AR Assistant for Pruning of Grapevines and Fruit Trees

MARIIA PODGUZOVA[1], SIMON HÄRING[1], JAMIU OJELEYE[1], STEPHAN KRAUß[2], DIDIER STRICKER[1,2]

[1] Department of Computer Science, University of Kaiserslautern-Landau (RPTU), 67663 Kaiserslautern, Germany
[2] Augmented Vision, German Research Center for Artificial Intelligence (DFKI), 67663 Kaiserslautern, Germany

Corresponding author: mariia.podguzova@dfki.de

Keywords: augmented reality, mobile devices, image-based 3d modelling, tracking, computer vision, deep learning, grapevine pruning, fruit tree pruning

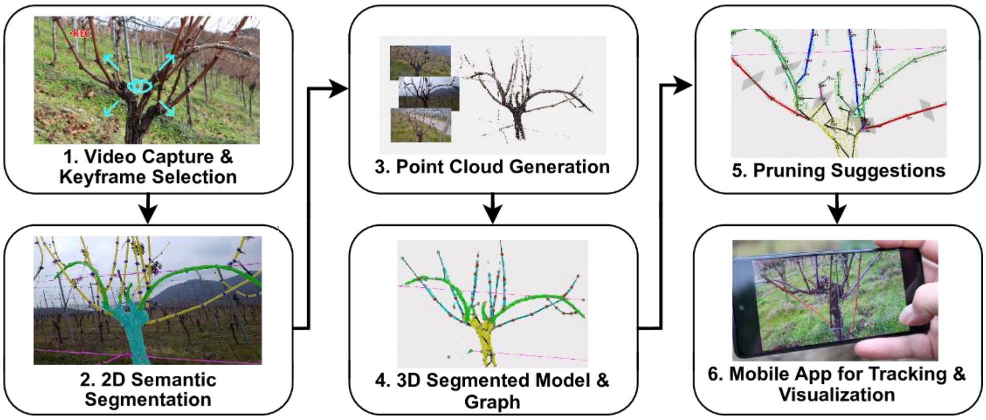

Figure 1. An overview of the pipeline that provides the user with pruning suggestions, given a video of the plant.

Introduction

Winter pruning in orchards is an essential but labor-intensive and time-consuming task, traditionally done manually to shape future growth by removing unwanted branches. Workers use various techniques to optimize yield, fruit quality or disease resistance. However, labor shortages and the need for professional training pose challenges for farmers. To address this, we offer support to make pruning more accessible to a broader segment of the workforce.

At the same time, computer vision in outdoor environments is complex due to varying lighting, weather, unique plant shapes, occlusions and similarity between foreground and background plants. Multiple studies have addressed the problem of pruning grapevines and other fruit trees [Amatya2016, Botterill2017, Gentilhomme2023, Tong2023]. These studies either focus on the entire automation pipeline [Botterill2017, Fourie2021] or separate steps, for instance, branch detection [Amatya2016, Zhang2018], reconstruction and skeletonization [You2022, Feng2024], and cut position localization [Marset2021]. The recent approaches [Fourie2021, Gentilhomme2023] prove the general feasibility of automated pruning systems but do not address real-world challenges like outdoor conditions and complex, occluded plant structures. Effective pruning systems for fruit trees require accurate spatial information. Several studies have highlighted challenges in capturing thin structures using laser scanners or 3D cameras, often requiring additional refinement or proper initial registration, which can be time-consuming [Tagarakis2013, Medeiros2017]. To achieve a balance between cost and benefit, affordable methods for 3D reconstruction need to be explored. In contrast to similar studies that use 3D sensors for apple trees [Majeed2018, Tong2023], we explore the potential of image-based approaches suitable for an augmented reality (AR) pruning assistant on a mobile device.

In this work, we address the mentioned challenges and present an AR assistant that enables inexperienced workers to carry out pruning for grapevines and reduces the size of the cut wounds, making the plants more resilient to fungal infections and promoting rich and healthy yield. We further apply this concept to other fruit trees, such as apple and peach trees, and highlight the improvements made in this direction. Our contributions can be summarized as follows. First, we present a pipeline that extracts 3D and semantic information from a video of a plant and outputs pruning suggestions using both traditional and deep-learning methods (Vid2Cuts). Second, we introduce a mobile AR application to display the results to the user. Third, we extend the pipeline for other fruit trees that are more challenging compared to grapevines due to their more complex 3D structure and larger size.

