Received: 7 April 2022 | Revised: 17 February 2023 | Accepted: 29 April 2023

DOI: 10.1002/rob.22194

RESEARCH ARTICLE

WILEY

NDT-6D for color registration in agri-robotic applications

Himanshu Gupta[1] | **Achim J. Lilienthal**[1,2] | **Henrik Andreasson**[1] | **Polina Kurtser**[1,3]

[1]Centre for Applied Autonomous Sensor Systems, Institutionen för naturvetenskap & teknik, Örebro University, Örebro, Sweden

[2]Perception for Intelligent Systems, Technical University of Munich, Munich, Germany

[3]Department of Radiation Science, Radiation Physics, Umeå University, Umeå, Sweden

Correspondence
Himanshu Gupta, Centre for Applied Autonomous Sensor Systems, Institutionen för naturvetenskap & teknik, Örebro University, Fakultetsgatan 1, Örebro 70182, Sweden.
Email: Himanshu.Gupta@oru.se

Funding information
Horizon 2020 Framework Programme

Abstract

Registration of point cloud data containing both depth and color information is critical for a variety of applications, including in-field robotic plant manipulation, crop growth modeling, and autonomous navigation. However, current state-of-the-art registration methods often fail in challenging agricultural field conditions due to factors such as occlusions, plant density, and variable illumination. To address these issues, we propose the NDT-6D registration method, which is a color-based variation of the Normal Distribution Transform (NDT) registration approach for point clouds. Our method computes correspondences between pointclouds using both geometric and color information and minimizes the distance between these correspondences using only the three-dimensional (3D) geometric dimensions. We evaluate the method using the GRAPES3D data set collected with a commercial-grade RGB-D sensor mounted on a mobile platform in a vineyard. Results show that registration methods that only rely on depth information fail to provide quality registration for the tested data set. The proposed color-based variation outperforms state-of-the-art methods with a root mean square error (RMSE) of 1.1–1.6 cm for NDT-6D compared with 1.1–2.3 cm for other color-information-based methods and 1.2–13.7 cm for noncolor-information-based methods. The proposed method is shown to be robust against noises using the TUM RGBD data set by artificially adding noise present in an outdoor scenario. The relative pose error (RPE) increased ~14% for our method compared to an increase of ~75% for the best-performing registration method. The obtained average accuracy suggests that the NDT-6D registration methods can be used for in-field precision agriculture applications, for example, crop detection, size-based maturity estimation, and growth modeling.

KEYWORDS

agricultural robotics, color pointcloud, in-field sensing, machine perception, RGB-D registration, stereo IR, vineyard

1 | INTRODUCTION

Automation in the agricultural domain is a fast-growing application of outdoor robotics mostly due to the lack of human labor and, as a result, increasing the cost of manual field operations such as harvesting, planting, pruning, and trimming (Oliveira et al., 2021). With recent advancements in the field of visual detection, three-dimensional (3D) reconstruction, and positioning using analytical and artificial intelligence-based methods, these labor-intensive tasks are being automated using robots (Bac et al., 2014; Bakker et al., 2006; Bawden et al., 2017). These systems have the potential to reduce costs and increase field productivity. They employ machine vision algorithms (Kamilaris & Prenafeta-Boldú, 2018; Tian et al., 2020) for the detection and positioning of target crops, mainly relying on color images.

For the calculation of the target crops' spatial position or morphological aspects, depth information is often required (Arad et al., 2020; Kurtser, Ringdahl, Rotstein, Berenstein, et al., 2020; Vit & Shani, 2018). For this purpose, the RGB-D camera is well suited as these sensors provide colored 2D images and 3D point clouds (Kurtser, Ringdahl, Rotstein, Berenstein, et al., 2020; Ringdahl et al., 2019) enclosed in a single rigid packaging capable of sustaining the harsh environmental conditions often encountered in these applications. The colored images can be used for color-based detection, plant growth monitoring, and ripeness estimation. The 3D point clouds can be used to estimate the physical crop size, shape, and target localization. Commercial-grade RGB-D sensors operating in field conditions have only recently become available in the market (Ringdahl et al., 2019; Vit & Shani, 2018). Till recently, point clouds were employed mostly exclusively for the navigation of the robot in the field using 2D and 3D LiDARs, an application often not requiring color data (Biber et al., 2012; Malavazi et al., 2018) or high pointcloud density.

With the penetration of the RGB-D sensors into the agricultural robotics domain, algorithms for in-field extraction of crop size, shape, ripeness, and position were developed. These algorithms rely mainly on previous work in indoor conditions where detailed 3D plant models can be extracted using hand-held 3D scanners (Schunck et al., 2021), acquiring data from multiple viewpoints. Despite previous work showing that employing multiple viewpoints can significantly improve precision (i.e., Harel et al., 2016; Kurtser & Edan, 2018b), most outdoor algorithms rely on pointclouds acquired from a single location. This can be attributed to a working assumption often voiced in the field that state-of-the-art registration algorithms generally fail to provide accurate registration results for the noisy outdoor sensory data acquired from RGB-D cameras and the dense and repetitive soft dynamic foliage present in the agricultural domain. While it was claimed in our previous work (Kurtser, Ringdahl, Rotstein, & Andreasson, 2020; Kurtser, Ringdahl, Rotstein, Berenstein, et al., 2020) that single frame detection can be sufficient for some applications, it is apparent that higher precision can be obtained from the registration of data from several viewpoints before analysis. For example, algorithms relying on a single frame RGB-D are often more prone to additional error due to a significant number of overflowing points near the boundary of the objects. This problem can be solved by combining consecutive point clouds using registration algorithms and 3D reconstruction of the registered point cloud. Similarly, single frame RGB-D only provides one object surface, the one facing the camera, making correct estimation of volume and location biased.

Beyond very close range applications such as plant morphological modeling and localization, registration of multiple pointclouds originating from consecutive frames acquired in field conditions from commercial grade RGB-D cameras can also potentially replace or supplement LiDARs in close range navigation. Enriched maps generated from the aggregation of consecutive pointclouds acquired by LiDARs for navigation purposes with dense close range information can lead to a variety of possible applications such as field monitoring and acquisition of measures such as yield.

Since pointcloud registration in field conditions is a well researched field in many domains, in this paper, we aim to investigate the reasons for the failure of state-of-the-art registration algorithms given the field conditions in which agricultural robots are to operate. We perform this by comparing a range of commonly used registration algorithms on a data set acquired in commercial vineyard conditions (GRAPES3D data set [Kurtser, Ringdahl, Rotstein, Berenstein, et al., 2020]). Once the weaknesses are identified, we propose our registration method, which is shown to be more robust in these conditions. We show our algorithm's robustness using a benchmark data set for RGB-D registration and SLAM methods, TUM RGBD data set (Sturm et al., 2012)) using the evaluation metrics of the data set.

1.1 | Contribution

Given the outlined need for RGB-D data registration in this specific setting, our contribution is as follows:

1. We introduce a novel registration method (NDT-6D) that successfully registers the collected data and shows to be more robust to sensory noise than the state-of-the-art registration methods including a supplementary code release.[1]
2. We present evaluation results for the current state-of-the-art registration methods on prototypical agri-robotics RGB-D data collected from a mobile robot in a vineyard setting. We compare these results to the evaluation of the same algorithms on a typical indoor benchmark data set.
3. We evaluate in detail the contribution of color cues for scan registration in the agricultural setting.
4. We provide an evaluation methodology that focuses on measures specifically relevant to agri-robotics applications.

[1]Code available here (last accessed Oct 2021): https://github.com/hgupta01/ndt-6d.git

The rest of the paper is structured as follows. First, we provide an overview of the use of RGB-D data in the agri-robotics domain and the challenges in data registration and multi-view analysis. Next, we provide an overview of the current methods of point cloud registration to which we compare our work, as well as some standard notations. In Section 4, we first introduce the empirical data used to evaluate the various registration algorithms, followed by a description of the suggested NDT-6D method. Finally, we present and discuss the detailed results obtained from applying the registration algorithms to the mentioned data sets in the results section, followed by a short conclusion.

2 | LITERATURE OVERVIEW

2.1 | Visual sensors in agricultural robotics

The most common sensors employed in agricultural robotics operating in field conditions are imaging cameras (Bac et al., 2014; Kamilaris & Prenafeta-Boldú, 2018). Factors such as robustness, low cost, low weight and size, and the fact that humans rely greatly on vision to perform manual crop monitoring and manipulations, are all contributors to the widespread use of the RGB camera in the applications of crop monitoring. Detection of the crop, as well as diseases and pests in field conditions, are most often achieved using a color camera placed facing the foliage (Al-Hiary et al., 2011; Bac et al., 2014; Kamilaris & Prenafeta-Boldú, 2018; Singh & Misra, 2017). The algorithms developed are aimed to detect abnormalities and foreign objects from imagery data. The main obstacles affecting performance in detection directly on foliage are most often high occlusion rates and variable lighting conditions. Some of the solutions proposed are often multi or hyperspectral cameras (Dale et al., 2013) thermal imaging(Vadivambal & Jayas, 2011) and light resilient adaptive algorithms (Arad et al., 2019; Zemmour et al., 2017).

Despite the clear advantages of relying on imagery data for most operations of crop monitoring, not all field operations can rely solely on color data or spectral data. Specifically, in agricultural robotics, the physical dimension and location of the detected crop can be crucial for continuous operation. The somewhat popular examples in agri-robotics include operations requiring physical manipulation of the plant and therefore requiring localization of the target in world coordinates for actions such as harvesting (Arad et al., 2020; Bac et al., 2014), weeding (Bakker et al., 2006; Bawden et al., 2017) and pruning (Botterill et al., 2017). Up till recently, most commercially available range measuring sensors, combined with RGB cameras failed to provide the necessary sensory data quality to be implemented in outdoor conditions, and therefore technical solutions such as visual servoing (e.g., for harvesting - [Arad et al., 2020; Barth et al., 2016]) or assumption of constant distance to target (e.g., for top-down weeding [Tillett et al., 2008]) are often employed. With the recent developments in commercial-grade RGB-D sensors the acquisition of acceptable quality colored point clouds is now possible in outdoor conditions as well (Ringdahl et al., 2019; Vit & Shani, 2018). These capabilities open the door for close sensing applications for monitoring the crop's physical size and location. Applications such as growth monitoring, maturity estimation based on physical size, and phenotypic features extraction were so far tested mostly in the laboratory and postharvest controlled photo chamber conditions (Hacking et al., 2019; Kirk et al., 2020; Nandi et al., 2016). The availability of such sensors is now enabling in-field size- based phenotypes acquisitions (Kurtser, Ringdahl, Rotstein, Berenstein, et al., 2020; Milella et al., 2019; Vit & Shani, 2018). All of these operations require depth sensors.

