TOWARD A LOW-COST, MULTISPECTRAL, HIGH ACCURACY MAPPING SYSTEM FOR VINEYARD INSPECTION

S. Beniaouf[a], R. Mabillard[b], A. Gressin[a,*]

[a] University of Applied Sciences Western Switzerland (HES-SO / HEIG-VD), 1400 Yverdon-les-Bains, Switzerland,
Insit Institute - (sami.beniaouf,adrien.gressin)@heig-vd.ch
[b] IG group SA, Switzerland - romain.mabillard@ig-group.ch

Commission III, WG III/10

KEY WORDS: Precision agriculture, Low-cost, Multispectral, Mobile mapping, Positioning.

ABSTRACT:

Confronted with the climate change challenge and the territorial constraints, agriculture has to modernize itself. The use of georeferenced data and remote-sensing imagery is a major step in this direction. This precision mapping of crops requires powerful and accurate acquisition systems, while remaining financially attractive. The development of multispectral sensors and low-cost GNSS makes it possible to consider systems that will be able to map at the plant scale. However, these positioning systems do not yet guarantee a precise overlap of data acquired at different times. Thus, we propose in this paper a method to register terrestrial image data, acquired on vineyard plots. Our method seeks to avoid image registration problems, such as illumination changes, by detecting the vine stocks, reconstructing them in 3D, and registering them individually. The 3D detection method is based on an image-based object detection method (Faster R-CNN) and a structure-from-motion reconstruction of object-masked images. The results that we obtained on a vineyard plot, allowed us to validate the method, with a precision of less than 10 cm, making it possible to map the vine by stock.

1. INTRODUCTION

Precision agriculture is a major topic in the world today, aiming to increase crop production while reducing the use of phytosanitary products, and thus reducing the environmental impact of agriculture. The use of positioning and geographic information systems was cited as one of the 10 most useful technologies in this field (Crookston, 2006). In particular, remote sensing, multi-spectral imagery, deep learning, and positioning have together significant potential (Martos et al., 2021).

Usually, crop analysis can be performed at different scales, ranging from the use of satellite sensors to the drone (Khaliq et al., 2019). A large part of the literature proposes to use UAV coupled with lightweight multispectral sensors, which allow covering large areas at a low-cost (Maes and Steppe, 2019). In particular, in the field of vineyards, the recent development of extraction methods by deep learning allows precise mapping of specific vine diseases (Kerkech et al., 2020). However, the use of UAVs is subject to weather and legislative hazards, which makes flights often complicated or impossible.

In addition, the use of onboard cameras on ground machines allows for even more accurate crop mapping (Figure 1), without the hazards of the drone. In particular, the fruits are usually located at the bottom of the plant and are only visible in oblique images taken from the ground. Some studies are already underway to use such data for the detection of vineyard diseases (Rançon et al., 2019). Nevertheless, such a system requires precise positioning solutions that are often expensive or difficult to implement (GNSS-RTK, IMU sensor, and Kalman Filter...). These positioning solutions are even mandatory for systems coupling LiDAR sensor (Moreno et al., 2020).

Several image-based positioning solutions exist, such as structure from motion (Rupnik et al., 2017) or visual-slam algorithms

(Taketomi et al., 2017), both are based on founding matching points between images. However, these methods suffer from a positioning drift, especially in vegetation contexts that make matching points calculations difficult. Either need external observation such as GNSS or Ground Control Point (GCP) to avoid such drift and ensure global positioning accuracy (Lhuillier, 2011). An alternative to avoid these drift issues in SLAM algorithms is loop detection, for instance by detecting and matching the scene layout (Baligh Jahromi et al., 2018).

In recent years, advances in deep learning object detection have led to the introduction of new methods based on semantic matching (Duan et al., 2020). For instance, (Wang and Zell, 2018) proposes to match only the interest points coming from the same type of object in each image in order to gain in robustness. Similarly, (Hu et al., 2019) apply Faster R-CNN convolution neural network model to add semantic information over feature point, allowing them to introduce a Bag-of-Words based similarity measure.

