POLITECNICO DI TORINO
Repository ISTITUZIONALE

Autonomous Navigation in Rows of Trees and High Crops with Deep Semantic Segmentation

(Article begins on next page)

29 April 2024

Autonomous Navigation in Rows of Trees and High Crops with Deep Semantic Segmentation

Alessandro Navone[1], Mauro Martini[1], Andrea Ostuni[1], Simone Angarano[1] and Marcello Chiaberge[1]

Abstract— Segmentation-based autonomous navigation has recently been proposed as a promising methodology to guide robotic platforms through crop rows without requiring precise GPS localization. However, existing methods are limited to scenarios where the centre of the row can be identified thanks to the sharp distinction between the plants and the sky. However, GPS signal obstruction mainly occurs in the case of tall, dense vegetation, such as high tree rows and orchards. In this work, we extend the segmentation-based robotic guidance to those scenarios where canopies and branches occlude the sky and hinder the usage of GPS and previous methods, increasing the overall robustness and adaptability of the control algorithm. Extensive experimentation on several realistic simulated tree fields and vineyards demonstrates the competitive advantages of the proposed solution.

I. INTRODUCTION

In recent years, precision agriculture has pushed the boundaries of technology to optimize crop production, improve the efficiency of farming operations, and reduce waste [1]. Modern farming systems must be able to extract synthetic key information from the environment, take or suggest optimal decisions based on that information, and execute them with high precision and timing. Deep learning techniques have shown great potential in realizing these systems by analyzing data from multiple sources, allowing for large-scale, high-resolution monitoring, and providing detailed insights for both human and robotic agents. The most recent advancements in deep learning also provide competitive advantages for real-world applications, such as model optimization for fast inference on low-power embedded hardware [2], [3] and generalization to unseen data [4], [5], [6]. At the same time, progress in service robotics has enabled autonomous mobile agents to embody AI perception systems and work in synergy with them to accomplish complex tasks in unstructured environments [7].

In particular, row-based crops are among the most studied applications (they constitute more than 75% of all planted acres of cropland across the USA [8]). In this scenario, research spans localization[9], path planning [10], navigation [11], monitoring[12], harvesting [13], spraying [14], and vegetative assessment [15], [16]. A particularly challenging situation occurs when standard localization methods, like GPS, fail to reach the desired precision due to unfavorable weather conditions or line-of-sight obstruction. That is the

Fig. 1. The proposed SegMin and SegMinD algorithms allow to precisely guide an autonomous mobile robot through a dense tree row solely using an RGB-D camera. A pear crop row in Gazebo is shown in the picture.

case, for example, of dense tree canopies, as shown in a simulated pear orchard in Figure 1.

Previous works have proposed position-agnostic vision-based navigation algorithms for row-based crops. A first vision-based approach was proposed in [17] using mean-shift clustering and the Hough transform to segment RGB images and generate the optimal central path. Later, [18] achieved promising results using multispectral images and simply thresholding and filtering on the green channel. Recently, deep-learning approaches have been successfully applied to the task. [19] proposed a classification-based approach in which a model predicts the discrete action to perform. In contrast, [20] proposed combining a segmentation model and a proportional controller to align the robot to the center of the row. Finally, a different approach was tested in [11] with an end-to-end controller based on deep reinforcement learning. Although these systems proved effective in their testing scenarios, they have only been applied in simple crops where a full view of the sky favors both GPS receivers [21] and vision-based algorithms [22].

This work tackles a more challenging scenario in which dense canopies partially or totally cover the sky, and the GPS signal is very weak. We design a navigation algorithm based on semantic segmentation that exploits visual perception to estimate the center of the crop row and align the robot trajectory to it. The segmentation masks are predicted by a deep learning model designed for real-time efficiency and trained on realistic synthetic images. The proposed navigation algorithm improves on previous works being adaptive to different terrains and crops, including dense

[1] Department of Electronics and Telecommunications, Politecnico di Torino, 10129, Torino, Italy. {firstname.lastname}@polito.it

canopies. We conduct extensive experimentation in simulated environments for multiple crops. We compare our solution with previous state-of-the-art methodologies, demonstrating that the proposed navigation system is effective and adaptive to numerous scenarios.

The main contributions of this work can be summarized as follows:

- we present two variants of a novel approach for segmentation-based autonomous navigation in tall crops, designed to tackle challenging and previously uncovered scenarios;
- we test the resulting guidance algorithm on previously unseen plant rows scenarios such as high trees and pergola vineyards.
- we compare the new method with state-of-the-art solutions on straight and curved vineyards, demonstrating an enhanced general and robust behavior.

