Collaborative UAV-UGV Environment Reconstruction in Precision Agriculture

Ciro Potena¹, Raghav Khanna², Juan Nieto², Daniele Nardi¹, and Alberto Pretto¹

Abstract—In air-ground multi-robots applications, where both Unmanned Aerial Vehicles (UAVs) and Unmanned Ground Vehicles (UGVs) operate in a coordinate way, the ability to obtain a unified environment representation of the target area is an essential requirement. However, a global registration of heterogeneous ground and aerial maps is a challenging task, especially for agricultural scenarios: the visual appearance of such kind of environment is rather homogeneous, it is difficult to find and exploit distinctive 3D geometrical structures, while the maps built using robots of different types show differences in both size and resolution and, possibly, scale errors.

In this paper, we tackle the cooperative UAV-UGV environment modeling problem in farming scenarios. We propose a novel maps registration pipeline that leverages a digital multi-modal environment representation which includes a vegetation index map and a Digital Surface Model (DSM). Using such map representation, we cast the data association problem between maps built from UAVs and UGVs as a multi-modal, large displacement dense optical flow estimation. The data association is then used to estimate an initial, non-rigid alignment between the maps that also compensates the (directional) scale discrepancies between them. A final refinement is then performed, by exploiting only meaningful parts of the registered maps.

We compare our method with standard registration techniques showing better alignment performances and better generalization properties over different misalignments and scale errors.

I. INTRODUCTION

The cooperation between aerial and ground robots offers undoubted advantages to many applications, thanks to the complementarity of characteristics and functionalities provided by these robots [17]. For instance, a UGV can carry high payloads, can perform targeted actions on the environment and it can operate for long periods of time, while a UAV allows rapid surveying and inspection of large areas [18], and can share information about areas of interest with the UGV. Air-ground cooperation is especially useful in Precision Agriculture scenarios, where the areas of interest are usually very large: a UAV can perform an initial survey of the field to localize areas where to deploy one ore more UGVs for detailed inspections or targeted interventions. Moreover, if both robots share the same environment representation, it becomes possible to update it with more detailed/high resolution map portions, or with semantic information coming from the processing of detailed views of the terrain [22]. On the other hand, building maps



Fig. 1: We assume that both a UGV and a UAV can generate a colored point cloud of the cultivated field (left). The proposed method aims to accurately merge these maps by means of an affine transformation that registers the UGV submap into the UAV aerial map (right), taking into account also the possible scale discrepancies.

using both these robots presents several challenges. UGVs and UAVs usually perceive the environment from very different points-of-views, while the agricultural fields are rather homogeneous both in the visual and in the geometric appearances. Geolocation information provided by reference sensors such as GPSs and Attitude and Heading Reference Systems (AHRSs) are often affected by not negligible errors [15], and cannot be directly exploited to solve sensor fusion problems. All those issues make standard multi-robot localization and mapping pipelines often ineffective. For similar reasons, combining 3D environment reconstructions (e.g., Fig. 1 (left)), independently built by aerial and ground robots¹, is a rather complex task as well. Maps can be affected by missing data and global deformations and, due to the reference sensors inaccuracies, also geotagged maps are often affected by global location and orientation biases.

In this paper, we provide an effective solution for the cooperative mapping task for heterogeneous robots, by proposing a 3D map registration pipeline specifically designed for farming scenarios. We assume that both a UAV and a UGV can generate a colored, geotagged point cloud of a target farm environment (Fig. 1). Our first solution is to tackle the data association problem between maps by means of a dense, globally regularized matching approach. We leverage on the intuition that points belonging to a cloud locally share similar displacement vectors that associate such points with points in the other cloud. Thus, by introducing a smoothness² term in the dense, regularized matching, we penalize the

This work was supported by the EC under Grant H2020-ICT-644227-Flourish. ¹Potena, Nardi and Pretto are with the Department of Computer, Control, and Management Engineering "Antonio Ruberti", Sapienza University of Rome, Italy. Email: {potena, nardi, pretto}@diag.uniromal.it. ²Khanna and Nieto are with the Autonomous Systems Lab, ETH Zurich. Email: jnieto@ethz.ch, raghav.khanna@mavt.ethz.ch

¹Maps are built, for instance, using sequences of geotagged images.

