# Design and Implementation of Computer Vision based In-Row Weeding System

Xiaolong Wu, Stéphanie Aravecchia, Cédric Pradalier

# Design and Implementation of Computer Vision based In-Row Weeding System

Xiaolong Wu, Stéphanie Aravecchia, and Cédric Pradalier*

*Abstract*— **Autonomous robotic weeding systems in precision farming have demonstrated their full potential to alleviate the current dependency on herbicides or pesticides by introducing selective spraying or mechanical weed removal modules, thus reducing the environmental pollution and improving the sustainability. However, most previous works require fast weed detection system to achieve real-time treatment. In this paper, a novel computer vision based weeding control system is presented, where a non-overlapping multi-camera system is introduced to compensate the indeterminate classification delays, thus allowing for more complicated and advanced detection algorithms, e.g. deep learning based methods. The suitable tracking and control strategies are developed to achieve accurate and robust in-row weed treatment, and the performance of the proposed system is evaluated in different terrain conditions in the presence of various delays.**

## I. INTRODUCTION

Feeding a growing population and protecting the environment are two of the main challenges for the next few decades. Precision agriculture, also known as smart farming, addresses both challenges notably by targeting treatment only to plants that need it. Indeed, today's crop cultivation practices rely heavily on the uniform application of herbicides for weed control. In this scope, developing flexible robotic systems is essential to widespread precision agriculture. Within the EU-project Flourish [1], an autonomous weed control system has been designed, which can manage weeds either chemically with a precise spot-spraying of herbicide, or mechanically with a stamper tool destroying the weeds.

To precisely target the weeds, the first step is actually to separate the weeds from the value crops. Vision-based classification approaches have proven effective in this domain, is even capable enough to detect the stem of the weeds [2]. Such methods, using Convolutional Neural Network (CNN), introduce a new challenge which is the indeterminate delay between the time the image is captured to perform the classification and the time the results of the classification are available to trigger the actuators.

Our proposed system, in Fig. 1, addresses this by introducing a non-overlapping multi-camera tracking system that could reliably retrace the 3D locations of weeds using intra- and inter-camera template-based tracking. The adverse effect of the delays induced by the CNN-based classifier can be efficiently alleviated, and the predictive control layer is designed to provide an accurate prediction of trigger timing and locations for weed treatment tools. A mechanical stamp tool, as well as a selective sprayer, are implemented in

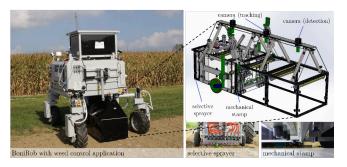*All authors are with the UMI2958 GeorgiaTech-CNRS, Metz, France. `firstname.lastname@georgiatech-metz.fr`

Fig. 1: Left: the Flourish BoniRob with our proposed weed control hardware module mounted at the bottom. Upper-Right: a CAD view of the module including the actuators and sensors. Bottom-Right: the example pictures of the selective sprayer and the mechanical stamping tool.

our weed control system to achieve fast and accurate weed treatment.

This paper makes the following contributions: 1) A novel self-sustained weed control system is designed, implemented, and extensively evaluated in the field for in-row weed removal; 2) a 3D multi-camera multi-object tracking algorithm using direct or indirect template matching is developed to provide fast and accurate weeds tracking across non-overlapping cameras.

## II. RELATED WORK

Robotic system for weed control has been in the field of research for a few decades [3]. Several methods to remove weeds adapted to robotics have been developed, either chemically or mechanically, for in-row weeding. Spraying of herbicides is a common weed-control method in traditional agriculture. Therefore, many robotics platforms performing selective spraying have been developed [4]. Significant effort has been made towards the in-row weeding [5] [6], however, the mechanical weed removal in the close-to-crop area is still challenging. One solution is to stamp the weeds with a metal rod with a diameter of 1 cm, proven effective in [7].