The next sections are organized as follows: Section II presents the proposed deep-learning-based control system for vision-based position-agnostic autonomous navigation in row-based crops, from the segmentation model to the controller. Section III describes the experimental setting and reports the main results for validating the proposed solution divided by sub-system. Finally, Section IV draws conclusive comments on the work and suggests interesting future directions.

II. METHODOLOGY

This work proposes a real-time control algorithm with two variants to navigate high-vegetation orchards and arboriculture fields and improve the approach presented in [20]. The proposed system avoids exploiting the GPS signal, which can lack accuracy due to signal reflection and mitigation due to vegetation.

The working principle of the proposed control algorithms is straightforward and exploits only the RGB-D data. Both the proposed solutions consist of four main steps:

1) Semantic segmentation of the input RGB frame.
2) Processing of the output segmentation mask using depth frame data.
3) Searching for the direction which leads the mobile platform towards the end of the row.
4) Generating linear and angular velocity commands to input the mobile robot.

Nonetheless, the two proposed methods differ only for steps 2 and 3 in employing the depth frame data and in the generation of the path which the robot should follow. In contrast, the segmentation technique 1 and the command generation 4 are carried out similarly. A schematic representation of the proposed pipeline is described in Figure 2.

As in [20] a first step, an RGB frame $\mathbf{X}_{rgb}^t \in \mathbb{R}^{h \times w \times c}$ and a depth map $\mathbf{X}_d^t \in \mathbb{R}^{h \times w}$ are acquired by a camera placed on the front of the mobile platform at each instant t, where h and w are the width of the frame and c is the number of channels. The received RGB data is then fed to a segmentation neural network model H_{seg}, which outputs a binary segmentation mask bringing the semantic information of the input frame.

$$\hat{\mathbf{X}}_{seg}^t = H\left(\mathbf{X}_{rgb}^t\right) \tag{1}$$

where $\hat{\mathbf{X}}_{seg}^t$ is the estimated segmentation mask. Moreover, the segmentation masks of the last N time instants $\{t - N, \ldots, t\}$ are fused to obtain more robust information.

$$\hat{\mathbf{X}}_{CumSeg}^t = \bigcup_{j=t-N}^{t} \hat{\mathbf{X}}_{seg}^j \tag{2}$$

where $\hat{\mathbf{X}}_{CumSeg}^t$ is the cumulative segmentation mask and the operator $\bigcup$ represent the logical bitwise *OR* operation over the last N binary frames.

Additionally, the depth map $\mathbf{X}_d^t$ is now used to consider the segmented regions between the camera position and a given depth threshold d_{th} to remove useless information given by far vegetation, which is irrelevant to control the robot's movement.

$$\hat{\mathbf{X}}_{segDepth}{}_{j=0,\ldots,w}^{i=0,\ldots,h}(i, j) = \begin{cases} 0, & \text{if } \hat{\mathbf{X}}_{CumSeg(i,j)}^t \cdot \hat{\mathbf{X}}_{d(i,j)}^t > d_{th} \\ 1, & \text{if } \hat{\mathbf{X}}_{CumSeg(i,j)}^t \cdot \hat{\mathbf{X}}_{d(i,j)}^t \leq d_{th} \end{cases} \tag{3}$$

where $\hat{\mathbf{X}}_{segDepth}$ is the resulting intersection between the cumulative segmentation frame and the depth map cut at a distance threshold d_{th}.

Henceforth the proposed algorithm forks in two variants, *SegMin* and *SegMinD*, respectively described in II-A and II-B.

A. SegMin

The first variant improves the approach proposed in [20]. After processing the segmentation mask, a sum over the column is performed to obtain a histogram $\mathbf{h} \in \mathbb{R}^w$, quantifying how much vegetation is present on each column. Hereafter, a moving average on a window of n elements is performed over the array to smooth the values and make the control more robust to punctual noise derived from the previous passages. Ideally, the minimum of this histogram x_h corresponds to the regions where less vegetation is present and, therefore, identifies the desired central path inside the crop row. If more global minimum points are present (i.e., there is a region where no vegetation is detected), the mean of the considered points is considered to be the global minimum and, in consequence, the continuation of the row.

