{f_0^{t+1}, \dots, f_M^{t+1}\}$ extracted from the data, we want to find a set of associations $\mathcal{A} = \{a_0, \dots, a_K\}$ with $a_k = (f_i^t, f_j^{t+1})$. In our setting, each fruit f_i has three different features: a position $\mathbf{p}_i \in \mathbb{R}^3$ in the global reference frame of the aligned point clouds, an RGB color $\mathbf{c}_i \in \mathbb{R}^3$, and the fruit's radius $\mathbf{r}_i$. Given the association $\mathcal{A}$, we can derive the changes in color $\Delta\mathbf{c}_i$ and radius $\Delta\mathbf{r}_i$.

Under natural operational conditions, identifying and associating fruits at different points in time is challenging as the environment is non-rigid. Classic features such as color and radius change substantially even over short periods, see Fig. 2. Thus, the matching results using only such features can be suboptimal. Additionally, between two data collection campaigns, some fruits may be harvested, and new fruits may develop, increasing the layer of complexity to our problem. In the following, we express the point clouds in the same reference frame. In our implementation, we achieve the alignment of $\{\mathcal{P}^t, \mathcal{P}^{t+1}\}$ using a RANSAC [14] scheme combined with G-ICP [32]. In the following sections, we describe our fruit extraction setup, present our relational histogram descriptor using fruit neighborhoods, and the association approach based on the Hungarian method to match fruits over time.

A. Fruit Descriptor

Our main observation is that fruits belonging to the same plant undergo similar transformations during plant movement or growth; thus, our objective is to generate a feature able to cope with the morphological changes of the fruits over time. We define a descriptor that encompasses the relationship between each fruit's neighbors.

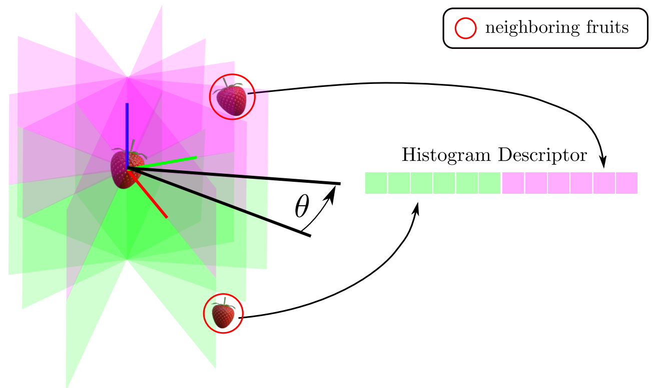

Fig. 3: A conceptual overview of the fruit descriptor generation. Given a fruit f_i, we determine the angle of neighboring fruits f_j relative to f_i and determine if the relative z coordinate is above 0, which determines in which part the neighboring fruit is counted, i.e., green or purple.

To make the notation more concise, we omit below the time index from the set of fruits. Let f_i be the fruit for which we want to compute our fruit descriptor $\mathbf{d}_i \in \mathbb{R}^{2S}$ capturing the neighborhood of the fruit organized in an angular discretization into sectors. The term

$$S = \left\lceil \frac{2\pi}{\theta} \right\rceil \tag{1}$$

is the number of sectors determined by the aperture of the predefined angle θ. Let $(x_i, y_i, z_i)^\top$ be the center of the i^{th} fruit f_i.

For each fruit f_j belonging to the set of neighbors of fruit f_i, we now determine the corresponding histogram bin index b_j as follows:

$$b_j = \left\lceil \frac{\hat{\theta}_j^i}{2\pi} \right\rceil + \mathbb{I}\left\{\hat{z}_j^i > 0\right\} S, \tag{2}$$

where $\mathbb{I}\{\cdot\}$ is the indicator function that returns 1 if its argument is true and 0 otherwise. Furthermore, $\hat{\theta}_j$ and $\hat{z}_j$ are the relative angles and relative z-coordinate of the j^{th} fruit with respect to fruit f_i, respectively:

$$\theta_j^i = \operatorname{arctan2}(x_j - x_i, y_j - y_i) \tag{3}$$

$$z_j^i = z_j - z_i. \tag{4}$$

Given the bin indexes b_j of each neighboring fruit f_j, we now increment the corresponding histogram entry, see Fig. 3. After adding the counts of all neighboring fruits, we normalize the histogram $\mathbf{d}_i$ by the length of the corresponding vector $\|\mathbf{d}_i\|_2$, which equals the number of fruits belonging to the neighborhood. In our experiment, we set its size to 27.

