Estimating 4D Data Associations Towards Spatial-Temporal Mapping of Growing Plants for Agricultural Robots

Luca Lobefaro   Meher V. R. Malladi   Olga Vysotska   Tiziano Guadagnino   Cyrill Stachniss

Abstract— **Our world is non-static, and robots should be able to track its changing geometry. For tracking changes, data associations between 3D points over time are key. In this paper, we investigate the problem of associating 3D points on plant organs from different mapping runs over time while the plants grow. We achieve a high spatial-temporal matching performance by combining 3D RGB-D SLAM, visual place recognition, and 2D/3D matching exploiting background knowledge. We showcase our approach in a real agricultural glasshouse used to grow sweet peppers, using RGB-D observations from a mobile robot traversing the environment. Our experiments suggest that with our approach, we can robustly make data associations in highly repetitive scenes and under changing geometries caused by plant growth. We see our approach as an important step towards spatial-temporal data association for robotic agriculture.**

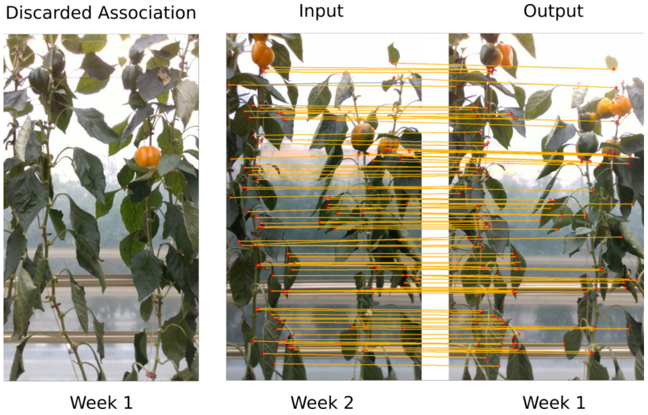


Fig. 1: Temporal associations between 3D point clouds re-projected on the images for visualization (we show only a subset of them for clarity). Our method works in presence of visual changes due to plant deformation and strong visual aliasing of the plants. It is able to match the images of the same plant taken in different weeks. At the same time, it can discard images that look similar but depict different plants, thanks to the combination of sequential place recognition and conditioned feature matching.

I. Introduction

Monitoring and tracking changing geometries over time is a common task for autonomous systems, as our world is not static. Whenever intelligent systems should model or understand how the world evolves, monitoring changes over time becomes relevant. In agriculture, measuring plant development over time is a key element in phenotyping and central for decision management or making breeding decisions.

Whenever the growth of a plant should be monitored over time, data associations between individual parts of the plant need to be estimated. These associations are the basis for computing time-aligned 3D models. Computing correct data associations is one of the most challenging problems in mapping and SLAM—it is well known that the SLAM problem simplifies dramatically with perfect data associations. Obtaining correct data associations, however, is challenging. This is especially true for highly repetitive environments and those undergoing continual changes. If both aspects come together, this task becomes even more challenging.

We aim to investigate means for computing data associations in agricultural environments such as glasshouses or an orchard monitored by a mobile robot. We explicitly focus on the data association problem and aim to find ways to associate plant organs over time. An example of such a data association over time is depicted in Fig. 1, in which associations between plants have been created. We do not address the complete 4D SLAM problem in the sense of building a spatial-temporal 4D model, as it is not fully clear how a 4D model should look like in the domain of growing plants. We, however, target realistic robotic data acquisitions in commercial glasshouses. We explicitly exclude artificial setups such as high-precision laser scanning in the lab, as done in prior work, including our own ones [6], [16], [22].

The main contribution of this paper is a practical, novel approach toward spatial-temporal mapping of plants in the agricultural domain. Our approach establishes data association between highly repetitive plants, even if they undergo changes over time. Our approach combines 3D RGB-D SLAM to build local models of plants while the robot traverses the glasshouse. When mapping the space again, we employ visual place recognitions exploiting sequences of images. Based on consistent image pairs, we can create the correspondences between observed plants across time, considering that stem locations are fairly static. Given the image data associations and plant associations, we then associate plant surface points across time, resulting in links

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, and with the Lamarr Institute for Machine Learning and Artificial Intelligence, Germany. Olga Vysotska is with Microsoft Mixed Reality & AI Lab Zurich, Switzerland; the contributions to this work have been done while being with the University of Bonn.

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 European Union's Horizon Europe research and innovation programme under grant agreement No 101070405 (DigiForest).