B. SegMinD

The second proposed approach consists of a variant of the previous algorithm, devised for wide rows with tall and thick canopies, which in the previous case would generate an ambiguous global minimum due to the constant presence of vegetation above the robot. This variant multiplies the previously processed segmentation mask for the normalized inverted depth datum.

$$\hat{\mathbf{X}}_{depthInv}^t = \hat{\mathbf{X}}_{segDepth}^t \bigcap \left(1 - \frac{\mathbf{X}_d^t}{d_{th}}\right) \tag{4}$$
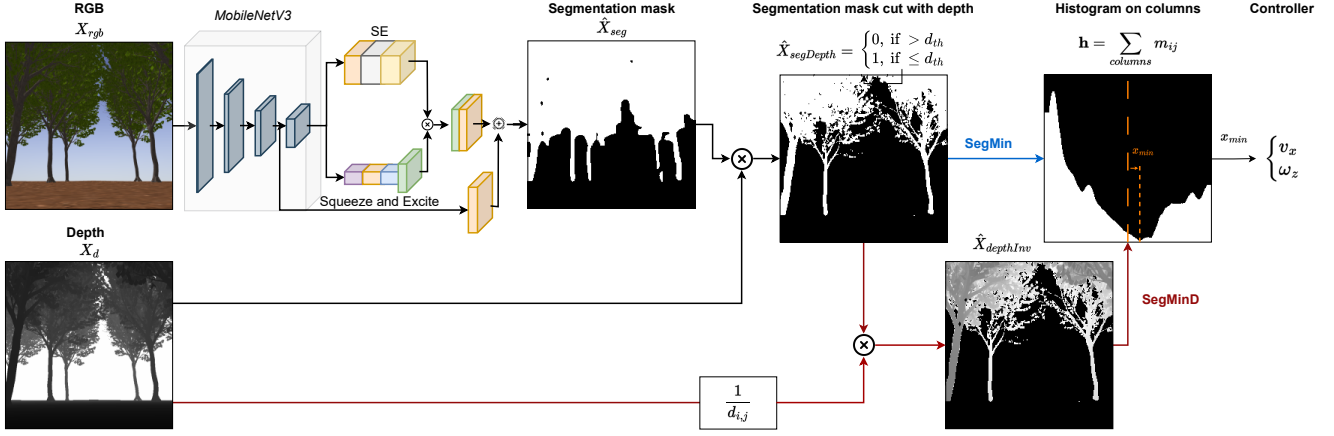
Fig. 2. Scheme of the overall proposed navigation pipeline. The RGB image is fed into the segmentation network, thus the predicted segmentation mask $\hat{\mathbf{X}}_{seg}^{t}$ is refined using the depth frame to obtain $\hat{\mathbf{X}}_{segDepth}^{t}$. The blue arrow refers to the SegMin variant, and red arrows refer to the SegMinD variant to compute the sum histogram over the mask columns. Images are taken from navigation in the tall trees simulation world.

where $\hat{\mathbf{X}}_{depthInv}^{t}$ is the result of the element-wise multiplication, represented by $\bigcap$, between the binary mask $\hat{\mathbf{X}}_{segDepth}^{t}$ and the depth frame $\hat{\mathbf{X}}_{d}^{t}$ normalized over the depth threshold d_{th}. As in the previous case, the sum over the column is performed to obtain the 1D array $\mathbf{h}$ and, later on, the smoothing through a moving average. The introduced modification allows the closer elements to exert a greater influence on identifying the row direction.

C. Segmentation Network

We adopt the same network used in previous works on real-time crop segmentation [20], [6]. The model consists of a MobilenetV3 backbone for feature extraction and an efficient LR-ASPP segmentation head [23]. In particular, the LR-ASPP leverages effective modules such as depth-wise convolutions, channel-wise attention, and residual skip connections to provide an effective trade-off between accuracy and inference speed. The model is trained with a similar procedure to [6] on the AgriSeg dataset[1]. Further details on the training strategy and hyperparameters are provided in Section III.

D. Robot heading control

The objective of the controller pipeline consists in keeping the mobile platform at the center of the row, which, in this work, is considered equivalent to keeping the row center in the middle of the camera frame. Therefore, as defined in the previous step, the minimum of the histogram should be centered in the frame width. The distance d from the center of the frame and the minimum is defined as:

$$d = x_h - \frac{w}{2} \tag{5}$$

The linear and angular velocities are then generated through custom functions similarly as in [24].

$$v_x = v_{x,max} \left[1 - \frac{d^2}{\left(\frac{w}{2}\right)^2} \right] \tag{6}$$

$$\omega_z = -k_{\omega_z} \cdot \omega_{z,max} \cdot \frac{d^2}{w^2} \tag{7}$$

where $v_{x,max}$ and $\omega_{z,max}$ are respectively the maximum achievable linear and angular velocities and k_{ω_z} is the angular gain which regulates the speed of the response. In order to avoid abrupt changes in the robot's motion, the final velocities $\bar{v}_x$ and $\bar{\omega}_z$ commands are smoothed with an Exponential Moving Average (EMA) as:

$$\begin{bmatrix} \bar{v}_x^t \\ \bar{\omega}_z^t \end{bmatrix} = (1 - \lambda) \begin{bmatrix} \bar{v}_x^{t-1} \\ \bar{\omega}_z^{t-1} \end{bmatrix} + \lambda \begin{bmatrix} v_x^t \\ \omega_z^t \end{bmatrix} \tag{8}$$

where t is the time step and λ is a chosen weight.