In recent years, significant progress has been made towards machine vision based real-time weed detection using learning-based methods. In [8], a mixture of small components of a fine-tuned deep conventional neural network (DCNN) is utilized to achieve accurate weed-crop classification, which could achieve a 90% precision with a 1.07-1.83Hz processing time. In [9], an encoder-decoder cascaded conventional neural network (CNN) is utilized to perform the dense semantic weed-crop classification, which achieves
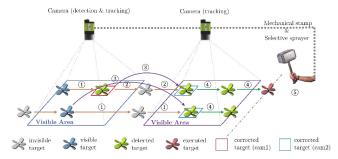
Fig. 2: Our proposed weed control system utilizes a multi-camera system to track the detected weeds using template matching techniques with uncertain detection delays, where the tracking can be naturally divided into intra- and inter-camera tracking layers. The tracked targets will be fed into a predictive control module to estimate the timing and position of treatment when approaching the weeding app.

a high classification precision but cannot guarantee a 2-5Hz real-time performance. In [10], a fully convolutional network (FCN) is formulated to encode the spatial information of plants in a row over sequentially acquired image sequence to perform the classification task, resulting in a stable 5Hz runtime performance . Other learning-based methods have been specialized on stem detection [11], [12], [2].

Concerning the tracking module, our proposed system implements a fixed-viewpoint down-looking camera and an entirely controlled lighting system, so that most of the state-of-the-art non-overlapping camera tracking methods are excessively complex for our application. For this reason, only applicable works are surveyed in this paper. In [13] [14] [15], Kalman filter based multi-camera tracking systems are developed to track multi-object in 2D image space. In [16], the particle filter based 3D model-based visual tracking algorithm utilizing lie group components is developed, where the both indirect and direct formulation are implemented to improve its robustness.

## III. SYSTEM DESIGN

In this section, the hardware description in Sec.III-A and software designs: the classification module in Sec. III-B, the multi-camera tracking module in Sec. III-C, and finally the control module in Sec. III-D, are briefly described.

### A. Hardware Description

*1) Mechanical design:* The weeding-unit, in Fig. 1, is the self-sustained module responsible for detecting and treating the weeds. The weeds are treated either mechanically by stampers or chemically by sprayers: 18 individually controlled stamps with 10mm-diameter bolts provide the mechanical weeiding, while 9 individually controlled nozzles deliver the chemical weeding. The actuators are controlled with a scalable Programmable Logic Controller (PLC using absolute timestamps, and the clocks (PC, PLC) are synchronized to fulfill the timing requirement for precision treatment.

*2) Perception system and calibration:* Three down-looking 4-channels JAI AD-130 GE cameras are mounting on our proposed system with the distance to the ground being approximatively 80cm. The ground resolution is approximatively 3 px/mm, and the field of view is 24 x 31 cm. To keep the same light conditions under the cameras the whole unit is covered by 3 mm acetal copolymer sheets, and artificial lights are mounted under to control the lighting. The RGB+NIR images are used for classification, and the RGB images only are used for tracking. Both the intrinsic and Inter-camera extrinsic parameters are estimated using OpenCV.

Three narrow-beam sonars (SRF235 Ultrasonic range finder), are mounted next to each camera to avoid doing the calibration every time the unit is pulled up and then down. The accuracy of the range measurement of these sonars is 1cm.

### B. Classification

The first step toward weed treatment is weed detection. In this project, the weed detection, published in [17], is a vision-based online classification of crops and weeds using RGB+NIR data. The runtime on the UGV computer (Intel i7 CPU, GeForce GTX-1080 GPU) goes up to 191ms from a mean of 124ms±14ms.

### C. Multi-Camera Tracking

The primary objective of our proposed multi-camera tracking algorithm is to keep updating the positions of all the detected weeds across multiple cameras with the existence of indeterminate classification delays to send the most possible weeds to the treatment layer.

*1) Multi-camera weed tracking using EKFs:* The proposed multi-camera multi-weed tracking algorithm is essentially a collection of EKFs with a cascade structure, which allows for the sequential update of object locations in Cartesian space. Each EKF in this framework could be seen as a single-camera model-based 3D visual tracking algorithm using edges or keypoints, which is so-called intra-camera tracking module. The inter-camera tracking module, in contrast, tries to project out-of-view objects on other camera planes, performing inter-camera template-based matching and eventually passes retraced object from one EKF to another.

