IEEE Robotics and Automation Letters (RAL) paper presented at the
2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
Madrid, Spain, October 1-5, 2018

Robust Long-Term Registration of UAV Images of Crop Fields for Precision Agriculture

Nived Chebrolu Thomas Läbe Cyrill Stachniss

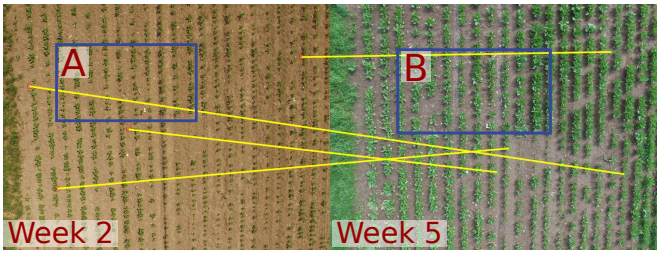
***Abstract*— Continuous crop monitoring is an important aspect of precision agriculture and requires the registration of sensor data over longer periods of time. Often, fields are monitored using cameras mounted on unmanned aerial vehicles (UAVs) but strong changes in the visual appearance of the growing crops and the field itself poses serious challenges to conventional image registration methods. In this paper, we present a method for registering images of agricultural fields taken by an UAV over the crop season and present a complete pipeline for computing temporally aligned 3D point clouds of the field. Our approach exploits the inherent geometry of the crop arrangement in the field, which remains mostly static over time. This allows us to register the images even in the presence of strong visual changes. To this end, we propose a scale-invariant, geometric feature descriptor that encodes the local plant arrangement geometry. The experiments suggest that we are able to register images taken over the crop season, including situations where matching with an off-the-shelf visual descriptor fails. We evaluate the accuracy of our matching system with respect to manually labeled ground truth. We furthermore illustrate that the reconstructed 3D models are qualitatively correct and the registration results allow for monitoring growth parameters at a per plant level.**

I. Introduction

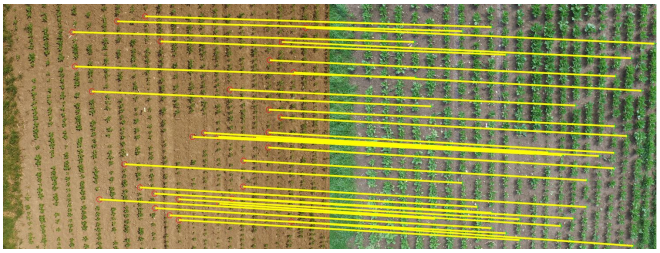
Automated crop monitoring is an important aspect of precision farming, because it allows the farmers to make informed decisions regarding when, where, and how much fertilizer or pesticide to apply in the field as well as to improve yield estimation. With the wide availability of commercial UAVs, it has become fairly easy to repeatedly acquire image data of the fields without any expert assistance. This has led to several new applications in the agricultural robotics community [2], [4], [11].

State-of-the-art image registration methods such as [1] are able to register images from a scene and compute a 3D model of the environment [6]. Typically, these methods rely on a visual descriptor such as SIFT, ORB, BRIEF or similar to perform the data association amongst the images. In crop farming, fields and crops are affected by strong visual changes, due to the weather, growing crops, and farm equipment such as tractors affecting the soil as shown in Fig. 2. Most registration methods are not able to cope well with these changes in appearance.

In this paper, we address the problem of registering UAV images of a field recorded over the crop season in the presence of large visual changes caused by crop growth and field management. The main idea of our approach is to take advantage of the fact that the position of crops as well as gaps between crops remains roughly the same over time, even if the visual appearance of the plants itself changes dramatically. A two-image matching example is depicted in Fig. 1. The first row shows SIFT-based correspondences. As it can be seen from the lines connecting the identified corresponding points, SIFT based association is rather poor. Our approach, however, finds better correspondences as seen in the second row.



(a) Failure in visual matching using SIFT



(b) Successful matching using our geometric descriptor

Fig. 1: Matching UAV images taken three weeks apart. Our method uses geometric cues to perform matching successfully whereas matching using SIFT fails in challenging conditions with large visual changes. Fig. 2 shows a zoomed-in view of the field rendering these changes better visible.

The main contribution of this paper is a novel method for registering images of a crop field taken using a UAV across the crop season. Our approach provides better correspondences between images under changing conditions caused by crop growth, weather, and field management. It copes with the visual aliasing problem in crop fields. We achieve this by presenting a descriptor that exploits crop and gap location information along the crop rows which is mostly invariant within the same field over time. This spatial information about crops and gaps is useful for matching images in this application domain.

We make the following three key claims. Using our approach, we are able to (i) match images taken from a UAV during multiple sessions over the field having large visual

All authors are with the University of Bonn, Institute of Geodesy and Geoinformation, Bonn, Germany.

This work has partly been supported by the EC under the grant number H2020-ICT-644227-Flourish.

Fig. 2: Zoomed-in view of the same area on the field three weeks apart. In addition to the vegetation growth, the texture of soil also changes dramatically over time. Texture rich regions such as the tire marks from the tractor in the left image are washed away in the rain while revealing other new objects like the stones embedded in the ground. Such strong changes make it very challenging for visual matching methods to work reliably.

differences across the crop season and thus can (ii) compute a 3D model of the field with a temporal dimension capturing the evolution of growing plants in the field. This model in turn allows us to (iii) monitor crop growth parameters such as leaf area over time. These three claims are supported by our experimental evaluation.

