



HAL
open science

Robotic weed control using automated weed and crop classification

Xiaolong Wu, Stéphanie Aravecchia, Philipp Lottes, Cyrill Stachniss, Cédric Pradalier

► **To cite this version:**

Xiaolong Wu, Stéphanie Aravecchia, Philipp Lottes, Cyrill Stachniss, Cédric Pradalier. Robotic weed control using automated weed and crop classification. 2020. hal-02484462

HAL Id: hal-02484462

<https://hal.science/hal-02484462v1>

Preprint submitted on 19 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Robotic Weed Control Using Automated Weed and Crop Classification

Xiaolong Wu

Georgia Institute of Technology
Department of Electrical and Computer Engineering
Atlanta, GA 30332, United States
xwu@gatech.edu

Stéphanie Aravecchia

UMI2958 GeorgiaTech-CNRS
Metz 57070, France
stephanie.aravecchia@georgiatech-metz.fr

Philipp Lottes

University of Bonn
Department of Photogrammetry
Bonn 53115, Germany
philipp.lottes@uni-bonn.de

Cyrill Stachniss

University of Bonn
Department of Photogrammetry
Bonn 53115, Germany
cyrill.stachniss@uni-bonn.de

Cédric Pradalier

UMI2958 GeorgiaTech-CNRS
Metz 57070, France
cedric.pradalier@georgiatech-metz.fr

Abstract

Autonomous robotic weeding systems in precision farming have demonstrated their full potential to alleviate the current dependency on agrochemicals such as herbicides and pesticides, thus reducing environmental pollution and improving sustainability. However, most previous works require fast and constant-time weed detection systems to achieve real-time treatment, which forecloses the implementation of more capable but time-consuming algorithms, e.g. learning-based methods. In this paper, a non-overlapping multi-camera system is applied to provide flexibility for the weed control system in dealing with the indeterminate classification delays. The design, implementation, and testing of our proposed modular weed control unit with mechanical and chemical weeding tools are presented. A framework that performs naive Bayes filtering, 3D direct intra- and inter-camera visual tracking, and predictive control, while integrating state-of-the-art crop/weed detection algorithms, is developed to guide the tools to achieve high-precision weed removal. The experimental results show that our proposed fully operational weed control system is capable of performing selective mechanical as well as chemical in-row weeding with indeterminate detection delays in different terrain conditions and crop growth stages.

1 INTRODUCTION

Over the past century, weed control has been a long-standing issue in the field of agriculture. The uniform application of herbicides has demonstrated its effectiveness at weed removal, however, also introducing environmental pollution, human health, and herbicide resistance concerns. Due to its adverse effects, governments and farmers seek to reduce the herbicide input (Hillocks, 2012) in agricultural activity. Precision farming provides a way to solve this challenge by involving weeding mechanisms (Weis et al., 2008) to perform the treatment on an individual plant-level or a small

weed cluster. In doing so, the use of agrochemicals can be drastically reduced or even eliminated. However, human-oriented precision weeding machinery usually requires inefficient and labor-intensive human resources, which cannot justify the economic benefits of herbicide savings.

Automated weed control, including weed detection and removal, has gained significant popularity in the community of precision farming over recent years (Fennimore et al., 2016) (Bechar and Vigneault, 2017), due to its great potential to improve the weeding efficiency while reducing the environmental and economic costs. Many robotic weed control systems have been proposed with focuses primarily on single tactics (Slaughter et al., 2008): selective chemical spraying (Lee et al., 1999), mechanical weeding (Pannacci et al., 2017), flaming (Datta and Knezevic, 2013), and electrical discharging (Vigneault and Benoît, 2001) (Blasco et al., 2002) (Xiong et al., 2017). However, the study (Kunz et al., 2018) indicates that a combination of tactics could maximize weeding performance, named integrated weed management (Chikowo et al., 2009) (Young, 2018). Such a system inherently enables the use of alternative weed destruction based on specific weed species in the field, thus further improving the weeding efficiency. The Flourish platform's weed control system, partially described in this paper, follows the same strategy that is capable of performing weed-specific management either chemically with a precise spot-spraying of herbicide, or mechanically with a stamper tool destroying the weeds.