²The smoothness is related to the matching parameters of neighboring elements.



Fig. 2: Pictures of the same portion of field seen from the UAV pointof-view (left) and from the UGV point-of-view (right). The local crop arrangement geometry, such as the missing crop plants, is generally not visible from the UGV point-of-view. The yellow solid lines represent an example of manually annotated correct point matches. It is important to underline the complexity required in obtaining correct data association, also from an human point-of-view. The fiducial markers on the filed have been used to compute the ground truth alignments between maps.

displacement discontinuities in the neighborhood of each point. This approach has been inspired by the Large diplacement Dense Optical Flow (LDOF) problem in computer vision and, to this end, we convert the colored point clouds into a more suited, multi-modal environment representation that allows to exploit two-dimensional approaches and to enhance both the semantic and the geometrical properties of the target map. Our map is devised as a grid, where each cell stores (i) the Excess Green index (ExG) and, as a DSM, (ii) the local surface height information (e.g., the height of the plants, soil, etc.). We then use the data provided by the GPS and the AHRS to extract an initial guess of the relative displacement and rotation between matched grid maps. Hence, we compute a dense set of point-topoint correspondences between matched maps, exploiting a modified version of a state-of-the-art LDOF system [14]. We tailored this algorithm to our environment representation by proposing a different cost function that involves both the ExG information and the local structure geometry around each cell. We extract the largest set of similar flows, to be used as point-to-point correspondences to infer a preliminary alignment transformation between the maps. In order to deal with directional scale errors, we use a non-rigid point-set registration algorithm to estimate an affine transformation. The final registration is obtained by running a robust pointto-point registration algorithm over the input point clouds, pruned from all points that do not belong to vegetation. An overview of the the proposed approach is reported in Fig. 3. We report a set of preliminary experiments (Sec. V) on data acquired by a UAV and a UGV on a real field in Eschikon, Switzerland. We show that the proposed approach is able to guarantee correct registrations for an initial translational error up to 4 meters, an initial heading misalignment up to 11.5 degrees, and a directional scale error of up to 20%. We also report a comparison with state-of-the-art pointto-point registration algorithms, showing that our approach outperforms them in all the experiments.

A. Related Work

The field of multi map registration is a recurrent and relatively relevant problem in literature, and several solutions have been presented, in both 2D ([1], [2], [24]) and 3D ([10], [16]) settings. Registration of point cloud based maps can also be considered as an instance of the more general point

set registration problem [4], [8]. The map registration problem is even more difficult when dealing with heterogeneous robots, where the 3D data is gathered from different pointsof-views and with different noise characteristics. Michael et al. [20] propose a collaborative UAV-UGV 2.5D mapping approach. A UGV, equipped with a LiDAR, builds an initial map of the environment. The scans are merged together by employing an ICP approach using a flat ground assumption. The UAV, equipped with a 2D LiDAR, is deployed only in specific locations and maps the environment by using a pose-graph Simultaneous Localization and Mapping (SLAM) algorithm. The maps are then fused together using an ICP algorithm that is initialized in the UAV starting location. In [9], Forster et al. fuse RGB-D data from the UGV and dense monocular reconstructions from the UAV. The registration is performed by using the global orientation provided by an IMU and a magnetometer, while for the xand y coordinates they employ a 2D local height map fitting procedure. Hinzmann et al. [12] deal with registering 3D LiDAR point clouds with sparse point clouds, by exploiting an initial guess provided by a GPS and different ICP variants. In [11], Gawel et al. present a global registration procedure for 3D LiDAR maps gathered by a UGV with visual feature maps acquired by a UAV. The proposed approach exploits the rough geometric structure of the environment and an attitude initial guess provided by an IMU. Zeng et al. [27] deal with the heterogeneous map-merging problem by matching a learnt descriptor. In [6], Dub et al. exploit line features instead of standard ones, showing better global localization performance. However, the proposed approach has been tested only with UGVs equipped with the same sensor setup. In summary, despite extensive literature addressing the problem of 3D map-merging for heterogeneous robots, all proposed methods make strong context-based assumptions.