II. Related Work

Recently, several works investigated into robotic applications in the context precision agriculture. Das *et al.* [4] and Bryson *et al.* [2] present various methods for automated monitoring for fields from ground and aerial vehicles. Lottes *et al.* [11] focus on distinguishing crops and weeds for targeted weeding while Kusumam *et al.* [8] detect and localize broccoli heads for selective harvesting. Other works such as [10], [17] have investigated towards analyzing plant growth from multi-spectral images and point clouds.

Several existing methods address the problem of finding data associations amongst images having large differences in visual appearance. Visual localization and place recognition for long-term applications require robust image matching in presence of strong illumination and seasonal changes [14], [18]. A comprehensive survey of the visual place recognition techniques can be found in [13]. Most of these techniques are designed for autonomous driving applications and do not lend themselves to be used for finding matches in field images having a large baseline.

A large corpus of literature exists for matching point patterns in images and other synthetic data. Gold *et al.* [7] and Hancock *et al.* [3] propose different formulations for estimating correspondences from noisy point sets enabling them to deal with deformable objects in the image. Wolfson [19] proposes a hashing based method using invariant properties of transformations to retrieve the correct object from a large database of objects. Our use of geometric descriptor is similar to Moreau *et al.* [20] where they use affine invariant properties to construct a descriptor for tracking planar objects. While these works are not directly applicable for our scenario, we borrow ideas from them and design a new descriptor along with a robust matching procedure suitable for matching point patterns detected in nadir view UAV images.

A highly related work has been proposed by Dong *et al.* [5] that address the problem of matching images from a field across time for the purposes of crop monitoring. They use a SLAM system to fuse the measurements from different sensors such as camera, GPS, IMU, etc. to obtain a high quality estimate of the camera poses and the field structure. This information is used to reject outliers during the data association step and in turn to improve the overall robustness. As the matching still relies on visual information, it is still bound to fail when visual appearance changes dramatically, such as in situations like rain. In contrast, our method is able to deal with such situations since it uses the geometrical information which remains mostly static even if the appearance changes dramatically.

III. Our Approach

A. Assumptions and overview

In this section, we present our approach of matching UAV images of the field taken over multiple data acquisition sessions separated over time. We make the following assumptions regarding the setup:

- the UAV camera is mounted in a near nadir view and there is sufficient overlap between consecutive images;
- the field is roughly planar in a local region (i.e., our approach may not work in wine yards);
- a ground sampling distance so that plants span over several pixels in the image, but this ground sampling distance does not need to be known nor be constant;
- the crops are planted in rows, the row positions and plant spacing, however, is unknown (c.f. Fig. 1).

To register images over multiple sessions, i.e., different UAV flights over the crop season, into a common reference frame, we perform the registration based on four consecutive steps: (i) computing a point based geometric representation for the images exploiting the crop arrangement on the field. This leads to a detection of points, which remains mostly static over different sessions. (ii) We exploit this information to encode the local geometry around each detected point in the image using a scale invariant descriptor. (iii) We then compute point correspondences between overlapping images in a data association step. (iv) Finally, through bundle adjustment followed by a dense matcher, we compute the optimized camera poses and spatially aligned 3D point clouds of different sessions in a common reference frame. The comparison of the point clouds allow us on the one

hand to qualitatively check the registration accuracy and on the other hand to derive crop growth parameters which serve as an application example. In the following, we discuss these steps in more detail.

B. Step 1: Extract geometry information from UAV images

To capture the structure of the crop field that remains invariant over time, we need to identify the static aspects given the images. Once the crops are planted, they do not move and the stems/centers of the crops remain rather fixed over time. Therefore, the locations of the crop centers can be used as a static description of the field. The local constellations formed by these points can be seen as a geometric signature of a particular local region covered by an image. Our current implementation assumes that crops are planted in rows, which is the case for most crop fields as this simplifies the computation of features. The row arrangements is not supposed to be known beforehand but the existence of crop rows is assumed.

We can compute the crop centers using the following procedure which is also illustrated in Fig. 3:

1) Compute the vegetation mask exploiting the excess green index (ExG) given by

$$I_{ExG} = 2\,I_G - I_R - I_B \tag{1}$$

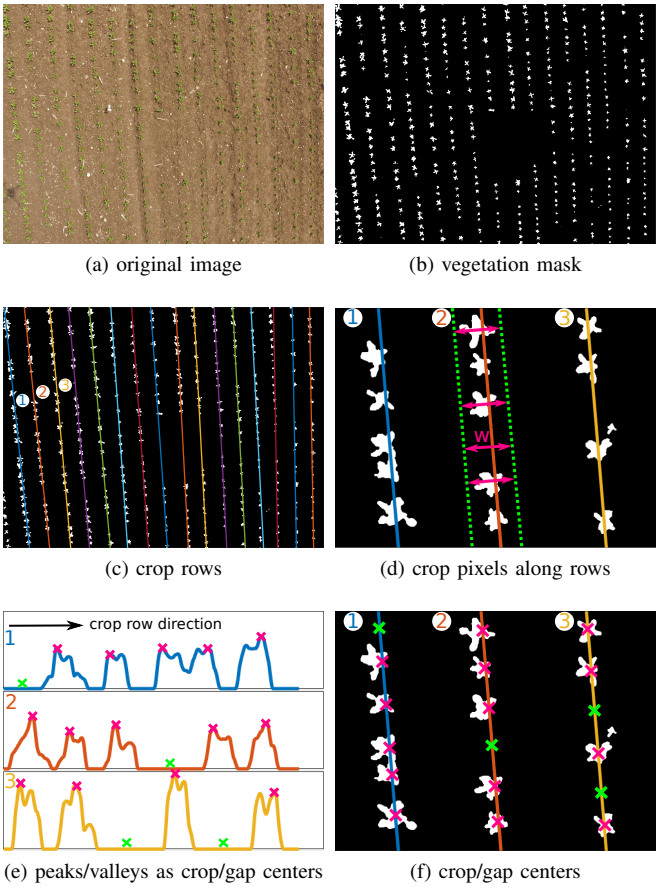
where I_R, I_G and I_B correspond to intensities of the red, blue and green channels of the original image. We then apply a threshold θ given by the Otsu's method [16] on I_{ExG} to get a binarized image (Fig. 3b).
2) Find the lines through the vegetation pixels using the Hough transform for finding crop rows (Fig. 3c).
3) Compute a histogram of vegetation pixels perpendicular to the direction of the detected rows. The width w is taken to be half the inter crop row distance (Fig. 3d).
4) Find the peaks of this histogram to identify the potential centers of the crops (Fig. 3e).