Smart weeding machines rely on the performance of the machine vision system to detect weeds. However, the environmental uncertainties, including illumination condition and color variance of leaves or soil, affect the performance of the machine vision system, thus upper-bounding the weed control accuracy. Along with the flourish of artificial intelligence, significant progress has been made towards learning-based weed detection methods. Such methods, using Convolutional Neural Network (CNN), have proven to provide more reliable crop/weed detection results. For a single-camera system, the high variance of detection delays could directly lead to target missing or limited time for subsequent weed tracking and actuation procedures, thus introducing significant uncertainties into weed control applications. The Flourish system addresses this challenge by introducing a non-overlapping multi-camera system, where the weeds are classified in the first camera and tracked in subsequent cameras, thus providing more time for weed detection without compromising the weeding performance.

Non-overlapping multi-camera tracking can be generally divided into intra- and inter-camera tracking, where the intra-camera part tracks the weeds while mapping their 3D locations in real-time and the inter-camera part seeks to retrace the previously detected weeds. In the field of intra-camera tracking, the indirect method (Klein and Murray, 2007) (Strasdat et al., 2010) (Mur-Artal and Tardós, 2017) has dominated the research field for a long time since the self-recognized features can provide considerable robustness to both the photometric noise and geometric distortion in images. However, recent works have shown that the direct methods (Newcombe et al., 2011) (Pizzoli et al., 2014) (Engel et al., 2014) (Engel et al., 2018) could provide more accurate and robust motion estimation due to their high flexibility of image information usage in comparison to the indirect approaches using certain types of features. To this end, our intra-camera tracking implements direct formulation to achieve better tracking precision and robustness.

However, the conventional direct tracking algorithm cannot provide satisfying performance for inter-camera tracking due to the inter-camera lighting changes (Wu et al., 2019). In general, the direct formulation compares the intensity value of pixels over a local patch across images by making the brightness constancy assumption. However, the inter-camera lighting difference breaks such an assumption, thus resulting in significant performance degradation. To achieve good tracking performance under lighting changes, the illumination-invariant tracking algorithms propose to take the illumination differences into their optimization frameworks, which provides feasible ways to perform reliable inter-camera tracking.

This paper presents the design, implementation, and evaluation of our proposed multi-camera weed management system for in-row weed removal through mechanical intervention or selective chemical spraying. The paper focuses on two key developments: (1) the mechanical weed unit design consisting of a multi-camera perception system, two mechanical stamper modules, and one spot-spraying module and (2) the weed control framework integrating computer-vision-based weed detection, direct 3D multi-camera weed tracking, and predictive actuation. Fig. 1 shows the workflow of our proposed system undertaking weed management. Both the crops and weeds are first detected and classified by a state-of-the-art learning-based algorithm using camera images, where the delayed classification results are propagated and retraced in the latest image. All the labeled crops and weeds are tracked and associated with newly

