A Multisensor Platform for Comprehensive Detection of Crop Status: Results from two Case Studies

Stefan Rilling
Fraunhofer IAIS
Sankt Augustin, Germany
stefan.rilling@iais.fraunhofer.de

Michael Nielsen
Danish Technological Institute
Taastrup, Denmark
mlnn@teknologisk.dk

Annalisa Milella
CNR ISSIA
Bari, Italy
milella@ba.issia.cnr.it

Christian Jestel
Fraunhofer IAIS
Sankt Augustin, Germany
christian.jestel@iais.fraunhofer.de

Peter Fröhlich
AgriCircle AG
Rapperswil, Switzerland
peter@agricircle.com

Giulio Reina
University of Salento
Lecce, Italy
giulio.reina@unisalento.it

Abstract

The measurement of the growth state and health status of single plants or even single parts of the plants within a crop to conduct precision farming actions is a difficult task. We address this challenge by adopting a multi-sensor suite, which can be used on several sensor-platforms. Based on experimental field studies in relevant agricultural environments, we show how the acquired hyperspectral, LIDAR, stereo and thermal image data can be processed and classified to get a comprehensive understanding of the agricultural acreage.

1. Introduction

The defining characteristic of Precision Farming is the ability to perform agricultural operations on a narrow scale. In the case of agriculture, the ability to accurately locate the crops or leaves with problems and to accurately apply a local remedy without wasting resources or contaminating the environment is one of the key problems [4].

This work presents a unifying framework allowing incorporation of many different types of sensor data, methods for creating 3D maps and maximizing map accuracy to facilitate operations on a narrow scale with a smaller environment footprint, methods for combining this data to make relevant information easily visible to the farmer, and methods for incorporating real-time sensor data into historical data both to increase precision during applications and to provide fast automated safety responses.

978-1-5386-2939-0/17/$31.00

2. Multi-sensor System

2.1. Hyperspectral Imaging

Hyperspectral data was recorded with VIS-NIR Specim V10 spectrograph on a Basler ACA1920-155um 400-1000nm. Hyperspectral cameras are line-scan systems, therefore the spatial dimension is collected through vehicle movement. The VIS-NIR setup was 1920 pixel lines with 2nm digital resolution at 128Hz, optically limited to 8nm usable bands, and outputs 16-bit images, effective dynamic range is 72db. Calibration was done using fluorescent light with known peaks for y-axis alignment, and spectrolon white reference for sensitivity normalisation.

Figure 1 shows an image of fluorescent light, which contains horizontal lines where the peaks mainly from mercury are located. The Halcon operator *lines_gauss* was used to extract the location, and a linear regression of the average line rows and the mercury peaks was done to compute the first order linear expression that converts Y position into the corresponding wavelength.

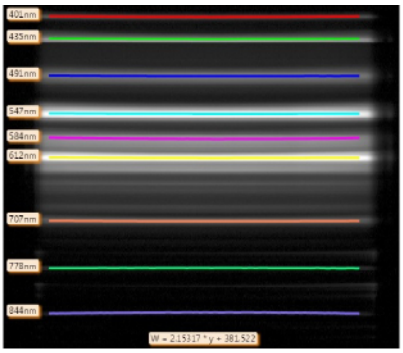


Figure 1. Calibration of the hyperspectral bands from flourescent light.

2.2. 3-D LIDAR Scanner

The LIDAR system in use is a *SICK LMS-100* laser scanner. The device scans a 2-D plane within an operational range of 0.5m to 20m with an Infrared Laser (905nm). The scanner is mounted on a rotating platform; a 3-D point cloud is acquired by a rotation of the 2-D scanning plane. Alternatively, the rotating platform can be locked; a 3-D point cloud is then acquired by moving the system around. The movement data is acquired using an *xsens MTi-G-710 GNSS/INS*. Both GPS data (lat/lon) and linear / angular velocities and accelerations are measured. The hardware components are driven by a *Raspberry Pi 3* single-board computer, which serves for controlling the scanning system and the data recording. The output of the system is the measured point cloud over time, and the position, orientation and acceleration values of the rotating platform and the sensing system.