III. EXPERIMENTS AND RESULTS

A. Simulation Environment

The proposed control algorithm was tested through the use of Gazebo[2] simulation software. The software was selected because of its compatibility with ROS 2 and can incorporate plugins that simulate sensors, such as cameras. A Clearpath Jackal model was utilized to assess the algorithm's effectiveness. The URDF file, available through Clearpath Robotics, contains all the necessary information regarding the mechanical structure and joints of the robot. During the simulation, an Intel Realsense D435i plugin was utilized, positioned 20 cm in front of the robot's center, and tilted 15° upwards. This positioning gave the camera a better view of the upper branches of trees.

The navigation algorithm was tested in four different custom simulation environments: a common vineyard, a pergola vineyard characterized by vine poles and shoots above the row, a pear field constituted by small size trees, and a high trees field where canopies of the trees are merged above the row. Each simulated field adopts a different terrain, miming the irregularity of uneven terrain. The detailed measurements of the simulation world are described in Table I.
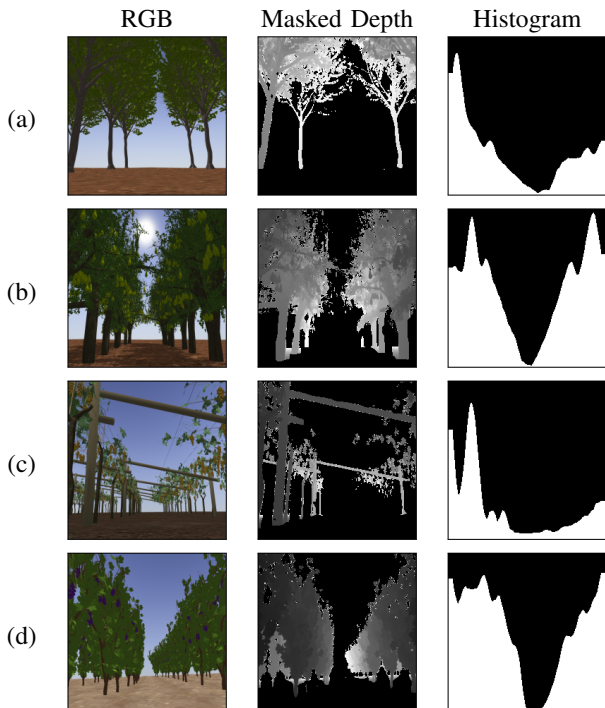
RGB | Masked Depth | Histogram
(a)
(b)
(c)
(d)

Fig. 3. Sample outputs of the proposed SegMinD algorithm for High Trees (a), Pear Trees (b), Pergola Vineyard (c), and Vineyard (d). Predicted segmentation masks are refined cutting values exceeding a depth threshold. The sum over mask columns provides the histograms used to identify the center of the row as its global minimum.

TABLE I

SIZE OF THE DIFFERENT SIMULATED CROPS, REFERRING TO THE AVERAGE VALUES OF THE DISTANCE BETWEEN ROWS, THE DISTANCE BETWEEN PLANTS ON THE ROW, AND THE HEIGHTS OF THE PLANTS.

Gazebo worlds	Rows distance [m]	Plant distance [m]	Height [m]
Common vineyard	1.8	1.3	2.0
Pergola vineyard	6.0	1.5	2.9
Pear field	2.0	1.0	2.9
High trees field	7.0	5.0	12.5

During the experimental part of this work, we consider frame dimensions equal to $(h, w) = (224, 224)$, which is the same size as the input and the output of the neural network model, with the number of channels $c = 3$. The maximum linear velocity has been fixed to $v_{x,max} = 0.5 m/s$, and the maximum angular velocity has been fixed to $\omega_{z,max} = 1 rad/s$. The angular velocity gain $\omega_{z,gain}$ has been fixed to 0.01, and the EMA buffer size has been fixed to 3. The depth threshold has been changed according to the various crops. In particular, it has been empirically fixed to 5 m in the case of vineyards, while it was increased to 8 m for pear trees and pergola vineyards and 10 m for tall trees according to the average distance from the rows in the diverse fields.