Figure 2. Comparison of grapevines (left), peach trees (center) and apple trees in high-density orchards (right).

Vid2Cuts

We implemented a multi-step pipeline to extract pruning suggestions from smartphone videos of a grapevine plant [Häring2024]. Figure 1 provides an overview of our approach. First, we automatically select a set of non-blurry frames, that are well spread throughout the video. This ensures good coverage of the entire plant from different angles. Next, we use a deep neural network [Pan2023] to identify different parts of the grapevine in the selected images via semantic segmentation. Using this information, we can also separate foreground from background in the images and use Meshroom [Griwodz2021] to generate a 3D point cloud of the plant. We combine the 2D semantic masks with the 3D information, by projecting the points into the masks via the intrinsic and extrinsic camera parameters also estimated by Meshroom. Doing so, every point is assigned the most common class it was projected onto. Next, the segmented point cloud is simplified into an abstract graph model. Each vertex in the graph has a 3D position and is connected by an edge to another vertex if there is a direct connection between them on the plant. The pruning rules defined by winemakers are then applied to this graph. Using these, a pair of good fruit rods and a pair of cones for the following year are selected. They are highlighted blue and red respectively in the top-right of Figure 1. Based on this selection, a set of cutting planes is generated. Finally, the graph model and the pruning suggestions can be viewed on a smartphone in the field as an AR overlay on the camera feed.

Adaptation to Fruit Trees

We developed the first prototype for grapevines, which naturally exhibit a "flat" branching pattern. Moreover, the plant size is relatively small, allowing a view from one side to be sufficient for making pruning decisions. In contrast, fruit trees present more challenges due to their thin, long structures and numerous occlusions. Additionally, these trees have complex spatial structures that need to be captured from a 360-degree view on distinct levels due to their considerable height, see Figures 2 and 3. Furthermore, pruning rules for fruit trees differ significantly based on the type of fruit, the variety, and even the farming location. Therefore, we plan to offer a tool for experts to define new rule sets, making our system more flexible. This will require the extraction of high-level properties and sub-routines which can then be combined to define new rules.

Enhanced 3D Reconstruction

Traditional computer vision reconstruction approaches, such as Structure from Motion [Soatto1998], struggle to reconstruct thin branches crucial for pruning decisions for fruit trees.

The first challenge is precise camera localization. Feature extraction often fails due to the lack of texture on apple trees, requiring more overlapping images and increasing resource usage. We improved this by incorporating a hierarchical approach for camera localization [Sarlin2019]. The main advantage is leveraging a global approach for the visual place recognition [Berton2023], combined with local feature extraction and matching, employing SuperPoint [DeTone2018] and SuperGlue [Sarlin2020] for a more robust and efficient feature matching, reducing the need for many overlapping keyframes.

The second challenge involves reconstruction of thin, occluded branches. We replaced Meshroom [Griwodz2021] with a Gaussian splatting-based approach (GS) [Kerbl2023], optimized for thin structures. Figure 3 shows the 3D models obtained using Meshroom and the Gaussian-based reconstruction approach. Due to the significant height of the trees, the keyframes exhibit considerable variation in their backgrounds. This makes feature matching challenging, even with the background removed, as the structures are very thin. This can be seen in the results obtained from Meshroom, which selected only the lower part of the tree for matching. By improving camera pose estimation, we achieved more accurate poses and a complete reconstruction using Gaussian-based methods.