Our system employs a pre-calibrated multi-camera system, so that we can have the camera intrinsic matrices $K_1, K_2, \ldots, K_N$ , lens distortion parameter $L_1, L_2, \ldots, L_N$, and camera extrinsic matrix $T_2^1, T_3^2, \ldots, T_N^{N-1}$ as the prior. The acquired images are rectified, normalized, and converted into gray-scale before feeding into our tracking layer.

Our proposed EKF algorithm for intra-camera tracking is defined in terms of its states $x_{k,i}$ , control inputs $u_{k,i}$, measurements $z_{k,i}$, where $i$ represents the object index, $k$ indicates the discrete time. The state vector and control input for the $i^{th}$ object can be described as

$$x_{k,i} = [P_{k,i}, s_{k,i}]^T \quad u_{k,i} = [\xi_{k+1}^k, \Delta s_{k,i}]^T \quad (1)$$

where $\boldsymbol{P}_{k,i} = [x_{k,i}, y_{k,i}, z_{k,i}]^T$ represents the 3D coordinates of the centroid of $i^{th}$ object in camera frame, $\boldsymbol{\xi}_{k+1}^k \in \mathfrak{se}(3)$ represents the 6DOF camera motion vector from $k$ to $k+1$ camera frame in Lie Group, and $\boldsymbol{s}_{k,i} = [h_{k,i}, w_{k,i}]^T$ $\Delta \boldsymbol{s}_{k,i} = [\Delta h_{k,i}, \Delta w_{k,i}]$ represent the size and the predicted change of search scope of $i^{th}$ object in template matching procedure, respectively.

The measurement vector $\boldsymbol{z}_{k,i}$ of each object can be written as

$$z_{k,i} = [\boldsymbol{p}_{k,i}, d_{k,i}]^T \tag{2}$$

where $\boldsymbol{p}_{k,i} = [u_{k,i}, v_{k,i}]^T$ represents the measured 2D coordinates of the centroid of the $i^{th}$ object in image space after template matching, and $d_{k,i}$ represent the measured depth of $i^{th}$ object from sonar system.

Following the process and measurement equation in EKF framework, our proposed system can be formulated as following:

$$\boldsymbol{x}_{k+1,i} = f(\boldsymbol{x}_{k,i}, \boldsymbol{u}_{k,i}) + \boldsymbol{w}_{k,i} = \begin{bmatrix} \boldsymbol{R}(\boldsymbol{\xi}_{k+1}^k) \boldsymbol{P}_{k,i} + \boldsymbol{t}(\boldsymbol{\xi}_{k+1}^k) \\ \boldsymbol{s}_{k,i} + \Delta \boldsymbol{s}_{k,i} \end{bmatrix} + \boldsymbol{w}_{k,i} \tag{3}$$

$$\boldsymbol{z}_{k,i} = h(\boldsymbol{x}_{k,i}) + \boldsymbol{v}_{k,i} = \begin{bmatrix} \pi(\boldsymbol{K}\boldsymbol{P}_{k,i}) \\ d_{k,i} \end{bmatrix} + \boldsymbol{v}_{k,i} \tag{4}$$

where $\boldsymbol{R}(\boldsymbol{\xi}) \in SO(3)$ $\boldsymbol{t}(\boldsymbol{\xi}) \in \mathbb{R}^3$ are the rotation matrix and the translation vector, $\pi(\cdot)$ represents a projection function that $\boldsymbol{p}_{k,i} = \pi(\boldsymbol{P}_{k,i}) = [x_{k,i}/z_{k,i}, y_{k,i}/z_{k,i}]^T$, and $\boldsymbol{w}_{k,i}$ $\boldsymbol{v}_{k,i}$ are Gaussian process and measurement noise with normal probability distribution, respectively.