We observed that instead of crop centers, the missing crops, i.e., the gaps within the rows, provide an even more distinctive representation than the crop centers itself. This is particularly the case in the later growth stages, in which nearby crops often overlap. Therefore, we *use the gaps instead of the crop centers as the points representing the geometry in the field* based on the images.

To exploit the gaps instead of the crop centers, we follow the same procedure as for the crop centers, but with the difference that the gaps correspond to the valleys in the histogram computed in Step 3 (marked with green crosses in Fig. 3e). Multiple missing crops occurring consecutively are represented by a single gap point at the center of the valley. Further steps of the method are agnostic to the choice of points or how these points are calculated. Fig. 4 illustrates an example with the extracted crop centers and gaps overlaid on the original image.

C. Step 2: Scale-invariant local geometry descriptor

Given the points identified in Sec. III-B, we aim at encoding the local geometry around each point as a descriptor



(a) original image (b) vegetation mask

(c) crop rows (d) crop pixels along rows

(e) peaks/valleys as crop/gap centers (f) crop/gap centers

Fig. 3: Steps for computing crop and gap centers from the image.

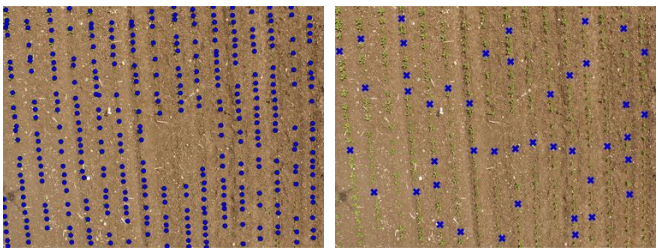


Fig. 4: Extracted points (left for crops, right for gaps) for the same image.

vector in order to facilitate image matching. We exploit the nadir-view assumption of the UAV and thus can assume that images taken during different flights may differ in scale, translational offset, and rotation in the image domain. No affine transformation needs to be considered because of the nadir images. To estimate these transformation parameters, we need a descriptor, which is scale-invariant in addition to be invariant to translation and rotation. We construct an own descriptor for each point P using the ratios of distances and relative angles between the k nearest neighboring points of P in order to meet this criteria. The number of neighbors to consider is a user-defined parameter. The smaller the k, the less expressive/unique is the descriptor of P and the larger k, the more sensitive is the description with respect to outlier points. In our implementation and all experiments,

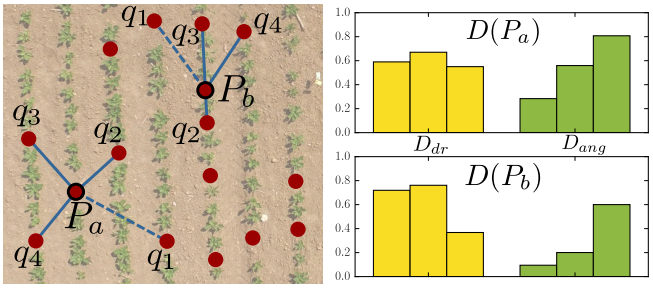


Fig. 5: Computing a scale-invariant descriptor for a point using local geometry. Left: descriptor computation for two gap points P_a & P_b with k = 4. Right: visualizes the corresponding descriptors.

we use $k = 4$, i.e., we consider the four nearest points to P for the computation. In this work, we chose the value of k empirically by evaluating the number of matches obtained for different values of $k \in [3, 8]$ and found that $k = 4$ gave us the best results for our datasets. Consider Fig. 5 for an illustration of how to compute the descriptor D for a given point P, which is defined by the following computations:

1) Given the k nearest points $Q_k = \{q_1, \ldots, q_k\}$ to P, we compute the so-called reference point R as the point in Q_k with the largest distance to P in the image:

$$R = \underset{q \in Q_k}{\operatorname{argmax}} \|P - q\|. \tag{2}$$

Without loss of generality, we assume that q_1 is the reference point R in Q_k and that Q_k is ordered according to the anti-clockwise angle between the line $\overline{Pq_1}$ (dashed line in Fig. 5) and the lines $\overline{Pq_i}$ with $i = 2, \ldots, k$. Computing all the elements of the descriptor in this order makes the descriptor rotation invariant.