2.2 | Mapping and data fusion in orchard and vineyard settings

Several recent projects in autonomous monitoring of vineyards and orchards have focused on the need to fuse and aggregate information collected from ground mobile robots in a form of a semantically enriched map. VineScout (Fernández-Novales et al., 2021) autonomous ground vehicle equipped with an IR sensor was used to monitor grapevine water status. The information is aggregated into maps of the entire vineyard. More classic simultaneous mapping and localization (SLAM) algorithms were tested in vineyard conditions by the Bacchus project (Papadimitriou et al., 2022) with the aim of generating navigation maps. Wang, Tang, and Whitty (Wang et al., 2020) generated maps of flower density in apple orchards using a ground robot equipped with RGB and RGBD sensors. Despite the aggregation of the semantic data in a form of a map using geolocation extracted from the GPS unit, the authors do not register the pointclouds from the RGBD camera but perform single-frame detection.

2.3 | Multi-view and point cloud registration in the agricultural automation domain

The use of multiple viewpoints planning for an eye-in-hand robotic configuration or drone field monitoring is a widely discussed issue in agri-robotic vision applications (Barth et al., 2016; Bulanon et al., 2009; Hemming et al., 2014; Kurtser & Edan, 2018a, 2018b; Zaenker et al., 2021, 2020). The discussion often focuses on target visibility due to the high occlusion levels requiring multiple viewpoints to overcome the problem. The sensor viewpoint planning methods often focus on the need to plan the sensing strategy under time constraints and expected information content. These methods often do not register the point clouds but rather plan the next optimal viewpoint. Attempts to register RGB-D point clouds in agricultural settings, acquired from on-ground robots often focus on grasping pose calculation for fruit harvesting (Guo et al., 2020; Lehnert et al., 2016) or growth modeling (Alenya et al., 2011; Chebrolu et al., 2020).

To the best of the authors' knowledge, these applications have been tested so far only in indoor laboratory conditions and do not

deal with issues of data registration under challenging illumination, occlusions, and plant density. Point cloud registration in field conditions has been so far implemented exclusively in navigation and mapping applications of the mobile agri-robot, acquiring 3-D point clouds using laser scanners and LiDARs, which are more resilient to outdoor illumination conditions (Gao et al., 2018; Shalal et al., 2013). Therefore, registration methods applied so far in field conditions have mostly overlooked the possible added value of color information for registration purposes.

An exception to this is the work of Dong, Roy, and Isler (Dong et al., 2020), who performed tree row mapping using registration of pointclouds acquired from an RGB-D camera. To register the pointcloud, the authors proposed a tailor-made algorithm that relies on domain knowledge in the form of semantic constraints, such as the presence of tree trunks and their expected orientation.

In our previous work, we have shown how the acquisition of colored-point clouds can be used for both detections (Kurtser, Ringdahl, Rotstein, & Andreasson, 2020a) and volume estimation (Kurtser, Ringdahl, Rotstein, Berenstein, et al., 2020) of grapes in vineyard conditions with the goal of yield prediction. In both previous applications, we have employed single-frame non-registered point clouds under the assumption that state-of-the-art registration algorithms generally fail to provide accurate registration results for both the noisy outdoor sensory data acquired from RGB-D cameras and the dense and repetitive feature lacking soft and dynamic foliage present in the agricultural domain.

In this paper, we aim to challenge this assumption through the evaluation of several state-of-the-art registration algorithms and propose our own registration method.

3 | POINT SET REGISTRATION

In this section, we review the state-of-the-art registration methods that are used in this work. We start by defining the registration problem mathematically and discuss the registration methods briefly.

Registration of two point clouds $\mathcal{X}$ and $\mathcal{Y}$ means finding the transformation matrix $\mathcal{T}$ that aligns the point clouds. It is an iterative optimization problem in which registration loss is minimized. The registration problem can be mathematically expressed as in the following equation:

$$\mathcal{T}^* = arg\min_{\mathcal{T}} \mathcal{L}(\mathcal{T}(\mathcal{X}), \mathcal{Y}), \tag{1}$$

where $\mathcal{T}$ is the rigid transformation matrix ($\mathcal{T}(x) = \mathcal{R}x + t$), represented using rotation matrix $\mathcal{R} \in SO(3)$ and translation vector $t \in R^3$. $\mathcal{L}$ is the registration loss function and depends on the registration algorithm used.

In this paper, we build upon the well-known and most widely used registration algorithms—iterative closest point (ICP) registration (Korn et al., 2014) and Normal Distribution Transform (NDT) registration (Stoyanov et al., 2012).

Since Besl and McKay (1992) first used the term ICP, several variations were proposed. However, according to the review performed by Pomerleau et al., (Pomerleau et al., 2015), the main variation in ICP algorithms include variations in: (1) transformation functions; (2) data filters; (3) distance functions.

In the described above application in agri-robotics, the scans are not expected to scale significantly. As a result, the ICP variations presented focus only on rigid transformation functions that include translation and rotation changes only. Data filters in the case of point clouds are used to reduce noise by doing feature enhancement (e.g., calculating point normal, extracting corner or surface points) and feature reduction (e.g., point density reduction, ground removal). Besides the basic point cloud data filters, image-based data filtering methods are also used in this work, described in Section 4. In this work, we used the Euclidean distance function with point-to-point and point-to-plane distance for ICP point-2-point and ICP point-2-plane registration respectively.

Published variations of NDT algorithms are more scarce and often conceptually do not vary significantly from the below method. Nevertheless, some variations are available (Das & Waslander, 2014; Magnusson et al., 2009; Stoyanov et al., 2012; Valencia et al., 2014). Therefore, we chose the most common ones that rely on point-to-distribution and distribution-to-distribution distances.

Finally, some recent efforts in Deep Learning based registration methods have gained popularity by training networks for feature extraction and registration (Villena-Martinez et al., 2020). These methods appear promising in data-abundant applications but can be expected to require large amounts of data, a common bottleneck in the agricultural robotics domain (Kamilaris & Prenafeta-Boldú, 2018; Kurtser, Ringdahl, Rotstein, & Andreasson, 2020). The designed networks often rely on the same conceptual approach as ICP and NDT of searching key points using images and could be viewed as an extension to the SIFT-ICP method we evaluate.

3.1 | ICP registration

The ICP registration loss function can be defined as the sum of the squared distance between the entities in the source cloud ($\mathcal{X}$) to the corresponding entity in the target point cloud ($\mathcal{Y}$). Here, the entity could be a point, a plane, or a shape, and the corresponding entity is usually the nearest neighbor of the transformed entity or point in the target point cloud. In work by Tavares et al. (2020), a detailed description of the ICP registration loss functions is presented. The generalized ICP loss function as per Tavares et al. (2020) can be defined as the sum of the distance between the matching features in point clouds and can be written as in Equation 2.

$$\mathcal{L}(\mathcal{X}, \mathcal{Y}) = \sum_i w_i d(\mathcal{X}, C(\mathcal{Y}))^2, \tag{2}$$

where d is the distance function, w_i is the optional weight for the entity pair and C defines the corresponding entity of source point

cloud $\mathcal{X}$ in target point cloud $\mathcal{Y}$. When the entities are points, the loss function is the sum of Euclidean distance between the corresponding points (Besl & McKay, 1992), and correspondence is the nearest neighbor based on Euclidean distance. In Chen and Medioni (1992), point-to-plane correspondence is established, where the loss function is defined such that the distance between the point in the source point cloud is decreased along the normal of the corresponding plane of points in the target point cloud.

3.2 | NDT registration

NDT registration is a point set registration algorithm which uses NDT maps. The NDT maps are constructed by dividing the point cloud into grids called NDT cells. For each NDT cell, normal distribution $\mathcal{N}(\mu, \Sigma)$ is calculated using the points that fall in the grid. There are two types of NDT registration algorithms, point-2-distribution (P2D) and distribution-to-distribution (D2D). In NDT registration, the cost function is minimized iteratively with respect to the rigid transformation matrix $\mathcal{T}$. The NDT P2D registration cost function between a point cloud $\mathcal{X}$ and NDT map $\mathcal{M}_{\mathcal{Y}}$ (where $\mathcal{M}$ is the NDT map of point cloud $\mathcal{Y}$) is defined as the negative likelihood of point in $\mathcal{X}$ belonging to the NDT cells in map $\mathcal{M}_{\mathcal{Y}}$. The cost function for NDT P2D registration is given in Equation 3

$$f_{p2d}(\mathcal{X}, \mathcal{Y}, \mathcal{T}) = \sum_{x \in X} \sum_{\mu, \Sigma \in \mathcal{M}_Y} -c_1 e^{\left(-\frac{c_2}{2}(\mathcal{T}(x)-\mu)^T \Sigma^{-1}(\mathcal{T}(x)-\mu)\right)}, \quad (3)$$

where, c_1 and c_2 are positive regularization factor mentioned by Magnusson et al. (2007), x iterate over the points in point cloud $\mathcal{X}$ and μ, Σ iterate over the NDT cells of map $\mathcal{M}_{\mathcal{Y}}$. To make the cost calculation computationally less expensive, NDT cell parameter (μ and Σ) closest to the transformed point is used.

NDT D2D registration cost function is defined between two NDT maps (source map $\mathcal{M}_{\mathcal{X}}$ and target map $\mathcal{M}_{\mathcal{Y}}$) and represents the dissimilarity between the maps. There are two types of NDT D2D registration cost function, the first cost function Equation (4) is defined as the sum of the L_2 distances between NDT cells of source and target map (Andreasson & Stoyanov, 2012). The second cost function Equation 5 is based on fuzzy logic (Liao et al., 2022).

$$f_{d2d}(\mathcal{M}_{\mathcal{X}}, \mathcal{M}_{\mathcal{Y}}, \mathcal{T}) = -c_1 \sum_{i=1}^{N_{M_X}} \sum_{j=1}^{N_{M_Y}} e^{\left(-\frac{c_2}{2} d_{ij}\right)}, \quad (4)$$

$$f_{fuzzy}(\mathcal{M}_{\mathcal{X}}, \mathcal{M}_{\mathcal{Y}}, \mathcal{T}) = \sum_{i=1}^{N_{M_X}} \frac{1}{\sum_{j=1}^{N_{M_Y}} \frac{1}{d_{ij}}}, \quad (5)$$

where,

$$d_{ij} = \mu_{ij}^T \Sigma_{ij}^{-1} \mu_{ij},$$
$$\mu_{ij} = T(\mu_i) - \mu_j, \Sigma_{ij} = R^T \Sigma_i R + \Sigma_j.$$

3.3 | TEASER++

Both ICP and NDT-based registration algorithms are powerful registration tools that have been employed actively in the current literature, but in recent times more robust and fast registration algorithms have been introduced. A representative example of this is the TEASER++ algorithm (Yang et al., 2020). This algorithm is specifically designed to provide robust pointcloud registration in the presence of large amounts of outlier correspondences, a condition expected in the outdoor agricultural data set. The registration algorithm utilizes the correspondences (Fast Point Feature Histograms (FPFH) point features used for the color point cloud in the paper and also used in this work) between points and uses a graph-based method of finding the maximum clique to reject most of the outliers. In addition, the registration cost function is decoupled for translation, rotation, and scale estimation and based on Truncated Least Squares (TLS) cost, which is robust to a large fraction of outlier correspondences.