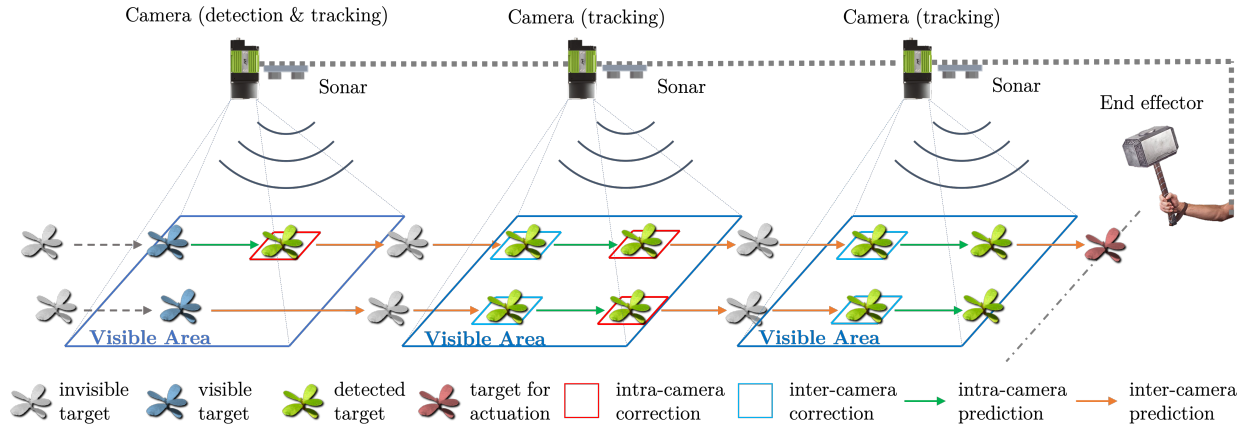


Figure 1: Our proposed weed control system uses a non-overlapping multi-camera system to track the detected weeds using direct matching techniques with high-variance detection delays. For the small delays, the weed (up-row) can be retraced in the detection camera and tracked until treated; for large delays, the weed (down-row) can only be retraced in the next camera or even later, which typically fails in single-camera weed control systems. The weeds are tracked across cameras and finally fed into a predictive control module to estimate the timing and position of treatment when approaching the weeding tools.

classified plants until moving out of the scope of the first camera. Each plant is finally fed into a biased naive Bayes filter to rule out the false positives further, and only the correct weeds are tracked across cameras until treated. Both the direct intensity-based intra-camera tracking and illumination-robust inter-camera tracking are developed to achieve high-precision pose estimation of tracked weeds, and a standard predictive control layer is implemented to perform weed intervention by using mechanical or/and chemical tools based on the size of weeds from the perception system. Results from the field tests demonstrate that our proposed weed management system can successfully distinguish weeds and crops and removes the weeds with an accuracy of over 90%, which is independent of the classification delays, terrain conditions, and crop growth stages.

This paper extends our previous work presented in (Wu et al., 2019) by proposing two essential modifications to boost the intra- and inter-camera tracking accuracy and robustness. (1) Instead of implementing an EKF-based object tracker using 2D-2D template matching, we developed a non-keyframe simultaneous tracking and mapping algorithm that allows for the 3D-2D object matching strategy, which improves both the intra- and inter-camera tracking performance under high depth-variance environments, such as fields with matured crops. To support this claim, the newly collected dataset that contains images with high-density matured sugar beets is introduced to test our proposed weed management system under different crop growth stages. (2) An illumination-robust cost is involved in our inter-camera tracking layer to compensate for the object appearance changes due to the inevitable lighting condition difference across cameras, thus improving the inter-camera tracking precision of our proposed system.

The subsequent sections of the paper are organized as follows: Sec.2 reviews the state-of-the-art works in the field of weed control system, learning-based weed classification, and 3D tracking; Sec.3 describes the integrated mechanical design of our weed control unit and the intrinsic and extrinsic calibration procedures of our non-overlapping camera system; Sec.4 details our proposed weed control framework incorporating weed detection and classification, multi-camera tracking, and predictive control; Sec.5 contains the experimental results in real field under different test conditions; Sec.6 concludes with a discussion on the lessons learned.

2 RELATED WORKS

In this section, both the integrated robotic weed control systems in Sec. 2.1 and the main components of our proposed weed control framework, such as learning-based weed detection and classification module in Sec. 2.2 and the intra-

and inter-camera tracking algorithms in 2.3 are surveyed to support our design choices.