2) The descriptor D will be $2(k-1)$-dimensional and consists of two parts of equal size $D = (D_{dr}, D_{ang})$.

3) The first half D_{dr} of the descriptor vector D consists of distance ratios from P to the individual point, normalized by $\|P - q_1\|$:

$$D_{dr} = \left[\frac{\|P - q_2\|}{\|P - q_1\|}, \frac{\|P - q_3\|}{\|P - q_1\|}, \ldots, \frac{\|P - q_k\|}{\|P - q_1\|} \right] \tag{3}$$

We chose distance ratios in the descriptor because they remain invariant to scale.

4) The second half D_{ang} of the descriptor vector D consists of the angles that each point in Q_k has with respect to $\overline{Pq_1}$, normalized by 2π:

$$D_{ang} = \left[\frac{\angle(q_1, P, q_2)}{2\pi}, \frac{\angle(q_1, P, q_3)}{2\pi}, \ldots, \frac{\angle(q_1, P, q_k)}{2\pi} \right]$$

The $\angle(q_1, P, q_i)$ refers to the angle between the lines $\overline{Pq_1}$ and $\overline{Pq_i}$. An example illustrating the descriptor vector computation for two points P_a and P_b is shown in Fig. 5.

D. Step 3: Data association amongst images

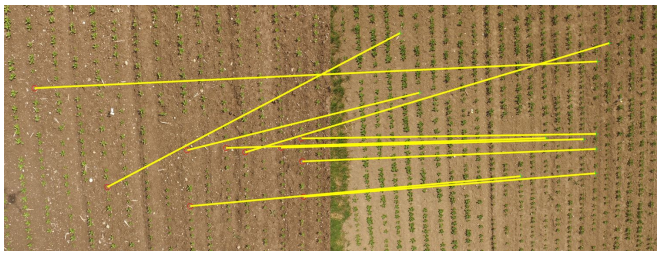
For each image, we compute the set of feature descriptors, one descriptor per detected point in the image. Our data association consists of three steps, the first two steps of the data association are rather standard. First, we compute a pair-wise matching of the descriptors of I_1 and I_2 and compare them using the L_2 norm. In the same spirit as done by Lowe [12] for SIFT matching, we reject those matches that have a high distance under the L_2 norm as well as those where the $\frac{L_{\text{best}}}{L_{\text{second}}} > 0.8$, where L_{best} and L_{second} are the scores for best and the second best match for a descriptor respectively. Second, we compute similarity transformations in a RANSAC loop to identify and remove outliers from the set of corresponding points.

The third step deviates from standard data association approaches. Given that the crop arrangement on the field is highly repetitive, i.e., has a high visual aliasing, a comparably large number of correspondences get eliminated by Lowe's ratio test. In this step, we consider to re-add correspondences in case they are compatible with the transformation found by RANSAC. Thus, the first two steps provide the initial alignment from a potentially quite small set of correspondences, which is typically free of gross errors. Then, we refine the alignment estimate by re-adding those correspondences, which are consistent with the initial guess. These are locally distinct but potentially ambiguous with respect to the descriptor globally and thus were eliminated before. In order to ensure high-quality, one-to-one correspondences, we use the Hungarian method [15] for data association in this recovery step. This step allows us to recover more correspondences that were not obtained directly by descriptor matching by making use of the transformation estimated in RANSAC step. The Hungarian method has a complexity of $\mathcal{O}(n^3)$ and thus is computationally expensive, but given that the number of possible associations with low distances that are compatible with the transformation is typically not too large, this does not turn out to be a computational bottleneck in practice. Fig. 6 depicts the correspondence between two images after each step of the matching.

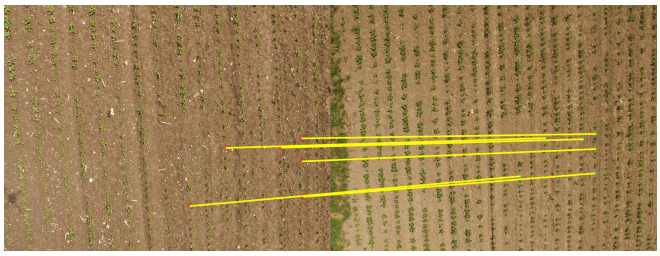
E. Step 4: Point cloud computation using bundle adjustment

In this last step, we perform a pairwise matching between all overlapping images, both spatially and in time across different sessions. Here, we have two options. If we have a low quality GPS information available, we can generate a candidate set of overlapping images from that. This allows for increasing the speed as only a subset of the images must be tested for correspondence. If no GPS information is available, all image pair combinations are tested.

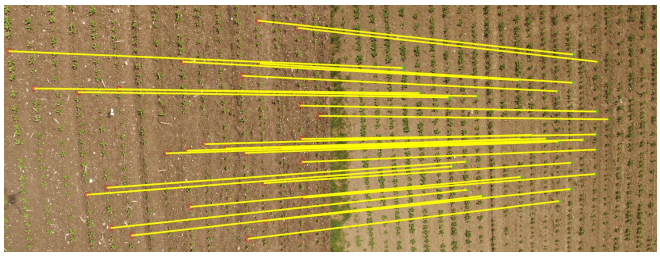
We compute the possible matches between all the potentially overlapping images and feed them into a bundle adjustment procedure [9]. This algorithm combines the pairwise matches to object points with multiple observations and generates approximate values for the camera poses and 3D object points, which serves as an initial guess for the subsequent optimization. After the adjustment, we obtain a set of optimized camera poses in a common reference frame. For each session separately, we can then compute a dense point cloud using these poses. Any dense matcher can be used here, we applied the patch-based multi-view stereo reconstruction technique (PMVS) by Furukawa and