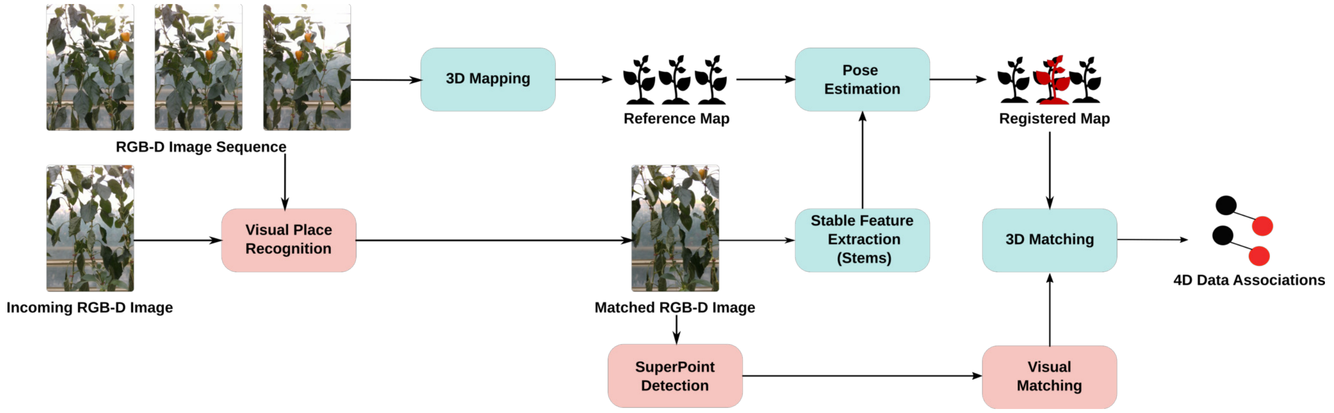


Fig. 2: Our proposed pipeline for 4D data associations combining 3D SLAM for mapping a single mission, visual place recognition based on image sequences, scan registration, and conditioned feature matching.

between plant parts over time. This sequential approach allows for generating robust 4D associations over time and space. We show the capabilities of our method using a real robot equipped with consumer RGB-D cameras operating in an agricultural glasshouse used to grow sweet peppers.

II. Related Work

Robotics for plant phenotyping has recently gathered attention in the scientific community. Das et al. [8] present methods for automated crop monitoring using a custom sensor suite. Riggio et al. [21] illustrate the use of RGB images for yield estimation in vineyards exploiting prior structural information. A central point in performing automatic phenotyping is considering the evolution of plants' traits over time. Chebrolu et al. [5], [7] propose robot localization and long-term registration using aerial images for precision agriculture. Chebrolu et al. [6] exploit a skeleton structure of the plant to find stable associations between different growth stages. The main idea of finding stable features to perform plant registration is an inspiration for our work, which, instead of using a skeleton, operates directly on images and raw point clouds.

A standard approach to 3D point matching is ICP by Besl and McKay [4]. Huang et al. [14] assume rigid scene and offer a recent survey on point cloud matching and registration. However, most of these methods require a good prior about the relative pose between the two point clouds to provide suitable matches. This is often not given in agricultural environments due to the non-rigid deformation that natural structures are affected over time.

Visual place recognition (VPR) can be a powerful tool to find images taken from the same location and is essential in highly repetitive environments. Classical approaches use local features such as SURF [2] or SIFT [15] to perform image matching. However, these methods can be susceptible to scene changes. Arandjelovic et al. [1] introduce a low-dimensional global neural descriptor called NetVLAD, now a standard for visual place recognition techniques. Recently, fully deep learning-based approaches have also been proposed in this field. Masone et al. [17] provide a comprehensive overview of deep learning approaches for visual place recognition.

Milford et al. [18] propose to exploit sequential information of the input data for VPR in autonomous driving and transfer the single image matching into a sequence matching problem using multiple images. Vysotska et al. [25] follow the idea of sequence-to-sequence alignment and formalize a sequential matching as an online informed graph search. This method constitutes one of the core modules of our pipeline, as it is robust to substantial visual changes. Additionally, our approach exploits 3D information to cope with highly repetitive natural environments commonly observed in the agricultural domain when monitoring rows of similar-looking plants.

Currently, there is no method for temporal matching rows of plant point clouds that uses only RGB-D information to keep track of the evolution of traits in time. Dong et al. [10] provide a method for robust data association of both, spatial and temporal domains, exploiting a factor graph for 4D reconstruction. Nevertheless, their approach uses external information, such as IMU and GPS information, while operating in an open field and observing numerous static elements. In contrast, Magistri et al. [16] perform a non-rigid registration of plants by finding temporal matches between skeletal structures. Extracting plant skeletons requires high-resolution 3D data, complete views, and isolation of plants. Thus, these are too strong assumptions when operating in a commercial glasshouse. Furthermore, Magistri et al. [16] work only with single plant instances, whereas our method is well-suited for entire rows of plants. Closely related to our application is the work by Riccardi et al. [20] that introduce a temporal fruit tracking and matching system, which uses an expensive automotive LiDAR sensor and keeps track of fruits. It tracks strawberries without considering other plant components leading to an easier problem as the berries are fairly distinct. While their approach handles a sequence of observations, it does not target the application of matching temporally distant sequences.