B. Segmentation Network Training and Evaluation

We train the crop segmentation model using a subset of the AgriSeg segmentation dataset [6]. In particular, for the High Tree and Pear crops, we train on Generic Tree splits 1 and 2, and on Pear; for Vineyards, we train on Vineyard
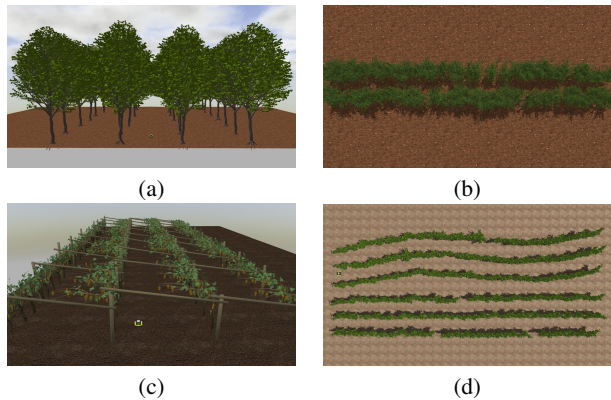

(a) | (b)
(c) | (d)

Fig. 4. Gazebo simulated environments used to test the SegMin approach in relevant different crops rows: wide rows composed of high trees (a), a narrow pear tree row (b), a pergola vineyard with asymmetric rows (c), straight and curved vineyard rows (d). In the last case, the tests were carried out in the second row from above and the second row from below.

and Pergola Vineyard (note that the testing environments are different from the ones from which the training samples are generated). In both cases, the model is trained for 50 epochs with Adam optimizer and learning rate 3×10^{-4}. We apply data augmentation by randomly applying cropping, flipping, greyscaling, and random jitter to the images. Our experimentation code is developed in Python 3 using TensorFlow as the deep learning framework. We train models starting from ImageNet pretrained weights, so the input size is fixed to (224×224). All the training runs are performed on a single Nvidia RTX 3090 graphic card.

C. Navigation Results

The overall navigation pipeline of SegMin and its variant SegMinD are tested in realistic crops fields in simulation using relevant metrics for visual-based control without precise localization of the robot, as done in previous works [20], [11]. The camera frames are published at a frequency of 30 Hz, while the inference is carried out at 20 Hz, and the controllers publish the velocity commands at 5 Hz. The evaluation has been performed using the testing package of the open-source PIC4rl-gym[3] in Gazebo [25]. The selected metrics aim at evaluating the effectiveness of the navigation (clearance time) as well as the precision, quantitatively comparing the obtained trajectories with a ground truth one through Mean Absolute Error (MAE) and Mean Squared Error (MSE). The ground truth trajectories have been computed by averaging the curve obtained by interpolating the plants' poses in the rows. For the asymmetric pergola vineyard case, the row is intended as the portion of the pergola without vegetation on top, as shown in Figure 4 (c). The response of the algorithms to terrain irregularity and rows geometry is also studied, including in the test significant kinematic information of the robot. The cumulative heading average $\gamma[rad]$ along the path is considered, together with the mean linear velocity $v_{avg}[m/s]$ and the standard deviation of the angular velocity $\omega_{stddev}[rad/s]$ commands predicted to keep

[3]https://github.com/PIC4SeR/PIC4rl_gym

TABLE II

Navigation results obtained in different test fields with the SegMin, SegMinD, and previous work SegZeros segmentation-based algorithms. The metrics test the effectiveness of the navigation (clearance time) and its precision with Mean Absolute Error (MAE) and Mean Squared Error (MSE) between obtained and ground truth path. The cumulative heading average $\gamma[rad]$, the mean linear velocity $v_{avg}[m/s]$, and the standard deviation of the angular velocity $\omega_{stddev}[rad/s]$ commands provide relevant kinematic information of the robot while navigating. SegZeros is not applicable in the case of high trees, pear trees, and pergola vineyards since the sky may be covered by vegetation.