B. Fruit Matching

We use the Hungarian algorithm [21] to solve the fruit matching problem, initially designed to assign a set of n jobs to a set of n machines but later generalized to assignments of sets with different cardinality [35]. The Hungarian method computes the optimal, i.e., minimal-cost solution given a fixed cost function in $O(n^3)$. More precisely, given a cost matrix $\mathbf{C}$, it computes the optimal solution of the following minimization problem:

$$\mathbf{X} = \underset{\mathbf{X}}{\operatorname{argmin}} \|\mathbf{C} \circ \mathbf{X}\|_F \tag{5}$$

$$\text{s.t.} \quad \sum_i^n \mathbf{X}_{i*} = 1 \quad \text{and} \quad \sum_j^m \mathbf{X}_{*j} \leq 1, \tag{6}$$

where $\circ$ refers to the Hadamard product, i.e., the element-wise matrix multiplication, and $\|\cdot\|_F$ is the Frobenius norm of a matrix. The term $\mathbf{X}$ is a selection matrix in which $\mathbf{X}_{i,j} = 1$ if the match (i, j) contributes to the solution; otherwise, $\mathbf{X}_{i,j} = 0$. Intuitively, the term $\|\mathbf{C} \circ \mathbf{X}\|_F$ in Eq. (5) computes the sum of all individual assignment costs that contribute to the solution specified through the selection matrix $\mathbf{X}$. Additionally, the first constraint in Eq. (6) ensures that each node in the first scan is connected to only one node in the second scan. The second constraint in Eq. (6) guarantees that each keypoint in the second scan is connected to at most one node in the first scan.

We start by creating a cost matrix $\mathbf{C}_a \in \mathbb{R}^{N \times M}$ where columns represent fruits belonging to $\mathcal{F}^t$ and rows represent fruits belonging to $\mathcal{F}^{t+1}$. With this setting, using the Hungarian algorithm, we obtain a number of associations equal to the cardinality of the smaller set, $|\mathcal{A}| = \min(|\mathcal{F}^t|, |\mathcal{F}^{t+1}|)$. This is suboptimal given that fruits in $\mathcal{F}^t$ may not be present in $\mathcal{F}^{t+1}$ due to a harvesting or pruning/thinning action. Thus, we define a second cost matrix $\mathbf{C}_u \in \mathbb{R}^{N \times N}$ where each element stores the same constant value u, representing the cost of the fruit not being assigned. Our complete cost matrix $\mathbf{C} \in \mathbb{R}^{N \times M+N}$ is then defined as:

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_a & \mathbf{C}_u \end{bmatrix}. \tag{7}$$

This strategy allows the Hungarian method to compute solutions with unassigned fruits in $\mathcal{F}$ and $\mathcal{F}^{t+1}$.

We propose to encode in $\mathbf{C}_a$ the absolute position, the radius of each fruit, and the descriptors defined in the previous section:

$$\mathbf{C}_a = \alpha\,\mathbf{P} + \beta\,\mathbf{D} + \gamma\,\mathbf{S}, \tag{8}$$

where:

$$\mathbf{P}_{i,j} = \|\mathbf{p}_i^t - \mathbf{p}_j^{t+1}\|, \tag{9}$$

$$\mathbf{D}_{i,j} = \|\mathbf{d}_i^t - \mathbf{d}_j^{t+1}\|, \tag{10}$$

and

$$\mathbf{S}_{i,j} = \|r_i^t - r_j^{t+1}\|. \tag{11}$$

The element $\mathbf{P}_{i,j}$ represents the Euclidean distance between the position of fruit $f_i \in \mathcal{F}^t$ and $f_j \in \mathcal{F}^{t+1}$. Similarly, the element $\mathbf{D}_{i,j}$ represents the Euclidean distance between the descriptors of fruit $f_i \in \mathcal{F}^t$ and $f_j \in \mathcal{F}^{t+1}$, and the element $\mathbf{S}_{i,j}$ represents the absolute difference between the radius of fruit $f_i \in \mathcal{F}^t$, and $f_j \in \mathcal{F}^{t+1}$. Computing the optimal assignment using the Hungarian method has a time complexity of $O(n^3)$, where n is the number of fruits that need to be matched. In our scenario, it typically takes around 150 ms to match fruits between two points in time. Thereby, the Hungarian method considers the possible data association between approximately 700 fruits.

Parameter	α	β	γ	u
Value	0.15	0.62	0.93	2.7

TABLE I: Parameters found by Optuna and used in Eq. (8) for our approach. α weighs the position term, β the descriptor term, γ the radius term, and u is the constant penalty cost for unassigned fruits.

Features	**Precision [%]**	**Recall [%]**	**F-score [%]**
FPFH [30]	51.97	52.45	52.21
Spin Images [19]	58.90	58.89	58.89
Our	**66.34**	**64.11**	**65.21**

TABLE II: 4D fruit matching using TLS data. Our approach is, on average, 6 % and 13 % better in all three metrics compared to Spin Images and FPFH.

IV. Experimental Evaluation

The primary focus of this work is to build a pipeline able to match individual fruits over time, allowing the tracking of fruits' traits. We show that our system can match fruits over time and that our descriptor benefits the task, enabling us to monitor fruit traits such as color and radius over time.

A. Dataset and Metrics

We collected our dataset using a high-precision laser scanner in a commercial strawberry greenhouse near Bonn, Germany. We collected the dataset in 3 different sessions, where we scanned weekly the same environment, resulting in roughly 50 million points and more than a thousand individual fruits. We selected a specific row of strawberries of 5 meters long as our restricted dataset space. We manually labeled around 700 fruits per each time section, divided the environment into two non-overlapping sets, and manually labeled their fruit correspondences. We used the smaller set, representing 20% of the scene, to find the best parameters for our descriptor Eq. (8) using Optuna [2]. Then we used the remaining part of the scene to run all the experiments. To evaluate the matching performances, we compute precision, recall, and f-score utilizing the definition of true positive, false positive, and false negative.

B. 4D Matching Performances

To compare our proposed solution with classic 3D geometric histogram descriptors [3], we specifically select Spin Images [19] and Fast Point Feature Histogram (FPFH) [30] as geometric descriptors due to their proven performance for 3D object matching [19], [11], [28], 3D point classification [3], and place recognition [30], [1]. As reference frame for computing the histogram descriptors, we used the global coordinate frame and a kD-tree [4] from Open3D for the nearest neighbor search. For Spin Image, we used 20 bins,

Fig. 4: Four examples show the matching results between two different time sections of fruits that belong to the same area. The green lines represent the correct matches between corresponding fruits, while the wrong ones are depicted in red. Each example is a partial visual representation of the matching results from the entire experiment.

which resulted in a descriptor of length 400. For FPFH, we use 33 bins. We report our approach's parameters of Eq. (8) in Tab. I. As shown in Tab. II, our proposed fruit matching pipeline outperforms both baselines in terms of the three metrics. In particular, our approach is, on average, 6 percent points and 13 percent points better in all three metrics compared to Spin Images and FPFH, respectively. This is because our descriptor captures the distribution of neighboring fruits, which helps in the presence of morphological changes in the fruits. In Fig. 4, we can observe a qualitative evaluation of the performances of our method in four different examples extracted from the dataset space.

C. Ablation Study

We run an additional set of experiments to verify the impact on the final results of each term in Eq. (8). We report the results of this study in Tab. III. As expected, we notice that the global positions alone are insufficient to achieve good performance. This can be seen in the 45.7% of recall, which is by far the lowest score in the experiments.