The algorithm is also supplemented with easily implementable code, which makes it a great candidate for comparison. Given the complexity of the algorithm and space considerations in this article, we refer the interested reader to the original paper (Yang et al., 2020).

3.4 | Introduction of color information

The ICP and NDT methods described above do not use color information, and the cost function is solely based on the geometrical information of the points. With the introduction of RGB-D data, the cost functions can be adjusted either by using the color information of every single point or by using image features to find the correspondences. Korn et al. (2014) find the correspondences between the point clouds using color points and register them using the ICP registration. In Huhle et al. (2008), colored NDT cells are used for registrations which are defined as Gaussian mixture models (GMM) in color space and corresponding weighted spatial means and covariance. The color NDT registration is derived from the NDT P2D registration cost function, and the cost is calculated as the weighted negative likelihood of spatial point in the GMM of NDT cells, where the weights are likelihood in the color space. Andreasson and Stoyanov (2012) used SURF image features to find the correspondence between points in two RGB-D frames, and registration is done using the NDT D2D registration method. Our method is derived from the work of Korn et al. (2014) and Stoyanov et al. (2012), by introducing a novel approach to utilizes the color and geometric info for finding the correspondences between the colored NDT cells. The NDT D2D registration cost function is used for minimizing the distance between correspondences.

A schematic representation of the inclusion of color information in point cloud registration is presented in Figure 1. The pipeline from RGB-D images to the registered point cloud includes getting the

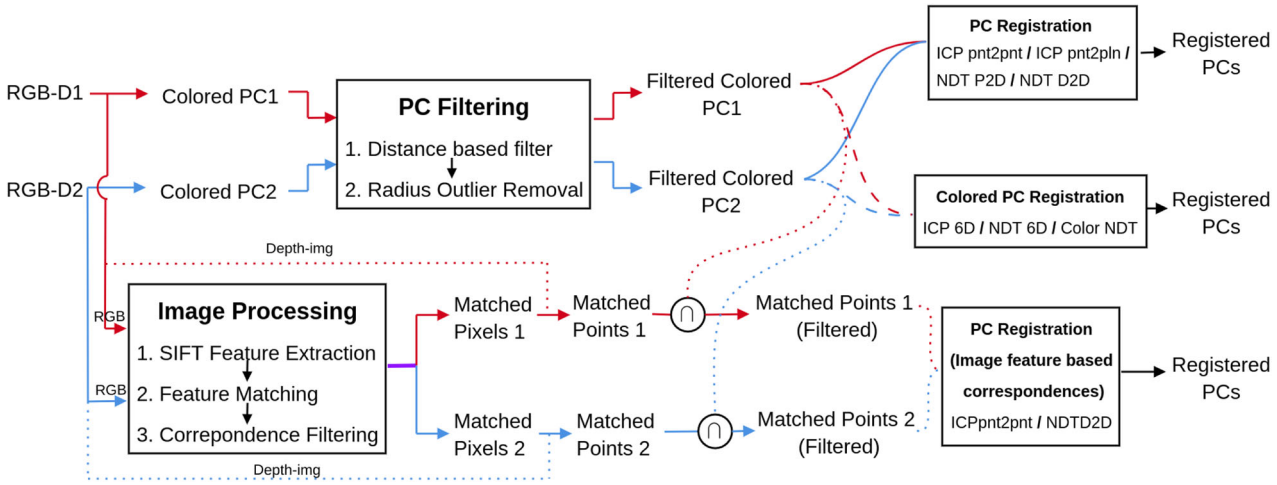

FIGURE 1 Flowchart for the scene registration using RGB-D images. Registration using 3D points and/or color information and Registration using image features and 3D points.

point cloud from RGB-D images, filtering the point cloud, getting the correspondences based on either distance or image features, and then performing the registration.

4 | METHODS AND MATERIALS

4.1 | Evaluation data sets

To evaluate the registration methods outlined above we use 2 data sets—GRAPES3D data set (Kurtser & Edan, 2018b), and TUM RGBD data set (Sturm et al., 2012). The data sets were chosen given the following criteria. (1) The data set must be acquired by a commercial grade RGB-D camera; (2) It must include a view towards the target with significant variability in distance to the targets and other objects in the scene; (3) Consecutive frames must be acquired with significant overlap (i.e., acquisition frequency must be reasonable to the speed of motion of the camera); (4) The pointcloud and color images should be acquired in a feature-rich environment. The criteria were chosen to adhere to the possible application of the proposed algorithms - pointcloud registration for better crop monitoring in orchards, vineyards, and greenhouse conditions.

The proposed data sets were both collected using RGB-D cameras to acquire colored pointcloud. GRAPES3D data sets represent the environmental vineyards and orchard conditions in which the algorithms are intended to be used. The TUM data set provides valuable benchmark data for deeper insights into algorithm performance in noise-free indoor conditions. The high-quality data of the TUM data set provides the additional ability to compare registration algorithms subject to artificial noise to gain insights into the stability and robustness of the registration algorithms.

Unfortunately, the number of RGB-D data sets publicly available for benchmarking is limited at this point and does not adhere to the criteria mentioned above. Well-established benchmarks such as SugarBeats 2016 (Chebrolu et al., 2017) and Rosario (Pire et al., 2019) are acquired for the use cases of aerial crop monitoring in open fields. While these data sets include a detailed ground truth using GPS-RTK, the acquisition protocol differs significantly. The pointclouds are typically acquired in a top-down viewpoint, which occludes illumination disturbances and provides a rather constant distance to target measure in field conditions. Additionally, the acquisition is often performed at a low frequency with limited overlap between frames. While the TUM data set was not acquired in agricultural conditions, the acquisition protocol used is adhering to the criteria mentioned above and provides additional insights. In Appendix B we provide additional results of applying the algorithms to the SugraBeats2016 data set (Chebrolu et al., 2017) and further explain the limitations in translating the suggested methods to aerial crop monitoring applications.

4.1.1 | GRAPES3D

The RGB-D point clouds were collected using an Intel Realsense D435 camera mounted on Greenhouse Spraying Robot (GSR) platform in two different conditions, a *controlled outdoor setup* with potted grape plants and a *commercial vineyard setup*. The Realsense D435 has field of view of 87°(±3°) × 58°(±1°) × 95°(±3°) and has active stereo depth resolution of 1280 × 720. The data is collected by teleoperating the GSR robot in a straight line with the camera mounted in two different configurations: (1) facing the growing row at 90°; and (2) facing the growing row at 45° horizontal angle with respect to the moving direction. The data set contains bag files with color images, depth images, and camera info. The color and depth

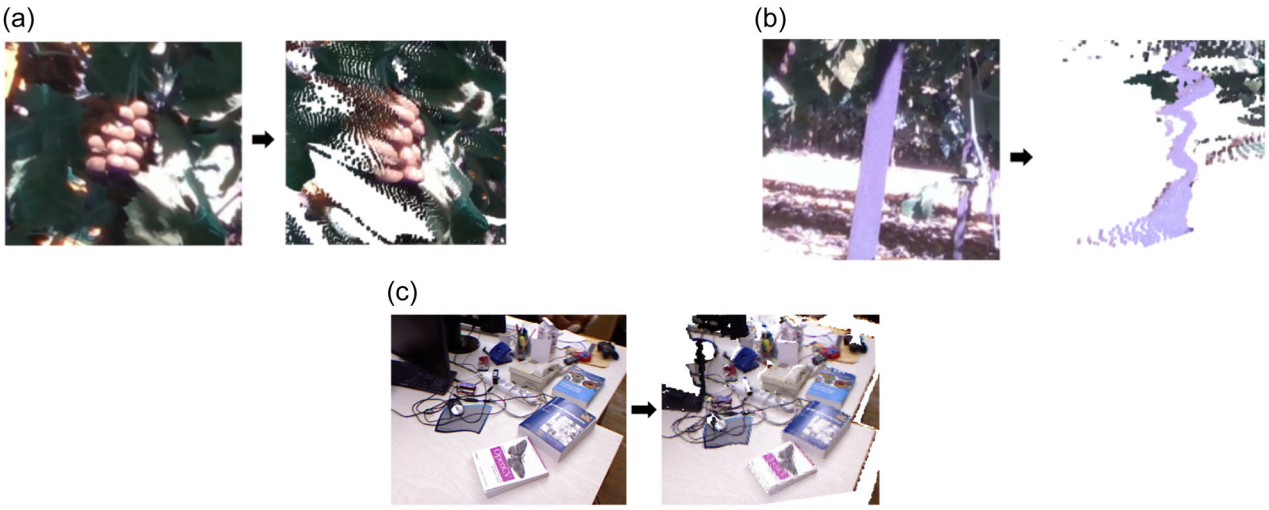

FIGURE 2 Outlier points example from the GRAPES3D data set and pointcloud example for TUM RGBD. (a) Overflowing points at the edge of an object, (b) wavy surface points for a flat surface, and (c) noise-free pointcloud example from the TUM data set.

images are extracted from bag files and aligned using *rs-covert* tool of the librealsense library.[2] These aligned RGB-D images are converted to 3D colored point cloud using the following equation:

$$p = \left(\frac{u - c_x}{f_x}d, \frac{v - c_y}{f_y}d, d\right)^T \tag{6}$$

$$d = \mathcal{I}_d(u, v)/s,$$

where, p is the 3D point in camera frame corresponding to the pixel coordinates (u, v), d is the depth value at the pixel coordinates (u, v) in depth image $\mathcal{I}_d$, s is the scale factor, c_x and c_y are the pixel center of the camera sensor, and f_x and f_y are the focal length of camera.

4.1.2 | TUM data sets

The sequences of the TUM RGBD data set used in this work were collected using a Microsoft Kinect sensor in indoor scenarios like offices and rooms. A high-accuracy motion-capture system with eight high-speed tracking cameras was used to collect the ground-truth trajectory. The data set includes scans with significantly less noise than expected in scans collected in outdoor agricultural conditions, as seen in Figure 2. As a result, the evaluation using the TUM data set aims to provide insights into our method's robustness to noise. Specifically, we evaluate the following scenarios—(1) noisy blurred images and (2) sparse pointclouds. We used "freiburg1_desk," "freiburg1_room," and "freiburg1_xyz" data sequences as these are long sequences with features present in each scan for registration methods and SIFT feature matching.