(a) Initial descriptor matching



(b) After RANSAC step



(c) After recovery step

Fig. 6: Stages of data association procedure. The details regarding each stage is discussed in Sec. III-D

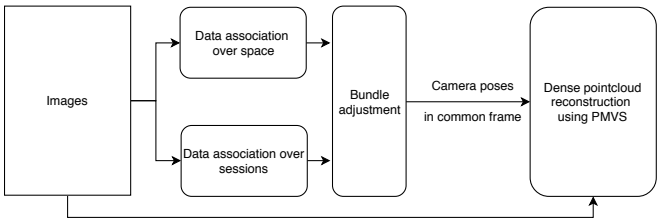


Fig. 7: 3D reconstruction pipeline showing steps for computing temporally aligned point clouds from different sessions.

Ponce [6]. The individual point clouds from each session are already aligned to a common reference frame since the used poses are the result of a common adjustment in the previous step. The complete pipeline is illustrated in Fig. 7.

IV. Experimental Evaluation

The experiments in this section are designed to illustrate the capability of our image registration approach for field monitoring tasks in agriculture and to support the claims made in the introduction of the paper.

A. Data

We recorded several datasets [1] of sugar beet crops spanning over multiple weeks for two different fields, referred as A and B here. For the field A, we recorded the datasets across four sessions using a DJI MATRICE 100 UAV. The flight altitude for each session is between 8 m to 12 m above the ground. We recorded the images using the Zenmuse X3 camera with an image resolution of 4000×2250 pixels having a ground sampling distance of 4 mm per pixel at a height of 10 m. For the field B, we used a DJI PHANTOM 4 UAV across two sessions recorded almost one month apart. The UAV was equipped with a GoPro camera set up to take an image every second at a resolution of 3840×2880. The flight altitude for the two sessions varied between 10 m and 18 m above the ground having a ground sampling distance of 9 mm per pixel at 15 m height. As the GoPro uses a wide angle lens, we first undistort the images before applying the registration pipeline. The average plant sizes in the fields range from 5 cm to 20 cm in diameter across the crop season. Furthermore, the images were taken under different weather and soil conditions. Tab. I provides an overview. The most challenging datasets are Ses. 2-3 (A) and Ses. 1-2 (B) due to the large time gap of 3-4 weeks between them whereas Ses. 3-4 (A) is the easiest being only 5 days apart.

[1]The datasets used in the paper can be downloaded from here: www.ipb.uni-bonn.de/data/uav-sugarbeets-2015-16/

TABLE I: Overview of the datasets

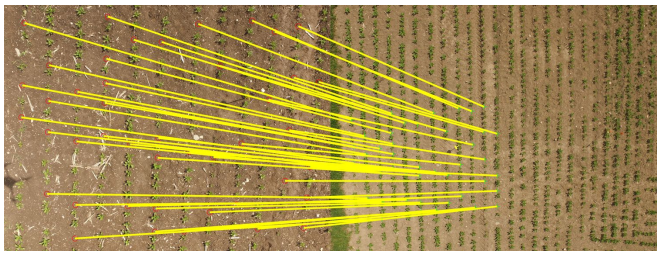
Field	Ses.	Date	# of images	crop size	weather
A	1	May 20	45	7 cm	cloudy
	2	May 27	175	10 cm	sunny
	3	June 17	121	15 cm	overcast
	4	June 22	140	20 cm	cloudy
B	1	May 8	99	5 cm	sunny
	2	June 5	95	15 cm	cloudy

TABLE II: Matching statistics across the crop season

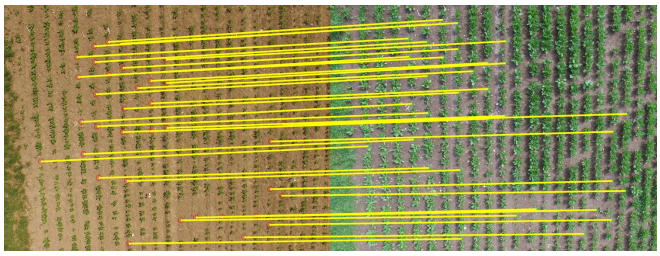
Field	Ses.	pts per img pair	# of matches			res err (px)
			Lowe test	RANSAC	Recovery	
A	1-2	58	27	11	42	4.21
	2-3	55	20	7	38	4.38
	3-4	57	24	9	40	4.35
B	1-2	74	39	15	56	4.91

B. Matching images across the crop season

The first experiment is designed to show that our approach is able to match images across the crop season having large difference in visual appearance. We perform matching between images within individual sessions and then across sessions. As described in Sec. III, we compute the gap points and construct our geometric descriptor for each of the images. We compute descriptors with $k = 4$ neighboring points to encode the local geometry. Tab. II summarizes the overall statistics for matching images across the sessions. It lists the average number of common gap points per image pair, the number of correspondences after the Lowe-ratio test, RANSAC, and recovery steps as well as the average residual error. We observe that around 30% of the initially matched points survive the RANSAC step and correspondences for roughly 70% of the points are re-established in the recovery step. Overall for field A, we observe an average residual error of 4.3 pixels, which corresponds to a ground distance