2.1 Robotic Weed Control System

Nowadays, selective chemical spraying is still one of the most promising mechanisms to kill weeds. (Lee et al., 1999) proposed a prototype selective spraying system using computer vision and a precise herbicide application to remove the weeds. Although the system merely provided a 47.6% successful spraying rate, the reduction of herbicide use could reach up to 97%. (Lamm et al., 2002) tested the prototype described above with a computer vision algorithm that is capable of distinguishing grass-like weeds in the cotton field and sprayed 88.8% weeds and 21.3% crops with 0.45 m/s operation speed. To improve the accuracy of the system. (Steward et al., 2002) developed an advanced control strategy that achieved a 91% hit accuracy with an operation speed ranging from 0.9 to 4.0 m/s. (Nieuwenhuizen et al., 2010) designed a weed control system using a five-nozzle sprayer with 0.2m working width for efficiency, and it achieves an average 14mm error in the longitudinal direction and 7.5mm error in the transverse direction with 0.8m/s driving speed. To enable in-row weeding, (Søgaard and Lund, 2005) and (Søgaard and Lund, 2007) investigated the first prototype using a 126-mm-wide micro-spray and adopted a plan-then-go strategy to allow for a time-consuming image processing algorithm to generate high-precision target position estimation, which achieved 2.8mm average error with 0.2m/s driving speed. Later, (Midtby et al., 2011) demonstrated a system using a micro-sprayer revised by an inkjet printer head, which sprayed only weeds with a 94% success rate. (Urdal et al., 2014) proposed a drop-on-demand weed control system that focused on the fluid dynamics and electronics design of the droplet dispensing without experimental evaluation. Both the micro-sprayer and the droplet system was designed to improve the treatment precision of individual weeds, however, also sacrificing the region coverage. In (Underwood et al., 2015) and (Reiser et al., 2017), a fine spray nozzle attached to a mobile manipulator was developed to perform high-precision spraying without compromising the working area. Finally, (Blue River Technology, 2018) is commercializing a solution with robotic nozzles targeting weeds in cotton fields.

More and more studies have been concentrated on mechanical weeding applications, which eliminates the chemicals on the field (Pannacci et al., 2017). (Tillett et al., 2002) mounted automatically guided hoes on the rear of the tractor to perform the inter-row weeding at high speed, relying on real-time crop row detection. (Griepentrog et al., 2007) (Nørremark et al., 2008) introduced a real-time kinetics global positioning system (RTK-GPS) and a pre-recorded crop map to perform the localization and guide the hoes to perform intra-row weeding with centimeter-level accuracy. (Åstrand and Baerveldt, 2002) (Van der Weide et al., 2008) (Tillett et al., 2008) (Gobor, 2013) incorporated a computer system to guide the rotary hoes to achieve selective in-row weed removal: the experimental results showed a successful removal rate up to 53%. (Naio, 2018) is commercializing a small autonomous robot, with multiple tools, performing intra-row weeding with a hoe or a brush and in-row weeding with a spring, this last tool being adapted only to specific grown-up plants. To improve the treatment of the close-to-crop area, (Langsenkamp et al., 2014) (Michaels et al., 2015) proposed an alternative tool that has been evaluated in the greenhouse and the real-field condition: a metal rod with a diameter of 10mm hitting the growth center of the weed. This tool has proven effective to inhibit dicot weeds growth (broadleaves) but by design this tool fails to remove weeds with a fibrous root system such as grass, as monocot weeds cannot be damaged by damaging the leaves. This tube-stamp tool has been successfully implemented on the BoniRob robot platform with a visual servoing strategy: high-speed image processing from a camera implemented next to the weeding tool allowed to position it precisely, with a millimeter-level accuracy.

The robotic weed control system with multiple weeding tools named integrated weed management system (Chikowo et al., 2009) (Young, 2018), has also been reported (Kunz et al., 2018), aiming at maximizing the weed treatment efficiency and success rate using a combination of tactics. (Bawden et al., 2017) proposed a heterogeneous weed management system that is capable of selectively applying a mechanical or chemical control method based on the species of the weeds. (Chang and Lin, 2018) developed a small-scale agricultural robot that can automatically weed and perform variable rate irrigation within a cultivated field. Our proposed weed control system also implements such a strategy, which installs the mechanical stampers and chemical sprayer onboard for smart weed management.

2.2 Learning-Based Weed Detection and Classification

As explained before, significant progress has been made towards learning-based weed and crop detection methods. In (Lottes et al., 2016), we developed a system that performs vegetation detection, object-based feature extraction, random forest classification, and smoothing through a Markov random field to accurately classify sugar beet and weed. The object-based approach neglecting Markov smoothing achieves 1-2Hz in runtime performance, and a more accurate keypoint-based version runs merely at 0.60-1.26Hz. In (Lottes and Stachniss, 2017), we exploit the domain-specific knowledge of crop the spatial arrangement to perform semi-supervised online visual crop and weed classification with 5.23-8.06Hz runtime performance. In (McCool et al., 2017), a mixture of small components of a fine-tuned deep convolutional neural network (DCNN) is utilized to achieve accurate weed-crop classification, which could deliver a 90% precision with 1.07-1.83Hz processing time. In (Sa et al., 2018), an encoder-decoder cascaded convolutional neural network (CNN) is utilized to perform the dense semantic weed-crop classification, which achieves a high classification precision but cannot guarantee a stable real-time performance (2-5Hz). In (Lottes et al., 2018b), we formulate a fully convolutional network (FCN) to encode the spatial information of plants in a row over sequentially acquired image sequence to perform the classification task, resulting in a better runtime performance (5Hz).