Localization and mapping in an agricultural scenario is a topic that is recently gathering great attention in the robotics community: Weiss *et al.* [25] discuss the use of MEMS based 3D LiDAR sensors for plant detection and mapping in contrast to traditional vision-based approaches; English *et al.* [7] proposed a vision based crop-row detection and following system; in [15], we proposed a multi-cue positioning and mapping system for a UGV moving in a cultivated field. Most of these systems, however, deal with a single robot, and the problem of fusing maps built from multiple robots is usually not adequately addressed.

Registering 3D maps in an agricultural environment, in some respects, is even more difficult than in other environments: the environment is homogeneous, poorly structured and it usually gives rise to strong sensor aliasing. For these reasons, the standard approaches mentioned above cannot directly be applied in practice.

A relevant work that deals with some of the challenges mentioned above has been proposed by Dong *et al.* [5]. They address the problem of matching images from the same field across time, by using a SLAM system to fuse measurements gathered by heterogeneous sensors such as cameras, GPSs, and an IMUs. This data is also used to reject outliers in



Fig. 3: Overview of the the proposed approach. Starting from the left side, we show: (i) the input colored point clouds gathered by the UAV and UGV, respectively, where in the UAV map we highlight through a dashed circle the initial search area given by the GPS prior; (ii) the obtained multi-modal grid maps that, for visualization purposes, we report as colored images; (iii) the output from the LDOF system: we represent with green solid lines the inliers, while the remaining solid lines are outliers; (iv) the selected data associations; (v) the aligned point clouds according to the initial affine transform (for visualization purposes, we colored in blue and red the UGV and UAV point clouds, respectively); (vi) the final registration result after the refinement step.

the data association process, resulting in a higher overall robustness. On the other hand, since the data association is mainly handled by standard visual features, the proposed algorithm cannot manage the drastic viewpoint change when matching aerial and ground maps. A similar problem has been tackled by Chebrolu et al. [3]. In this paper, the authors deal with the image registration from a nadir aerial point-ofview by exploiting the almost static geometry of the crop arrangement in the field. The idea is to use a scale invariant feature descriptor that encodes the local plant arrangement geometry. The results show that the data association obtained with the proposed descriptor allows to successfully register images taken over time. Unfortunately, data from UAV and UGV are acquired from very different points-of-views, so the local crop geometry may be lost in one of the two views: this fact may prevent a direct application of this method.

II. PROBLEM STATEMENT AND ASSUMPTIONS

Given two maps of a farmland \mathcal{M}_A and \mathcal{M}_G (Fig. 3, first column), both represented by 3D colored point clouds and built from data gathered from a UAV and a UGV, respectively, our goal is to find a transformation $F : \mathbb{R}^3 \to \mathbb{R}^3$ that allows to accurately align them. \mathcal{M}_A and \mathcal{M}_G can be generated, for instance, by means of image-based 3D reconstruction techniques applied to a sequence of geotagged images of a field acquired by the two robots. Our method makes the following assumptions about the two maps:

- 1) They can have different spatial resolutions but they refer to the same field, with some overlap among them;
- 2) They were built in the same time slot;
- They are approximately geotagged, possibly with noisy locations and orientations;
- They can be affected by local inconsistencies, missing data, and deformations such as directional scale errors.

The last hypothesis implies the violation of the rigid-body transformation assumption between the two maps: for this reason, we represent *F* as an affine transformation that allows anisotropic (i.e., non-uniform) scaling between the maps. Since we cannot recover the ground-truth scale factor of the reconstructed field, we assume that \mathcal{M}_A is not affected by any scale inconsistencies, so we look for a transformation

that aligns \mathcal{M}_G with \mathcal{M}_A by correcting the scale errors of \mathcal{M}_G with respect to \mathcal{M}_A . The latter is an acceptable assumption, since the map created by the UAV is usually larger than \mathcal{M}_G , and generated by using less noisy GPS readings, so the scale drift effect tends to be canceled.