For tracking applications, the established object motion model serves as a coarse initialization for later template-based matching, thus reducing search scope and processing time to achieve fast implementation. For our application, we utilize a simplified VO algorithm, assuming a flat ground surface, to generate $\boldsymbol{\xi}_{k+1}^k$, so that we can achieve high computational efficiency without compromising tracking accuracy. After propagation, template-based matching is performed for every object in the field of view with a search region defined by a projected center $\boldsymbol{p}_{k,i} = \pi(\boldsymbol{K}\boldsymbol{P}_{k,i})$ and a window size $\boldsymbol{s}_{k,i}$ in the EKF state vector. If successful, the corrected template center $\boldsymbol{p}_{k,i}$ combined with average depth $d_{k,i}$ measured from the sonar system, are used as new measurements, feeding into EKF correction process; otherwise, the control input set $\Delta \boldsymbol{s}_{k,i} > 0$ to be positive to increase the search scope for the next iteration.

If the object is already out of scope from the current camera, the inter-camera tracking passes the disappeared object to next EKF. The inter-camera tracking algorithm is, in essence, a simple template-based matching algorithm. Firstly, the out-of-view object is projected from current camera $m$ to next camera $n$, which can be described as

$$\boldsymbol{p}_n = \pi(\boldsymbol{K}_n \boldsymbol{T}_n^m (R(\boldsymbol{\xi}) \boldsymbol{P}_m + \boldsymbol{t}(\boldsymbol{\xi}))) \tag{5}$$

where $\boldsymbol{\xi}$ represents the estimated camera motion between two image time-stamp from unsynchronized camera $m$ and $n$. If the predicted center of object $\boldsymbol{p}_n$ enters the visible area

of camera $n$, an inter-camera template matching is carried out. If successfully matched, the state vector of this object is removed from the EKFs in camera $m$ and inserted into those from camera $n$ ; otherwise, we increase the search region of this object in the EKF in camera $m$ and try to match it at the next iteration.

It should be noted that we implement both the indirect (keypoints) and direct (edges) method – denomination from [16] – in motion propagation and correction for the intra- and inter-camera tracking modules. A detailed evaluation and comparison of their performances are presented in Sec. IV-B.

*2) Delayed initialization and pose propagation:* To deal with the unpredictable delays, a time-stamp indexed frame buffer is preserved in the memory to ease the later image searching and the pose propagation. Each frame in the buffer, indexed by the image capture time, contains the received gray-scale image, the estimated camera pose from the VO, and the average pixel depth from the sonar system. When a classification result arrives at the tracking layer, the corresponding frame is extracted from the buffer given its time-stamp. The 2D appearance-based templates of detected plants are extracted, and the centroid of the object is back-projected on Cartesian space to initialize the state vectors in the first EKF. Then their positions are propagated through the intra- and inter-camera tracking modules given inter-frame poses in the buffer. When the object has transformed to the latest camera frame and successfully matched to current image template, the distances between the new object to all known ones are calculated. If the shortest distance falls below a certain threshold, these two objects will be considered to be the same, and the state vectors are combined.

*3) Biased naive Bayesian classifier:* To protect the valuable plants among the weeds, an incremental naive Bayesian classifier with biased probabilistic model is utilized to filter out the false positives from weed detector. Consider conditional probability model $P(C_i|\boldsymbol{l}_i)$, with sequentially arrived classified label $\boldsymbol{l}_i = l_{i,1}, l_{i,2}, \ldots, l_{i,n}$ from detector and $C_i$ as the output label from our proposed classifier, the naive Bayesian filter formula can be written as $P(C_i|\boldsymbol{l}_i) \propto P(\boldsymbol{l}_i|C_i) P(C_i)$, where $P(C_i)$ is a class prior probability of an object belonging to weed or plant, and $P(C_i|\boldsymbol{l}_i)$ is conditional probability that is pre-defined with a probability model defined as follows:

| $P(\boldsymbol{l}_i, |C_i)$ | $l_i = weed$ | $l_i = plant$ |
|---|---|---|
| $C_i = weed$ | 0.8 | 0.2 |
| $C_i = plant$ | 0.1 | 0.9 |

It should be noted that the probability model is biased because destroying a valuable plant is always considered to be a much more serious problem than just failing to remove a weed from the user's point of view.

We follow a standard incremental implementation of this filter described in [18], and the classification results is finalized when all the classification results have been delivered to our tracking algorithm. At this time, the objects classified as plants are deleted from the corresponding EKF and are not tracked anymore.