By aiming at treating the plant's stem location with high precision intervention methods such as mechanical stamping, the detection system also has to provide the exact stem location to the robotic system. Besides the semantic segmentation algorithms, learning-based stem localization algorithms have also been developed to facilitate weed intervention. In (Haug et al., 2014), a keypoint-based random forest is trained to predict the stem regions of plants. In (Kraemer et al., 2017), the FCN based method is formulated to perform this task. In (Lottes et al., 2018a), we developed an end-to-end CNN to jointly learn the class-wise stem location and the pixel-wise semantic segmentation of weeds and crops.

2.3 Intra- and Inter-Camera Tracking

Monocular simultaneous localization and mapping (SLAM) and visual odometry (VO) are two major computer vision approaches to solve the camera tracking and environmental mapping issue, which can be generally divided into two categories: indirect and direct methods. Indirect methods, (Klein and Murray, 2007) (Mur-Artal and Tardós, 2017), pre-process acquired an image to extract a certain type of features, which is utilized to recover the camera pose and scene structure by minimizing geometric error. Unlike indirect approaches relying on features, direct methods directly use intensities of the image to estimate camera motion and scene structure rather than features. For direct dense mapping, (Stühmer et al., 2010) (Newcombe et al., 2011) (Pizzoli et al., 2014), the photometric error with spacial geometry prior formulations are developed to reconstruct a dense depth map and corresponding camera motion accurately. However, the real-time performance of such methods can only be guaranteed by introducing modern, powerful GPU due to their intensive computational load. To ease the computational demand, the semi-dense methods, (Engel et al., 2013) (Engel et al., 2014), reconstruct the inverse depth map through pixel-wise small-baseline stereo comparisons instead of joint optimization, enabling real-time CPU implementation. More recently, a direct sparse formulation (Engel et al., 2018) was proposed to jointly optimize motion and structure by omitting the smoothness prior, which makes real-time performance feasible. In recent years, direct approaches have attracted more attention due to their higher accuracy and robustness in both tracking and reconstruction, compared with indirect methods. The main advantage of the direct formulation is that the scene points are not required to be self-recognizable, thus allowing for more complete usage of information in an image. As a result, such characters of direct formulation makes it quite suitable for our intra-camera tracking application.

Weed tracking across non-overlapping cameras, named inter-camera tracking, mainly concentrates on object occlusion, cross-camera appearance ambiguities, and illumination changes. However, our proposed system implements fixed-viewpoint down-looking cameras with an artificial lighting system, so that only the illumination difference between cameras needs to be carefully treated. In recent years, a number of illumination-robust VO and SLAM systems have been proposed, implementing various models or descriptors to alleviate the adverse effect of external lighting changes, to achieve robust tracking across cameras. To gain robustness against the global illumination changes, either the median value of pixel residuals (Meilland et al., 2011) (Gonçalves and Comport, 2011) (Bloesch et al., 2015) (Greene et al., 2016) or an affine brightness transfer function (Klose et al., 2013) (Engel et al., 2015) is estimated to compensate