To the best of our knowledge, we propose in this paper the first method to perform 4D matching of plants between incoming RGB-D image streams and a prior map acquired at a different point in time, without using any positional information despite aliasing due to repetitive structures of plants in a glasshouse.

III. Our Approach

We propose a pipeline that computes data associations between temporally distant point clouds recorded from plants in a glasshouse. In particular, we perform the matching between a map of the environment recorded at a previous data collection and a point cloud generated from an RGB-D stream captured in a different data recording session. This allows us to keep track of the non-rigid changes that plants undergo during growth without relying on any additional information about the pose of the robot.

In this section, we provide the details of our approach, starting from how we constructed the 3D representation of the first data acquisition used as a reference model in our pipeline for the next run (Sec. III-A). Then, given the new incoming RGB-D image stream, we exploit the approach proposed by Vysotska et al. [25] to perform visual place recognition and understand which part of the reference model we are currently observing. The output of this module is image correspondences that we can turn into a rough estimate of the robot's position in the prior model (Sec. III-B). We further refine the pose estimate using scan alignment. As plants undergo non-rigid changes in structure, we perform the alignment on temporally stable features as explained in Sec. III-C. Finally, in Sec. III-D, we present the temporal association algorithm, where we use a combination of descriptor matching and 3D nearest neighbor search on the aligned plants' point clouds which is feasible given the prior steps that constrain the associations. We sketch our proposed pipeline in Fig. 2.

We perform our measurements using the robotic platform introduced by Smitt et al. [24], which has three vertically mounted RGB-D cameras that provide side-view. The robot moves along the plant rows through the glasshouse. We use only the middle camera to perform visual localization in Sec. III-B because it captures most information about the plants. Otherwise, we use all three cameras to generate the point clouds used in our approach.

A. 3D Mapping of Plant Rows

In this section, we explain how we obtain the initial 3D representation of the environment at one point in time.

Let $\mathcal{S}_t$ be a sequence of images captured with an RGB-D sensor at time t, which is an ordered set of images $S_t^i = \{I_t^1, I_t^2, \ldots, I_t^n\}$, where n is the number of images acquired in the session at time t. We exploit the wheel odometry to associate to $\mathcal{S}_t$ a set of corresponding initial poses $\mathcal{T}_t = \{\mathsf{T}_t^1, \mathsf{T}_t^2, \ldots, \mathsf{T}_t^n\}$, where $\mathsf{T}_t^i \in \mathbb{R}^{4\times 4}$ represent the pose in the world coordinate frame of the session at time t as homogeneous matrices.

Given the RGB-D images and an initial guess of the poses, we can generate a colored 3D point cloud by exploiting the pixel-wise depth d_i of each RGB-D image using the classic pinhole camera unprojection [13]. Following the word by Smitt et al. [24], we maintain only the 3D points belonging to the measured crop row, filtering out all points outside the relevant depth range.

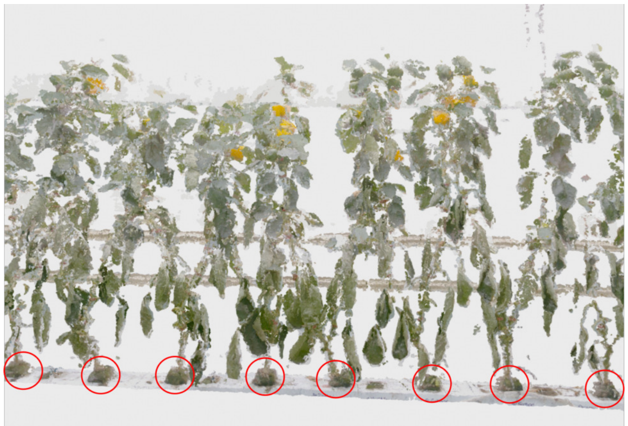


Fig. 3: An example point cloud map produced by our approach. We show in the red circles the stems of the plants that we use as temporally stable features for the registration.

Then, we apply the RGB-D odometry algorithm proposed by Park et al. [19] to refine the pose using the observations. With this approach, both photo consistency and geometry constraints are considered to update the poses T_t^i of each image I_t^i.

