2023 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor) | 979-8-3503-1272-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/MetroAgriFor58484.2023.10424392

First Step Towards Embedded Vision System for Pruning Wood Estimation

Bernardo Lanza
DIMI, University of Brescia
Brescia, Italy
0009-0005-3561-754X
bernardo.lanza@unibs.it

Cristina Nuzzi
DIMI, University of Brescia
Brescia, Italy
0000-0001-5530-6136
cristina.nuzzi@unibs.it

Davide Botturi
DIMI, University of Brescia
Brescia, Italy
0000-0003-1831-5994
davide.botturi@unibs.it

Simone Pasinetti
DIMI, University of Brescia
Brescia, Italy
0000-0002-5098-6395
simone.pasinetti@unibs.it

***Abstract*—This paper focuses on the development and evaluation of a portable vision-based acquisition device for vineyards, equipped with a GPU-accelerated processing unit. The device is designed to perform in-field image acquisitions with high-resolution and dense information. It includes three vision systems: the Intel® RealSense™ depth camera D435i, the Intel® RealSense™ tracking camera T265, and a Basler RGB DART camera. The device is powered by an Nvidia Jetson Nano processing board for both simultaneous data acquisition and real-time processing. The paper presents two specific tasks for which the acquisition device can be useful: wood volume estimation and early bud counting. Acquisition campaigns were conducted in a commercial vineyard in Italy, capturing images of vine shoots and buds using the prototype device. The wood volume estimation software is based on image processing techniques, achieving an RMSE of $2.1\ cm^3$ and a mean deviation of $1.8\ cm^3$. The buds detection task is obtained by fine-tuning the YOLOv8 model on a purposely acquired custom dataset, achieving a promising F1-Score of 0.79.**

***Index Terms*—Measurement science, vineyard monitoring, vision systems, pruning wood estimation, bud detection, precision farming, deep learning, image processing, embedded systems.**

I. Introduction

Nowadays, grapevine farming and wine production are focusing on improving yield quality [1]. This trend started in the late ‘80s and involves highly trained and specialized employees like agronomists and enologists. Continuous studies and research are being conducted to scientifically enhance wine quality through innovative farming methods [2] and advancements in the wine fermentation process [3]. The pursuit of improved product quality necessitates also a more precise management of individual grapevines [4]. For instance, in the production of Amarone wine, vines are planted closer together to maximize their number while minimizing the grape-to-vine ratio [5]. Numerous techniques are being explored to enhance the management of individual vines [4], thereby driving the adoption of new technologies. On the other hand, we are facing the effects of climate change [6], combined with the need for a sustainable farming-oriented approach. Climate change has resulted in more frequent and intense meteorological events with significant impacts on micro and macro biological aspects [7].

In this context, sustainable vineyard management becomes crucial, requiring intelligent dosing of water, fertilizer, and pesticides to preserve the biological environment [8]. Producers are becoming involved in how technology can improve quality and contain the cost of an environmental-safe production. Yield and vegetative prediction are the main challenges in this endeavor, and they rely heavily on in-field data, which are difficult and time-consuming to acquire [9]. In this context, optical sensors provide the capacity to collect a substantial amount of data from the environment, facilitating the extraction of the predominant vineyard feature: color. Color not only characterizes the state of the plants but also discriminates among their various organs, making it a valuable source of information for detailed analysis. The ideal solution for agronomists is a device able to scan the field at a close distance, thus acquiring high-resolution images dense with information in which each part of the plant is big enough to actually detect and understand its health and growing status. Therefore, there is a rising need for autonomous ground vehicles (AGVs) in contrast with aerial vehicles capable of performing various tasks in the field alongside monitoring, such as disbudding, collection, and pruning, which heavily rely on imaging sensors for environmental mapping and localization [10]. Vision systems mounted on different vehicles offer reliable measurements during motion if appropriately designed and tuned. Images acquired through those systems have to be processed in order to obtain reliable information about the vineyard status. Frameworks such as $OpenCV$ have been introduced to mathematically interpret these light-derived data types, extracting geometric and chromatic characteristics. In recent times, machine learning techniques have emerged, enabling the conversion of these features into qualitative information that is well-suited for agronomic assessment and decision-making support.