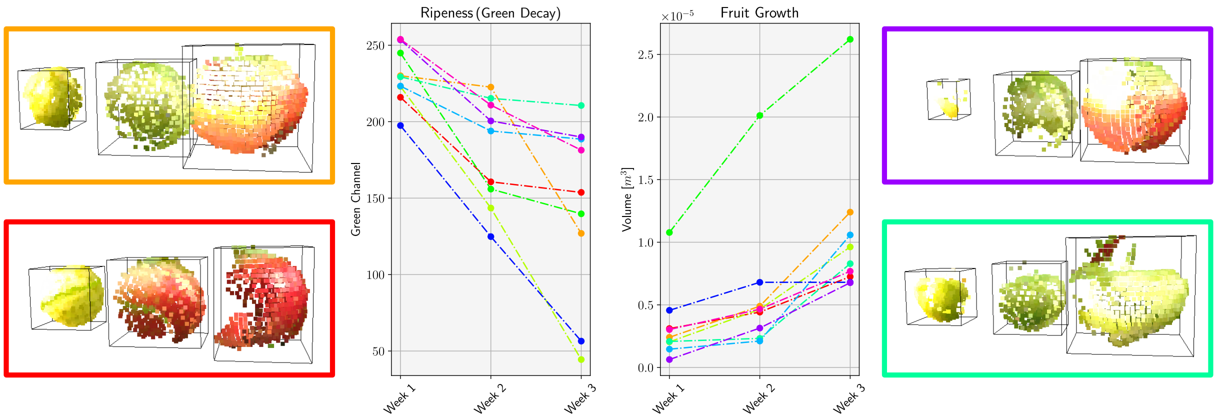

Fig. 5: Growth tracking of individual strawberries. In the middle, we plot the evolution of phenotypic traits from nine matched fruits, such as color (as a proxy for ripeness) and volume of individual fruits. On the sides, we show four different point clouds of the same fruit over time out of the nine described in the plots, where the rectangles' color matches the line color.

Interestingly, by adding the radius term, we see a significant increase in the results. Notably, we are able to reach an F-score above 60% only using our proposed descriptor. The results highlight the importance of our descriptor in robustly tracking fruits over time. We expected such behavior, given that this term describes the relative position of each fruit in its neighborhood. This supports our initial idea that neighboring fruits have similar deformations. In fact, fruits close to each other most likely belong to the same plant branch, which is one of the most critical factors in terms of the temporal displacement of the fruits. When using radius information with our descriptor, we see a slight gain in performance, showing our descriptor's robustness.

D. Monitoring Traits of Individual Fruits

We showcase an example of fruits' growth estimation to support the claim that our approach allows monitoring fruit traits such as color and radius over time. For each fruit, we compute the color by averaging its points' RGB values and estimating its volume by approximating it with a sphere. Given the computed associations between the three sessions of our dataset, we can track the aforementioned traits over time. In Fig. 5, we plot the evolution of each fruit's green component and volume. We show the green component of the fruits' color since unripe strawberries are green and turn red when they are ready to be harvested. On the side, we show the point clouds representing the same fruit scanned on different weeks to appreciate the development of individual fruits. The colors of the rectangles match the line colors in the plot. Note that we plot only a tiny subset of the fruit's tracking for better visualization.

Features	Precision [%]	Recall [%]	F-score [%]
positions	56.22	45.70	50.42
positions + radius	61.16	57.97	59.52
positions + descriptor	**66.99**	62.88	64.87
positions + descriptor + radius	66.34	**64.11**	**65.21**

TABLE III: Ablation Study. The results highlight the importance of our descriptor in robustly tracking fruits over time.

V. Conclusion

In this paper, we address the problem of finding corresponding fruits between point cloud data collected at different points in time. We propose a 4D matching pipeline that is robust to structural changes of the fruit, such as color, radius, and fruit position. This enables accurate tracking of fruit traits over time and monitoring the growth state. We implemented and evaluated our approach on real-world greenhouse data collected over three weeks and provided comparisons with traditional geometric descriptors. The experiments show that our proposed solution has superior fruit tracking performance and can robustly associate fruits over multiple weeks.