In this paper, we propose to show how an existing low-cost multispectral UAV-based camera (namely Parrot Sequoia) can be turned into a terrestrial mobile mapping system, with high precision positioning capability thanks to an automatic registration system using vine stocks detected by deep learning.

2. EQUIPMENT AND DATA ACQUISITION

2.1 Multispectral camera

Our acquisition was made with the Sequoia multispectral camera from Parrot, which has 4 mono-band sensors (Green, Red, Near-Infrared, and Red Edge), and an RGB camera (Table 1). It is also interesting to notice that the RGB sensor is provided with a rolling shutter, while the other sensors (monochromatic) use global shutters. The rolling shutter will cause significant distortions if the camera moves too quickly, for example in the case of vibration.

---
*Corresponding author

Figure 1: Comparison of a drone view (Kerkech et al., 2020) and a ground view of vineyards. One can notice the fruits under the plant on the terrestrial image.

The camera comes with a Sunshine Sensor which records the luminance received by the sun during the capture of each image to correct the differences in lighting between the images due to the passage of clouds or others and to have equalized reflectances. Such equalized reflectance is not used yet but would be very useful for applications such as disease detection.

The Sunshine Sensor also includes a GNSS and an Inertial Measurement Unit (IMU), providing a global location with metric precision. Because of its low accuracy, the IMU does not provide interesting information for data orientation and will not be used here.

Sensor	Size (px)	Bands (nm)	Shutter
RGB	4608 x 3456	-	Rolling
Green	1280 x 960	550 ± 40	Global
Red	1280 x 960	660 ± 40	Global
Red edge	1280 x 960	735 ± 10	Global
Near infrared	1280 x 960	790 ± 40	Global

Table 1: Multispectral sequoia camera (Parrot) sensors specifications.

To use this camera on a terrestrial platform, a 3D printed box was designed, allowing to mount the camera looking towards the vine with the Sunshine Sensor on top facing the sun (Figure 2).

2.2 Data acquisition

A first calibration dataset has been acquired on the calibration site of the HEIG-VD. This site is composed of 18 targets, distributed in 3D and measured with a total station. Images were taken from

Figure 2: Sensor mounted on a tractor using a 3D printed adapter.

7 viewpoints, located on two lines at two different distances from the calibration site. From each of the positions, 4 views have been shot by varying the orientation of the camera (Figure 3), resulting in 28 views (each composed of 1 RGB image and 4 mono-band images).
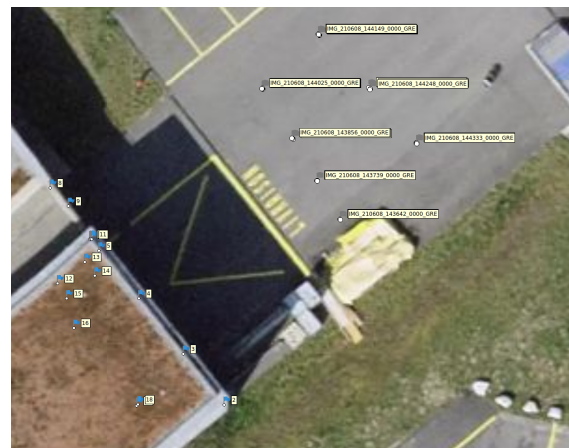
Figure 3: Calibration site at the HEIG-VD composed of 18 GCPs.

Experiments were carried out on a vineyard in the region of Yverdon-les-Bains, Switzerland, which was chosen as the study site because of its proximity, the availability and interest of the owner for the project, and the good condition of the vineyards (Figure 4). The vineyard has a surface of about 2.5 ha.

On this experimental site, several acquisitions have been made on different dates. In this paper, the following two acquisitions were used:

1) Reference dataset was acquired, together with several targets measured with a GNSS-RTK;

2) Target dataset was acquired at the same place, a few weeks later, without targets.