Ongoing research endeavors are focused on benchmarking algorithms for detecting the plant’s hydric state and leaf diseases [11]. The Televitis research group based in La Rioja, Spain, is actively testing various imaging approaches in vineyards [12], employing diverse techniques to estimate pruning mass [13]. Moreover, the Eden library group is developing deep learning datasets to detect different types of vine diseases [14] and is also working on a vision-based device for infield image acquisition [15]. However, the research community is not focusing enough on the problem of in-field acquisitions

979-8-3503-1272-0/23/$31.00 ©2023 IEEE

and dealing with unstructured environments, while this is a problem typically addressed by commercial companies.

With this work, we developed a first prototype of a portable vision-based acquisition device equipped with a GPU-accelerated processing unit. The device can be used for in-field reliable image acquisitions, also mapping the saved images in the vineyard thanks to odometry data acquired jointly. Moreover, we envisioned a couple of tasks for which this device can be extremely useful for agronomists: (i) the estimation of wood volume (vine branches and shoots), and (ii) early bud counting. Wood volume estimation can be useful to analyze the actual growth of plants during the previous year, enabling tailored fertilization practices [16] and helping to evaluate if pruning strategies are effective, or even detect diseases affecting the lignification of the vine (e.g. Flavescence dorèe [17]). Early bud detection and tracking contribute to generating a vigor map of the vineyard, representing the growth of upcoming vine shoots that will produce leaves, flowers, and grapes, also giving an estimation of the future yield of the whole vineyard way before the production of fruits, which is a practice sometimes adopted in viticulture.

II. Materials

A. Acquisition device

In this paper we describe and validate our optical acquisition device shown in Fig. 1. A total of 3 vision systems were mounted and fixed on a professional tripod using custom-made 3D-printed supports. The cameras adopted are (A) an Intel® RealSense™ depth camera D435i, (B) an Intel® RealSense™ tracking camera T265, (C) a Basler RGB DART camera (model daA2500-14uc), equipped with C-mount optics of focal length 8 mm, and iris aperture set to F 1.4 to obtain a depth of field in sharp focus on the shoots and out of focus on the background. To ensure a fast acquisition and image-saving frame rate, an NVIDIA Jetson Nano processing board was used as the acquisition device (device D in Fig. 1). With its multi-processing libraries and dedicated video card, the Jetson Nano enabled the simultaneous processing of multiple camera streams. Additionally, this GPU-based device offers the advantage of low power consumption and can be powered by an external battery (e.g. power banks or solar panels). It is worth noting that the depth camera D435i datasheet [18] claims that the depth resolution is less than 2% at a distance of 2 m and, more generally, the depth accuracy is between 2.5 mm to 5 mm at 1 m distance from the object, showing accuracy drifts that increase with distance [19]. However, in comparison to other RGB-D devices available in the market, the Intel® RealSense™ D435i stands out as not only cost-effective but also highly robust for outdoor measurements. Additionally, it boasts low power consumption, making it an ideal choice for integration into mobile embedded platforms. Furthermore, during our data acquisition process, we maintained controlled background conditions while intentionally allowing uncontrolled natural sunlight to illuminate the vine shoots. This approach ensured that we captured real-world variations in lighting conditions, thus enhancing the robustness and authenticity of our experimental setup.

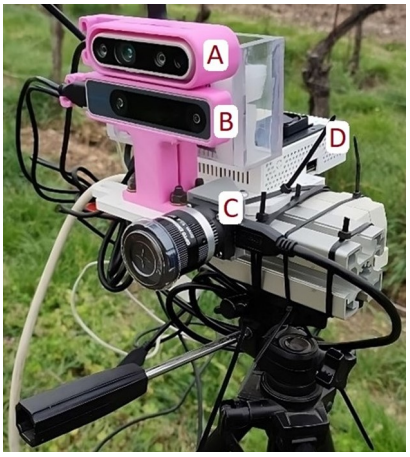

Fig. 1: Image of the proposed acquisition set-up. (A) depth camera D435i, (B) tracking camera T265, (C) Basler RGB camera with optics, (D) Nvidia Jetson Nano.