### D. Control Design

Considering an extrinsic calibrated camera-tool system, we already have the homogeneous transformation matrix from last camera frame to the first row of the stamping tool $\boldsymbol{T}_{r1}^{cam}$. The $i^{th}$ object position $\boldsymbol{P}_{cam,i}$ in camera frame and the estimated velocity $\boldsymbol{v}_{cam}$ from VO system can be transformed to the coordinate system of the first-row stamping tool, so that the triggering timing $t_{pred}$ and predicted horizontal displacement $x_{r1,i}^{prime}$ can be calculated by solving the first two rows of the formulated linear system equation:

$$\boldsymbol{P}'_{r1,i} = \boldsymbol{T}(\boldsymbol{R}_{r1}^{cam}\boldsymbol{v}_{cam}t_{pred} + \boldsymbol{R}_{r1}^{cam}\boldsymbol{v}_{cam}t_{delay})\boldsymbol{T}_{r1}^{cam}\boldsymbol{P}_{cam,i} \quad (6)$$

where $\boldsymbol{R}_{r1}^{cam}$ is the rotation matrix extracted from $\boldsymbol{T}_{r1}^{cam}$, $T(\cdot)$ is the transformation function using 6-DOF motion prediction vector as input, $t_{delay}$ is the measured processing delay, and the $\boldsymbol{P}'_{r1,i} = [x_{r1,i}^{prime}, 0, z_{r1,i}^{prime}, 1]^T$ is the predicted position of object.

It should be noted that the accuracy achieved by our proposed open-loop predictive control strategy can be easily affected by the whole treatment process delay, introduced by PLC latency and execution tool dynamics. With our high-performance PLC, the accumulated processing delay can be simplified as a constant delay parameter $t_{delay}$, which is measured and validated by experiments in the laboratory.

## IV. EXPERIMENTS

The experimental evaluation section is designed to illustrate the performance of our proposed weed control system in various real field conditions with indeterminate delays.

### A. Field Robot

The BoniRob is a multi-propose field robot designed by BOSCH DeepField Robotics, which provides a slot to install various tools for specific tasks. In recent years, a variety of agriculture applications have been successfully developed and validated using this robot, such as selective spraying, mechanical weed control, and weed monitoring. In this work, our proposed weed control system is mounted on BoniRob, operating in various real field condition to evaluate our proposed method.

### B. Evaluation of Multi-Camera Tracking

The proposed multi-camera tracking performance is extensively evaluated regarding accuracy, robustness, and runtime performance in the real field. We both compare the indirect and direct methods in our proposed framework, as well as evaluate the effect of terrain conditions, the density of plants, and various delay introduced by classification. A simple vegetation detection classifier provides the ground truth centroid positions of the plants.

*1) Different terrain condition:* The experiments presented here are designed to analyze the intra- and inter-camera tracking accuracy and robustness quantitatively in various field conditions, where a 0.2m/s vehicle speed and a simple vegetation detection algorithm with a 0.138s average classification delay are chosen. In Fig. 3, a detailed comparison

between the direct and indirect tracking method in four typical terrain conditions is presented.

From the comparison of the two methods in our proposed tracking framework, we can generalize that (1) the direct method shows marginally better tracking precision in motion propagation phase than that of indirect method in all terrain conditions, (2) the indirect method holds a larger convergence radius, with a similar tracking accuracy, compared with direct approach in the context of template-based matching, (3) both methods works slightly better in intra-camera tracking than that in inter-camera tracking even though the lighting condition is fully controlled, and (4) the overall tracking accuracy is mainly limited by the template matching precision rather than that of motion propagation, which can better explain why all VO algorithms showed similar tracking precision in the end in Sec. **??**. It should be clarified that keeping motion propagation error small is still critical for system robustness since it is tightly related with the tracking error after a blind area between cameras, which define the initial displacement of template matching.