In future work, we will investigate ways to remove the terrestrial laser scanner from the pipeline, as TLS scanning slows down data collection and often requires a specialized operator. Instead, we consider automatizing the mapping process by relying solely on sensors commonly mounted on a robot. For instance, we could first use a SLAM system to estimate the robot's odometry and a course 3D map of the environement, subsequently detect and track strawberries throughout consecutive image frames, and finally compute their global 3D position. This approach poses different challenges as commercial range sensors either have low accuracy in outdoor environments (RGB-D cameras) or lack color information (robotics LiDAR) which is a crucial cue in the agricultural context. Moreover, each mentioned step might produce a certain amount of error propagating throughout the entire pipeline and compromising the final result. Nonetheless, despite the complexity of this solution, this allows collecting significant amounts of data that farmers can analyze and exploit in different scenarios, such as predicting the necessary resources for harvesting, the prompt localization of sick plants, or finding the best growth conditions.

Acknowledgments

We thank the staff from the Institute of Geodesy and Geoinformation Bonn for supporting us with the data collection and the team of the Campus Klein Altendorf for support and providing the glasshouse facility.

References

[1] 3D Object Recognition in Cluttered Scenes with Local Surface Features: A Survey. *IEEE Trans. on Pattern Analalysis and Machine Intelligence (TPAMI)*, 36(1), 2014.

[2] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proc. of the Intl. Conf. on Knowledge Discovery and Data Mining*, 2019.

[3] J. Behley, V. Steinhage, and A. Cremers. Performance of Histogram Descriptors for the Classification of 3D Laser Range Data in Urban Environments. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2012.

[4] J. Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517, 1975.

[5] 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, 2009.

[6] 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.

[7] 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.

[8] 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.

[9] N. Chebrolu, F. Magistri, T. Läbe, and C. Stachniss. Registration of Spatio-Temporal Point Clouds of Plants for Phenotyping. *PLOS ONE*, 16(2):e0247243, 2021.

[10] 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.

[11] B. Drost, M. Ulrich, N. Navab, and S. Ilic. Model globally, match locally: Efficient and robust 3D object recognition. In *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2010.

[12] T. Duckett, S. Pearson, S. Blackmore, B. Grieve, W. Chen, G. Cielniak, J. Cleaversmith, J. Dai, S. Davis, C. Fox, P. From, I. Georgilas, R. Gill, I. Gould, M. Hanheide, A. Hunter, F. Iida, L. Mihalyova, S. Nefti-Meziani, G. Neumann, P. Paoletti, T. Pridmore, D. Ross, M. Smith, M. Stoelen, M. Swainson, S. Wane, P. Wilson, I. Wright, and G. Yang. Agricultural robotics: The future of robotic agriculture. *arxiv preprint:1806.06762v2*, 2018.

[13] F. Fiorani and U. Schurr. Future scenarios for plant phenotyping. *Annual Review of Plant Biology*, 64:267–291, 2013.

[14] M. Fischler and R. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM*, 24(6):381–395, 1981.

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

[16] J.A. Gibbs, M. Pound, A. French, D. Wells, E. Murchie, and T. Pridmore. Active vision and surface reconstruction for 3d plant shoot modelling. *IEEE/ACM Trans. on Computational Biology and Bioinformatics*, 17(6):1907–1917, 2019.

[17] F. Görlich, E. Marks, A.K. Mahlein, K. König, P. Lottes, and C. Stachniss. UAV-Based Classification of Cercospora Leaf Spot Using RGB Images. *Drones*, 5(2), 2021.

[18] 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.

[19] A. Johnson and M. Hebert. Using spin images for effcient object recognition in cluttered 3D scenes. *IEEE Trans. on Pattern Analalysis and Machine Intelligence (TPAMI)*, 21(5):433–449, 1999.

[20] J. Kierdorf, I. Weber, A. Kicherer, L. Zabawa, L. Drees, and R. Roscher. Behind the leaves: Estimation of occluded grapevine berries with conditional generative adversarial networks. *Frontiers in Artificial Intelligence*, 5:830026, 2022.