3. PROPOSED METHOD

The main idea of our method is to match low-accuracy data from a low-cost sensor with high-accuracy reference data previously acquired (for example in the leaf-off condition in winter). Such

Figure 4: Vineyard chosen as a case study (red box).

high accuracy dataset can come from a GNSS-RTK sensor, or by introducing manual GCPs.

Since major changes in luminosity and vegetation states do not allow the application of conventional methods (such as key-points extraction and matching), we propose to extract objects of interest (namely vine stock) and register such objects in 3D (Figure 5).

Due to the low accuracy of GNSS data and the bundle adjustment process, a non-rigid registration method is necessary (Figure 12-Top).
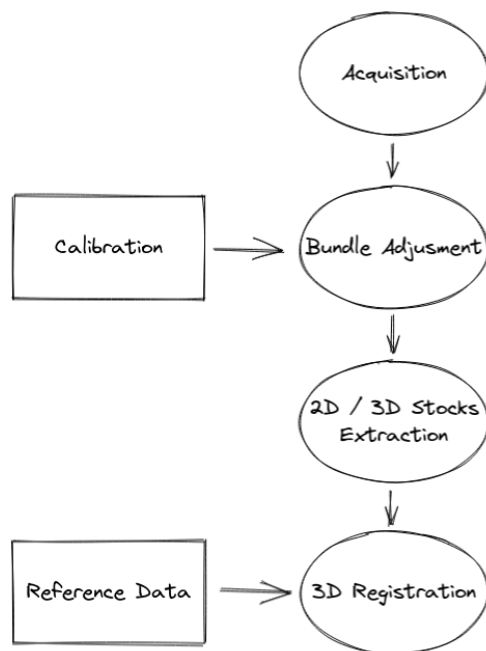
Figure 5: Overview of our non-rigid registration method.

3.1 Camera calibration

Before starting the acquisitions, the camera was first calibrated on a known site previously presented. This calibration allowed us to determine the parameters of each sensor of the Parrot Sequoia camera (focal length, principal point, radial (Ki), and tangential (Pi) distortion parameters), as well as the lever arms and boresight matrix of the sensors in respect to each other. All these computations were performed with the commercial software Metashape Agisoft.

The resulting residuals are shown in Table 2. The image residuals are a bit high (about 2 pixels), while one could expect subpixel values, this can be explained in part by the presence of a slight blur in the images and the weak quality of the monochromatic images.

X (cm)	Y (cm)	Z (cm)	XYZ (cm)	Image (pixel)
1,84	1,42	0,94	2,51	1,72

Table 2: Terrestrial and images residuals on GCPs after camera calibration.

3.2 Bundle adjustment

Once the acquisitions were made, the data has been separated by vine row. Then, each line was computed separately: a bundle adjustment was performed by specifying the camera's parameters previously determined during the calibration (intrinsic calibration, boresight...). Fixing the parameters helps to robustify the results and reduce errors in these difficult vegetation conditions. At this step, a sparse cloud was obtained. This sparse cloud is composed of matching keypoint from the bundle adjustment, roughly georeferenced from GNSS data for the target dataset (with an accuracy of the order of one meter). The reference dataset has been treated similarly, by adding GCPs in the bundle adjustment, in order to guarantee centimetric georeferencing.

3.3 2D/3D stocks extraction

The next step is to extract the stocks of vines by a deep learning method. This step is done in two stages (Figure 6):

1) Stocks are extracted in 2D on each individual image;

2) 2D objects from the images are reprojected in 3D.

Objects extraction was done with the Detectron2 library (Wu et al., 2019). Nearly 500 examples of vine stocks were manually labeled on RGB images from the Sequoia camera, collected from different vineyard plots in Switzerland. Then, a vine detection algorithm has been trained from a pre-trained model. The model used (Mask R-CNN), is based on ResNet-50 with Feature Pyramid Network (FPN).

An average precision of 76% was obtained on the training dataset. Such a model made it possible to extract the stocks from the images of our two datasets presented above (Figure 7).