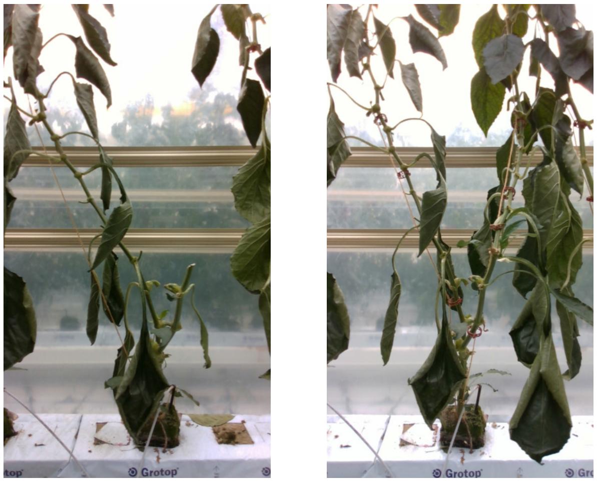
This straightforward mapping system is computationally highly efficient and can run on a single CPU core. Fig. 3 shows an example of the aggregated point cloud map produced with this approach using the refined poses.

B. Sequence Alignment over time using Visual Localization

Let us now consider a new recording session in the same environment but at a future time $t+1$ e.g., one week later. We have again an ordered set of images $\mathcal{S}_{t+1} = \{I_{t+1}^1, I_{t+1}^2, \ldots, I_{t+1}^m\}$, where m is the number of images acquired in this new session. Our objective is to associate each image I_{t+1}^j to an image I_t^i in the reference sequence $\mathcal{S}_t$, if such a match exists. Once such an image correspondence is known, we can transfer the pose T_t^i to I_{t+1}^j obtaining a reasonable estimate of the robot's current location in the map recorded at time t. This gives an initial global alignment of the missions.

As the environment in which we are operating is changing due to the growth of the plants and is highly repetitive, we have to exploit the sequential nature of the recorded data. We use the approach of Vysotska et al. [25] for robust image sequence alignment over time. In the following, we briefly discuss the method of Vysotska et al. [25] as it is a core component to make our pipeline robust and align missions even under drastic appearance changes.

For each image I_t^i, we compute a global visual feature vector. More specifically, we use NetVLAD descriptor [1] as it is robust to illumination changes, which we often experience in our data. As a result, we obtain a low-dimensional representation $\boldsymbol{d}_t^i \in \mathbb{R}^D$ for each image I_t^i. We



(a) Query

(b) Reference

Fig. 4: Example of visual place recognition result. (a) A frame from the query sequence, (b) the result of the visual place recognition algorithm. The correct match is found even with many missing parts of the plant.

use the cosine distance to determine the similarity $\text{sim}(I_t^i, I_t^j)$ between images I_t^i and I_t^j:

$$\text{sim}(S_t^i, S_t^j) = \frac{{\boldsymbol{d}_t^i}^\top \boldsymbol{d}_t^j}{\|\boldsymbol{d}_t^i\| \|\boldsymbol{d}_t^j\|}. \qquad (1)$$

We perform a search in a data association graph to obtain the best possible match. More specifically, we use a directed acyclic graph $G = (\mathcal{X}, \mathcal{E})$ as a data structure. Following the notation of Vysotska et al. [25], $\mathcal{X}$ are the nodes, and $\mathcal{E}$ are the edges. Each node represents the fact that an image I_t^i is compared to an image in I_{t+1}^j, and each edge represents a possible movement of the robot between image recordings.

To model that the robot can move at different speeds or the cameras can have different frame rates, each node in the data association graph is connected to K edges. For larger values of K, we have more nodes connected with the current one, so more images in the sequence are considered as possible next match. In our approach, we adopt $K = 5$ as in the original work of Vysotska et al. [25].

Another problem is that all the plants share similar appearances, making it challenging to differentiate plants. In our approach, we exploit the idea of lazy data associations originally proposed by Hähnel et al. [12] in the context of SLAM.

We allow for committing late on a specific data association through a graph search procedure. To enable an online execution, we build the graph incrementally while searching for possible matches. The key idea is to use a heuristic about image matching cost to enable a fast and effective search in the data association graph. We refer to [25] for details.

The result of this approach is a sequence-consistent set of matched images with pairs (i, j), $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$. From that, we can determine the pose alignment over time $\mathcal{T}_{t+1}$ by setting $\mathsf{T}_{t+1}^j = \mathsf{T}_t^i$ as the initial guess.