Analyzing the effect of the terrain roughness and the density of plants on our tracking framework, we can observe that the motion propagation precision and convergence radius of template matching is correlated with the number of good features or edge pixels in image, where both the rough terrain and the plants can provide such features and edges for motion estimation and template matching. In contrast, the degree of ground flatness plays a minor role in tracking with the down-looking camera setup.

*2) Different classification delay:* In this paper, one of our major claims is that our system could work with a much higher range of classification delays without compromising overall tracking accuracy and efficiency compared with previously proposed weed control systems. To support our claim, the effect of classification delays on our proposed tracking system are studied in term of the tracking accuracy and runtime performance. In this evaluation, a simple vegetation detection algorithm with a fake classification delay generator is utilized to examine our tracking performance. The vehicle speed for this test is set to be 0.1m/s, thus allowing to test classification with more significant delays. In Fig. 4, we can find that the higher classification delay would request higher computational effort, which is caused by the template matching for every three propagation to ensure the convergence, but wouldn't compromise the overall tracking precision.

*3) Different vehicle speed:* To examine the system capabilities, the tracking accuracy, both before and after template matching, are evaluated in different vehicle velocities in the real field, with the simple vegetation detection algorithm. It should be noted that the maximum allowed speed could only reach up to $0.4m/s$ for safety reasons, where the results for higher velocities are simulated by sampling recorded images to study the limitation of our proposed system. In Fig. 4, the motion propagation error increases exponentially after a certain speed, which is attributed to the reduced overlap area between sequential images. However, the tracking error after
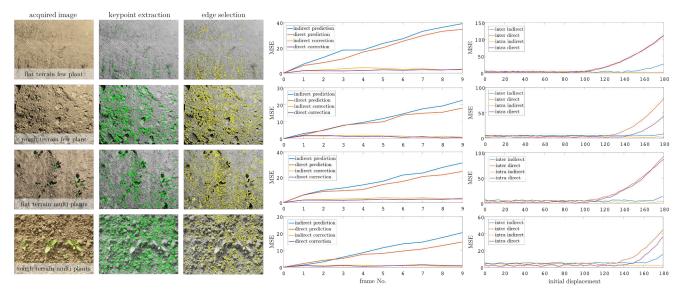
Fig. 3: Left: Sample images acquired in four typical real field condition and their keypoints and edges extraction results are presented. Middle: the image space tracking error with and without template matching correction, using direct and indirect methods, are plotted versus the tracked frame number. Right: the intra- and inter-camera tracking error using direct and indirect approaches are plotted versus initial displacement of the reference template to study the convergence radius of template matching.
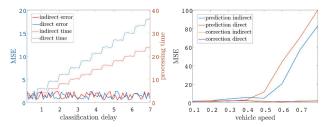


Fig. 4: Left: the MSE [pixel] and processing time [*ms*] versus various simulated classification delays [*s*] in the delayed initialization phase are plotted, where the template matching is performed after every three propagation to maintain the real-time performance. Right: the overall tracking MSE before and after template matching versus vehicle speeds are plotted to study the effect of vehicle speed on our proposed tracking system.



Fig. 5: Left: sample pictures of leaves after mechanical weed removal. Right: sample pictures after selective spraying.

template matching is not highly affected due to the adaptive searching area selection strategy described in Sec. III-C.1.

### C. Evaluation of Predictive Control

The proposed predictive control module is evaluated on both the mechanical weeding tool and the selective sprayer in a real-world field with flat and rough terrain areas. The simple vegetation detection algorithm utilized to generate classification results.

For tests of the mechanical weed removal, real leaves with an average radius of 10mm are chosen as targets, and we manually count the successful stamping rate after execution as the performance metric. The stamping evaluation is performed in short paths to control the test condition in an outdoor environment better. For short-path tests, we use 5-10

targets per meter and repeat such test for 10 times for each test speed.