[2]https://github.com/IntelRealSense/librealsense- last accessed October 2022.

Additionally, due to the availability of ground-truth information in the TUM data sets, we are able to provide registration error measures compared to ground truth as described in Section 4.4.

4.2 | Preprocessing

In Grapes3D data sets, the point cloud obtained from the RGB-D image contains many outlier points which need to be removed. The outlier points in a point cloud are points that do not belong to the surface of any object and occur due to noise in the sensor or ambient noise/parameters, as seen in Figure 2. Some common reasons for noise/outlier points in the point cloud generated using an RGB-D camera are different viewing angles, light intensities, different reflection properties of the objects, vibration, or jerk in the camera position. Since the RGB-D camera is calibrated for near objects, we observed a reduction in depth accuracy with increased object distance. Therefore the point cloud must first be filtered based on the distance from the camera origin. Other outlier points, like overflowing points near the edges, wavy surfaces, or points due to sensor noise, can be removed (not completely) using analytical methods like radius outlier removal or statistical outlier rejection. In this work, the radius outlier removal method is used to filter the outliers. According to this method, the points which satisfy the condition in Equation (7) are filtered from the point cloud, where $\mathcal{N}$ is the neighborhood function that returns the number of points in the radius r of point p and $n_{\min}$ is the threshold for a minimum number of points. The different stages of the filtered point cloud are shown in Figure 3.

$$\mathcal{N}(p, r) < n_{\min}. \tag{7}$$

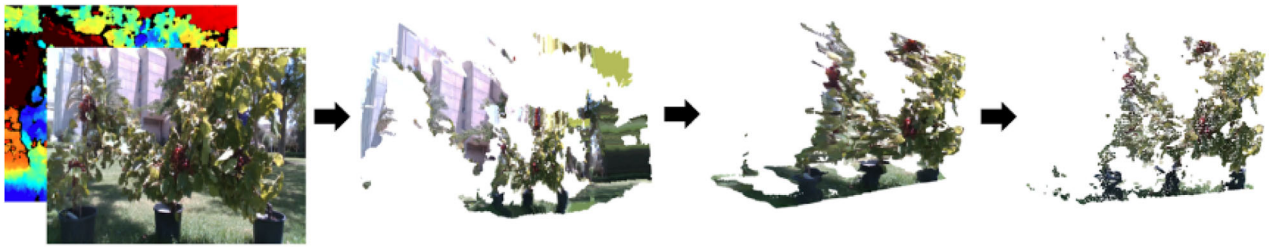

FIGURE 3 Filtered point cloud at different filtration stages. From left to right (i) RGB-D image, (ii) unfiltered point cloud, (iii) filtered point cloud based on distance filter, and (iv) filtered point cloud based on radius outlier filter and voxel grid downsampling.

From the aligned RGB-D images, the point cloud is calculated using Equation (6) with scale factor s = 800, and intrinsic camera parameters. The point clouds are filtered using the same criteria for each registration algorithm (1) distance-based filter: points at a distance of more than 3 m are rejected; (2) radius based filter: the point is rejected if the number of neighboring points in a radius of 0.01 m is less than 20 (3) voxel grid filter: if downsampling is required (in case of NDT P2D registration), the grid size of 0.02 m is used. The same preprocessed clouds are used for all compared methods.

To evaluate the methods under different colorspaces, three versions of the data sets were generated in RGB, HSV, and Lab colorspaces.

For the TUM data set, Equation (6) was used to convert the RGBD scans into pointclodus; the camera parameters (focal length and image center) and scale factor used for conversion were taken from Sturm et al. (2012). For this data set, no outlier removal preprocessing procedure was done because pointclouds are inherently less noisy and more feature-rich, as seen in Figure 2. The only pre-processing was to artificially introduce noise in pointclouds to evaluate the robustness of the algorithms. These degradation procedures included generating data sets with (1) downsampled pointclouds with a voxel size of 1 cm, and (2) applying a Gaussian blur kernel with window size 5 × 5.

4.3 | Proposed registration method

The new method (NDT-6D) that is proposed in this work is based on Stoyanov et al. (2012) and Korn et al. (2014). The method uses NDT D2D registration but introduces color to the method by calculation of NDT cells mean using both point and color information (Equation [8]). For covariance calculation using Equation (9), only geometric information is used. Colors are represented in Lab color space as used in the work of Korn et al. (2014).

$$\mu_{6d} = \left[\mu_p^T, \mu_c^T\right]^T, \tag{8}$$

$$\Sigma = \frac{1}{n}\sum_i^n (x_i - \mu_p)^T (x_i - \mu_p). \tag{9}$$

The cell means μ_{6d} are used only to find the cell correspondences between NDT maps using k-nearest neighbor methods. And registration is done by using only geometric information of the cell (μ_p, and Σ). For optimization of the NDT D2D registration cost function, an auto-differentiation library Ceres (Agarwal et al., 2022) was used. Hence, the derivatives of the NDT6D cost function were not calculated for optimization.

Our method's novelty lies in how we incorporate color information in NDT registration. While NDT D2D employs the same registration cost function, it neglects the color information in the point cloud and relies solely on its shape information. In contrast, our approach utilizes the color information of NDT cells to establish correspondences between the source and target point clouds, similar to Korn et al. (2014), without increasing the algorithm's complexity. This is different from the previous method of incorporating color information in NDT registration, which involved calculating color GMMs for NDT cells and was more complex. By building on the works of NDT D2D registration and ICP-6D, we propose a new NDT-based registration method that effectively integrates color information.

4.4 | Evaluation method and measures

In this work, state-of-the-art registration methods are tested and evaluated on a complex agricultural data set, GRAPES3D, and a non-agriculture indoor data set, TUM RGBD data set. The point clouds are pre-processed as described in the previous sub-section. The registration methods evaluated in this work are divided into three main groups. The first group contains the registration methods that only used geometric information for registration. The second group of the registration method uses color point information for registration, and the third group uses image features.

1. 3D/Geometric point info (Group1)—ICP point-to-point, ICP point-to-plane, NDT point-2-distribution, NDT distribution-to-distribution, and TEASER++.
2. Color + 3D/geometric point info (Group2)—Generalized ICP-6D, NDT-6D (ours), and color NDT registration

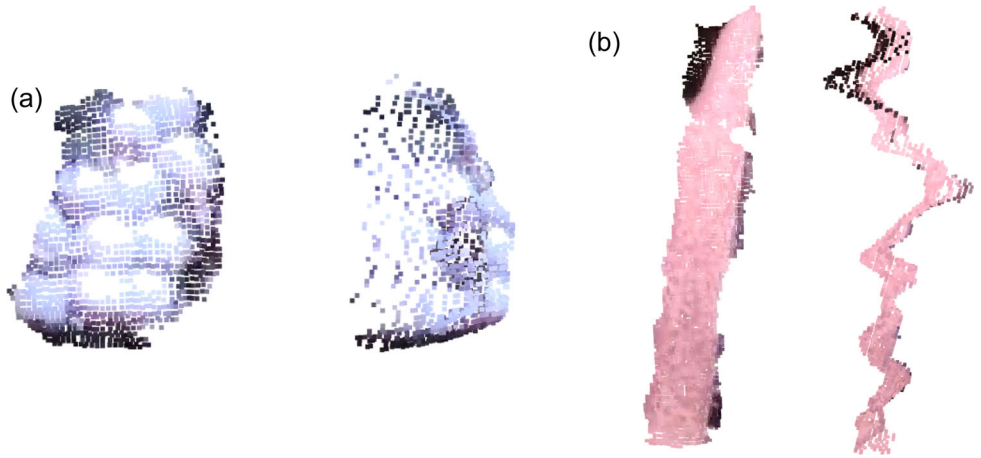

FIGURE 4 Example of extracted labeled data for registration evaluation. (a) Grape cluster and (b) supporting beam.

3. Image features + 3D/geometric point info (Group3)—SIFT Correspondence+ICP point-2-point, and SIFT Correspondence + NDT distribution-2-distribution

The registration methods in group1, group2, and group3 were tested on the GRAPES3D data set to test the performance of different registration methods in an agri-robotics scenario. In addition, the best-performing registration methods are further evaluated for robustness against noise using the TUM RGBD data set.

We performed pairwise registration of every consecutive scan for all registration. In the case of all ICP-based registration methods, we downsampled the pointclouds with a voxel grid of 0.015 m, and the grid size to create an NDT map for all evaluations was 0.15 m. For the TEASER++ algorithm, the noise parameter used was 25 mm, which provides the balance between robustness and fast iteration.

Due to the lack of ground truth information in the GRAPES3D data set, we choose to evaluate the registration performance on the GRAPES3D data set by calculating the root mean square error (RMSE) between the identifiable and static objects like grape clusters, pots, or beams (Figure 4). This method allows greater insight into the registration capabilities of small objects of interest, a feature required for acquiring physical crop measures, such as grape cluster volume.

The object of interest is marked by drawing a bounding box in the color image, followed by extracting the corresponding point cloud using color-based segmentation. To minimize errors and bias of a manual label, we use the practice outlined by Kurtser and Edan (2018b) for the extraction of a grape cluster from a region of interest marked manually. The steps include (1) Conversion of the image into NDI color space; (2) Clustering using K-means in NDI color space; (3) Clustering using K-means in 3D space (i.e., using points) for the NDI clusters. We used grape clusters for the evaluation by marking one prominent cluster in every fifth scan and extracting the pointcloud. The total number of grape clusters extracted from the whole data set was 194.

Given $\mathcal{X}$ the source point cloud, $\mathcal{T}(\mathcal{X})$ is the transformed point cloud in the reference frame of target point cloud $\mathcal{Y}$, the following measures are extracted:

1. Root mean square error between the labeled points $\in\mathcal{T}(\mathcal{X})$ and labeled points $\in\mathcal{Y}$.
2. Euclidean distance between the center of mass of the labeled points $\in\mathcal{T}(\mathcal{X})$ and labeled points $\in\mathcal{Y}$.

In the case of the TUM RGBD data set, the ground truth trajectory is given and used to evaluate different registration methods. We have reported the RMSE of relative pose error (RPE) using the tool provided with the data set.

5 | RESULTS AND DISCUSSION

5.1 | Results on GRAPES3D data set

The registration is performed in a pairwise manner, and the registration results are compared using evaluation measures described in the previous section. The registration results are given in Table 1.

As can be interpreted from the measures collected in Table 1, registration of colored point cloud (RBG-D data) is better when either the color information or image features are used to find the point correspondences between the two scans. The differences are statistically significant ($p < 0.0001$) as shown in the three-way ANOVA followed by a posthoc analysis in Appendix A.