III. DATA ASSOCIATION

In order to estimate the transformation F that aligns the two maps, we need to find a set of point correspondences, i.e. a set of matches $m_{A,G} = \{(p,q) : p \in \mathcal{M}_A, q \in \mathcal{M}_G\}$ between \mathcal{M}_A and \mathcal{M}_G , that represent points pairs that roughly belong to the same 3D position. As introduced before, conventional matching methods based on sparse matching of local features are unlikely to provide effective results due to the amount of repetitive and non-distinctive patterns spread in an agricultural environment. Instead, our method addresses the data association problem with a dense approach, inspired by the fact that points that are close to each other in \mathcal{M}_A should be matched with points close to each other in \mathcal{M}_G . This problem reminds the dense optical flow problem for RGB images: in this context, global methods (e.g., [13]) aim to build correspondences pixel by pixel between a pair of images by minimizing a cost function that, for each pixel, involves a data term that measures the point-wise similarity and a regularization term that fosters smoothness between nearby flows (i.e., nearby pixel to pixel associations).

A. Multi-Modal Grid Map

We aim to estimate $m_{A,G}$ by computing a "dense flow" that associates points in \mathcal{M}_A with points in \mathcal{M}_B . Obviously, methods designed for RGB images are not directly applicable to colored point clouds: we introduce here a multi-modal environment representation that allows to exploit such methods while enhancing both the semantic and the geometrical properties of the target map. A cultivated field is basically a globally flat surface populated by plants. A DSM³ can well approximate the field structure geometry, while a vegetation index can highlight the meaningful parts of the field and the visual relevant patterns: in our environment representation,

 $^{{}^{3}}A$ DSM is a raster representations (i.e., a rectangular grid) of the height of the objects on a surface.

we exploit both these intuitions. We generate a DSM from the point cloud; for each cell of the DSM grid, we also provide an ExG index. The ExG index enhances the green color channel in RGB images to highlight the presence of the vegetation. In practice, we transform a colored point cloud \mathcal{M} into a two dimensional grid map $\mathcal{J}: \mathbb{R}^2 \to \mathbb{R}^2$ (Fig. 3, second column), where for each cell we provide the surface height and the ExG index, with the following procedure:

- We select a rectangle that bounds the target area by means of minimum-maximum latitude and longitude;
- The selected area is discretized into a grid map J of w×h cells, by using a step of s meters. In practice, each of the w×h cells represents a square of s×s meters. Each cell is initialized with (0,0) pairs.
- 3) Remembering that \mathcal{M} is geotagged (see Sec. II), we can associate each 3D point of \mathcal{M} to one cell of \mathcal{J} .
- 4) For each cell with associated at least one 3D point:
 (a) We compute the height as the average of the z coordinates of the 3D points that belong to such cell;
 (b) we compute the the ExG index as the average of the the ExG indexes of the 3D points that belong to such cell, where for each point p we have:

$$ExG(p) = 2p_g - p_r - p_b; \tag{1}$$

with p_r , p_g and p_b the RGB components of the point.

B. Multi-Modal Large diplacement Dense Optical Flow

We transform both the \mathcal{M}_A and \mathcal{M}_G point clouds to be aligned into the corresponding multi-modal grid representations \mathcal{J}_A and \mathcal{J}_G , as described in the previous section. In the ideal case, with perfect geotags, each cell of \mathcal{J}_G is already in the correct position, and could be associated with the corresponding cell of \mathcal{J}_A that share the same coordinates. In other words, in the ideal case the "flow" that associates cells between the two maps is zero. Unfortunately in the real case, due to the inaccuracies of both the geotags and the 3D reconstruction, a non zero, potentially large offset is introduced in the associations. This offset is locally consistent but not constant for each cell, due to the directional scale errors. To estimate the offset map, we employ a modified version of the Coarse-to-fine PatchMatch (CPM) framework described in [14]. CPM is a recent LDOF system that provides cutting edge estimation results also in presence of very large displacements, and it is more efficient than other state-of-the-art methods with similar accuracy.