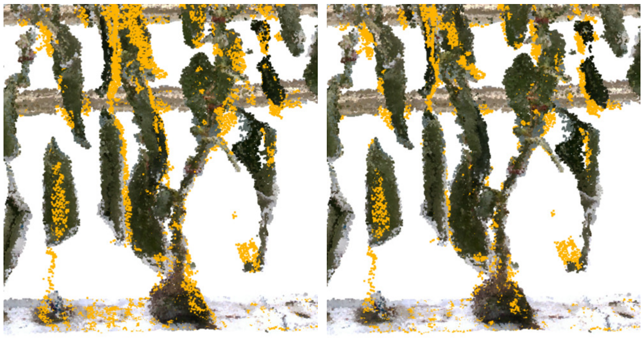


Fig. 5: Illustration of the point cloud registration using our matching based on the stable stem area. We show in yellow the query point cloud. On the left, the initial alignment using the pose computed by visual localization. On the right, the refined result obtained by registering the plants' stem areas.

Using only visual information, we can perform localization of the robot's poses in the new mission to the previously mapped environment, even with substantial visual changes and aliasing. Fig. 4 shows that our approach can find matches between images despite missing parts of the plant and different leaves density.

C. Plant Registration using Stable Features

Consider now an incoming image I_{t+1}^j from the sequence $\mathcal{S}_{t+1}$ and the corresponding associated poses T_{t+1}^j. As described in Sec. III-A, we can produce a colored 3D point cloud using the depth information and bring it into the same frame of the reference map. Fig. 5 (left) shows an example of a point cloud obtained from a single image unprojected into the reference map using only the positional information obtained with visual localization described in Sec. III-B. To achieve a plant-to-plant alignment, we propose to extract points that do not change between sessions to perform an initial point cloud registration of the plants. This is useful to correct the pose obtained with the visual localization approach to obtain a good overlapping between the current measurement and the target map and thus correct the set of poses T_{t+1}.

We adopt the plants' stems (see Fig. 3) as temporally stable features. In particular, they are extracted by simply applying a threshold on the z-axis to the point clouds. In this way, we can register plants without considering moving parts such as leaves and fruits, which can considerably change shape and position between the two sessions.

We perform point cloud registration using the plane-to-plane ICP approach originally proposed by Segal et al. [23]. Fig. 5 (right) shows the result of the registration algorithm, where the plants overlap along stable stem areas. As expected, we do not have a perfect overlapping of leaves and fruits due to changes in the plant geometry.

D. Temporal Matching across Whole Plants using Visual Feature Correspondances

To obtain a good association between different parts of the plants, we need a representation that is stable under non-rigid changes that the plant can undergo between different recording sessions. We exploit visual information to determine correspondences between point clouds while considering the alignment done so far. In particular given the incoming image S^j_{t+1} and the associated T^i_{t+1} from the sequence recorded at time t, we can compute local descriptors to determine which part of the pictures is relevant. We propose to use SuperPoint, a neural local feature extractor developed by Detone et al. [9] that is robust to visual changes.

Once we have the local descriptors for both images, we perform a matching between them, finding keypoints that represent the 3D points on the plants. Then, we apply a RANSAC schema [11] to filter outliers, using the corresponding Homography as a discriminator.

From the images, these matches are converted into 3D point correspondences by unprojecting each keypoint into the corresponding point cloud. For each unprojected point, the nearest neighbor inside the map is found. The nearest neighbor search is performed efficiently by computing a KD Tree representation of the point clouds as proposed by Bentley et al. [3]. Only matches for which the Euclidean distance is under a threshold of $0.5m$ are considered. Furthermore, matches are filtered using the same threshold on the z-axis used in the mapping step. This geometric filtering is possible thanks to the registration of the point clouds, which are now in a common reference frame. Fig. 7 shows the matches computed by our pipeline before and after the geometric filtering. By exploiting the 3D information, our system can discard most of the associations in the background and focus on semantically meaningful parts of the plant.

IV. Experimental Evaluation

The main focus of this work is the development of a method to perform temporal data associations between individual plants' organs in order to be able to keep track of their changes over time. The tracked points are 3D points on the plants' surfaces and stem from the plant recorded at different acquisitions while plants grow.

We present our experiments to show the capabilities of our method in finding correspondence on a real agricultural glasshouse dataset, with highly repetitive scenes and non-rigid changes. We want to point out how our approach is capable to address these results using only a single image and a previously built 3D map, without any prior knowledge about the pose of the incoming measurement.

A. Data Collection

To collect our data, we used the robotic platform described by Smitt et al. [24] operating in a glasshouse for growing sweet pepper in Bonn, Germany. In particular, 3 RGB-D cameras have been utilized to capture sweet peppers plants in an advanced growth stage. The sensors are Intel RealSense D435i, where infrared information is used to obtain depth information in addition to RGB data.