5.1.1 | Impact of color information

Of all the registration methods evaluated, ICP point-2-point registration performed the worst. The ICP point-2-plane

TABLE 1 Summary of registration results for algorithms not relying on color information, point-by-point color information, or color information from extracted image features.

		Controlled outdoors				Commercial vineyard			
	Viewing angle	90°		45°		90°		45°	
Color Information	Registration	RMSE	COM	RMSE	COM	RMSE	COM	RMSE	COM
NA	ICP Pt2pt (Besl & McKay, 1992)	0.101	0.129	0.11	0.107	0.137	0.152	0.064	0.074
	ICP Pt2Pl (Chen & Medioni, 1992)	0.071	0.094	0.072	0.069	0.116	0.125	0.04	0.047
	NDT P2D (Magnusson et al., 2007)	0.013	0.016	0.04	0.029	0.038	0.023	0.013	0.015
	NDT D2D (Stoyanov et al., 2012)	0.014	0.02	0.02	0.023	0.012	0.02	0.015	0.02
	Teaser++ (Yang et al., 2020)	0.022	0.024	0.02	0.024	0.024	0.031	0.018	0.022
Point Color Info	ICP-6D (Korn et al., 2014)	0.012	0.014	0.015	0.014	0.011	0.015	0.011	0.012
	Color NDT P2D (Huhle et al., 2008)	0.019	0.027	0.023	0.028	0.021	0.032	0.016	0.024
	NDT-6D HSV (ours)	0.077	0.048	0.102	0.06	0.146	0.05	0.047	0.037
	NDT-6D RGB (ours)	0.012	0.016	0.016	0.016	0.011	0.014	0.012	0.014
	NDT-6D Lab (ours)	0.012	0.015	0.016	0.016	0.011	0.014	0.012	0.014
Image Features	SIFT+NDT D2D (Andreasson & Stoyanov, 2012)	0.011	0.013	0.016	0.014	0.01	0.013	0.015	0.015
	SIFT+ICP Pt2Pt	0.011	0.013	0.015	0.014	0.011	0.015	0.011	0.012

Note: Measures collected for each registered pair of point clouds include root mean square error (RMSE) and distance between the center of mass (COM). The table presents the average value over all frame pairs in four scenarios from the GRAPES3D data set—controlled outdoor environment versus commercial vineyard conditions, and the camera facing straight into the growing row at 90° and 45°. Best, green; ours, orange.

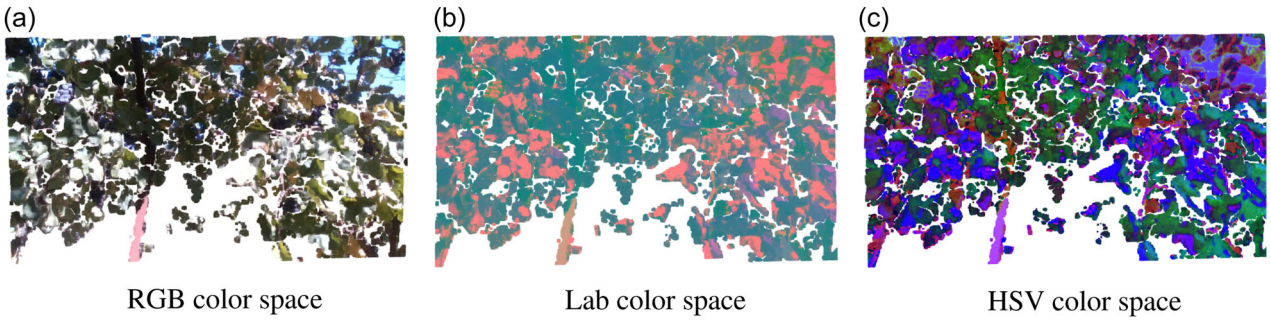

FIGURE 5 Pointcloud representation of (a) RGB, (b) Lab, and (c) HSV color spaces which shows similarity between RGB and Lab color spaces and dissimilarity from HSV color space.

(RMSE = 4–11 cm) also performed poorly, as no initial guess was provided for registration. The state-of-the-art pointcloud registration algorithms, like NDT point-2-distribution (RMSE = 1.3–4 cm) and NDT distribution-2-distribution (RMSE = 1.2–2 cm) have performed fairly well for a no color information based registration. The new and robust registration algorithm TEASER++ also did not perform well with RMSE close to 2 cm for all sequences. The reason for the poor performance of the above mentioned registration methods on this data set is the scene's complexity, with very few defined features and the abundance of noisy points in the point cloud. This conclusion is also supported by the fact that the simplified scene in the controlled outdoor data set yields a slightly lower average error (RMSE = 1.4–11 cm) compared with the commercial vineyard setting (RMSE = 1.2–13.7 cm).

The impact of scene complexity and noisy points on the registration can be mitigated by using the color information either on a point-by-point basis or as image features for searching corresponding points in the two scans making the registration more robust. In the case of using point color information, the ICP-6D and NDT-6D have performed similarly (RMSE = 1.1–1.6 cm) with better performance on the commercial vineyard data set, while Color NDT P2D registration was worst in the group.

5.1.2 | Impact of color-space

For NDT-6D, three color spaces (RGB, Lab, and HSV) were evaluated. RGB and Lab have performed better out of these color spaces

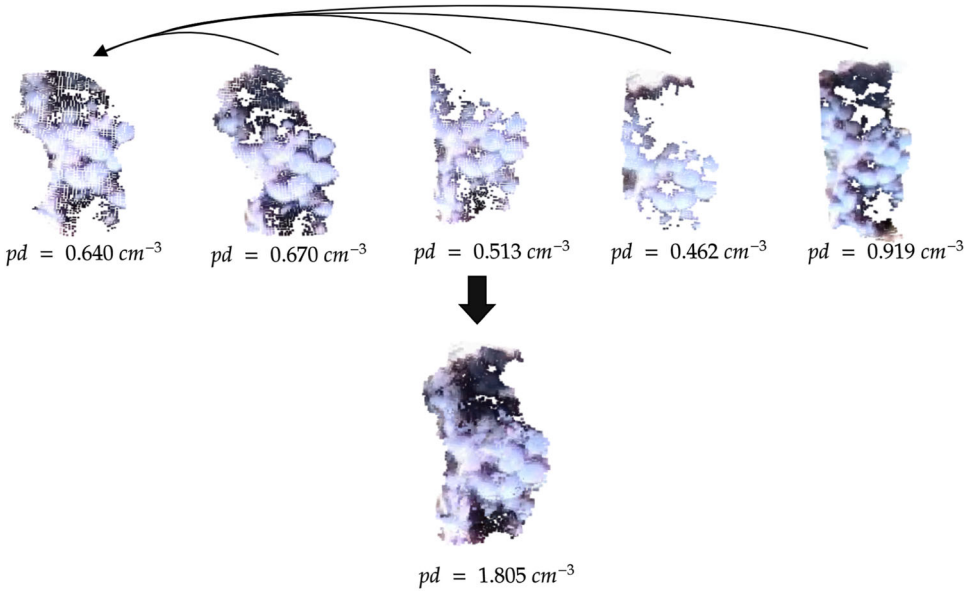

FIGURE 6 Point cloud of grape clusters for the individual frame and registered frames.

compared to HSV. The better performance of RGB and Lab color spaces could be that the leaves are all represented in similar values. In contrast, in the HSV color space, there are significant distinctions between objects of the same class, as seen in Figure 5. As a result, a lot of additional noise is added to the registration procedure leading to less accurate registration results.

5.1.3 | Registration using image features

For image features-based registration, NDT D2D and ICP point-2-point registration methods are evaluated with prefixed/predefined point correspondences using the best matched SIFT features (feature distance < 0.5). Between these two methods, SIFT+ICP point-2-point performed better when the viewing angle was 45°, and SIFT+NDT D2D performed better when the viewing angle was 90°. For SIFT based registration method the RMSE = 1.0–1.6 cm. The SIFT-NDT registration has performed better for a viewing angle of 90° as a more detailed NDT map was created due to direct viewing of the plants.

5.1.4 | Overall performance

In general, the best performing algorithms yielded a registration error (RMSE and COM distance) of 1–1.6 cm, with NDT-6D, giving the second-best results with a difference of 1 mm in RSME compared to the best results. These differences are found to be statistically insignificant ($p.val > 0.1p < 0.1$, see Appendix A). The RMSE error originates not only in registration error but, among others, in the manual labeling mismatch (shape of the labeled area) and variation in sensory data acquisitions of the same object from different directions. Therefore, with increased accuracy in the acquisition process, including the acquisition of ground truth information, the registration error could be even lower.

Nevertheless, this registered point cloud can be used in various outdoor in-field agricultural robotics applications with the reported accuracy. For example, previous research on in-field single frame grape cluster size estimation, using the same datasets that were evaluated in this work (Kurtser, Ringdahl, Rotstein, Berenstein, et al., 2020), reported a fitting error of 2.5–3.5 cm when estimating grape cluster size. The fitting was performed based on a single frame with a low average point cloud density of 0.6408/cm^3 for the test case shown in Figure 6. With the implementation of a registration procedure, cloud density is expected to increase (two times more for the test case), which will increase fitting accuracy for in-field phenotyping procedures. The registered point cloud cluster is downsampled to 80% using a voxel filter of 2.5 mm. The volume to calculate the point density is the maximum cluster volume out of all considered frames.

5.2 | Results on TUM RGBD data set

From the GRAPES3D data sets, we established that using color information in the form of color features or image features significantly increases registration accuracy. Therefore for the TUM data set, we evaluate only the methods employing color that generated the best results. We evaluate the algorithms on the data set with added noise similar to the ones experienced in the grapes3D data sets, including (1) Gaussian blurring kernel size 5 × 5; and (2) pointcloud downsampling (DS) with 1 cm voxel size. Table 2 summarizes the RMSE of relative position error, translation error eT (m), and rotation error eR (°) on the three data set sequences.

TABLE 2 Summary of registration results on TUM RGBD data set.