Although the proposed acquisition device is compact and easy to use, not every camera was employed at the same time.

To perform vine shoots volume estimation measurements only the depth camera D435i was used, which captures depth information at a resolution of 1280x720 pixels, together with its RGB sensor, which provides high-resolution color images at 1920x1080 pixels. Additionally, we utilized the tracking camera T265 to precisely localize in space each acquisition. However, the bud detection task requires color images of higher quality due to the low dimension of the buds compared to the background noise. Therefore, to capture a multitude of bud images the Basler camera was employed instead of the D435i depth camera. In addition, several images were also taken using a smartphone (RedMi Note 11 Pro, camera with sensor size $1/1.52$", resolution $12,000 \times 9,000$ px, iris aperture ranging from F 1.9 to F 2.4), some taken when it was mounted on the acquisition device, and some close-ups taken manually. This strategic combination allowed us to capture diverse sets of images, each exhibiting distinct chromatic and optical characteristics. The acquisitions were always coupled with the odometry data obtained by the tracking camera T265. By using the tracking camera, a CSV file containing the localization, orientation, and velocity information for each frame was also generated alongside the raw image data. Odometric data were acquired at a rate of $1,500$ FPS, allowing us to localize all the image frames and generate a vineyard map of the measurement locations.

B. Acquisition field and experimental campaigns

The Masi Agricola winery (N $4531'36.1596''$, E $1051'43.1496$") generously provided us access to one of their vineyards dedicated to the cultivation of Corvina grapes following the Guyot vine training system [20]. In the Valpolicella area (Verona – Italy) *Vitis vinifera cv. Corvina* is the main grape variety cultivated for the production of the

famous Amarone red wine. For both research purposes (shoots volume estimation and bud detection) the acquisitions were conducted in the same field and at the same time, leveraging slightly different protocols. Data collection took place in winter, before vine pruning, allowing for the presence of numerous vine shoots resulting from the previous spring and summer's vegetative phase. Data acquisition was conducted during daylight hours, encompassing various ambient lighting conditions.

Fig. 2: Image of the acquisition area in the field.

1) Shoots acquisition campaign: To extrapolate meaningful pixels from the vineyard surroundings without using AI segmentation algorithms, a background-free set-up was designed and implemented on the field by fixing white sheets behind the target vine as shown in the left part of Fig. 2. This experimental set-up enables easy segmentation of vine images using computer vision algorithms, resulting in the extraction of relevant pixels.

2) Buds acquisition campaign: In this case, it was not required to cover the background thanks to the settings of the Basler RGB camera that produced a short depth of field. Since a high amount of images needed to be saved for AI training purposes, the idea was to acquire pictures while moving in a continuous fashion. Hence, instead of taking single images, a full video of the whole movement along the vineyard was recorded.

III. Proposed Methodology

A. Shoots volume estimation procedure

The procedure applies image processing algorithms to the RGB image to filter data from the corresponding depth image using standard imaging techniques [21]. This results in a point cloud (PC) further processed to obtain sub-cylinders (SM) that approximate its volume. The final volume of the branch sample is obtained as the sum of the SM's volumes. Refer to Fig. 3 and Fig. 4 for the complete procedure.