[21] H. Kuhn. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97, 1955.

[22] C. Lehnert, D. Tsai, A. Eriksson, and C. McCool. 3d move to see: Multi-perspective visual servoing for improving object views with semantic segmentation. In *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2018.

[23] 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.

[24] 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.

[25] F. Magistri, E. Marks, S. Nagulavancha, I. Vizzo, T. Läbe, J. Behley, M. Halstead, C. McCool, and C. Stachniss. Contrastive 3d shape completion and reconstruction for agricultural robots using rgb-d frames. *IEEE Robotics and Automation Letters (RA-L)*, 7(4):10120–10127, 2022.

[26] E. Marks, F. Magistri, and C. Stachniss. Precise 3D Reconstruction of Plants from UAV Imagery Combining Bundle Adjustment and Template Matching. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2022.

[27] S. Nuske, S. Achar, T. Bates, S. Narasimhan, and S. Singh. Yield Estimation in Vineyards by Visual Grape Detection. In *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2011.

[28] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin. Matching 3D models with shape distributions. In *Proc. of the Int. Conf. on Shape Modeling and Applications*, 2002.

[29] J. Rodriguez-Sanchez and C. Li. Autonomous ground system for 3d lidar based field phenotyping. *Earth and Space Science Open Archive*, page 1, 2021.

[30] R. Rusu, N. Blodow, and M. Beetz. Fast point feature histograms (fpfh) for 3d registration. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, pages 3212–3217, 2009.

[31] I. Sa, Z. Ge, F. Dayoub, B. Upcroft, T. Perez, and C. McCool. DeepFruits: A Fruit Detection System using Deep Neural Networks. *Sensors*, 16(8):1222, 2016.

[32] A. Segal, D. Haehnel, and S. Thrun. Generalized-ICP. In *Proc. of Robotics: Science and Systems (RSS)*, 2009.

[33] C. Smitt, M. Halstead, T. Zaenker, M. Bennewitz, and C. Mccool. PATHoBot: A Robot for Glasshouse Crop Phenotyping and Intervention. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2021.

[34] 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)*, 2018.

[35] C. Stachniss. *Exploration and Mapping with Mobile Robots*. PhD thesis, University of Freiburg, Department of Computer Science, 2006.

[36] C. Stachniss, J. Leonard, and S. Thrun. *Springer Handbook of Robotics, 2nd edition*, chapter Chapt. 46: Simultaneous Localization and Mapping. Springer Verlag, 2016.

[37] S. Sun, C. Li, P.W. Chee, A.H. Paterson, C. Meng, J. Zhang, P. Ma, J.S. Robertson, and J. Adhikari. High resolution 3d terrestrial lidar for cotton plant main stalk and node detection. *Computers and Electronics in Agriculture*, 187(C):106276, 2021.

[38] J.P. Underwood, C. Hung, B. Whelan, and S. Sukkarieh. Mapping almond orchard canopy volume, flowers, fruit and yield using lidar and vision sensors. *Computers and Electronics in Agriculture*, 130:83–96, 2016.

[39] O. Vysotska and C. Stachniss. Effective Visual Place Recognition Using Multi-Sequence Maps. *IEEE Robotics and Automation Letters (RA-L)*, 4:1730–1736, 2019.

[40] A. Walter, R. Finger, R. Huber, and N. Buchmann. Opinion: Smart farming is key to developing sustainable agriculture. *Proc. of the National Academy of Sciences*, 114(24):6148–6150, 2017.

[41] M. Watt, F. Florani, B. Usadel, U. Rascher, O. Muller, and U. Schurr. Phenotyping: new windows into the plant for breeders. *Annual Review of Plant Biology*, 71:689–712, 2020.

[42] 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.

[43] 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(2):3599–3606, 2021.

[44] T. Zaenker, C. Lehnert, C. McCool, and M. Bennewitz. Combining local and global viewpoint planning for fruit coverage. In *Proc. of the Europ. Conf. on Mobile Robotics (ECMR)*, 2021.