Noise	Method		Room	Desk	XYZ
WO	SIFT+ICP	eT	0.033	0.033	0.018
		eR	1.235	1.385	0.018
	SIFT+NDT	eT	0.011	0.014	0.007
		eR	0.756	1.026	0.571
	ICP6D	eT	0.026	0.027	0.012
		eR	1.23	1.565	0.828
	NDT6D	eT	0.014	0.018	0.01
		eR	0.869	1.298	0.686
Blur	SIFT+ICP	eT	0.041	0.059	0.039
		eR	10.06	9.307	1.733
	SIFT+NDT	eT	0.022	0.021	0.012
		eR	1.251	1.269	0.76
	ICP6D	eT	0.026	0.026	0.011
		eR	1.198	1.459	0.78
	NDT6D	eT	0.016	0.02	0.01
		eR	0.932	1.325	0.675
DS	SIFT+ICP	eT	0.043	0.058	0.04
		eR	6.55	3.741	1.91
	SIFT+NDT	eT	0.025	0.026	0.015
		eR	1.403	1.392	0.943
	ICP6D	eT	0.019	0.019	0.007
		eR	0.971	1.242	0.654
	NDT6D	eT	0.016	0.02	0.009
		eR	0.935	1.328	0.677

Note: The root mean square error (RMSE) for relative pose error (eT [m] and eR [°]) for pointcloud without noise (WO), point cloud with blur (Blur), and downsampled point pointcloud (DS). Best, green; ours, orange.

We can see from Table 2 that NDT-based methods have performed better than ICP-based methods for each sequence with or without additive noise (RMSE = 0.7–1.4 cm). We can also see that in the scenario of no added noise, NDT6D has shown comparable but slightly worse results than SIFT+NDT. The results are statistically significant ($p < 1e{-}9$), as shown in Appendix A. With the addition of noise, NDT6D remained relatively robust, increasing RMSE by ~14% compared with the case without added noise. In comparison, the increase in RMSE for the best-performing registration method (SIFT+NDT) is more than 75% for added noise. Our proposed method, NDT-6D, outperforms the SIFT+NDT method for blur noise, especially in the room scenario. In the case of downsampling noise, SIFT+NDT accuracy was significantly degraded, with NDT6D and ICP6D presenting the best results.

6 | CONCLUSIONS

In conclusion, point cloud registration is a critical component for autonomous agricultural robotics in field conditions, enabling in-field crop manipulation and phenotyping activities. This paper has highlighted how the state-of-the-art methods for registering uncolored point clouds collected in the above-mentioned conditions are limited, and the importance of colored point clouds or complementary registered image data has been emphasized.

Most current methods utilizing color information rely on image features. This paper has proposed an alternative approach, NDT-6D, which directly integrates color into the distance metrics between NDT cells. The proposed method demonstrated comparable registration accuracy to current state-of-the-art color-based methods under highly accurate pointclouds while maintaining higher consistency in accuracy under noisier point clouds. Additionally, it has shown the capability to register point clouds under challenging conditions where image feature extraction is difficult.

These findings suggest that implementing the proposed algorithms in current commercial grade RGB-D cameras could potentially improve the accuracy of in-field agricultural operations relying on point cloud data, such as plant manipulation and crop and plant size estimation for growth modeling. However, state-of-the-art algorithms have limitations under conditions where the scene lacks features, as demonstrated in the aerial crops use case.

The successful use of RGB-D sensors for registration implies that, in certain conditions, the technology could replace LiDARs for autonomous navigation. Nevertheless, further testing is necessary to validate this proposition.

ACKNOWLEDGMENTS

This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 858101.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Himanshu Gupta http://orcid.org/0000-0001-9364-7994
Achim J. Lilienthal http://orcid.org/0000-0003-0217-9326

REFERENCES

Agarwal, S., Mierle, K. & Team, T.C.S. (2022) *Ceres Solver*. Available at: https://github.com/ceres-solver/ceres-solver

Alenya, G., Dellen, B. & Torras, C. (2011) 3D modelling of leaves from color and ToF data for robotized plant measuring. In: *2011 IEEE International Conference on Robotics and Automation*. pp. 3408–3414.

Al-Hiary, H., Bani-Ahmad, S., Reyalat, M., Braik, M. & Alrahamneh, Z. (2011) Fast and accurate detection and classification of plant diseases. *International Journal of Computer Applications*, 17(1), 31–38.

Andreasson, H. & Stoyanov, T. (2012) Real time registration of RGB-D data using local visual features and 3D-NDT registration. In: *SPME Workshop at International Conference on Robotics and Automation (ICRA)*. Vol. 1.

Arad, B., Balendonck, J., Barth, R., Ben-Shahar, O., Edan, Y. & Hellström, T. et al. (2020) Development of a sweet pepper harvesting robot. *Journal of Field Robotics*, 37(6), 1027–1039.

Arad, B., Kurtser, P., Barnea, E., Harel, B., Edan, Y. & Ben-Shahar, O. (2019) Controlled lighting and illumination-independent target detection for real-time cost-efficient applications. The case study of sweet pepper robotic harvesting. *Sensors*, 19(6), 1390.

Bac, C.W., van Henten, E.J., Hemming, J. & Edan, Y. (2014) Harvesting robots for high-value crops: State-of-the-art review and challenges ahead. *Journal of Field Robotics*, 31(6), 888–911.

Bakker, T., van Asselt, K., Bontsema, J., Mü ller, J. & van Straten, G. (2006) An autonomous weeding robot for organic farming. In: *Field and Service Robotics* 25. pp. 579–590.

Barth, R., Hemming, J. & van Henten, E.J. (2016) Design of an eye-in-hand sensing and servo control framework for harvesting robotics in dense vegetation. *Biosystems Engineering*, 146, 71–84.

Bawden, O., Kulk, J., Russell, R., McCool, C., English, A., Dayoub, F., Lehnert, C. & Perez, T. (2017) Robot for weed species plant-specific management. *Journal of Field Robotics*, 34(6), 1179–1199.

Besl, P.J. & McKay, N.D. (1992) Method for registration of 3-D shapes. In: *Sensor fusion IV: control paradigms and data structures*. Vol. 1611, pp. 586–606.

Biber, P., Weiss, U., Dorna, M. & Albert, A. (2012) Navigation system of the autonomous agricultural robot bonirob. In: *Workshop on Agricultural Robotics: Enabling Safe, Efficient, and Affordable Robots for Food Production (Collocated with IROS 2012), Vilamoura, Portugal*.

Botterill, T., Paulin, S., Green, R., Williams, S., Lin, J., Saxton, V., Mills, S., Chen, X. & Corbett-Davies, S. (2017) A robot system for pruning grape vines. *Journal of Field Robotics*, 34(6), 1100–1122.

Bulanon, D., Burks, T. & Alchanatis, V. (2009) Fruit visibility analysis for robotic citrus harvesting. *Transactions of the ASABE*, 52(1), 277–283.

Chebrolu, N., Läbe, T. & Stachniss, C. (2020) Spatio-temporal non-rigid registration of 3d point clouds of plants. In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*. pp. 3112–3118.

Chebrolu, N., Lottes, P., Schaefer, A., Winterhalter, W., Burgard, W. & Stachniss, C. (2017) Agricultural robot dataset for plant classification, localization and mapping on sugar beet fields. *The International Journal of Robotics Research*, 36(10), 1045–1052.

Chen, Y. & Medioni, G. (1992) Object modelling by registration of multiple range images. *Image and Vision Computing*, 10(3), 145–155.

Dale, L.M., Thewis, A., Boudry, C., Rotar, I., Dardenne, P., Baeten, V. & Pierna, J.A.F. (2013) Hyperspectral imaging applications in agriculture and agro-food product quality and safety control: a review. *Applied Spectroscopy Reviews*, 48(2), 142–159.

Das, A. & Waslander, S.L. (2014) Scan registration using segmented region growing NDT. *The International Journal of Robotics Research*, 33(13), 1645–1663.

Dong, W., Roy, P. & Isler, V. (2020) Semantic mapping for orchard environments by merging two-sides reconstructions of tree rows. *Journal of Field Robotics*, 37(1), 97–121.

Fernández-Novales, J., Saiz-Rubio, V., Barrio, I., Rovira-Más, F., Cuenca-Cuenca, A., SantosAlves, F., Valente, J., Tardaguila, J. & Diago, M.P. (2021) Monitoring and mapping vineyard water status using non-invasive technologies by a ground robot. *Remote Sensing*, 13(14), 2830.

Gao, X., Li, J., Fan, L., Zhou, Q., Yin, K., Wang, J., Song, C., Huang, L. & Wang, Z. (2018) Review of wheeled mobile robots' navigation problems and application prospects in agriculture. *IEEE Access*, 6, 49248–49268.

Guo, N., Zhang, B., Zhou, J., Zhan, K. & Lai, S. (2020) Pose estimation and adaptable grasp configuration with point cloud registration and geometry understanding for fruit grasp planning. *Computers and Electronics in Agriculture*, 179, 105818.

Hacking, C., Poona, N., Manzan, N. & Poblete-Echeverría, C. (2019) Investigating 2-D and 3-D proximal remote sensing techniques for vineyard yield estimation. *Sensors*, 19(17), 3652. Available from: https://www.mdpi.com/1424-8220/19/17/3652; https://doi.org/10.3390/s19173652

Harel, B., Kurtser, P., Van Herck, L., Parmet, Y. & Edan, Y. (2016) Sweet pepper maturity evaluation via multiple viewpoints color analyses. In: *CIGR-AgEng Conference, Aarhus, Denmark, 26-29 June, 2016*. pp. 1–7.

Hemming, J., Ruizendaal, J., Hofstee, J.W. & Van Henten, E.J. (2014) Fruit detectability analysis for different camera positions in sweet-pepper. *Sensors*, 14(4), 6032–6044.

Huhle, B., Magnusson, M., Straßer, W. & Lilienthal, A.J. (2008) Registration of colored 3D point clouds with a kernel-based extension to the normal distributions transform. In: *2008 IEEE International Conference on Robotics and Automation*. pp. 4025–4030.

Kamilaris, A. & Prenafeta-Boldú, F.X. (2018) Deep learning in agriculture: a survey. *Computers and electronics in agriculture*, 147, 70–90.

Kirk, R., Mangan, M. & Cielniak, G. (2020) Feasibility study of in-field phenotypic trait extraction for robotic soft-fruit operations. *UKRAS20 Conference: "Robots into the real world" Proceedings* (pp. 21–23). https://doi.org/10.31256/Uk4Td6l

Korn, M., Holzkothen, M. & Pauli, J. (2014) Color supported generalized-ICP. In: *2014 International Conference on Computer Vision Theory and Applications (VISAPP)*. Vol. 3, pp. 592–599.

Kurtser, P. & Edan, Y. (2018a) Statistical models for fruit detectability: spatial and temporal analyses of sweet peppers. *Biosystems Engineering*, 171, 272–289.

Kurtser, P. & Edan, Y. (2018b) The use of dynamic sensing strategies to improve detection for a pepper harvesting robot. In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. pp. 8286–8293.

Kurtser, P., Ringdahl, O., Rotstein, N. & Andreasson, H. (2020) Pointnet and geometric reasoning for detection of grape vines from single frame rgb-d data in outdoor conditions. In: *3rd Northern Lights Deep Learning Workshop, Tromsö, Norway 20-21 January, 2019*. Vol. 1, pp. 1–6.