1) 2D mask generation: The original RGB image of the sample is convered to Grayscale. To enhance the image quality, we apply a Histogram Stretch algorithm. A thresholding operation generates a binary mask M, which is further refined

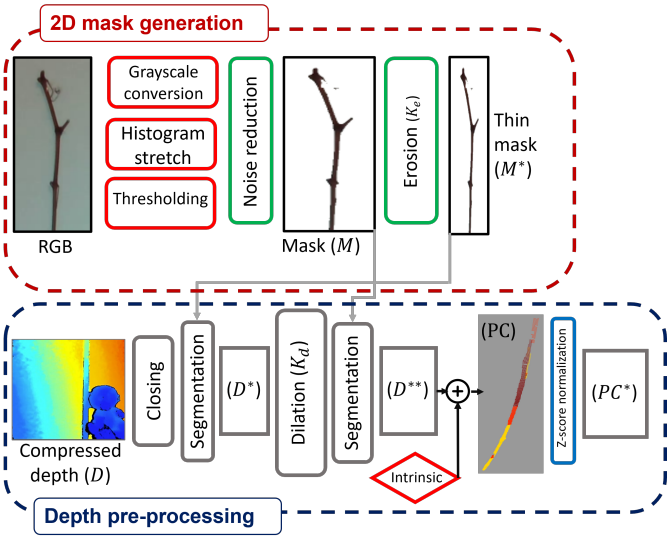

Fig. 3: RGB-D image processing pipeline.

through an erosion morphological operation with kernel K_e. This operation produces another thinner mask M^*.

2) Depth pre-processing: The compressed depth image D is refined by applying a Closing operation to reduce data noise. Then, the refined mask M^* is superimposed on the depth image, thus obtaining a depth image D^* without border effects that may happen due to the branch shape and light illumination noise (e.g., multipath error, occlusions). Since the resulting D^* may have holes due to reflective surfaces or occlusions, and the filtering step applied using M^* greatly reduced its border, we apply a Dilation procedure with kernel $K_d > K_e$. This fills out the holes and enlarges the perimeter of the sample. However, even if this procedure fills out the internal holes, it may distort the original shape of the sample. Therefore, we superimpose mask M to filter out wrong points and retain the original shape of the sample. The resulting pre-processed depth is called D^{**}.

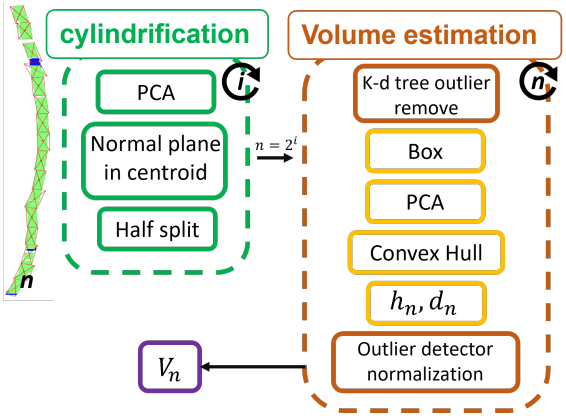

Fig. 4: Volume estimation procedure

3) Point Cloud generation: With the pre-processed depth data D^{**} in hand, we generate the corresponding PC by applying the intrinsic data of the IR sensor of the RGB-D

camera obtained from camera calibration (image center c_x, c_y, focal length f_x, f_y, distortion parameters). To correct any scattered behavior that may be present due to image compression, we apply a Z-score normalization procedure to the resulting PC.

4) Cylindrification: To compute the sample's volume, the idea is to divide it in sub-cylinders (SM) to reduce the approximation error. Therefore, the division of the original branch is performed through an iterative process. First, a PCA analysis is conducted on the original PC to extract its principal components. These are used to identify the *cutting direction* which is the longest of the three (first principal component). The PC is cutted in half along this direction, obtaining two halves of the original PC. This procedure is repeated i times, where $i = \log_2 \frac{h_{max}}{h_{min}}$. The resulting i is approximated to the closest integer value. As a result, after the cylindrification procedure, we end up with $n = 2^i$ SM. Afterwards, the SMs are filtered using a *K-d tree* to remove outliers belonging to the background. To apply it, considering that the camera has a minimum resolution of $1mm$, in this work we defined the radius of the *K-d tree* searching area equal to $2mm$ and the number of points belonging to the area at least equal to 6, thus avoiding pointy shapes that are actually noise and not real data.