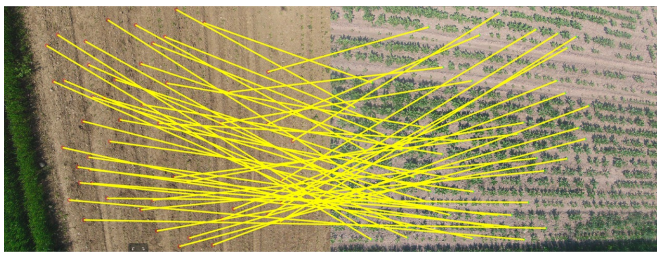


(a) Field A: Ses. 1 - Ses. 2 (1 week apart)



(b) Field A: Ses. 2 - Ses. 3 (3 weeks apart)

(c) Field A: Ses. 3 - Ses. 4 (5 days apart)



(d) Field B: Ses. 1 - Ses. 2 (4 weeks apart)

Fig. 8: Matching between image pairs from consecutive sessions.

of less than 2 cm. We have similar residual errors for field B at 4.9 pixels. While this accuracy does not match up to the usual sub-pixel accuracy of visual matching methods such as SIFT applied in non-changing environments, it is still a very good performance given the fact that physical growth of the plants and their changing appearance limits the accuracy with which the crop centers or the gaps can be detected. Fig. 8 shows example results from consecutive sessions for both fields. In all examples, visual matching using SIFT fails to find any reasonable set of correspondences.

C. Comparison against SIFT, ORB, and BRIEF

This experiment is designed to compare the matching performance of our approach against visual matching procedures using different descriptors. We perform the comparison between overlapping image pairs between each consecutive session for both the fields. In addition to the standard SIFT matching procedure using the default detector, we also compute the SIFT descriptor at the gap points computed by the detector in our approach. The intuition for doing this is that the gap regions are the least affected regions due to the movement of the tractor etc. on the field. Therefore, it provides the possibility of matching the texture of soil in these regions across different sessions. Tab. III provides a comparison of the matching performance using standard SIFT, SIFT at gap points, ORB, BRIEF and our approach. The table lists the percentage of image pairs matched successfully, maximum matches found for an image pair, and the inlier percentages for the matches computed by the three approaches. We consider image pairs having at least 4 matches resulting in a correct transformation as a successful match. For both fields A and B, we see that for most challenging datasets, i.e, Ses. 2-3 (A) and Ses. 1-2 (B), visual matching using the SIFT descriptor only matches between 9% to 16% of the image pairs successfully. Even for the successfully matched image pairs, the number of matches are very few and the percentage of inlier matches is only around 10% indicating that the matches are not reliable. The percentage of successful matches obtained with ORB and BRIEF is even worse. For example, they are not able to match any pairs form the dataset Ses. 1-2 (B). The SIFT descriptor computed at the gap points slightly improves the percentage of successful matches for Ses. 2-3 (A) while providing no improvement for other cases. However, for the relatively simpler dataset, i.e., Ses. 3-4 (A), both the visual approaches perform well as these images were captured only five days apart and are visually very similar. In comparison, our approach consistently matched around 85% of the image pairs with higher inlier percentages for each of the sessions including the challenging datasets of Ses. 2-3 (A) and Ses. 1-2 (B). This is because of the fact that our approach exploits the the geometry rather than relying on the visual appearance of the field.

TABLE III: Evaluation against visual descriptor matching

Field/ Ses.	SIFT / SIFT-gaps / ORB / BRIEF / Our approach		
	% pairs matched	max matches	% inlier
A/1-2	26/15/10/0/**89**	10/8/10/–/**42**	19/29/15/–/**41**
A/2-3	16/40/5/0/**85**	4/12/4/0/**38**	10/34/8/–/**35**
A/3-4	84/75/80/65/**86**	**103**/22/75/70/40	**67**/65/65/55/38
B/1-2	9/15/0/0/**87**	4/9/–/–/**56**	7/21/–/–/**39**

TABLE IV: Evaluation against ground truth

Field	Ses.	% of est matches	res. err (cm) (est/ref)	registration. err (trans/rot/scale)
A	1-2	91.67	1.47/0.98	3.19 px/0.38°/0.31%
	2-3	84.86	1.75/0.77	4.54 px/0.60°/0.42%
	3-4	85.19	1.74/0.86	4.07 px/0.42°/0.32%
B	1-2	87.22	3.93/2.16	3.94 px/0.47°/0.35%

D. Ground truth evaluation

This experiment is designed to evaluate the accuracy of our the matching results against the ground truth. To perform this

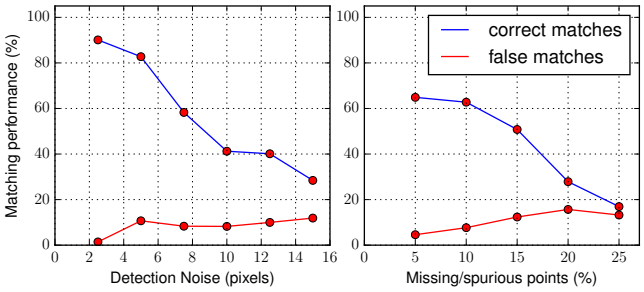


Fig. 9: Descriptor robustness under varying noise levels assessed in terms of percentage of correct matches (true positives) and false matches (false positives + false negatives).