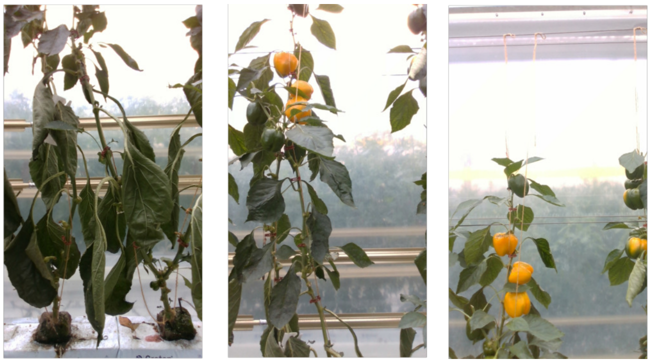


Fig. 6: From left to right: images acquired from camera 1, camera 2 and camera 3.

TABLE I: Quantitative results for our approach.

Total matches	Avg. matches per image	Total Precision	Avg. Precision per image
16869	168	0.957	0.954

We have a FOV overlap of about 20% between the three cameras, which are placed one on top of the other. Fig. 6 shows the images acquired from the three cameras, and it is easy to visualize the point of view utilized.

In particular, we collect two datasets, one week apart from each other. The first, used as the reference, was collected on the 19th of September, 2022. The second, on the 26th of September 2022. The plants' shapes have changed between the two missions; leaves have grown, fruits have ripened and changed color, and some have even been harvested.

B. Quantitative Evaluation

Evaluating the 3D points matches of non-rigid objects is difficult, especially with an unlabelled dataset if no plant organ instance is available. For this reason, we performed the evaluation projecting each 3D match on the original images. We can then evaluate the quality of each match by manually inspecting the corresponding image points to check if they represent the same entity, a long manual procedure.

Because of the contained speed of the robot during the recording, we have a low displacement between consecutive frames. For this reason, we considered one frame every 10 in the evaluation to ensure a sufficient difference in the portion of the field that is observed. As a result, we consider a total of 100 frames, each made by three images, for a total of $16,869$ associations that we manually analyzed in our evaluation.

We measure the precision of our method for temporal matching by computing the number of correct associations among all reported ones. In Tab. I, we show the number of associations and the corresponding precision i.e., the number of correct matches over the total reported associations, both as average per image and as a cumulative metric. In all the images, the result of the visual localization was always correct, with an accuracy of 100%. This provides us with a good initial guess for the plant registration algorithm. Our pipeline achieves more than 95% precision on the temporal matches.

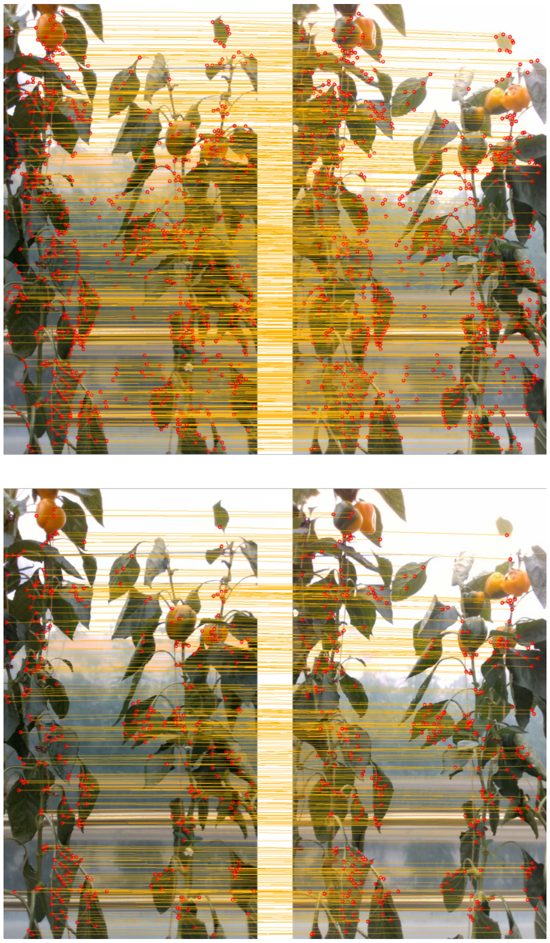


Fig. 7: On the top, the associations found by matching only SuperPoint descriptors. On the bottom, the result obtained by filtering the matches exploiting the 3D information.

C. Qualitative Evaluation

As only providing the precision in a quantitative evaluation is one part of the story — but the only one that this type of evaluation allows us to do — we also provide a set of qualitative example images illustrating the matches in Fig. 9. Furthermore, in Fig. 8, we show qualitatively the result of a baseline that matches SIFT descriptors from a query image to the reference image sequence to find temporal matches. As expected, this vanilla approach cannot find the corresponding image in the reference sequence. As a consequence, it fails to find good correspondence between the two plants, whereas our system can solve both problems and allows us to find good 3D temporal matches, as we can see in Fig. 9.