To project previously detected 2D objects in 3D, we applied masks from the object detection on each image. Then, masked images are used to perform a 3D reconstruction by dense correlation in Metashape Agisoft (Figure 6). Thus, only points from vine stock are reconstructed in 3D: which allowed us to have a 3D reconstruction of all vine stocks in the current vine row (Figure 8-Top) and Figure 12-Top). One can notice in this Figure that one stock is well adjusted (red circle), but that the others are less and less so, which is due to the poor quality of the georeferencing of the target dataset.

For the following, the vine stocks will have to be individualized, in order to be able to adjust them separately. Thus, after the 3D extraction of stocks, we applied the DBSCAN method to segment the 3D point cloud into a set of points per stock. The DBSCAN algorithm groups the points from near to near in the same cluster if they are closer than a given distance (epsilon).
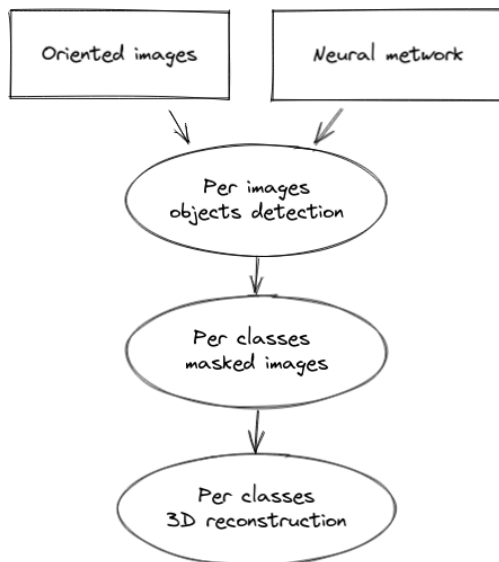
Figure 6: Details of 2D/3D object extraction method.

Figure 7: Stocks extraction on one individual image. One can notice that tests have also been conducted on the extraction of grape bunches.

This method allows us to easily segment the stocks by choosing a value of epsilon a little smaller than the average distance between the objects (here we chose eps = 30 cm). Lastly, stocks were numbered in ascending order along the vine line (on color per stock in Figure 8-Center).

3.4 3D registration

A non-rigid 3D registration is done in this step (Figure 10): the idea is to estimate many small rigid registrations, by founding matching vine stocks from the reference and target dataset, and computing 3D transformation (rotation and translation) for each of them.

Even if the GNSS system of the camera provide meter accuracy positioning, since the orientation from the IMU is not used, and the GNSS points are aligned along the vine line, the reconstructed 3D point cloud from the target dataset may have offset greater than 10 meters with the reference data (Figure 12-Top). Indeed, the reconstructed points can be placed in any direction around the acquisition axis, so in the worst case, one cloud could be three or four meters to the left of the axis and the other at the same distance but to the right.

Thus, to solve the problems of significant offset between the reference and the target acquisitions (greater than the distance between two vine stocks), vine stocks from the target data are
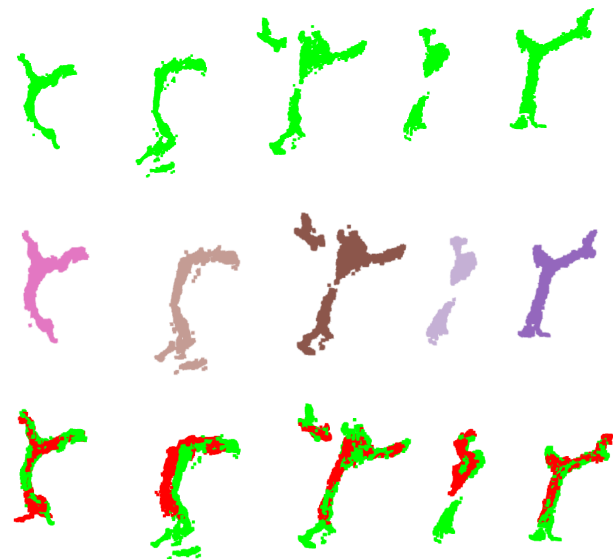
Figure 8: (Top) 3D reconstructed stocks, (Center) individualized objects (one color per stock), (Bottom) and the result of the 3D registration (greeen : reference, and red : target datasets).