For efficiency issues, CPM looks for the best correspondence of some *seeds* rather than every pixel: the seeds are a set of points regularly distributed whithin the image. Given two images $\mathcal{I}_0, \mathcal{I}_1 \in \mathbb{R}^2$ and a collection of seeds $S = \{s_n\}$ at position $\{p(s_n)\}$, the goal of this framework is to determine the flow of each seed $f(s_n) = M(p(s_n)) - p(s_n) \in \mathbb{R}^2$, where $M(p(s_n))$ is the corresponding matching position in \mathcal{I}_1 for the seed s_n in \mathcal{I}_0 . The flow computation for each *seed* is performed by a coarse-to-fine random search strategy by minimizing the cost function:

$$f(s_n) = \underset{f_{s_i}}{\operatorname{argmax}}(C(f(s_i))), s_i \in \{s_m\}$$
(2)

where $C(f(s_i))$ denotes the match cost between the patch centered at $p(s_m)$ in \mathcal{I}_0 and the patch centered in $p(s_m) + f(\cdot)$ in \mathcal{I}_1 . For a comprehensive description of the flow estimation pipeline, we refer the reader to [14].

Our goal is to use the CPM algorithm to compute the flow between \mathcal{J}_A and \mathcal{J}_G . To exploit the full information provided by our grid maps (see Sec. III-A), we modified the CPM matching cost in order to take into account both the height and ExG channels. We split the cost function in two terms:

$$C(f(s_i)) = \alpha \cdot C_{SIFT}(f(s_i)) + \beta \cdot C_{FPFH}(f(s_i))$$
(3)

 $C_{SIFT}(f(s_i))$ is the SIFT [19] based match cost as in the original CPM algorithm: in our case the SIFT descriptors have been computed from the ExG channel of \mathcal{J}_A and \mathcal{J}_G . $C_{FPFH}(f(s_i))$ is a match cost computed using the height channel. We chose the Fast Point Feature Histograms (FPFH) [23] descriptor for this second term: the FPFH descriptors are robust multi-dimensional features which describe the local geometry of a point cloud, in our case they are computed from the organized point cloud⁴ generated from the height channel of \mathcal{J}_A and \mathcal{J}_G . The parameters α and β are the weighting factors of the two terms. As in [14], the patch-based matching cost is chosen to be the sum of the absolute difference over all the 128 and 32 dimensions of the SIFT and FPFH flows, respectively, at the matching points. With the proposed cost function, we take into account both the visual appearance and the local 3D structure of the plants.

Once we have computed the dense flow between \mathcal{J}_A and \mathcal{J}_G (Fig. 3, third column), we extract the largest set of similar flows up to a distance threshold t_f ; these flows define a set of point-to-point matches $m_{A,G}$ that will be used to infer a preliminary alignment (Fig. 3, fourth column).

IV. NON-RIGID REGISTRATION

The estimation of the non-rigid transformation between the clouds is addressed in two steps. A preliminary affine transformation \hat{F} is computed by solving a non-rigid registration problem with known point-to-point correspondences. We extract two organized point clouds from \mathcal{J}_A and \mathcal{J}_G and, using the matches $m_{A,G}$ computed in Sec. III-B, we compute \hat{F} by solving an optimization problem with cost function the sum of the squared distances between corresponding points (Fig. 3, fifth column). To estimate the final registration, we firstly select from the input colored point clouds \mathcal{M}_A and \mathcal{M}_G two subsets, \mathcal{M}_A^{veg} and \mathcal{M}_G^{veg} , that includes only points that belong to vegetation. The selection is performed by using an ExG based thresholding operator over \mathcal{M}_A and \mathcal{M}_G . This operation enhances the morphological information of the vegetation, while reducing the size of the point clouds to be registered. We finally estimate the target affine transformation F by exploiting the Coherent Point Drift (CPD) [21] point set registration algorithm over the point clouds \mathcal{M}_A^{veg} and \mathcal{M}_{G}^{veg} , using \hat{F} as initial guess transformation.