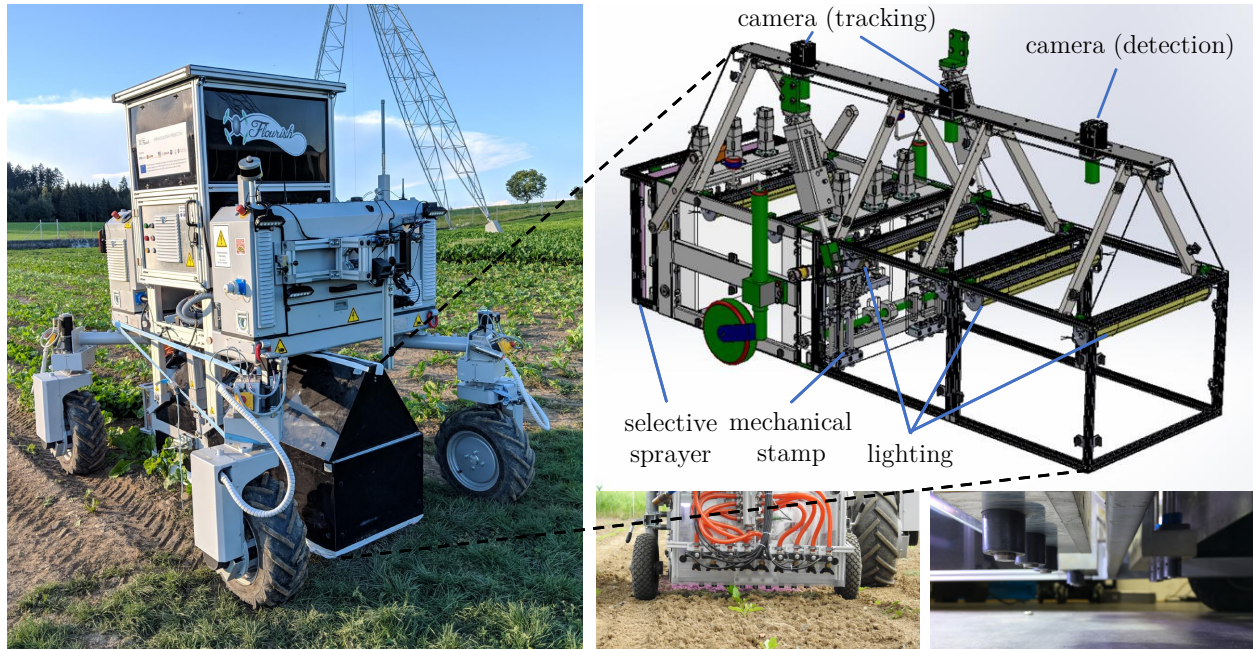


Figure 2: Left: the robotic platform Flourish BoniRob with our proposed weed control unit mounted at the bottom. Upper-Right: a CAD view of the unit including the actuators and sensors. Bottom-Right: the pictures of the selective sprayer and the mechanical stamping tool.

for the induced adverse effect in the optimization. For local lighting changes, (Dai et al., 2017) proposes to use image gradients, rather than pixel intensities, to formulate the direct energy function, thus gaining local lighting invariance. (Crivellaro and Lepetit, 2014) relies on the dense computation of a deliberately designed local descriptor to obtain a clear global minimum in energy function while preserving convergence basin by convolving with a low-pass filter. The methods based on the census transform (Alismail et al., 2016) (Alismail et al., 2017) use a binary descriptor to achieve local illumination invariance during the motion estimation. Based on the analysis from our previous work (Wu and Pradalier, 2019), we found the affine-model-based approach shows both good accuracy and large convergence basin, which can be used to formulate our inter-camera tracking algorithm.

3 WEED CONTROL UNIT

The weed control unit of the Flourish system, Fig. 2, incorporates an integrated weed removal module developed in Sec. 3.1 and a multi-camera perception system developed in Sec. 3.2. The main design objectives of this weed control unit are high weed throughput, precise treatment, and flexibility. To reach these goals, the weed control unit uses multiple actuators and operates them while driving. The weeds are treated mechanically with two ranks of stampers or chemically with one rank of sprayers. The weeds are detected and tracked with three cameras with non-overlapping fields of view. As a first prototype, the module is designed with a width of 400mm , which can be extended to cover more horizontal space in future iterations. The decision on which tool is used on which weed in this first prototype is only based on a size criterion: large weeds are sprayed while small weeds are stamped.