Test Field	Method	Clearance time [s]	MAE [m]	MSE [m]	Cum. γ_{avg} [rad]	$v_{avg}[m/s]$	$\omega_{stddev}[rad/s]$
High Trees	SegMin	**40.409 ± 0.117**	0.265 ± 0.005	0.084 ± 0.003	0.079 ± 0.001	0.487 ± 0.000	0.054 ± 0.002
	SegMinD	40.440 ± 0.515	**0.174 ± 0.006**	**0.036 ± 0.002**	0.048 ± 0.002	0.484 ± 0.006	0.063 ± 0.019
Pear Trees	SegMin	**42.058 ± 1.228**	0.034 ± 0.012	**0.002 ± 0.001**	0.013 ± 0.002	0.483 ± 0.003	0.108 ± 0.054
	SegMinD	42.259 ± 1.912	**0.031 ± 0.017**	**0.002 ± 0.002**	0.016 ± 0.004	0.477 ± 0.009	0.026 ± 0.004
Pergola Vineyard	SegMin	**40.859 ± 0.386**	**0.077 ± 0.011**	**0.011 ± 0.003**	0.030 ± 0.022	0.479 ± 0.003	0.174 ± 0.021
	SegMinD	41.135 ± 0.329	0.097 ± 0.052	0.015 ± 0.014	0.029 ± 0.011	0.475 ± 0.004	0.204 ± 0.032
Straight Vineyard	SegMin	**50.509 ± 0.305**	**0.105 ± 0.003**	**0.014 ± 0.001**	0.033 ± 0.002	0.487 ± 0.000	0.079 ± 0.011
	SegMinD	50.629 ± 0.282	0.110 ± 0.005	0.018 ± 0.003	0.026 ± 0.009	0.486 ± 0.001	0.088 ± 0.005
	SegZeros	53.695 ± 1.029	0.138 ± 0.025	0.024 ± 0.010	0.027 ± 0.004	0.457 ± 0.008	0.089 ± 0.008
Curved Vineyard	SegMin	53.321 ± 0.249	0.115 ± 0.008	0.017 ± 0.002	0.036 ± 0.008	0.487 ± 0.001	0.088 ± 0.021
	SegMinD	**51.444 ± 1.030**	**0.093 ± 0.005**	**0.012 ± 0.001**	0.015 ± 0.004	0.484 ± 0.007	0.065 ± 0.008
	SegZeros	71.048 ± 27.132	0.108 ± 0.044	0.019 ± 0.009	0.045 ± 0.008	0.395 ± 0.127	0.114 ± 0.039

the robot correctly oriented. The mean value of ω is always close to zero due to the consecutive correction of the robot orientation.

The complete results collection is reported in Table II. For each metric, an average value and the standard deviation are indicated since all the experiments have been repeated over 3 runs on a 20 m long track in each crop row. The proposed method demonstrates to solve the problem of guiding the robot through tree rows with thick canopies (high trees and pears) without a localization system, as well as in peculiar scenarios such as the pergola vineyards. The identification of plant branches and wooden supports hinders the usage of previously existing segmentation-based solutions that were based on the assumption of finding a free passage solely considering the zeros of the binary segmentation mask [20]. We refer to this previous method as SegZeros in the comparison of the results that we tested using the same segmentation neural network.

The SegMin approach based on histogram minimum search demonstrates to be a robust solution to guide the robot through tree rows. The introduction of the depth inverse values as a weighting function allows SegMinD to further increase the precision of the algorithm in following the central trajectory of the row in complex cases such as wide rows (high trees) and curved rows (curved vineyard). The different sum histograms obtained with SegMin and SegMinD are directly compared in Figure 5, showing the sharper trend and the global minimum isolation obtained, including the depth values. Moreover, the novel methods show competitive performance also with standard crop rows where a free passage to the end of the row can be seen in the mask without the disturbance of canopies. The histogram minimum approach significantly reduces the navigation time and the trajectory precision in vineyard rows (straight and curved) compared to the previous segmentation-based baseline method. The
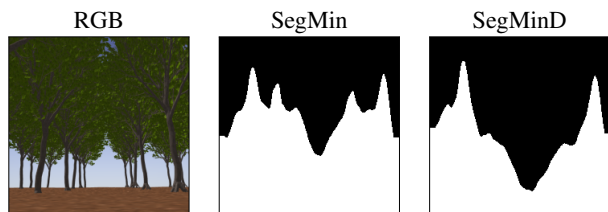
RGB   SegMin   SegMinD

Fig. 5. Comparison of the two histograms obtained using the two different algorithms, given the RGB frame on the right. It can be noticed how SegMinD offers a narrower and less ambiguous global minimum point.

search of plant-free zero clusters in the map results in being less robust and efficient, leading the robot to undesired stops during the navigation and to an overall slower and more oscillating behavior. Moreover, the standard deviation of the angular velocity is coherent with the obtained results, being smaller in the cases when the trajectory is more accurate, and the cumulative heading shows larger values when the algorithms are more reactive.