⁴An organized point cloud is a cloud that reminds a matrix like structure.



Fig. 4: Average success registration rate curves for different initial transformations between the clouds *Row A*, *Row B* and *Row C*, and the cloud built by the UAV (*Soybean Field*). The initial scale error incrementally increases from the left to the right column with percentages of 0%, 10%, 20%, 25%, respectively. In the first row, the *x* axis reports the increasing translational error, for fixed rotational and scale errors. In the second row, the *x* axis reports the increasing rotational error, for fixed translational and scale errors. It is important to point out that the success registration rate of the Go-ICP method is only reported for the cases without an initial scale error (first column) since this approach only deals with rigid transformations.

TABLE I: This table reports the average accuracy among all the successful registrations between each UGV cloud and the UAV cloud, each column refer to a specific initial scale error.

soybean row ID	approach	registration err. (trans/ros/scale) scale error 0%	registration err. (trans/ros/scale) scale error 5%	registration err. (trans/ros/scale) scale error 10%	registration err. (trans/ros/scale) scale error 15%	registration err. (trans/ros/scale) scale error 20%	registration err. (trans/ros/scale) scale error 25%
А	Ours	0.03 m/0.05°/-	0.03 m/0.03°/1.9%	0.04 m/0.04°/2.1%	0.04 m/0.05°/2.0%	0.03 m/0.03°/2.1%	0.05 m/0.04°/2.3%
	ICP	0.02 m/0.08°/-	0.05 m/0.07°/2.3%	0.04 m/0.08°/2.2%	fail	fail	fail
	CPD [21]	0.02 m/0.03°/-	0.04 m/0.08°/2.2%	0.03 m/0.09°/2.3%	0.03 m/0.08°/2.4%	fail	fail
	Go-ICP [26]	0.02 m/0.06°/-	fail	fail	fail	fail	fail
В	Ours	0.02 m/0.04°/-	0.03 m/0.05°/2.0%	0.03 m/0.04°/1.9%	0.05 m/0.06°/2.1%	0.05 m/0.07°/2.3%	0.05 m/0.09°/2.2%
	ICP	0.04 m/0.05°/-	0.04 m/0.07°/2.1%	0.05 m/0.09°/2.4%	fail	fail	fail
	CPD [21]	0.03 m/0.03°/-	0.03 m/0.05°/2.0%	0.03 m/0.06°/2.2%	0.04 m/0.08°/2.3%	fail	fail
	Go-ICP [26]	0.02 m/0.07°/-	fail	fail	fail	fail	fail
С	Ours	0.04 m/0.02°/-	0.04 <i>m</i> /0.06°/2.1%	0.05 m/0.07°/2.3%	0.04 m/0.03°/2.2%	0.05 m/0.08°/2.4%	0.05 m/0.10°/2.5%
	ICP	0.03 m/0.07°/-	0.03 <i>m</i> /0.08°/2.3%	0.05 m/0.10°/2.4%	fail	fail	fail
	CPD [21]	0.04 m/0.06°/-	0.04 <i>m</i> /0.05°/1.9%	0.04 m/0.07°/2.1%	0.05 m/0.08°/2.4%	fail	fail
	Go-ICP [26]	0.03 m/0.07°/-	fail	fail	fail	fail	fail