3.1 Integrated weed removal module

Equally to usual agricultural machines, our proposed weed removal module shown in Fig. 2, is hung up on a parallelogram, whose total length is 1980mm . This technique ensures a horizontal position on every adjusted height of the unit, which is crucial to ensure high precision of the actuation. Two pneumatic cylinders are pushing the unit in topmost

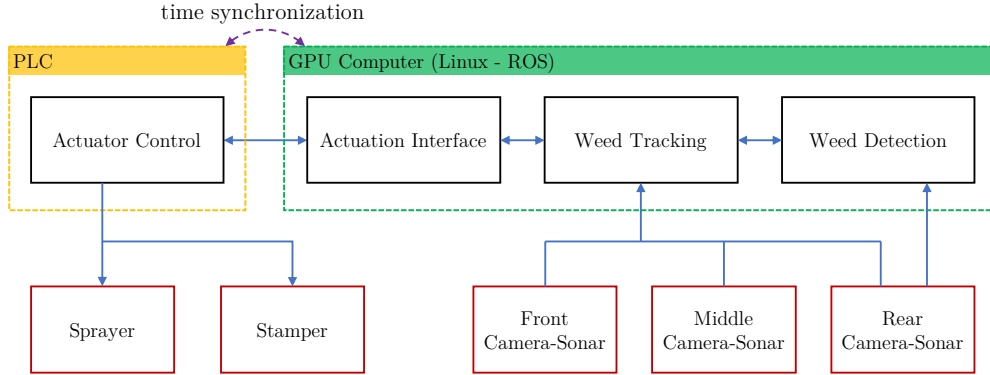


Figure 3: Architecture of computers, sensors and actuators dedicated to the weed treatment.

and lowest position. The exact height adjustment is done by two depth wheels, which can be adjusted smoothly by a spindle with crank. Working height is planned at 30 mm between soil and stamp underneath. Moreover, the mechanical construction is built robustly and there is only one degree of freedom for the actuators positioning, i.e. cross to the driving direction. The weeding tools are described below.

The stamping tool is composed of 18 stamps, arranged in two ranks. Each stamp consists out of a small double operating (stroke + return stroke) pneumatic cylinder, with a stroke of 100mm and a 10mm diameter stainless steel bolt. All stamps are individually controllable. The key specifications of the mechanical stamping tool are listed in Fig. 1.

Table 1: Specification of Actuator

average stamping time	40ms
maximum horizontal speed	1.53m/s
stamping accuracy	<1mm
positioning repeatability	<1mm

Separated from the stamps, through a protective plate, the spray application is positioned in the back. This application is assembled out of nine nozzles individually controlled by magnetic valves. The opening angle of the nozzles is 40°, and the mounting height is adapted accordingly to the required precision. For our in-row weeding application, each sprayer mounted in the lowest position can cover a 30mm-diameter area, which is sufficient to prevent valuable crops growing close to the weeds from being sprayed.

Both actuators are controlled with a scalable, programmable logic controller (PLC). Other tools that need computation, i.e weed detection and tracking, are done on a computer dedicated to the weed control (Intel i7 CPU, GeForce GTX-1070 GPU) running Linux and ROS. The PLC, a Rexroth IndraControl L75.1 interfaces all pneumatic components, motors and pumps to trigger the actuators. A wrapper is used to communicate with the ROS network. This PLC manages three lists of targets: one for the front stamping row, one for the rear stamping row and one for the sprayers, each target having a unique identifier. The weed treatment during movement is a time-critical part of the process because a small delay can already lead to a position error at centimeter-level that is large enough to miss a small weed. Exact triggering of the actuation requires the PLC to use the absolute time stamps. To fulfill the timing requirement for precision treatment the clocks (PC, PLC) are synchronized. This configuration is illustrated on Fig. 3.

3.2 Perception System and Calibration

In Fig. 2, three ground-facing global shutter cameras (JAI AD-130 GE) with 8mm fix length lens (Fujinon TF15-DA-8) are mounted on our weed control unit. The distance to the ground is approximately 800mm, and the field of view (FOV) of each camera can reach up to 240mm x 310mm. To protect the camera setup and the perception