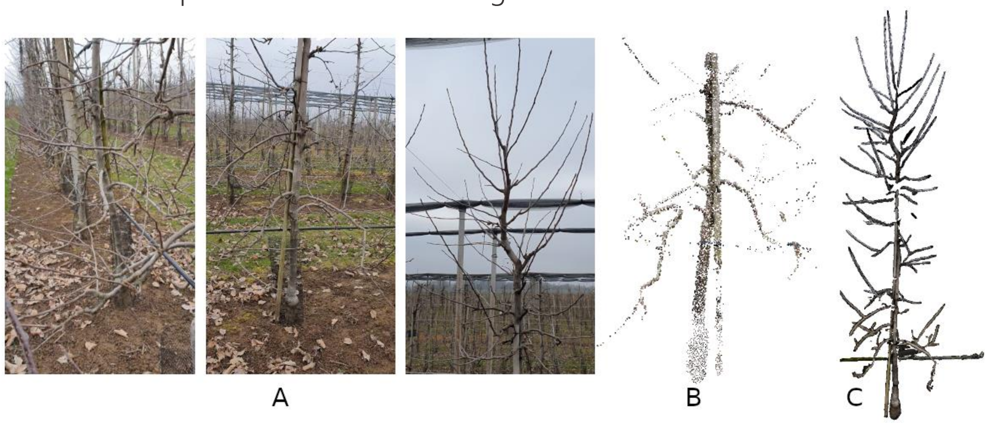

Figure 3. A: Examples of keyframes. Images 1 and 2 show other trees in the background, while image 3 has a smooth sky, complicating feature matching. B: Dense point cloud from Meshroom. C: Mesh using GS.

Tracking

In the initialization phase of our tracking system, traditional methods like template matching were initially used to match 2D nodes in live frames with precomputed 2D-3D correspondences from a pre-selected keyframe. However, this approach is limited due to its lack of invariance to scale, rotation, and sensitivity to illumination variations. To enhance robustness, we replaced template matching with a more advanced method. Specifically, we integrated SuperPoint as our 2D feature extractor, leveraging its strong capability of descriptor generation. For matching these features, we utilized LightGlue [Lindenberger2023], resulting in accurate camera pose estimation even under challenging scenarios.

To address the inadequacy of using a single pre-selected keyframe for larger fruit trees, we compute an optimal set of keyframes that collectively cover the tree's features, with precomputed 2D-3D correspondences for each keyframe. Furthermore, selecting an initial reference frame from these multiple keyframes for accurate initialization is a challenge. To resolve this, we match the features of each keyframe against the live frame. The keyframe with the highest number of matching features is then selected as the initial reference frame, ensuring a more reliable and accurate initialization process.

Conclusion

In this work, we presented our first prototype for generating pruning suggestions based on hand-held monocular video of a grapevine, along with a mobile AR application to visualize cut positions. The app is intended for vineyard use, enabling inexperienced workers to perform accurate pruning.

We addressed challenges encountered during development and extended our solution to larger and more complex fruit trees such as apple and peach trees. We enhanced 3D reconstruction as well as tracking in the AR mobile application making it more efficient and accurate for thin and long branches. The development of the extended AR mobile app for visualization of pruning suggestions for fruit trees is kept as future work.

Acknowledgement

This work was partially funded by the AI:Pruning project and partially funded by the EU project CORTEX^2 (GA: Nr 101070192).

References

Amatya, S., Karkee, M., Gongal, A., Zhang, Q. & Whiting, M. D., “Detection of cherry tree branches with full foliage in planar architecture for automated sweet-cherry harvesting,” Biosyst. Eng., pp. 3–15, 2016.

Berton, G., Trivigno, G., Caputo, B. & Masone, C. "EigenPlaces: Training Viewpoint Robust Models for Visual Place Recognition," Proceedings of the IEEE/CVF ICCV, 2023, pp. 11080-11090.

Botterill, T., Paulin, S., Green, R., Williams, S., Lin, J., Saxton, V., Mills, S., Chen, X. & Corbett-Davies, S., ‘‘A robot system for pruning grape vines,’’ J. Field Robot., pp. 1100–1122, 2017.

DeTone, D., Malisiewicz, T. & Rabinovich, A., "SuperPoint: Self-Supervised Interest Point Detection and Description," 2018 IEEE/CVF CVPRW, 2018, pp. 337-33712.

Feng, W., Jiao, M., Liu, N., Yang, L., Zhang, Z. & Hu, S., “Realistic reconstruction of trees from sparse images in volumetric space,” Computers & Graphics, vol. 121, 2024, 103953.