5) Volume estimation: For each SM, the algorithm computes the enclosing 3D box by using the min and max of each dimension. This results in the SM size along each axis (S_x, S_y, S_z). Due to the specifications of the cylindrification procedure (i.e., the definition of i and n), the highest of the three is the height h_n of the $n{-}th$ SM. Then, we project the 3D data of the SM on a plane obtained by applying a PCA on the points of the SM (ideally, this should be equal to projecting to plane XY). Therefore, we apply a Convex Hull to compute the area A_n of the SM and obtain the diameter as $d_n = \frac{A_n}{h_n}$. This method accurately approximates the mean diameter of branches with varying diameters. After obtaining h_n and d_n for each SM, we perform a statistical analysis to found incorrect data by checking the mean and standard deviation of both h and d. If an outlier value is found, it is replaced by h^* or d^* accordingly, computed as the mean of the preceding and following SM h or d. The total wood volume V_{tot} is then calculated using:

$$V_n = L_n \times \pi \times \frac{d_n^2}{4} \tag{1}$$

$$V_{tot} = \sum_{j=1}^{n} V_j \tag{2}$$

B. Bud detection

In order to accurately map the presence of new buds in vineyards, which serve as indicators of canopy architecture and vine health, we conducted fine-tuning of the YOLOv8 image detection model [22]. We have selected the YOLOv8n nano model, which is the lightest variant among the available models, to reduce computational power. This decision was made to ensure the feasibility of implementing our software on the device in the final real-time prototype. As the device is intended to be operational on a moving platform, uploading image data in real-time can be both time-consuming and energy-intensive. By enabling onboard image processing and updating only the numerical data, we can optimize the entire vineyard monitoring process and provide real-time information to the agronomist. This approach minimizes the need for data transfer and facilitates efficient monitoring operations. This process involved a reconfiguration of the neural network by fine-tuning the last layer to detect a single class and generate bounding boxes around the buds. We assembled a custom dataset by taking several photos of the early buds during our acquisition campaigns in winter. The final dataset contains a total of 200 images, 85% of them taken using the Basler RGB camera, 10% using the smartphone camera (RedMi Note 11 Pro), and the remaining 5% were images downloaded from the internet without referring to a particular published dataset (see Section **??** for cameras details). This variation in data sources allowed us to incorporate a wide range of optical features into our dataset, enhancing the model's ability to generalize across different scenarios. Image samples of the three sets are shown in Fig. 5. The YOLOv8 neural network operates with a maximum image resolution of 640 px. However, buds are relatively small, hence we leveraged the capabilities of its embedded augmentation library [22], which includes image manipulation and cropping functions. As a result, the images in our dataset have a resolution of up to 1280 px, enabling the network to effectively process the augmented input data.

(a) (b) (c)

Fig. 5: Example images from our custom dataset taken using (a) the Basler DART color camera, (b) the RedMi Note 11 Pro smartphone camera, (c) downloaded from the internet.

IV. Preliminary Results

A. Shoots volume estimation results

Once the shoots volume estimation procedure was established, we proceeded to measure real vine shoots sample in a semi-controlled environment. We performed measurements at different distances to validate our model and identify the optimal distance to acquire the wood branch. 450 RGB-D images were taken at distances ranging from 600 mm to 1200 mm in laboratory conditions with sunlight illumination and controlled background. Fig. 6 shows the normal distribution of the errors with respect to the actual dimensions.

As expected, the estimation error exhibits a linear increase with distance, resulting in an acceptable uncertainty up to a distance of 1000 mm. We identify a secure range for future

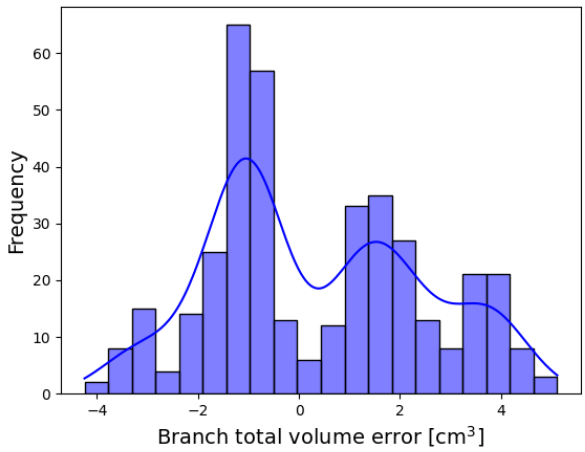

Fig. 6: Histogram distribution of the volume measurement errors with respect to the actual volume.

acquisition between 600 mm (lower sensor limit) and 1000 mm. For shoot volume measurements, we obtained a total RMSE of 2.1 cm^3 (9.7%) and a mean deviation of 1.8 cm^3 (8.3%).