Kurtser, P., Ringdahl, O., Rotstein, N., Berenstein, R. & Edan, Y. (2020) In-field grape cluster size assessment for vine yield estimation using a mobile robot and a consumer level RGB-D camera. *IEEE Robotics and Automation Letters*, 5(2), 2031–2038.

Lehnert, C., Sa, I., McCool, C., Upcroft, B. & Perez, T. (2016) Sweet pepper pose detection and grasping for automated crop harvesting. In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*. pp. 2428–2434.

Liao, Q., Sun, D. & Andreasson, H. (2022) Fuzzypsreg: strategies of fuzzy cluster-based point set registration. *IEEE Transactions on Robotics*, 38(4), 2632–2651. https://doi.org/10.1109/TRO.2021.3123898

Magnusson, M., Lilienthal, A. & Duckett, T. (2007) Scan registration for autonomous mining vehicles using 3D-NDT. *Journal of Field Robotics*, 24(10), 803–827.

Magnusson, M., Nuchter, A., Lorken, C., Lilienthal, A.J. & Hertzberg, J. (2009) Evaluation of 3D registration reliability and speed-a comparison of ICP and NDT. In: *2009 IEEE International Conference on Robotics and Automation*. pp. 3907–3912.

Malavazi, F.B., Guyonneau, R., Fasquel, J.-B., Lagrange, S. & Mercier, F. (2018) Lidar-only based navigation algorithm for an autonomous agricultural robot. *Computers and Electronics in Agriculture*, 154, 71–79.

Milella, A., Marani, R., Petitti, A. & Reina, G. (2019) In-field high throughput grapevine phenotyping with a consumer-grade depth camera. *Computers and Electronics in Agriculture*, 156, 293–306.

Nandi, C.S., Tudu, B. & Koley, C. (2016) A machine vision technique for grading of harvested mangoes based on maturity and quality. *IEEE Sensors Journal*, 16(16), 6387–6396.

Oliveira, L.F., Moreira, A.P. & Silva, M.F. (2021) Advances in agriculture robotics: a state-of-the-art review and challenges ahead. *Robotics*, 10(2), 52.

Papadimitriou, A., Kleitsiotis, I., Kostavelis, I., Mariolis, I., Giakoumis, D., Likothanassis, S. & Tzovaras, D. (2022) Loop closure detection and slam in vineyards with deep semantic cues. In: *2022 International Conference on Robotics and Automation (ICRA)*. pp. 2251–2258.

Pire, T., Mujica, M., Civera, J. & Kofman, E. (2019) The Rosario dataset: multisensor data for localization and mapping in agricultural environments. *The International Journal of Robotics Research*, 38(6), 633–641.

Pomerleau, F., Colas, F. & Siegwart, R. (2015) A review of point cloud registration algorithms for mobile robotics. *Foundations and Trends in Robotics*, 4(1), 1–104.

Ringdahl, O., Kurtser, P. & Edan, Y. (2019) Performance of rgb-d camera for different object types in greenhouse conditions. In: *2019 European Conference on Mobile Robots (ECMR)*. pp. 1–6.

Schunck, D., Magistri, F., Rosu, R.A., Cornelißen, A., Chebrolu, N. & Paulus, S. et al. (2021) Pheno4d: a spatio-temporal dataset of maize and tomato plant point clouds for phenotyping and advanced plant analysis. *Plos One*, 16(8), e0256340.

Shalal, N., Low, T., McCarthy, C. & Hancock, N. (2013) A review of autonomous navigation systems in agricultural environments. In *SEAg 2013: Innovative agricultural technologies for a sustainable future*. University of Southern Queensland.

Singh, V. & Misra, A.K. (2017) Detection of plant leaf diseases using image segmentation and soft computing techniques. *Information Processing in Agriculture*, 4(1), 41–49.

Stoyanov, T., Magnusson, M., Andreasson, H. & Lilienthal, A.J. (2012) Fast and accurate scan registration through minimization of the distance between compact 3D NDT representations. *The International Journal of Robotics Research*, 31(12), 1377–1393.

Sturm, J., Engelhard, N., Endres, F., Burgard, W. & Cremers, D. (2012) A benchmark for the evaluation of rgb-d slam systems. In: *Proceedings of the International Conference on Intelligent Robot Systems (IROS)*.

Tavares, A.C., Lawin, F.J. & Forssén, P.-E. (2020) Assessing losses for point set registration. *IEEE Robotics and Automation Letters*, 5(2), 3360–3367.

Tian, H., Wang, T., Liu, Y., Qiao, X. & Li, Y. (2020) Computer vision technology in agricultural automation—a review. *Information Processing in Agriculture*, 7(1), 1–19.

Tillett, N., Hague, T., Grundy, A. & Dedousis, A. (2008) Mechanical within-row weed control for transplanted crops using computer vision. *Biosystems Engineering*, 99(2), 171–178.

Vadivambal, R. & Jayas, D.S. (2011) Applications of thermal imaging in agriculture and food industry—a review. *Food and Bioprocess Technology*, 4(2), 186–199.

Valencia, R., Saarinen, J., Andreasson, H., Vallvé, J., Andrade-Cetto, J. & Lilienthal, A.J. (2014) Localization in highly dynamic environments using dual-timescale ndt-mcl. In: *2014 IEEE International Conference on Robotics and Automation (ICRA)*. pp. 3956–3962.

Villena-Martinez, V., Oprea, S., Saval-Calvo, M., Azorin-Lopez, J., Fuster-Guillo, A. & Fisher, R.B. (2020) When deep learning meets data alignment: a review on deep registration networks (drns). *Applied Sciences*, 10(21), 7524.

Vit, A. & Shani, G. (2018) Comparing RGB-D sensors for close range outdoor agricultural phenotyping. *Sensors*, 18(12), 4413.

Wang, X.A., Tang, J. & Whitty, M. (2020) Side-view apple flower mapping using edge-based fully convolutional networks for variable rate chemical thinning. *Computers and Electronics in Agriculture*, 178, 105673.

Yang, H., Shi, J. & Carlone, L. (2020) TEASER: fast and certifiable point cloud registration. *IEEE Transactions of Robotics*.

Zaenker, T., Lehnert, C., McCool, C. & Bennewitz, M. (2021) Combining local and global viewpoint planning for fruit coverage. arXiv preprint arXiv:2108.08114.

Zaenker, T., Smitt, C., McCool, C. & Bennewitz, M. (2020) Viewpoint planning for fruit size and position estimation. arXiv preprint arXiv:2011.00275.

Zemmour, E., Kurtser, P. & Edan, Y. (2017) Dynamic thresholding algorithm for robotic apple detection. In: *2017 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*. pp. 240–246.

How to cite this article: Gupta, H., Lilienthal, A.J., Andreasson, H. & Kurtser, P. (2023) NDT-6D for color registration in agri-robotic applications. *Journal of Field Robotics*, 40, 1603–1619. https://doi.org/10.1002/rob.22194

APPENDIX A: STATISTICAL ANALYSIS OF EXPERIMENTAL RESULTS

A.1 | GRAPES3D data set

The absolute registration errors produced for the GRAPES3D data set are presented in Figure A1. The figure presents the originally acquired errors and a log2 transformation of the errors, which normalizes their distribution.

Log transformed errors were tested with a three-way ANOVA (Table A1) accounting for registration algorithm (NDT D2D/NDT-6D in Lab color space/SIFT+NDT D2D), environment (Controlled outdoor environment/commercial Vineyard), and viewing angle (90°/45°). The residuals were found to adhere to the Normal distribution assumption (Kolmogorov–Smirnov test $p = 0.84$). A posthoc analysis of the statistically significant differences between algorithms (Table A2) shows the errors produced by the colorless D2D method are significantly ($p < 0.0001$) higher than the NDT-6D and SIFT+NDT methods.

A.2 | TUM data set

The absolute registration translation errors compared to IMU produced for the TUM data set are presented in Figure A2. The figure presents the originally acquired errors and a boxcox transformed errors with parameter $\lambda = -0.1$, which normalizes their distribution.

The transformed errors were tested with a three-way ANOVA (Table A3) accounting for registration algorithm (NDT-6D in/SIFT +NDT D2D/ICP6D), environment (room/desk/xyz), and noise (WO/Blur/DS). The residuals were found to adhere to the Normal distribution assumption (Kolmogorov–Smirnov test $p > 0.81$ for 500 randomly selected values). A posthoc analysis of the statistically significant differences between every pair of algorithms ($p < 1e{-}9$).

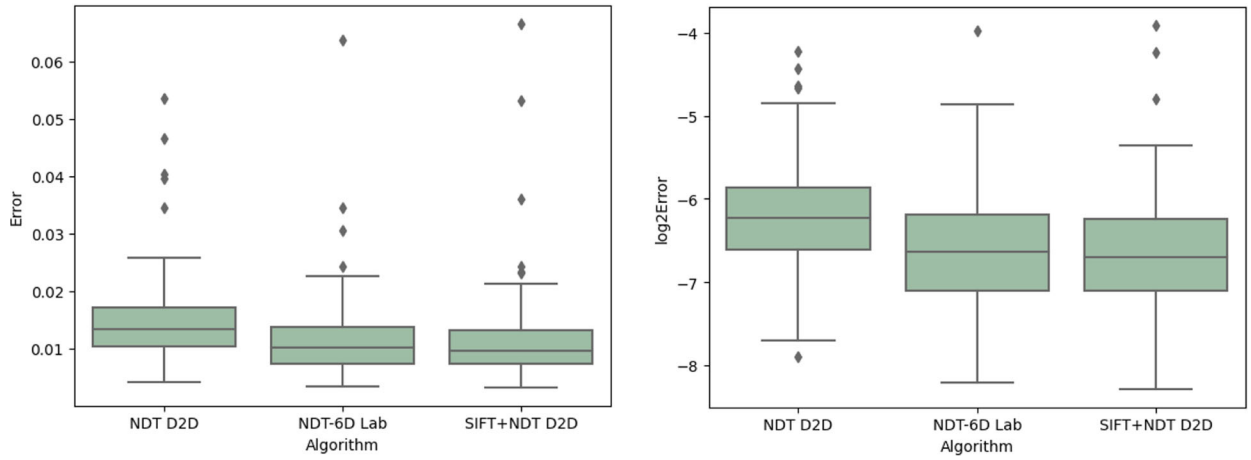

FIGURE A1 Distribution of registration errors for GRAPES3D data set before (top) and after (bottom) a log transformation as a function of registration algorithm. Results show log transformation is required to fit a normal distribution to the errors. Results also show higher error rates for the NDT D2D method compared to alternative.