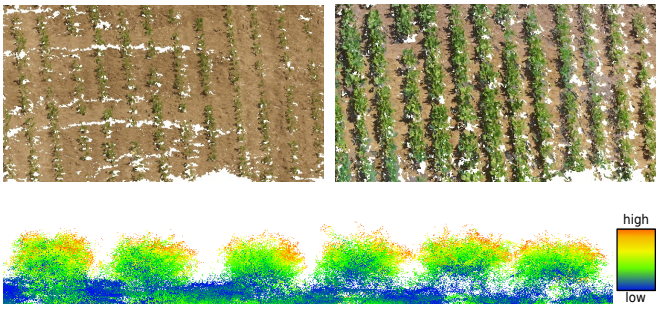


Fig. 10: Temporally aligned 3D point clouds. Top: 3D reconstruction for a portion of the field from the same viewpoint for Ses. 2 (left) and Ses. 3 (right). Bottom: cross section of a part of the point cloud from Ses. 3. The color of the point cloud represents the difference between the point clouds from the two sessions, i.e. Ses. 2 and Ses. 3. The portion close to ground does not change much between the sessions and therefore has a small difference indicated with blue color. In contrast, the top parts of the crops are colored green/yellow/red indicating bigger differences between the point clouds from the two sessions. This is due to the physical growth of the plant between the two sessions.

analysis, we compare our results with ground truth parameters for 10 image pairs between each sessions. All the evaluation parameters are summarized in Tab. IV. The ground truth parameters are computed based on control points, which have been provided manually. Using these control points, we compute the reference ground truth registration parameters under a similarity transform. We further manually establish unique correspondences between the image pair points under these registration parameters and consider them as the ground truth correspondences. We provide a measure of the quality of matching in terms of the percentage of correspondences estimated by our method as compared to the ground truth correspondences. On average, our method is able to recover up to 90% of all possible correspondences. We also compute the residual error based on the estimated correspondences and compare it to the residual error of the manually generated ground truth. For field A, we obtain an average residual of around 1.6 cm as compared to the ground truth residual close to 1 cm. This is only slightly worse than the ground truth results using manually measured control points, which indicates that the estimated parameters are correct. The residual error for field B is in the same range as that of field A. The absolute value of the error is higher only due to the lower ground resolution of 9 mm per pixel for this flight. Further, we evaluate the accuracy of the registration parameters by computing the average errors (translation, rotation, and scale) with respect to the ground truth parameters. We observe an average translation error of close to 4 pixel. We also obtain an average rotational error of 0.5°, and a scale error of less than 0.5% with respect to the ground truth parameters.

E. Descriptor performance under noisy detections

This experiment is designed to show the robustness of the descriptor under noisy detection conditions. We perform this analysis by simulating two kinds of noise, (i) a Gaussian noise affecting the location of the points, and (ii) missing/spurious detection of the points i.e., outliers or gross errors. We assess the performance of descriptors by computing the percentages of correct matches (true positives) and the false matches under varying levels of noise. The false matches includes both false positives, i.e., the points that are incorrectly matched and false negatives, i.e, the matches which were missed. For the noise of type (i), we vary the noise up to 15 pixel. The typical noise level for the gap detection procedure for our images is around 5 pixel. In Fig. 9, we observe that even for high noise levels (15 pixel), about 30% of the correspondences are identified correctly whereas the false matches are below 20% after the performing the Lowe's test. We observe a similar trend under missing/spurious points noise. We are able to identify up to 20% of the matches even when one fourth of the points are wrongly detected. These correspondences provide sufficient information for our data association procedure to match the images successfully. Furthermore, the RANSAC step eliminates the wrong correspondences resulting from incorrect descriptor matching and we finally recover only the consistent but initially ambiguous correspondences during the recovery step. This further supports the claim that we are able to perform matching robustly under substantial noise.

F. Time aligned 3D point clouds

This experiment is designed to show that our reconstruction pipeline allows us to compute temporally aligned 3D point clouds of the field (Sec. III-E) and thus support our second claim. The top two point clouds in Fig. 10 illustrate the result by rendering a portion of the field from the exact same camera position both for Ses. 2 and Ses. 3 respectively. This allows us to monitor the evolution and the changes on the field over time. To assess the quality of the alignment, we visualize the difference between the two aligned point clouds. The bottom part of Fig. 10 shows a cross-section view of the point cloud from Ses. 3, where the color signifies the difference between the point cloud from Ses. 2 and Ses. 3. The difference increases as the color changes from blue to red. We see that the alignment of the point clouds looks qualitatively correct as the space in-between the crop rows has a small difference indicated by the blue color. We also observe that the lower portions of the crops have a smaller difference as this portion overlaps with the crops from Ses. 2,

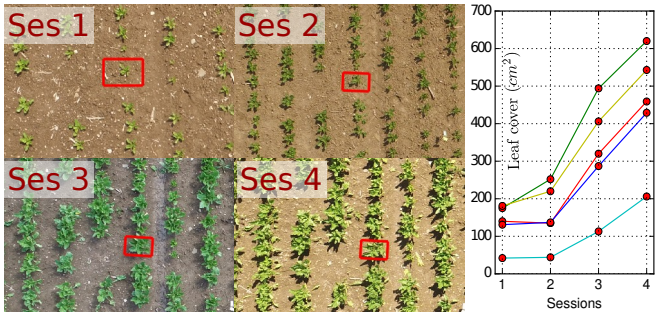


Fig. 11: Monitoring crop growth parameters. Left: Same crop identified in the bounding box over different sessions using our registration results. Right: plot of leaf cover over time at five different sites in the field.

whereas portions at the top have a larger difference reflecting the crop growth between the two sessions.