For experiments of selective spraying, we set up a webcam to monitor targets after spraying, taking advantage of the wet area is visible on the dry field surface. Due to this simple monitoring approach, the evaluation of selective spraying can be done with a full-row operation by counting the successful execution rate manually afterward from the recorded video. The example pictures of the test field and targets after successful treatment are presented in Fig. 5, and the quantitative experimental results are provided in Table I. The successful treatment rates regarding different vehicle speeds and terrain roughness are presented in the table, where the leaves with an average 10mm radius are chosen as targets. We don't have the experimental evaluation of mechanical weeding with higher speed than 0.2 m/s because the higher driving speed of the vehicle may introduce a huge tangential force on the stamper, which may lead to the malfunction of the stamping tool. From the table, we can observe that the successful treatment rate is almost invariant with speed in both flat and

TABLE I: Successful Treatment Rates

| vel. [m/s] | 0.05 | 0.1 | 0.2 | 0.3 | 0.4 |
|---|---|---|---|---|---|
| | *stamping* | | | | |
| *flat* | 112/113 | 99/102 | 90/91 | - | - |
| *rough* | 120/121 | 84/86 | 93/96 | - | - |
| | *spraying* | | | | |
| *flat* | 125/125 | 124/124 | 129/129 | 132/134 | 112/113 |
| *rough* | 121/121 | 111/112 | 118/119 | 131/133 | 129/129 |

TABLE II: The Weed Classification and Treatment Rate

| module | detection | b. classifier | tracker | treatment |
|---|---|---|---|---|
| | *spraying in real field* | | | |
| *0.1m/s* | 199/182 | 182/182 | 182/182 | 182/182 |
| *0.2m/s* | 193/182 | 181/182 | 181/182 | 181/182 |
| | *stamping in simulated test* | | | |
| *0.1m/s* | 92/58 | 57/58 | 57/58 | 57/58 |
| *0.2m/s* | 88/58 | 57/58 | 57/58 | 57/58 |



Fig. 6: Left: an in-row mechanical weeding and spraying test is conducted in the real-world field in Ancona, IT. The weeds after spraying can be seen in the picture, and the mechanical treatment results are not visible on the ground. Right: the performance of in-row mechanical weeding is presented in a simulated test environment in Bonn, GE. The faked targets made of leaves attached to modeling clay is utilized to visualize the execution results.

rough field ground.

### D. Evaluation of In-Row Weed Control

The purpose of the in-row weed control evaluation is to analyze the whole system performance, especially the classification, and final treatment rates. The overall performance evaluation, named in-row weed control, is performed in the real field and simulated environment. The reason why we introduce the extra experiment in a simulated environment is that (1) the extensive test on the real field with mechanical tool is not very welcome by farmers due to the potential failure to destroy of the plants, (2) after some trials we found that the stamping holes after real field weeding are not really visible on leaves and cannot easily be distinguished with other holes on ground, and (3) this extra-test on flat terrain at the same time tests our classification results and provides us with a baseline performance, which can help us to better understand the limitation of our proposed system.

The example pictures for in-row weeding test condition, both in the real field and simulated environment, with representative results, are shown in Fig. 6. The quantitative evaluation of the classification rate from the detector, Bayesian classifier, and tracker, as well as the successful treatment rate, from each layer of the proposed system, is recorded in Table II. The classification rate from the detector, Bayesian classifier, and multi-camera tracker, as well as the final treatment rates, regarding two typically set vehicle speed, are presented in the table, where the real weeds serve as targets in the real field, and the leaves serve as fake targets in simulated tests. From the table, we can find that our proposed Bayesian filter could effectively remove the false positives from the detector, and the proposed weed control system that integrates classifier, tracker and weeding machine could perform reliable treatment for in-row weeds with good precision. From the real field tests, we observe that the stamping solution for weed removal is, in general, less efficient than using the selective sprayer, which can be attributed to their hard-to-locate stem positions and the hard-to-compensate uncertainty induced by the vibration while driving on the unstructured terrain.

## V. CONCLUSION

A novel computer vision based weeding control system is designed, implemented and evaluated. A non-overlapping multi-camera system is introduced to compensate the indeterminate classification delays induced by modern CNN-based detection algorithms, and a 3D multi-camera multi-object 3D tracking algorithm using EKF is developed to provide high-precision tracking results across cameras. In our proposed framework, the direct or indirect template matching methods are implemented and compared to justify our system design. A biased naive Bayesian filter is designed to remove the false positives from the detector. To adopt an operation-while-driving strategy, both low- and high-level control strategies are deliberately designed for high-precision weed removal. The tracking and control performance of the proposed system is extensively evaluated in different terrain conditions regarding various classification delays and vehicle speeds, and the final in-row weed removal performance is also assessed to validate our claim that our system can provide accurate and reliable in-row weed removal service in the real field.