Nonetheless, the trajectories obtained with the SegMin, SegMinD and SegZeros algorithms are also visually shown in Figure 6 inside representative scenarios: a cluttered, narrow row with small pear trees, a wide row with high trees, and curved vineyards with state-of-the-art method SegZeros.

IV. CONCLUSIONS

In this work, we presented a novel method to guide to a service-autonomous platform through crop rows where a precise localization signal is often occluded by the vegetation. Trees rows represented an open problem in row crop navigation since previous works based on image segmentation or processing failed due to the presence of branches and canopies covering the free passage for the rover in the image. The proposed pipeline SegMin and SegMinD overcome this limitation by introducing a global minimum search on the
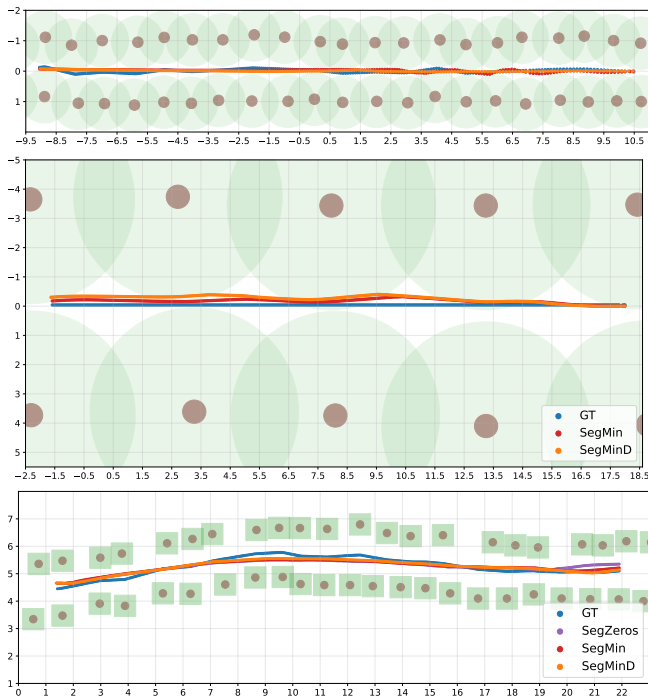
Fig. 6. Trajectories comparison between our proposed algorithms (SegMin and SegMinD) and the ground truth central path (GT): Pears (top), High Trees (center), Curved Vineyard (bottom). In the last graph, the trajectory generated with the SegZeros algorithm is also reported for comparison.

sum histogram over the mask columns. The experiments conducted demonstrate the ability to solve the navigation task in wide and narrow tree rows and, nonetheless, the improvement in efficiency and robustness provided by our method over previous works in generic vineyards scenarios.

Future works will see the test of the overall system in real-world tree rows, orchards, and vineyards to further validate the robustness of the solution with respect to sim2real gap problems and hardware resources.

ACKNOWLEDGMENTS

REFERENCES

[1] Z. Zhai, J. F. Martínez, V. Beltran, and N. L. Martínez, "Decision support systems for agriculture 4.0: Survey and challenges," *Computers and Electronics in Agriculture*, vol. 170, p. 105256, 2020.

[2] V. Mazzia, A. Khaliq, F. Salvetti, and M. Chiaberge, "Real-time apple detection system using embedded systems with hardware accelerators: An edge ai application," *IEEE Access*, vol. 8, pp. 9102–9114, 2020.

[3] S. Angarano, F. Salvetti, V. Mazzia, G. Fantin, D. Gandini, and M. Chiaberge, "Ultra-low-power range error mitigation for ultrawideband precise localization," in *Intelligent Computing: Proceedings of the 2022 Computing Conference, Volume 2*. Springer, 2022, pp. 814–824.

[4] M. Martini, V. Mazzia, A. Khaliq, and M. Chiaberge, "Domain-adversarial training of self-attention-based networks for land cover classification using multi-temporal sentinel-2 satellite imagery," *Remote Sensing*, vol. 13, no. 13, p. 2564, 2021.

[5] S. Angarano, M. Martini, F. Salvetti, V. Mazzia, and M. Chiaberge, "Back-to-bones: Rediscovering the role of backbones in domain generalization," *arXiv preprint arXiv:2209.01121*, 2022.

[6] S. Angarano, M. Martini, A. Navone, and M. Chiaberge, "Domain generalization for crop segmentation with knowledge distillation," *arXiv preprint arXiv:2304.01029*, 2023.

[7] A. Eirale, M. Martini, L. Tagliavini, D. Gandini, M. Chiaberge, and G. Quaglia, "Marvin: An innovative omni-directional robotic assistant for domestic environments," *Sensors*, vol. 22, no. 14, p. 5261, 2022.