V. EXPERIMENTS

In order to analyze the performance of our system, we collected 4 datasets in a soybean field in Eschikon (Switzerland): (i) one aerial sequence of GPS-AHRS tagged images (Soybean Field) gathered over the full field by a DJI Mavic Pro UAV, flying at an altitude of 10 meters (e.g., Fig. 2 (left)); (ii) three ground sequences of GPS-AHRS tagged images (Row A, Row B, and Row C) framing small portions of the same field, acquired moving by hand the UAV camera with a forward point-of-view, to simulate a UGV moving along the crop rows (e.g., Fig. 2 (right)). For each dataset, we obtained the 3D colored point clouds by using a professional photogrammetry software (e.g, Fig. 1). For each ground sequence, we provided a ground truth affine transformation that aligns the related cloud with the point cloud obtained with the UAV. In the following experiments, we compared our approach with 3 point-set registration algorithms: (i) a non-rigid ICP algorithm, (ii) the original CPD [21] method, and (iii) a recent state-of-the art, globally optimal, ICP algorithm (Go-ICP) [26]. We directly apply those methods after pre-aligning the input point clouds by using the initial guess provided by the GPS-AHRS couple, and after pruning from the clouds all the points belonging to the soil terrain, as done in Sec. IV.

A. Robustness Evaluation

This experiment is designed to show the robustness of the proposed approach under different initial transformations between the clouds to be aligned. We sample random heading, translation, and scale displacements between the point clouds, running the tested registration algorithms starting from the sampled displacements. The final registration is considered correct if all the following conditions are met:

$$e_t <= 0.05 \ m \quad e_r <= 0.1^\circ \quad e_s <= 2.5\%$$
 (4)

with e_t the translation error, e_r the rotation error, and e_s the (directional) scale error. The results are illustrated in Fig. 4. In all tests our method outperforms the other methods by a wide margin, generally ensuring correct registrations for errors up to 4 meters for the translation, 11.5 degrees for the rotation, and 20% for the scale, with acceptable results also for a scale error of 25%.

B. Accuracy Evaluation

This experiment is designed to evaluate the point cloud alignment accuracy for successful registrations, under different initial transformations between the aligned clouds (see Sec. V-A). The results are reported in Tab. I, some qualitative results are reported in Fig. 1 and Fig. 5. The accuracy



Fig. 5: (top) Qualitative result of the registration obtained by our method aligning the input point clouds *Row A* (in blue) and *Soybean Field* (in red); (bottom) the same result but showing the merged colored point cloud: the input clouds are indistinguishable, a sign that the alignment is correct.

provided by our method is in line with the other methods and, when the latter fail, ours continues to provide an accuracy that does not decrease as the initial alignment error increases.

C. Runtime Evaluation

We finally report the runtime results in Tab. II. Our method is far from being the fastest one (most of the time is spent on computing the FPFH features), on the other hand it is much more efficient than Go-ICP.

TABLE II: Runtime comparison between the tested methods.

	Ours	ICP	CPD [21]	Go-ICP [26]
Average Runtime [sec]	79.8	3.6	8.2	193.1

VI. CONCLUSIONS

In this paper we addressed the cooperative UAV-UGV environment reconstruction problem in agricultural scenarios by proposing an effective way to align 3D maps acquired from both aerial and ground points-of-views. We cast the data association problem as a LDOF problem among grid maps that encode both geometric and semantic information. The presented experiments support our claims, with our method outperforming other approaches in several tests.

The future work includes the improvement of runtime performance and a more exhaustive experimental evaluation.