grouped by 3, then each target group is aligned successively with the 3 closest groups of the reference data set (current one, previous and next ones in the row), see Figure 9.
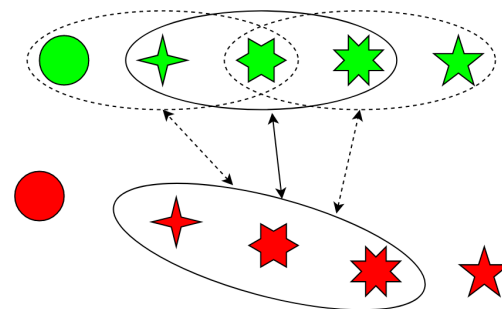
Figure 9: Neighborhood stock matching candidate details: reference data in green, and target data in red, each objects represents a vine stock.

An approximate transformation is founded by computing differences between reference and target objects' barycenters. Then, the fine matching process is based on the Iterative Closest Point algorithm (Chetverikov et al., 2002, Gressin et al., 2013). The best-fitting group from the reference data set is selected on the criterion of the highest number of matching points.

Computation of DBSCAN clustering and ICP registration have been computed with the open3D library (Zhou et al., 2018).

We thus obtain a list of correspondence between the newly acquired vine stocks and those of the reference data, and for each matching the associated 3D rigid transformation. By applying those 3D transformations to each vine stock of the target dataset, we are able to properly align both datasets.

4. RESULTS

As indicated in Section 2.2, the experiments were carried out on two datasets acquired at different dates, a reference, and a target dataset. At first, the data were processed by row of vines.
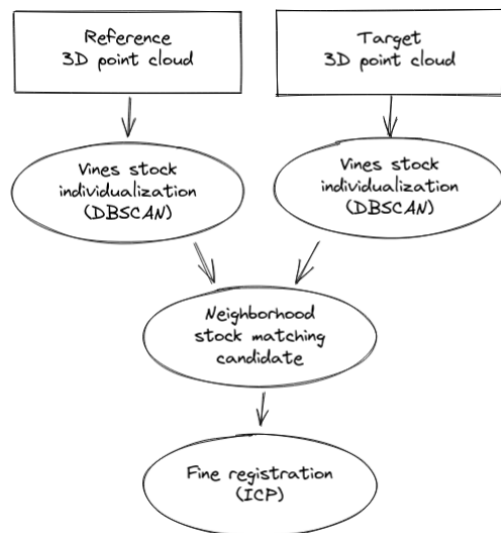
Figure 10: Details of the 3D non rigid registration method.

The data from the same row of vines were selected from both the reference and the target datasets, then processed separately to reconstruct each vine in 3D (Figure 12-Top): 201957 points and 16 clusters have been extracted from the target dataset, respectively 183401 points and 18 clusters from the reference dataset (Figure 12-Center). This difference in the number of stocks detected has two explanations: one of the two acquisitions covers a slightly larger area, and with the leaves some vines can be grouped by two, without impacting the results presented below. Thus, a fine and precise 3D reconstruction of the vines has been produced.

After the matching step, 16 individual rigid transformations have been computed (one for each vine stock extracted in the target dataset), all with inliers RMSE smaller than 1 cm (Figure 11). Details of vine stocks after registration can be seen in Figure 8-Bottom.
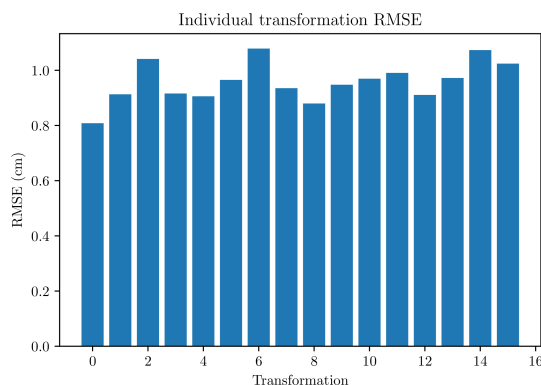
Figure 11: Individuals per vine stocks 3D registration RMSE.