G. Monitoring crop growth parameters

To support our third and last claim, we show in the following experiment that our registration results allows us to monitor growth parameters at a per plant level. We manually provide bounding boxes around crops in the first session and compute the locations of the new bounding boxes in the corresponding images from different sessions using our registration results. Fig. 11 shows an example where the same plant is identified through different sessions. To monitor the growth of the plant, we compute the total leaf area (from top view) for the plant in each of the sessions. We compute this area by first extracting a vegetation mask inside the bounding box using the excess green index (ExG) and compute the area under it. Fig. 11 shows the plot of the total leaf area for individual plants at five different sites on the field over all the sessions. As it is expected, we see a general trend of increasing leaf area with time. For the plant shown in our example , the leaf area increases from about $150\,cm^2$ in Ses. 1 to $430\,cm^2$ in Ses. 4. The growth on the sites is consistent with the BCCH growth scale index for sugar beets. This experiment illustrates that our registration results are accurate enough for monitoring growth parameters at per a plant level. However, it should be noted our main goal here is not to analyze crop growth, but to facilitate such analysis by registering images taken over time to a common coordinate frame. Other works such as [10], [17] address the issue of analyzing crop growth in more detail.

V. Conclusion

In this paper, we presented a novel approach to register UAV images of agricultural fields that show large variations in the visual appearance over the crop season. Our method utilizes the inherent geometry of the crop arrangement in the field by exploiting the negative information about missing crops, i.e., gaps in the crop rows and uses this information for matching. This allows us to successfully register images even in situations where matching based on common visual descriptors such as SIFT, ORB, or BRIEF fail. The experiments suggest that our approach provides a robust and efficient alignment, which in turn allows us to obtain temporally aligned 3D point cloud and to monitor individual plants. Our work is an important step for UAV-supported precision agriculture applications that require temporally aligned models of whole fields up to an individual plant level such as in-field phenotyping, continuous yield forecasting, or similar.

Acknowledgments

We thank Raghav Khanna, Frank Liebisch for assisting with the data acquisition campaign and ETH Zurich Crop Science group for providing access to the fields.

References

[1] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S.M. Seitz, and R. Szeliski. Building rome in a day. *Communications of the ACM*, 54(10):105–112, 2011.

[2] M. Bryson, A. Reid, F.T. Ramos, and S. Sukkarieh. Airborne vision-based mapping and classification of large farmland environments. *Journal of Field Robotics (JFR)*, 27(5):632–655, 2010.

[3] M. Carcassoni and E.R. Hancock. Spectral correspondence for point pattern matching. *Pattern Recognition*, 36(1):193–204, 2003.

[4] 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 IEEE on Automation Science and Engineering (CASE)*, pages 462–469, 2015.

[5] J. Dong, J.G. 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.

[6] Y. Furukawa and J. Ponce. Accurate, dense, and robust multiview stereopsis. *IEEE Trans. on Pattern Analalysis and Machine Intelligence (TPAMI)*, 32(8):1362–1376, 2010.

[7] S. Gold, A. Rangarajan, C.P. Lu, S. Pappu, and E. Mjolsness. New algorithms for 2d and 3d point matching: pose estimation and correspondence. *Pattern Recognition*, 31(8):1019–1031, 1998.

[8] K. Kusumam, T. Krajnk, S. Pearson, G. Cielniak, and T. Duckett. Can you pick a broccoli? 3d-vision based detection and localisation of broccoli heads in the field. In *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, pages 646–651, 2016.

[9] T. Läbe and W. Förstner. Automatic relative orientation of images. In *Proc. of the Turkish-German Joint Geodetic Days*, 2006.

[10] Y. Li, X. Fan, N.J. Mitra, D. Chamovitz, D. Cohen-Or, and B. Chen. Analyzing growing plants from 4d point cloud data. *ACM Trans. on Graphics*, 32(6):157, 2013.

[11] P. Lottes, R. Khanna, J. Pfeifer, R. Siegwart, and C. Stachniss. UAV-Based Crop and Weed Classification for Smart Farming. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2017.

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

[13] S. Lowry, N. Sunderhauf, P. Newman, J.J Leonard, D. Cox, P. Corke, and M.J. Milford. Visual place recognition: A survey. *IEEE Trans. on Robotics (TRO)*, 32(1):1–19, 2016.

[14] M. Milford and G.F. 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.

[15] J. Munkres. Algorithms for the assignment and transportation problems. *Journal of the Society of Industrial and Applied Mathematics*, 5(1):32–38, 1957.

[16] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Trans. on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.

[17] J. Pfeifer, R. Khanna, C. Dragos, M. Popovic, E. Galceran, N. Kirchgessner, A. Walter, R. Siegwart, and F. Liebisch. Towards automatic uav data interpretation for precision farming. In *Proc. of the International Conf. of Agricultural Engineering (CIGR)*, 2016.

[18] O. Vysotska and C. Stachniss. Lazy Data Association for Image Sequences Matching under Substantial Appearance Changes. *IEEE Robotics and Automation Letters (RA-L)*, 2016.

[19] H.J. Wolfson and I. Rigoutsos. Geometric hashing: an overview. *IEEE Computational Science and Engineering*, 4(4):10–21, Oct 1997.

[20] L. Yang, J.M. Normand, and G. Moreau. Local geometric consensus: A general purpose point pattern-based tracking algorithm. *IEEE Trans. Vis. Comput. Graph.*, 21(11):1299–1308, 2015.