2.3. Stereo and Thermal Imaging

The sensory set-up comprises: two Basler DART DaA1600-60uc for stereo vision with 40mm baseline and a thermal camera Micro Epsilon ThermalImager160. The stereo camera delivers images up to a resolution of 1600$\times$ 1200 at 60 Hz, which was reduced to 7 Hz in the experiment to reduce the computational burden. The ThermalImager can deliver images up to a resolution of 160$\times$120 at 120 Hz, reduced to 15 Hz. In order to register all sensor data with respect to a common reference frame, a calibration phase was carried out to estimate the intrinsic and extrinsic parameters of the sensors, using a set of images of a planar checkerboard. To make corners visible to the thermal camera, the checkerboard was placed under a heat lamp. Please, refer to [2] for a detailed description.

2.4. Sensor Platforms

Sensors were mounted on-board different platforms, including an unmanned ground vehicle (UGV), an aerial drone and a caterpillar vehicle (Figure 2). The robotic platforms were remotely controlled to operate in relevant agricultural settings while the sensory data were recorded for further processing off-line.
The UGV consists of the Husky A200 robotic platform, a non-holonomic four-wheel drive (4WD) skid-steer robot, featuring a payload of about 20 kg and a maximum travel velocity of 1 m/s. The vehicle control and the data acquisition system are based on ROS[1], a meta operating system able to provide services and hardware abstraction for robotics platforms and devices and a publish-subscribe messaging infrastructure designed to support the quick and easy construction of distributed computing systems.
The aerial platform is an Aibot X6 Hexacopter, featuring maximum payload of 2.0 kg and maximum speed of 40 km/h, and integrated GPS, RTK and inertial sensors. Finally, a commercial Niko caterpillar was employed.


Figure 2. The sensor platforms used. From left to right: Husky robot; Aibot X6 aerial drone; Niko caterpillar vehicle.

3. Agricultural Use Cases

One part of the field-tests was performed within the commercial vineyard Höcklistein in Switzerland. The grape variety planted there are *Räuschling*, *Sauvignon Blanc*, *Chardonnay*, *Pinot Noir* and *Merlot*. The diseases which could be found there were Downy mildew (Plasmopara viticola), powdery mildew (Erysiphe necator), acid rot, grey mould (Botrytis cinerea) and under-developed grapes resulting from water stress or too much water.
Another part of the field tests was performed at a private farm located in San Cassiano, Lecce, Italy, where grapes and olive trees are planted. Figure 3 shows an aerial view of the experimental farm. It includes a vineyard and an olive grove that are connected through a dirt road. The grapes had a bacterial disease called Lupa that is common in Salento. Furthermore, there was a lot of withering.

[1] http://www.ros.org/

4. Data Processing

4.1. Semantic Labeling of RGB Images

The ViDi Suite, a deep learning based industrial image analysis software system, relies on a collection of algorithms that allow training deep neural networks with limited amount of data. The implemented algorithms fall into three main categories: image filtering and preprocessing, supervised and unsupervised neural network training procedures, and smart and adaptive sampling procedures. The innovative sampling procedure developed at ViDi Systems were specially designed to deal with the biggest challenge imposed by industrial images: namely large images and limited amount of samples. This makes it fitting for our 4MP RGB images from the Salento campaign.

The Vidi segmentation tool was trained in three layers of semantic labelling: Grape clusters on the closest row, Withered Leaves, and Region of interests around each grape clusters into another layer that detected Lupa disease.

The tool was good at segmenting only the closest row, but few false positives appeared on the next row over. It would be possible to remove those using 3D information. Furthermore, there was sporadic small regions appearing on leaf, soil, trunk edges, which can be filtered by looking at the temporal tracking of the regions. For example, a filter can accept only regions that stay for three images in a row.

It is notable how the tool is able to understand the large variations on healthy grapes green, dark, shadows, and is still able to label Lupa on small parts of clusters, both green and dark.

More layers of semantic labelling can be added, such as the colour grading. The system could then track the ripening through detecting begin and end of the colour change. Furthermore, the grapes and disease ratio can be quantified, and we expect a performance increase by useing four channel images created from hyperspectral hypercubes (i.e. through PCA, ICA, four best feature selection by cross validation).

Figure 3. Experimental farm located in San Cassiano, Lecce, Italy. Aerial view taken from Google Earth (40°03′35.40″N, 18°20′50.98″E).

Figure 4. The same images showing results of the grape segmentation (left), and withered leaves (right).