The final result of the alignment of the target dataset on the reference data is shown in Figure 12-Bottom. On this dataset, we observed residuals after the alignment of less than 10 cm, thus opening up the possibility of mapping at the level of each vine stock.

In our current dataset, there were no unmatched stocks in each dataset. This is a limitation that will have to be taken into account in our further work.

5. CONCLUSION AND PERSPECTIVES

In this paper, we have proposed a method to register newly acquired images from a low-cost multispectral camera on an existing reference dataset, by automatically extracting the vine stocks and registering those objects.

The proposed method has been successfully tested on a small real data set acquired on a vineyard. The results obtained on this dataset showed a clear improvement of the georeferencing of the dataset, allowing to switch from a metric accuracy to an accuracy better than 10 cm, in a fully automatic manner.

Such precision makes it possible to consider the precise mapping of vineyards, in particular at the scale of the vine stock, allowing to answer questions such as: which vine stock is sick? or which vine stock produces which quantity of grapes?

In the future, we would like to evaluate the robustness of our method on a larger dataset: what will be the problems of scaling up? what is the sensitivity of our method to the change of illumination?

Although the method was developed for vineyards, it could be generalized to different types of objects, such as fruit trees, salad plants in agriculture, or urban materials for various applications, such as introducing global positioning on the visual-slam algorithm.

REFERENCES

Baligh Jahromi, A., Sohn, G., Jung, J., Shahbazi, M. and Kang, J., 2018. Layout slam with model based loop closure for 3d indoor corridor reconstruction. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences IV-2, pp. 25–32.

Chetverikov, D., Svirko, D., Stepanov, D. and Krsek, P., 2002. The trimmed iterative closest point algorithm. In: 2002 International Conference on Pattern Recognition, Vol. 3, pp. 545–548 vol.3.

Crookston, R. K., 2006. A top 10 list of developments and issues impacting crop management and ecology during the past 50 years. Crop Science 46(5), pp. 2253–2262.

Duan, C., Junginger, S., Huang, J., Jin, K. and Thurow, K., 2020. Deep Learning for Visual SLAM in Transportation Robotics: A review. Transportation Safety and Environment 1(3), pp. 177–184.

Gressin, A., Mallet, C., Demantké, J. and David, N., 2013. Towards 3d lidar point cloud registration improvement using optimal neighborhood knowledge. ISPRS Journal of Photogrammetry and Remote Sensing 79, pp. 240–251.

Hu, M., Li, S., Wu, J., Guo, J., Li, H. and Kang, X., 2019. Loop closure detection for visual slam fusing semantic information. In: 2019 Chinese Control Conference (CCC), pp. 4136–4141.