B. Bud detection results

We performed fine-tuning retraining the YOLOv8 neural network to improve its performance in detecting buds, particularly in dynamic image settings. Typical metrics to evaluate object detector performance are F1-Score, Precision, and Recall as common standard [23]. In vineyard images, false positives can easily arise due to the abundance of wood pixels, especially from the foreground rows extending toward the background rows. Due to their small size in relation to the image plane, frequent partial occlusion, and potential confusion with the background, we accepted the possibility of some missed detections. As a result, we devised a plan to explore the frequency of occluded buds in order to improve the Recall value.

Our validation procedure involved a small dataset consisting of 45 brand new images, each containing a minimum of three buds. To ensure an independent and robust validation, we carefully composed the dataset with 85% of the images sourced from the internet, 10% captured using a smartphone camera, and the remaining 5% acquired with the Basler RGB camera in the vineyard. By constructing a validation dataset that is specular (opposite) to the training dataset, we aimed to prevent the network from relying solely on camera-specific features, thus enhancing its generalization ability. The resulting model achieved an F1-Score of 0.79, Precision and Recall of 0.88 and 0.72 respectively. Based on our preliminary dataset, the obtained metrics show promise. The high precision achieved is consistent with our objective of minimizing false predictions. However, a lower recall value is acceptable due to the presence of occluded buds and their heterogeneity.

V. Conclusions

The obtained results for both the shoots' volume estimation and bud detection are promising, despite being at their initial stage of development. We establish an optimal configuration for our multi-sensor device to be used in future vineyard acquisitions, maximizing sensor sensitivity and minimizing measurement uncertainty. It is observed that the volume estimation depends greatly on the quality of the raw data (both color and depth). This is especially crucial for measurements of such thin objects as the vines' branches and shoots. Therefore, in future developments, our focus should be on the improvement of the volume estimation procedure leveraging better quality images taken from cameras such as the Basler DART color camera. However, we need to pay attention to other parameters when choosing the camera, such as the overall cost and integration with the embedded platform. Our upcoming winter plans involve the validation of this measurement system in an authentic vineyard setting. To achieve this, we will develop an intelligent segmentation approach capable of distinguishing the vine shoots from the background without relying on the use of white sheets. The bud estimation model, on the other hand, achieved an acceptable level of accuracy; however, it may be improved in the future by expanding the dataset with images taken with different backgrounds. In fact, we suspect that the reduced accuracy is due to the presence of unseen noise that strongly characterizes the images downloaded from the internet in contrast with the ones taken using our acquisition device and software. Future developments of the project as a whole will include the estimation of the leaf area index and grape bunches volume, in conjunction with leaf and grape object detectors.

References

[1] E. Pomarici, A. Corsi *et al.*, "The italian wine sector: Evolution, structure, competitiveness and future challenges of an enduring leader," *Italian Economic Journal*, vol. 7, 03 2021.

[2] E. Cataldo, M. Fucile, and G. B. Mattii, "A review: Soil management, sustainable strategies and approaches to improve the quality of modern viticulture," *Agronomy*, vol. 11, no. 11, 2021. [Online]. Available: https://www.mdpi.com/2073-4395/11/11/2359

[3] S. Maicas, "Advances in wine fermentation," *Fermentation*, vol. 7, no. 3, 2021. [Online]. Available: https://www.mdpi.com/2311-5637/7/3/187

[4] A. Matese and S. Di Gennaro, "Technology in precision viticulture: A state of the art review," *International Journal of Wine Research*, vol. 7, 05 2015.