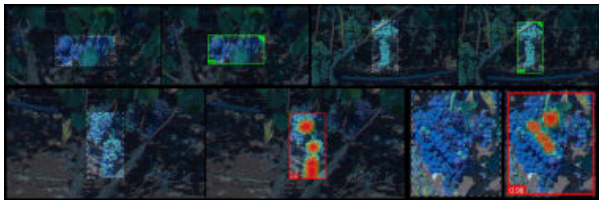


Figure 5. Dark and green healthy grapes (top) and Lupa infected grapes (bottom).

4.2. Semantic Labeling from Hyperspectral imaging

The semantic labeling was done under laboratory conditions with halogen light, because outdoor light is very complex. It was necessary to first validate that the sensor can see the relevant features. Multilayer perceptron neural networks with softmax based activation function are used for the classifications. Pixelbased followed by spatial filtering and interpretation. A linear neuron-based approach will be used for regression.

The classification procedure is as follows:

- Stitching the hyperspectral images into a hypercube.
- Whitebalance from white reference Bw = (B-Bl) / (W-Bl) (white balanced pixel Bw, pixel bands B, white reference W, black reference Bl
- Normalize each pixel P = (Bwn-mean(Bw)) / Std(Bw)
- Label areas of each class, including background class
- Feature selection through cross validation. Outputs a Score for the quality of classifier.
- Compile a new multichannel image from the selected bands only.
- Retrain with regularization parameters.

- Save list of selected bands, and trained classifier.

The classification was followed by morphology, connected components and minimum area filter to clean up noisy pixels.

First tests with hyperspectral classification in controlled conditions was done on cultivars: Chardonnay, Sauvignon Blanc, Pinot Noir, and Merlot.

Cross validation score was 0.98 within the green grapes, and 0.91 for all four. Naturally, the dark grapes perform lesser due to SNR. However, this is for single pixel accuracy. Table 1 shows for each cultivar the percentage of pixels classified as which cultivar. Applying regional percentage grouping, a probability for each class can be computed.

Table 1. Percentage of area of classified pixels into each cultivar.

Cultivar	S.B.	Chardonnay	Pinot Noir	Merlot
S.B.	86%	14%	0	0
Chard	6%	94%	0	0
P.N.	19%	5%	60%	16%
Merlot	25%	6%	18%	51%

Sauvignon blanc had most diseases which made it the best candidate to test the classification of the various diseases. One classifier dedicated to detecting which side of the clusters faced the sun, assuming the ripeness would be different on the sunny side. The score was low (0.75). A second classifier dedicated to classifying healthy, acid rot and botrytis, score 0.98. It worked well also on Chardonnay, but not on the red grapes. Finally, a classifier that detected magnesium stress on leaves, score 0.89, which is lower than usual due to the fact that during annotation it is difficult not to include healthy pixels in the sick training data.

4.2.1 Transfer to outdoor usage

As the spectrolon white reference was imaged with different orientations to capture the different mixtures of direct and ambient light , the hypothesis was that the trained classifiers could be augmented to fit those different light mixtures as suggested in [5]. However, as also concluded in [1], 900-1000nm is too noisy and the noise magnitude overrides the signal fully when normalizing the spectra. The most significant wavelength for magnesium stress was 946nm, so the hypotheses that classifiers could be transferred from the lab condition to outdoors was quickly rejected.

Instead, a simpler case to annotate grape, leaf, and trunks was attempted using 400-900nm in outdoor data. A classifier that annotates leaves, grapes and trunks was applied (c.f. figure 6). The selected wavebands in prioritized order were: 756nm, 493nm, 681nm, 735nm, 627nm, and 825nm. The resulting score was 0.97.

Figure 7 shows results where a classifier was trained to also detect acid rot. Selected features in order were: 868nm, 412nm, 717nm, 618nm, 545nm, 556nm, 681nm, 435nm, and 521nm. Score 0.90.