TABLE A1 Results of a three-way ANOVA test performed on log-transformed registration errors for the GRAPES3D data set as a function of registration algorithm (NDT D2D/NDT-6D in Lab color space/SIFT+NDT D2D), environment (controlled outdoor environment/commercial Vineyard), and viewing angle (90°/45°). Results show statistically significant differences between algorithms, angles, and environments.

	sum_sq	*df*	*F*	PR(>*F*)
Algorithm	12.02	2.0	14.80	6.25e−07
Angle	13.58	1.0	33.45	1.46e−08
Environment	2.42	1.0	5.97	1.49e−02
Algorithm: Angle	0.11	2.0	0.13	8.72e−01
Algorithm: Environment	0.14	2.0	0.17	8.38e−01
Angle: Environment	0.41	1.0	1.01	3.13e−01
Residual	162.77	401.0	–	–

TABLE A2 Results of a posthoc test performed on log-transformed registration errors for the GRAPES3D data set as a function of registration algorithm (NDT D2D/NDT-6D in Lab color space/SIFT+NDT D2D). Results show statistically significant differences between NDT D2D and color-based algorithms.

	NDT D2D	NDT-6D Lab	SIFT +NDT D2D	
NDT D2D	–	0.000016	7.68e−07	Outliers
NDT-6D Lab	–	–	0.47	

APPENDIX B: ANALYSIS OF SUGARBEETS2016 DATABASE

The SugarBeets2016 data set is a benchmark data set acquired using a mobile robot equipped with an RGB-D camera in sugarbeet fields and presents a use case for crop monitoring in arable crops. The data is acquired using a Kinect sensor mounted on a mobile platform, with the RGBD camera facing the ground and the distance from the ground is ~1.1 m. The data was collected with 1 Hz frequency, during which the robot moved ~0.3 m forward. The pointcloud resolution is relatively small and reasonably flat (Figure B2) compared with pointcloud in Grapes3D and TUM RGBD data sets due to the restricted viewing area of the camera which poses challenges to various registration algorithms (Figures B1 and B2).

In this section, we present the evaluation of the outlined algorithms for registering the collected pointclouds. The outlined registration algorithms include no-color (ICP, NDT-D2D), color-based (ICP-6D, our NDT-6D), and feature-based (SIFT+NDT D2D). Unfortunately, all the above algorithms failed to register consecutive scans, as the pose estimation from scan registration resulted in an identity transformation matrix or a minor deviation from it.

Below we outline the reasons for the failure of the various tested registration algorithms resulting in their limitations.

- No-color: Geometric-based registration algorithms such as NDT D2D and ICP Pt2Pl rely on matching 3D shapes or minimizing point-to-plane errors. However, in the SugarBeets2016 data set, the top-down viewpoint facing the ground leads to flat point clouds with fewer distinct 3D features. The high speed and low frame acquisition rate also result in limited overlap between consecutive point clouds, making registration difficult. Additionally, the choice of using a Kinect sensor for indoor applications results in ambient light penetrating between the wheels, reducing the point cloud density in certain areas and further decreasing the available 3D features for registration. The output of these algorithms is usually an identity matrix.

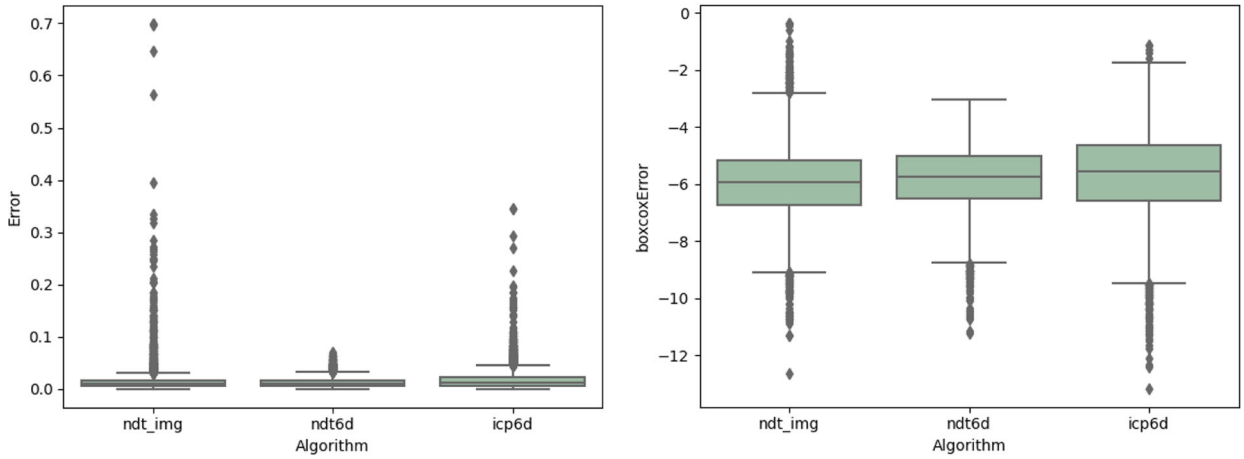

FIGURE A2 Distribution of registration errors for the TUM data set before (top) and after(bottom) a box-cox transformation with $\lambda = -0.1$ as a function of the registration algorithm. Results show box-cox transformation is required to fit a normal distribution to the errors. Results also show fewer outlier values and lower variance values in the error rate for NDT-6D.

TABLE A3 Results of a three-way ANOVA test performed on box-cox ($\lambda = -0.1$) transformed registration translation errors for the TUM data set as a function of registration algorithm (NDT-6D in/SIFT+NDT D2D/ICP6D), environment (room/desk/xyz), and noise (WO/Blur/DS). Results show statistically significant differences between algorithms.

	sum_sq	*df*	*F*	PR(>F)
Algorithm	227.67	2.0	79.79	3.28e−35
Environment	3821.73	1.0	2678.85	0.0e+00
Noise	201.12	2.0	70.48	3.32e−31
Algorithm: Environment	377.23	2.0	132.21	1.11e−57
Environment: Noise	21.94	2.0	7.69	4.58e−04
Algorithm: Noise	913.55	4.0	160.08	1.38e−134
Residual	2.22e4	1.60e4	NaN	NaN

FIGURE B1 Images from the SugarBeet2016 data set.

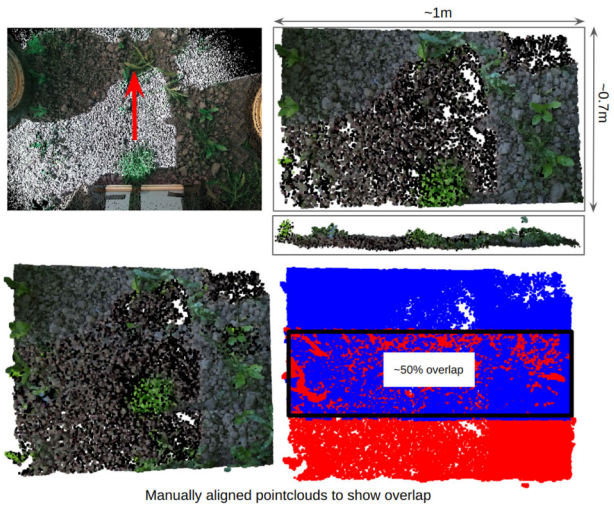

FIGURE B2 Visualization of single pointcloud from SugarBeet2016 data set and pointcloud overlap for consecutive pointclouds by manually registering the pointclouds.

- Color-based registration: In color-based registration methods such as ICP-6D and NDT-6D, color information is used to find point correspondences between point clouds. However, the Sugarbeets data set has significant areas of homogeneous color (the ground), which limits the number of distinct color features. Some frames may not contain any objects besides the soil and imaging artifacts, leading to errors such as points appearing on the robot or variations in color due to ambient light. The lack of variability in color information combined with the low overlap between consecutive point clouds and limited 3D features results in minor deviations from the identity matrix and high error rates compared to the ground truth.
- Image feature-based algorithms: Of the mentioned above limitations of color-based algorithms, the feature-based registration algorithm (SIFT+NDT D2D) shows the potential to register the SugarBeets pointcloud. In some consecutive frames, correct SIFT features can be observed (Figure B3 but

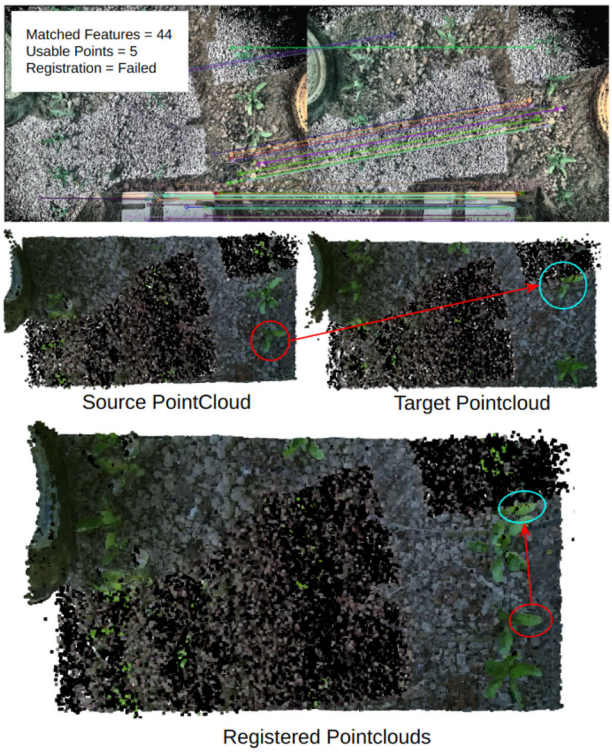

FIGURE B3 Visualization of the SIFT-NDT D2D registration for SugarBeet2016 data set showing matched SIFT features (top), source and target pointcloud (middle), and registered pointclouds. The registration of pointcloud failed even though correctly matched SIFT feature points are present because the noise in pointcloud results in an incorrect selection of matched pairs.

the number of valid points extracted from SIFT matches and filtered point clouds is significantly lower than other GRAPES3D and TUM RGBD datasets. As a result, The rate of erroneous outlier points is also higher, making their adverse effect on registration more notable. Although SIFT+NDT D2D has resulted in no valid registration or a highly erroneous registration matrix in most cases, it has worked for a few scan pairs with correct SIFT feature matches and valid points in the point cloud. An example of SIFT+NDT D2D registration is shown in Figure B3.

The aforementioned attributes of SugarBeets2016, arise mostly from the fact that the data set was not originally intended for use as a SLAM benchmark, limiting the applicability of current registration algorithms. Some of these attributes, such as errors caused by ambient light due to the use of a Kinect sensor, have been addressed in more recent RGBD sensors like the Intel Realsense D435, which was used to collect the GRAPES3D data set analyzed in this study. However, other attributes such as poor image and pointcloud features will persist in applications that follow the outlined protocol. Such applications require extensive development of registration algorithms to perform well under these conditions.