[5] S. Festini, "How amarone della valpolicella is made," January 2021. [Online]. Available: https://tasteverona.com/en/how-amarone-is-made/

[6] G. Jones, M. White *et al.*, "Climate change and global wine quality," *Climate Change*, vol. 73, pp. 319–343, December 2005.

[7] S. Francesca, S. Giosuè *et al.*, "Downy mildew (plasmopara viticola) epidemics on grapevine under climate change," *Global Change Biology*, vol. 12, pp. 1299 – 1307, May 2006.

[8] T. Thorat, B. Patle, and S. K. Kashyap, "Intelligent insecticide and fertilizer recommendation system based on tpf-cnn for smart farming," *Smart Agricultural Technology*, vol. 3, p. 100114, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S277237552200079X

[9] M. De la Fuente Lloreda, R. Linares *et al.*, "Comparison of different methods of grapevine yield prediction in the time window between fruitset and veraison," *Journal International des Sciences de la Vigne et du Vin*, vol. 49, 03 2015.

[10] M. Bergerman, S. M. Maeta *et al.*, "Robot farmers: Autonomous orchard vehicles help tree fruit production," *IEEE Robotics Automation Magazine*, vol. 22, no. 1, pp. 54–63, 2015.

[11] R. Oberti, M. Marchi *et al.*, "Automatic detection of powdery mildew on grapevine leaves by image analysis: Optimal view-angle range to increase the sensitivity," *Computers and Electronics in Agriculture*, vol. 104, pp. 1–8, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169914000611
[12] J. Tardaguila, M. Stoll *et al.*, "Smart applications and digital technologies in viticulture: A review," *Smart Agricultural Technology*, vol. 1, p. 100005, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2772375521000058
[13] B. Millan, M.-P. Diago *et al.*, "Vineyard pruning weight assessment by machine vision: towards an on-the-go measurement system," *OENO One*, vol. 53, pp. 333–345, 05 2019.
[14] N. Mylonas, I. Malounas *et al.*, "Eden library: A long-term database for storing agricultural multi-sensor datasets from uav and proximal platforms," *Smart Agricultural Technology*, vol. 2, p. 100028, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2772375521000289
[15] E. Library, "Meet our intelligent "viewer"," 2023. [Online]. Available: https://edenlibrary.ai/eden-viewer
[16] K. Fleming, D. Westfall, and M. Brodahl, "Evaluating farmer defined management zone maps for variable rate fertilizer application," *Precision Agriculture*, vol. 2, pp. 201–215, 10 2000.
[17] J. Chuche and D. Thiéry, "Biology and ecology of the flavescence dorée vector scaphoideus titanus: a review," *Agronomy for Sustainable Development*, vol. 34, no. 2, pp. 381–403, April 2014.
[18] Intel, "Intel® realsense™ camera d400 series product family datasheet rev. 01/2019." January 2019. [Online]. Available: https://www.intel.com/content/dam/support/us/en/documents/emerging-technologies/intel-realsense-technology/Intel-RealSense-D400-Series-Datasheet.pdf
[19] R. Shahmoradi, "Investigating the feasibility of using a realsense depth camera d435i by creating a framework for 3d pose analysis," July 2022. [Online]. Available: http://essay.utwente.nl/92232/
[20] A. G. Reynolds and J. E. V. Heuvel, "Influence of grapevine training systems on vine growth and fruit composition: A review," *American Journal of Enology and Viticulture*, vol. 60, no. 3, pp. 251–268, 2009.
[21] R. Szeliski, *Computer Vision: Algorithms and Applications*, 1st ed. Berlin, Heidelberg: Springer-Verlag, 2010.
[22] G. Jocher, A. Chaurasia, and J. Qiu, "Yolo by ultralytics," January 2023. [Online]. Available: https://github.com/ultralytics/ultralytics
[23] J. Lever, "Classification evaluation: it is important to understand both what a classification metric expresses and what it hides," *Nature Methods*, vol. 13, no. 8, pp. 603+, August 2016.