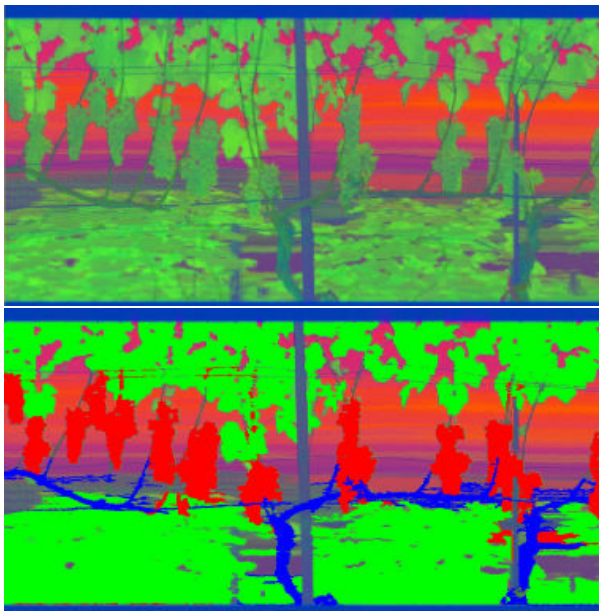


Figure 6. (top) the hyperspectral hypercube visualized as a color image using 3 first channels of PCA. (bottom) each pixel classified as green: leaf, red: grape, blue: trunk.

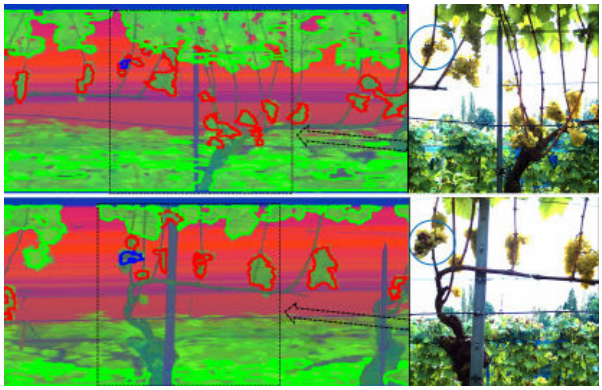


Figure 7. Detection of acid rot in the chardonnay vineyard. Small occurances like 1-2 grapes were not detected. Larger clusters is 4-5 infected grapes were detected, like those seen in the left side of the RGB images.

The currect obstacle in using this in practice is that whenever the light changes, such as a blue sky changing to overcast or rain, the NIR light disappears, so even with white referencing, the dynamic range is not large enough to compensate. See figure 8.

4.3. Stereo-Thermal Integration

This part of the research aims to produce a multi-modal map of the environment by combining stereo and thermal data. In detail, from each incoming stereo pair, first, a dense

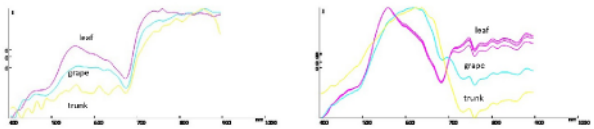
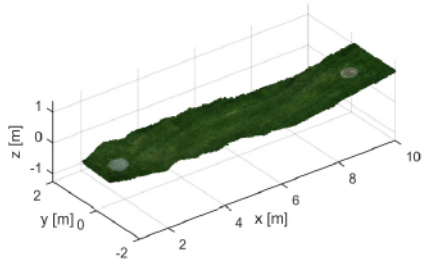
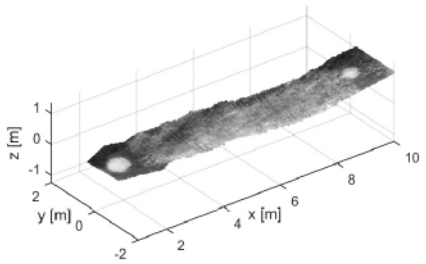


Figure 8. Problem with changing weather condition and surface normal dependency of white balance. Even when rebalancing with white reference, the spectra of leaf, grape and trunk is affected greatly by the change from blue sky (left) to overcast near rainy (right).

3D point cloud is reconstructed. Then, 3D points are projected over the thermal image, and their thermal values are extracted and assigned to the corresponding 3D points in the stereo map. This originates what we refer to as a 3D thermal cloud. Finally, based on vehicle motion information provided by visual odometry, subsequent stereo-thermal point clouds are stitched together, resulting in a multi-modal representation of the environment, which includes geometric, colour, and thermal information. As an example, Figure 9 shows the stereo map (a) and the corresponding thermal layer (b) for a test on grass. Two manhole covers are clearly visible along the path at about 3 and 9 m, due to their different appearance and reflectance properties.