### REFERENCES

[1] "The flourish project," http://flourish-project.eu/.
[2] P. Lottes, J. Behley, N. Chebrolu, A. Milioto, and C. Stachniss, "Joint stem detection and crop-weed classification for plant-specific treatment in precision farming," *arXiv preprint arXiv:1806.03413*, 2018.
[3] A. Bechar and C. Vigneault, "Agricultural robots for field operations. part 2: Operations and systems," *Biosystems Engineering*, vol. 153, pp. 110–128, 2017.
[4] "Ecorobotix," http://www.ecorobotix.com/e.
[5] "Naio dino large-scale vegetable weeding robot," https://www.naio-technologies.com/en/agricultural-equipment/large-scale-vegetable-weeding-robot/.
[6] N. Tillett, T. Hague, A. Grundy, and A. Dedousis, "Mechanical within-row weed control for transplanted crops using computer vision," *Biosystems Engineering*, vol. 99, no. 2, pp. 171–178, 2008.

[7] A. Michaels, S. Haug, and A. Albert, "Vision-based high-speed manipulation for robotic ultra-precise weed control," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*. IEEE, 2015, pp. 5498–5505.

[8] C. McCool, T. Perez, and B. Upcroft, "Mixtures of lightweight deep convolutional neural networks: applied to agricultural robotics," *IEEE Robotics and Automation Letters*, vol. 2, no. 3, pp. 1344–1351, 2017.

[9] I. Sa, Z. Chen, M. Popović, R. Khanna, F. Liebisch, J. Nieto, and R. Siegwart, "weednet: Dense semantic weed classification using multispectral images and mav for smart farming," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 588–595, 2018.

[10] P. Lottes, J. Behley, A. Milioto, and C. Stachniss, "Fully convolutional networks with sequential information for robust crop and weed detection in precision farming," *arXiv preprint arXiv:1806.03412*, 2018.

[11] S. Haug, P. Biber, A. Michaels, and J. Ostermann, "Plant stem detection and position estimation using machine vision," in *Proceedings of the International Workshop on Recent Advances in Agricultural Robotics (RAAR2014)*, 2014.

[12] F. Kraemer, A. Schaefer, A. Eitel, J. Vertens, and W. Burgard, "From plants to landmarks: Time-invariant plant localization that uses deep pose regression in agricultural fields," *arXiv preprint arXiv:1709.04751*, 2017.

[13] A. Chilgunde, P. Kumar, S. Ranganath, and W. Huang, "Multi-camera target tracking in blind regions of cameras with non-overlapping fields of view." in *BMVC*, 2004, pp. 1–10.

[14] X. Li, K. Wang, W. Wang, and Y. Li, "A multiple object tracking method using kalman filter," in *Information and Automation (ICIA), 2010 IEEE International Conference on*. IEEE, 2010, pp. 1862–1866.

[15] C.-T. Chu, J.-N. Hwang, K.-M. Lan, and S.-Z. Wang, "Tracking across multiple cameras with overlapping views based on brightness and tangent transfer functions," in *Distributed Smart Cameras (ICDSC), 2011 Fifth ACM/IEEE International Conference on*. IEEE, 2011, pp. 1–6.

[16] C. Choi and H. I. Christensen, "Robust 3d visual tracking using particle filtering on the special euclidean group: A combined approach of keypoint and edge features," *The International Journal of Robotics Research*, vol. 31, no. 4, pp. 498–519, 2012.

[17] P. Lottes and C. Stachniss, "Semi-supervised online visual crop and weed classification in precision farming exploiting plant arrangement," in *Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International Conference on*. IEEE, 2017, pp. 5155–5161.

[18] K. P. Murphy *et al.*, "Naive bayes classifiers," *University of British Columbia*, vol. 18, 2006.