Fourie, J., Bateman, C., Hsiao, J., Pahalawatta, K., Batchelor, O., Misse, P. E. & Werner, A., ‘‘Towards automated grape vine pruning: Learning by example using recurrent graph neural networks.’ Int. J. Intell. Syst., 2021.

Gentilhomme, T., Villamizar, M., Corre, J. & Odobez, J., "Efficient Grapevine Structure Estimation in Vineyards Conditions," in 2023 IEEE/CVF, ICCVW, 2023 pp. 712-720.

Griwodz, C., Gasparini, S., Calvet, L., Gurdjos, P., Castan, F., Maujean, B., De Lillo, G. & Lanthony, Y., “AliceVision meshroom: An open-source 3D reconstruction pipeline,” in Proc. 12th ACM MMSys, 2021, 241–247.

Häring, S., Folawiyo, S., Podguzova, M., Krauß, S. & Stricker, D. 2024. “Vid2Cuts: A Framework for Enabling AI-guided Grapevine Pruning.” IEEE Access Vol. 12, 5814-5836.

Kerbl, B., Kopanas, G., Leimkühler, & Drettakis, G., “3d gaussian splatting for real-time radiance field rendering,” ACM Trans. Graph., 2023.

Lindenberger, P., Sarlin, P. & Pollefeys, M., “LightGlue: Local Feature Matching at Light Speed”, ICCV, 2023.

Majeed, Y., Zhang, J., Zhang, X., Fu L., Karkee, M., Zhang, Q. & Whiting, M. D., ‘‘Apple tree trunk and branch segmentation for automatic trellis training using convolutional neural network based semantic segmentation,’’ IFAC-PapersOnLine, pp. 75–80, 2018.

Marset, W. V., Pérez, D. S., Díaz, C. A. & Bromberg, F., ‘‘Towards practical 2D grapevine bud detection with fully convolutional networks,’’ Comput. Electron. Agricult., 2021.

Medeiros, H.; Kim, D.; Sun, J.; Seshadri, H.; Akbar, S.A.; Elfiky, N.M.; Park, J. “Modeling dormant fruit tree for agricultural automation. J. Field Robot.” 2017, 1203–1224.

Pan, H., Hong, Y., Sun, W. & Jia, Y., “Deep dual-resolution networks for real-time and accurate semantic segmentation of traffic scenes,” IEEE Transactions on Intelligent Transportation Systems, 3448–3460, 2023.

Sarlin, P., Cadena, C., Siegwart, R. & Dymczyk, M., "From Coarse to Fine: Robust Hierarchical Localization at Large Scale," in 2019 IEEE/CVF CVPR, 2019, pp. 12708-12717.

Sarlin, P., DeTone, D., Malisiewicz, T. & Rabinovich, A., "SuperGlue: Learning Feature Matching with Graph Neural Networks," *2020 IEEE/CVF CVPR*, 2020, pp. 4937-4946.

Soatto, S. & Brockett, R., "Optimal structure from motion: local ambiguities and global estimates," Proceedings. 1998, IEEE Computer Society Conference on CVPR, 1998, pp. 282-288.

Tagarakis, A., Liakos, V., Chatzinikos, T., Koundouras, S., Fountas, S. & Gemtos, T. “Using laser scanner to map pruning wood in vineyards,” Precis. Agric. 2013, 633–639.

Tong, S., Zhang, J., Li W., Wang, Y. & Kang, F., “An image-based system for locating pruning points in apple trees using instance segmentation and RGB-D images,” Biosystems Engineering, 2023, pp. 277-286.

You, A., Grimm, C., Silwal, A. & Davidson J.R., ‘‘Semantics-guided skeletonization of upright fruiting offshoot trees for robotic pruning,’’ Comput. Electron. Agricult., 2022.

Zhang, J., He, L., Karkee, M., Zhang, Q., Zhang, X. & Gao, Z., ‘‘Branch detection for apple trees trained in fruiting wall architecture using depth features and regions-convolutional neural network (R-CNN),’’ Comput. Electron. Agricult., pp. 386–393, 2018.