Before registration

After stock extraction

After registration
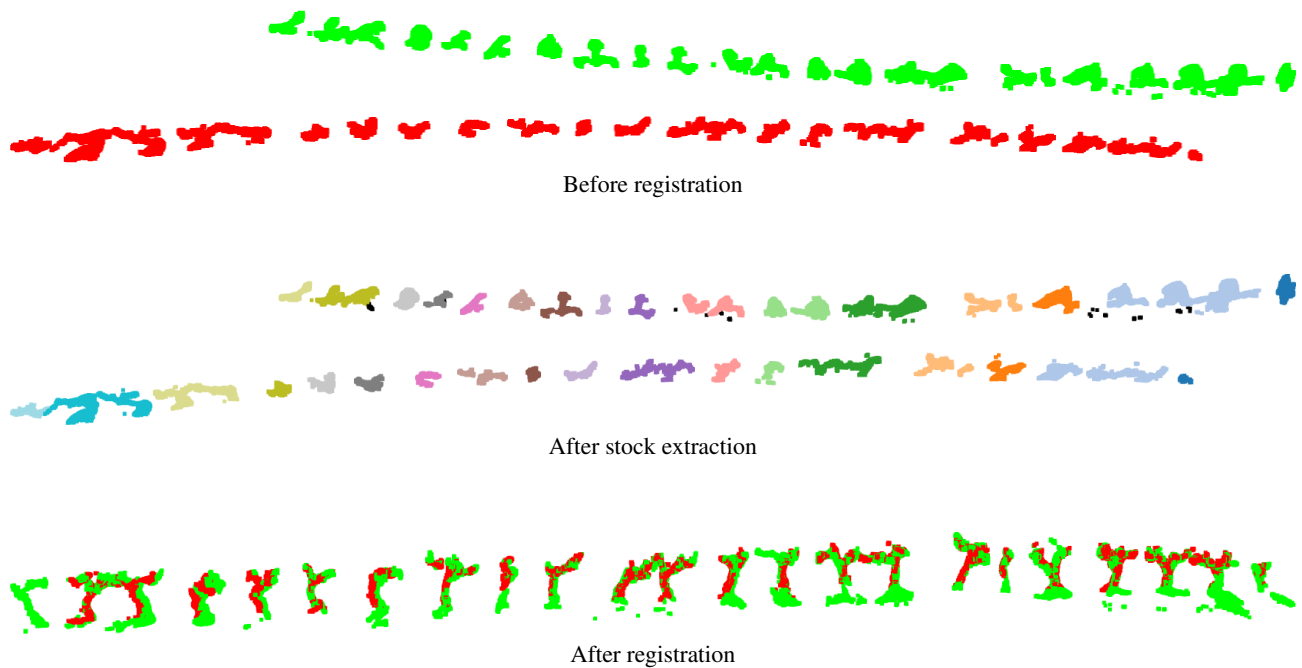
Figure 12: Result of the 3D vineyard registration (top) before registration, (center) after 3D stocks extraction (one color per detected stock), and (bottom) after registration : the reference dataset (green) and the target one (red).

Kerkech, M., Hafiane, A. and Canals, R., 2020. Vine disease detection in uav multispectral images using optimized image registration and deep learning segmentation approach. Computers and Electronics in Agriculture 174, pp. 105446.

Khaliq, A., Comba, L., Biglia, A., Ricauda Aimonino, D., Chiaberge, M. and Gay, P., 2019. Comparison of satellite and uav-based multispectral imagery for vineyard variability assessment. Remote Sensing.

Lhuillier, M., 2011. Fusion of gps and structure-from-motion using constrained bundle adjustments. In: CVPR 2011, pp. 3025–3032.

Maes, W. H. and Steppe, K., 2019. Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture. Trends in Plant Science 24(2), pp. 152–164.

Martos, V., Ahmad, A., Cartujo, P. and Ordoñez, J., 2021. Ensuring agricultural sustainability through remote sensing in the era of agriculture 5.0. Applied Sciences.

Moreno, H., Valero, C., Bengochea-Guevara, J. M., Ribeiro, A., Garrido-Izard, M. and Andújar, D., 2020. On-ground vineyard reconstruction using a lidar-based automated system. Sensors.

Rançon, F., Bombrun, L., Keresztes, B. and Germain, C., 2019. Comparison of sift encoded and deep learning features for the classification and detection of esca disease in bordeaux vineyards. Remote Sensing.

Rupnik, E., Daakir, M. and Pierrot Deseilligny, M., 2017. Micmac – a free, open-source solution for photogrammetry. Open Geospatial Data, Software and Standards 2(1), pp. 14.

Taketomi, T., Uchiyama, H. and Ikeda, S., 2017. Visual slam algorithms: a survey from 2010 to 2016. IPSJ Transactions on Computer Vision and Applications 9(1), pp. 16.

Wang, Y. and Zell, A., 2018. Improving feature-based visual slam by semantics. In: 2018 IEEE International Conference on Image Processing, Applications and Systems (IPAS), pp. 7–12.

Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y. and Girshick, R., 2019. Detectron2. https://github.com/facebookresearch/detectron2.

Zhou, Q.-Y., Park, J. and Koltun, V., 2018. Open3d: A modern library for 3d data processing. arXiv preprint arXiv:1801.09847.