References

- A. Birk and S. Carpin, "Merging occupancy grid maps from multiple robots," *Proceedings of the IEEE*, vol. 94, no. 7, pp. 1384–1397, July 2006.
- [2] J. L. Blanco, J. González, and J.-A. Fernández-Madrigal, "A robust, multi-hypothesis approach to matching occupancy grid maps," *Robotica*, vol. 31, pp. 687–701, 2013.
- [3] 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 (RA-L), vol. 3, no. 4, pp. 3097–3104, 2018.
- [4] H. Chui and A. Rangarajan, "A new point matching algorithm for non-rigid registration," *Comput. Vis. Image Underst.*, vol. 89, no. 2-3, Feb. 2003.
- [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] R. Dubé, D. Dugas, E. Stumm, J. I. Nieto, R. Siegwart, and C. Cadena, "Segmatch: Segment based loop-closure for 3D point clouds," *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2017.
- [7] A. English, P. Ross, D. Ball, and P. Corke, "Vision based guidance for robot navigation in agriculture," in *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2014.
- [8] A. W. Fitzgibbon, "Robust registration of 2D and 3D point sets," in British Machine Vision Conference, 2001, pp. 662–670.
- [9] C. Forster, M. Pizzoli, and D. Scaramuzza, "Air-ground localization and map augmentation using monocular dense reconstruction," in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems* (IROS), 2013.
- [10] C. Frueh and A. Zakhor, "Constructing 3D city models by merging ground-based and airborne views," in 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings., vol. 2, 2003.
- [11] A. Gawel, T. Cieslewski, R. Dub, M. Bosse, R. Siegwart, and J. Nieto, "Structure-based vision-laser matching," in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2016.
- [12] T. Hinzmann, T. Stastny, G. Conte, P. Doherty, P. Rudol, M. Wzorek, E. Galceran, R. Siegwart, and I. Gilitschenski, "Collaborative 3D reconstruction using heterogeneous UAVs: System and experiments," in *ISER*, ser. Springer Proceedings in Advanced Robotics, vol. 1. Springer, 2016, pp. 43–56.
- [13] B. K. P. Horn and B. G. Schunck, "Determining optical flow," Artif. Intell., vol. 17, no. 1-3, pp. 185–203, 1981.
- [14] Y. Hu, R. Song, and Y. Li, "Efficient coarse-to-fine patch match for large displacement optical flow," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [15] M. Imperoli, C. Potena, D. Nardi, G. Grisetti, and A. Pretto, "An effective multi-cue positioning system for agricultural robotics," *IEEE Robotics and Automation Letters (RA-L)*, vol. 3, no. 4, pp. 3685–3692, 2018.
- [16] J. Jessup, S. N. Givigi, and A. Beaulieu, "Robust and efficient multirobot 3D mapping with octree based occupancy grids," in *Proc. of the IEEE Intl. Conf. on Systems, Man, and Cybernetics (SMC)*, 2014.
- [17] R. Käslin, P. Fankhauser, E. Stumm, Z. Taylor, E. Mueggler, J. Delmerico, D. Scaramuzza, R. Siegwart, and M. Hutter, "Collaborative localization of aerial and ground robots through elevation maps," 2016.
- [18] R. Khanna, M. Möller, J. Pfeifer, F. Liebisch, A. Walter, and R. Siegwart, "Beyond point clouds-3D mapping and field parameter measurements using UAVs," in *IEEE ETFA*, 2015.
- [19] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Intl. Journal of Computer Vision (IJCV)*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [20] N. Michael, S. Shen, K. Mohta, Y. Mulgaonkar, V. Kumar, K. Nagatani, Y. Okada, S. Kiribayashi, K. Otake, K. Yoshida, K. Ohno, E. Takeuchi, and S. Tadokoro, "Collaborative mapping of an earthquake-damaged building via ground and aerial robots," *Journal* of Field Robotics (JFR), vol. 29, no. 5, pp. 832–841, Sept 2012.
- [21] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE Transactions on Pattern Analysis and Machine Intelli*gence, vol. 32, no. 12, pp. 2262–2275, 2010.
- [22] C. Potena, D. Nardi, and A. Pretto, "Fast and accurate crop and weed identification with summarized train sets for precision agriculture," in *Proc. of the 14th International Conference on Intelligent Autonomous Systems (IAS-14)*, 2016.
- [23] R. B. 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)*, 2009.
- [24] S. Saeedi, L. Paull, M. Trentini, and H. Li, "Multiple robot simultaneous localization and mapping," in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2011.
- [25] U. Weiss and P. Biber, "Plant detection and mapping for agricultural robots using a 3D lidar sensor," *Robotics and autonomous systems*, vol. 59, no. 5, pp. 265–273, 2011.
- [26] J. Yang, H. Li, D. Campbell, and Y. Jia, "Go-ICP: A globally optimal solution to 3D ICP point-set registration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 11, pp. 2241– 2254, 2016.
- [27] A. Zeng, S. Song, M. Nießner, M. Fisher, J. Xiao, and T. Funkhouser, "3DMatch: Learning local geometric descriptors from RGB-D reconstructions," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017.