[8] D. Bigelow and A. Borchers, "Major uses of land in the united states, 2012," *Economic Information Bulletin Number 178*, no. 1476-2017-4340, p. 69, 2017.

[9] W. Winterhalter, F. Fleckenstein, C. Dornhege, and W. Burgard, "Localization for precision navigation in agricultural fields—beyond crop row following," *Journal of Field Robotics*, vol. 38, no. 3, pp. 429–451, 2021.

[10] F. Salvetti, S. Angarano, M. Martini, S. Cerrato, and M. Chiaberge, "Waypoint generation in row-based crops with deep learning and contrastive clustering," in *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2022, Grenoble, France, September 19–23, 2022, Proceedings, Part VI*. Springer, 2023, pp. 203–218.

[11] M. Martini, S. Cerrato, F. Salvetti, S. Angarano, and M. Chiaberge, "Position-agnostic autonomous navigation in vineyards with deep reinforcement learning," in *2022 IEEE 18th International Conference on Automation Science and Engineering (CASE)*, 2022, pp. 477–484.

[12] L. Comba, A. Biglia, D. R. Aimonino, P. Barge, C. Tortia, and P. Gay, "2d and 3d data fusion for crop monitoring in precision agriculture," in *2019 IEEE international workshop on metrology for agriculture and forestry (MetroAgriFor)*. IEEE, 2019, pp. 62–67.

[13] C. W. Bac, E. J. van Henten, J. Hemming, and Y. Edan, "Harvesting robots for high-value crops: State-of-the-art review and challenges ahead," *Journal of Field Robotics*, vol. 31, no. 6, pp. 888–911, 2014.

[14] R. Berenstein, O. B. Shahar, A. Shapiro, and Y. Edan, "Grape clusters and foliage detection algorithms for autonomous selective vineyard sprayer," *Intelligent Service Robotics*, vol. 3, no. 4, pp. 233–243, 2010.

[15] G. Zhang, T. Xu, Y. Tian, H. Xu, J. Song, and Y. Lan, "Assessment of rice leaf blast severity using hyperspectral imaging during late vegetative growth," *Australasian Plant Pathology*, vol. 49, pp. 571–578, 2020.

[16] A. Feng, J. Zhou, E. D. Vories, K. A. Sudduth, and M. Zhang, "Yield estimation in cotton using uav-based multi-sensor imagery," *Biosystems Engineering*, vol. 193, pp. 101–114, 2020.

[17] M. Sharifi and X. Chen, "A novel vision based row guidance approach for navigation of agricultural mobile robots in orchards," in *2015 6th International Conference on Automation, Robotics and Applications (ICARA)*, 2015, pp. 251–255.

[18] J. Radcliffe, J. Cox, and D. M. Bulanon, "Machine vision for orchard navigation," *Computers in Industry*, vol. 98, pp. 165–171, 2018.

[19] P. Huang, L. Zhu, Z. Zhang, and C. Yang, "An end-to-end learning-based row-following system for an agricultural robot in structured apple orchards," *Mathematical Problems in Engineering*, vol. 2021, 2021.

[20] D. Aghi, S. Cerrato, V. Mazzia, and M. Chiaberge, "Deep semantic segmentation at the edge for autonomous navigation in vineyard rows," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 3421–3428.

[21] M. S. N. Kabir, M.-Z. Song, N.-S. Sung, S.-O. Chung, Y.-J. Kim, N. Noguchi, and S.-J. Hong, "Performance comparison of single and multi-gnss receivers under agricultural fields in korea," *Engineering in Agriculture, Environment and Food*, vol. 9, no. 1, pp. 27–35, 2016.

[22] S. Zaman, L. Comba, A. Biglia, D. R. Aimonino, P. Barge, and P. Gay, "Cost-effective visual odometry system for vehicle motion control in agricultural environments," *Computers and Electronics in Agriculture*, vol. 162, pp. 82–94, 2019.

[23] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, *et al.*, "Searching for mobilenetv3," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1314–1324.

[24] S. Cerrato, V. Mazzia, F. Salvetti, and M. Chiaberge, "A deep learning driven algorithmic pipeline for autonomous navigation in row-based crops," *CoRR*, vol. abs/2112.03816, 2021.

[25] M. Martini, A. Eirale, S. Cerrato, and M. Chiaberge, "Pic4rl-gym: a ros2 modular framework for robots autonomous navigation with deep reinforcement learning," *arXiv preprint arXiv:2211.10714*, 2022.