In Fig. 7, we qualitatively show that exploiting the 3D information is essential to solve the data association problem in highly changing environments, such as glasshouses. In particular, we compare purely visual feature matching with our 3D data association pipeline. As we can see in the pictures, we can filter most of the false positive matches by relying on geometric information. This is possible thanks to our registration algorithm that exploits stable temporal features and allows us to have both plant point clouds in the same reference frame.

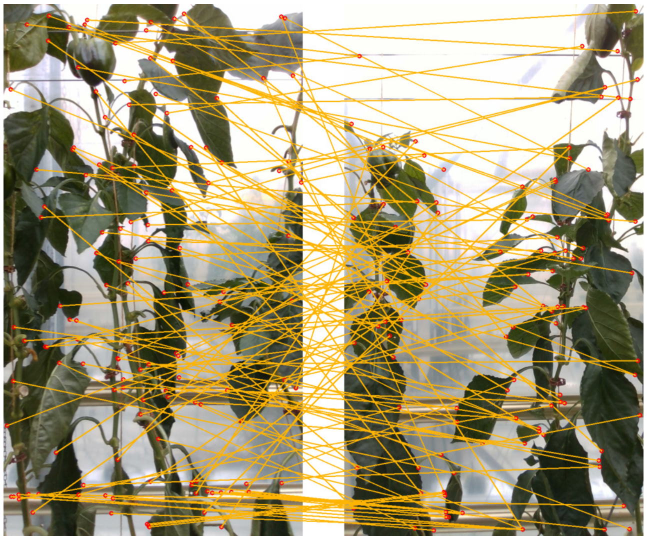


Fig. 8: Example of using a nearest neighbor search on SIFT descriptors to perform visual place recognition for a query image. Due to the highly repetitive structure of the glasshouse rows, this approach fails in finding the corresponding reference image.

V. Conclusion

In this paper, we presented a novel approach for 4D data association in a real agricultural glasshouse. Our approach operates in highly repetitive scenes under changing geometries caused by plant growth. Our method combines 3D RGB-D SLAM, visual place recognition, and background knowledge to address the data association problem. This allows us to successfully obtain temporal matches between 3D point clouds over time, using only a single RGB image with depth information to keep track of the evolution of plants' shapes in time.

We implemented and evaluated our approach on a real agricultural glasshouse dataset and supported the claims made by this paper. The experiments suggest that spatial-temporal data association is possible for natural and dynamic environments by means of a single RGB-D camera. Our work is an important step toward the 4D reconstruction of agriculture fields, especially suitable for phenotyping and breeding decisions. Plants' geometry evolution tracking using consumer RGB-D cameras can potentially enhance autonomous in-field operations for robots in agriculture.