(a)



(b)

Figure 9. Stereo-Thermal combination: Visual (a) and thermal layer (b) for a test on grass. Two manhole covers can be clearly seen, owing to their different appearance and reflectance.

4.4. LIDAR Data Processing and Classification

The acquired point cloud data has to be processed in order to combine it with the data recorded by the other parts of the S3-CAV sensor suite, such as the hyperspectral sensor system.

4.4.1 Point Cloud Generation and Filtering

The recorded point cloud data is transformed from the local scanner coordinate frame to the global coordinate system frame the known movement path of the sensor platform. This movement path is determined using a generalized extended Kalman Filter [3] which fuses the recorded GPS and acceleration data. A further extraction of the relevant row of grapevines is done geometrically by moving an extraction volume (i.e. a sphere or cylinder) along the movement path and selecting all the points inside the filter volume. The point cloud data is represented in a kd-tree to accelerate this process.

4.4.2 Classification of Trunks and Leaves

The extracted row of grapevines is then further processed to classify foliage and trunks using geometrical information in combination with the remission values measured by the LIDAR scanner. The threshold values for the classification were determined manually by comparing the resulting point cloud to real-world images. This classified point cloud is then stored in a 3D regular grid with the extends of the cloud's axis-aligned bounding box and a defined resolution. Each box-shaped grid cell is then classified as "foliage" or "trunk" by the number of corresponding points within the cell. Figure 10 shows the result of this process.

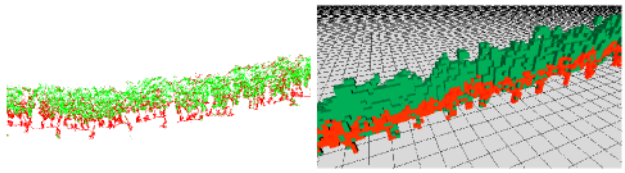


Figure 10. The left image shows the filtered and classified point cloud data. The foliage is coloured in green, the trunks are coloured red. The right image shows the decomposition into small boxes.

This 3D box representation is used for the estimation of the foliage volume of a plant, which is an important measurement for the application rates of plant protection products.

5. Conclusions and Future Work

We demonstrated a sensor suite for semantic labelling of crop status, in situ using deep learning and hyperspectral imaging. The system can also map results and measure crop growth using a georeferenced 3D representation.

The major obstacles in automatically creating robust maps for a Farm Management Information System (FMIS) are changing lights between calibrations and the dynamic

range of cameras. Future work will address this through imaging a green reference for auto exposure, and combining the hyperspectral method with deep learning. Furthermore, the data will be compressed into GeoTiffs and ISOXML path data with links to the full 3D representation.

The translation of the resulting data to ISOBUS-based application maps which can be exported directly to machinery should bridge the technological gaps in farmer adoption hindrances.

The financial support of the FP7 ERA-NET ICT-AGRI 2 through the grant Simultaneous Safety and Surveying for Collaborative Agricultural Vehicles (Id. 29839) (S3-CAV) is gratefully acknowledged.

References

[1] H. Aasen, J. Bendig, A. Bolten, S. Bennertz, M. Willkomm, and G. Bareth. Introduction and preliminary results of a calibration for full-frame hyperspectral cameras to monitor agricultural crops with UAVs. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-7, 2014 ISPRS Technical Commission VII Symposium*, XL(October):8, 2014.

[2] A. Milella, M. Nielsen, and G. Reina. Sensing in the visible spectrum and beyond for terrain estimation in precision agriculture. In *Proceedings of 11th European Conference on Precision Agriculture (ECPA)*, Edinburgh, UK, 2017.

[3] T. Moore and D. Stouch. A generalized extended kalman filter implementation for the robot operating system. In *Proceedings of the 13th International Conference on Intelligent Autonomous Systems (IAS-13)*. Springer, July 2014.

[4] G. Reina, A. Milella, R. Rouveure, M. Nielsen, R. Worst, and M. R. Blas. Ambient awareness for agricultural robotic vehicles. *Biosystems Engineering*, 146:114 – 132, 2016. Special Issue: Advances in Robotic Agriculture for Crops.

[5] L. Windrim, R. Ramakrishnan, A. Melkumyan, and R. J. Murphy. Hyperspectral CNN classification with limited training samples. *CoRR*, abs/1611.09007, 2016.