References

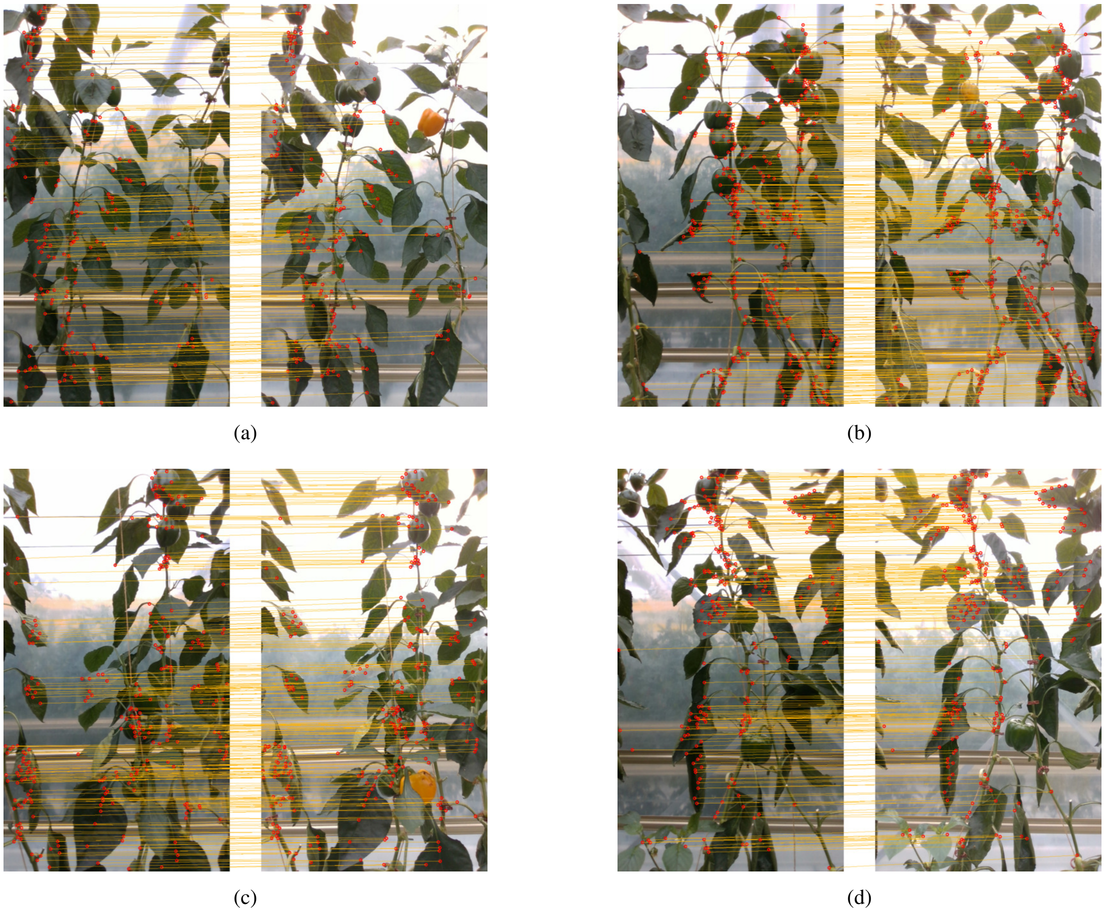
[1] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic. NetVLAD: CNN Architecture for Weakly Supervised Place Recognition. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[2] H. Bay, A. Ess, T. Tuytelaars, and L.V. Gool. Speeded-up robust features (SURF). *Journal of Computer Vision and Image Understanding (CVIU)*, 110(3):346–359, 2008.

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

[4] P. Besl and N. McKay. A Method for Registration of 3D Shapes. *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, 14(2):239–256, 1992.

[5] 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.



(a)

(b)

(c)

(d)

Fig. 9: Qualitative results of our approach. In (a) we show an example where a fruit has been picked between the two sessions. In (b) we show an example of different fruit growth stages. In (c) and (d) we have different illumination conditions and different positions of leaves.

[6] 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.

[7] N. Chebrolu, P. Lottes, T. Laebe, and C. Stachniss. Robot Localization Based on Aerial Images for Precision Agriculture Tasks in Crop Fields. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2019.

[8] J. Das, G. Cross, C. Qu, A. Makineni, P. Tokekar, Y. Mulgaonkar, and V. Kumar. Devices, systems, and methods for automated monitoring enabling precision agriculture. In *Proc. of the International Conf. on Automation Science and Engineering (CASE)*, 2015.

[9] D. DeTone, T. Malisiewicz, and A. Rabinovich. SuperPoint: Self-Supervised Interest Point Detection and Description. In *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017.

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

[12] D. Hähnel, W. Burgard, B. Wegbreit, and S. Thrun. Towards lazy data association in slam. In *Proc. of the Intl. Symposium on Robotic Research (ISRR)*, pages 421–431, Siena, Italy, 2003.

[13] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004.

[14] X. Huang, G. Mei, J. Zhang, and R. Abbas. A Comprehensive Survey on Point Cloud Registration. *arXiv preprint*, arXiv:2103.02690, 2021.

[15] D. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *Intl. Journal of Computer Vision (IJCV)*, 60(2):91–110, 2004.

[16] 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.

[17] C. Masone and B. Caputo. A survey on deep visual place recognition. *IEEE Access*, 9:19516–19547, 2021.

[18] M. Milford and G. Wyeth. SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2012.

[19] J. Park, Q. Zhou, and V. Koltun. Colored Point Cloud Registration Revisited. In *Proc. of the IEEE Intl. Conf. on Computer Vision (ICCV)*, 2017.

[20] A. Riccardi, S. Kelly, E. Marks, F. Magistri, T. Guadagnino, J. Behley, M. Bennewitz, and C. Stachniss. Fruit Tracking Over Time Using High-Precision Point Clouds. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2023.

[21] G. Riggio, C. Fantuzzi, and C. Secchi. A low-cost navigation strategy for yield estimation in vineyards. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2018.

[22] D. Schunck, F. Magistri, R. Rosu, A. Corneließen, 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.

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

[24] 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.

[25] O. Vysotska and C. Stachniss. Lazy Data Association For Image Sequences Matching Under Substantial Appearance Changes. *IEEE Robotics and Automation Letters (RA-L)*, 1(1):213–220, 2016.