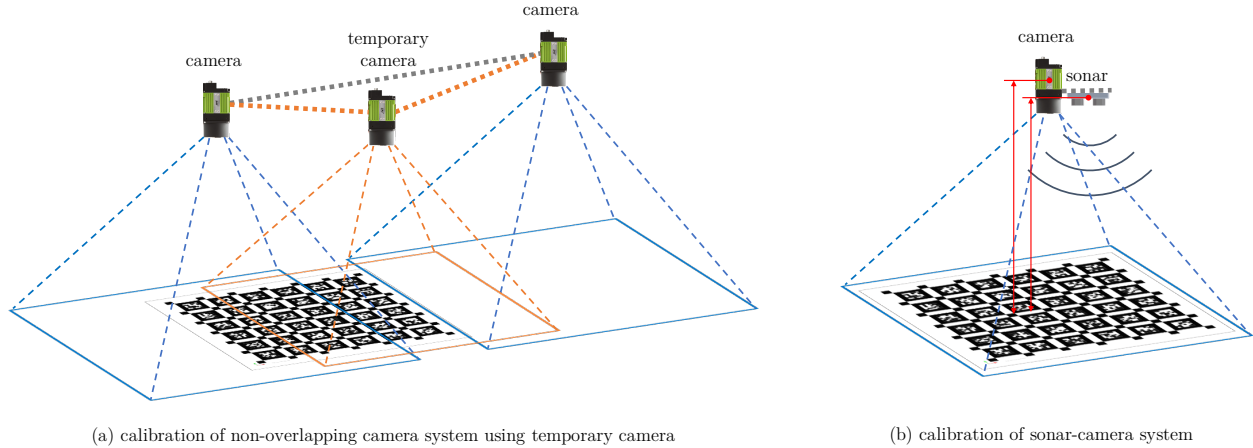


Figure 4: The schematic of sensor calibration. Left: the calibration of non-overlapping cameras using overlapping calibration software by introducing assistant camera in between. Right: the calibration of the camera and sonar height difference uses a marker field perpendicular to the central axes of the sensors.

system from natural light sources, the weed control unit is covered by $3mm$ acetal copolymer sheets, and artificial lights are mounted under to control the illumination. As in most state-of-the-art weed detection algorithms, we use the normalized difference vegetation index (NDVI) (Lottes et al., 2016), which utilizes both color (RGB) and the near-infrared (NIR) information. To do so, an RGB+NIR camera is used for the weed detection setup, whereas a RGB-only camera is used for the tracking setup, since the additional NIR information comes typically at a high cost and is not required for the tracking part.

To facilitate the high-precision estimation of the geometric properties of the weeds, all the intrinsic and extrinsic parameters of the non-overlapping camera system are calibrated. The intrinsic calibration for each camera is performed using standard OpenCV monocular calibration nodes (Bradski and Kaehler, 2000), where a pinhole camera model with radical tangential distortion parameters is estimated. The extrinsic calibration of our non-overlapping cameras is conducted using open-source overlapping camera calibration tool Kalibr (Furgale et al., 2013), where an assistant camera is temporarily incorporated into the system to generate overlapping areas between the cameras, as shown in Fig. 4 (a).

Besides, three narrow-beam sonars (SRF235 Ultrasonic Range Finder) are mounted next to each camera to help recover the absolute scale of the monocular VO estimates. The accuracy of the range measurement of these sonars is around $10mm$. The marker field calibration (Siltanen et al., 2007) between the camera and its corresponding sonar is conducted, where only the height difference between the two sensors is of interest, as shown in Fig. 4 (b). A set of artificial markers is attached on the flat ground surface inside the FOV of both camera and sonar, and the camera and sonar axis are placed perpendicular to the ground surface, thus allowing for the ground height estimation from both sensors. The measurements of heights from both sensors are processed and fed into a Kalman filter (KF), where the difference between two estimated heights is obtained as a calibration parameter.

4 WEED CONTROL SYSTEM

The Flourish project’s weed control system is a self-contained module that is responsible for detecting, tracking, and treating weeds in real time. In Fig. 5, the framework of our proposed weed control system is illustrated, and the process can be described as follows: The classification module implements a state-of-the-art learning-based computer vision algorithm to detect and discriminate crop and weeds from acquired images in Sec. 4.1. At the same time, the intra-camera tracking estimates the camera poses and 3D scene map using direct methods in Sec. 4.2. After receiving *delayed* classification results and scene structures, the object initializer and updater creates the templates of

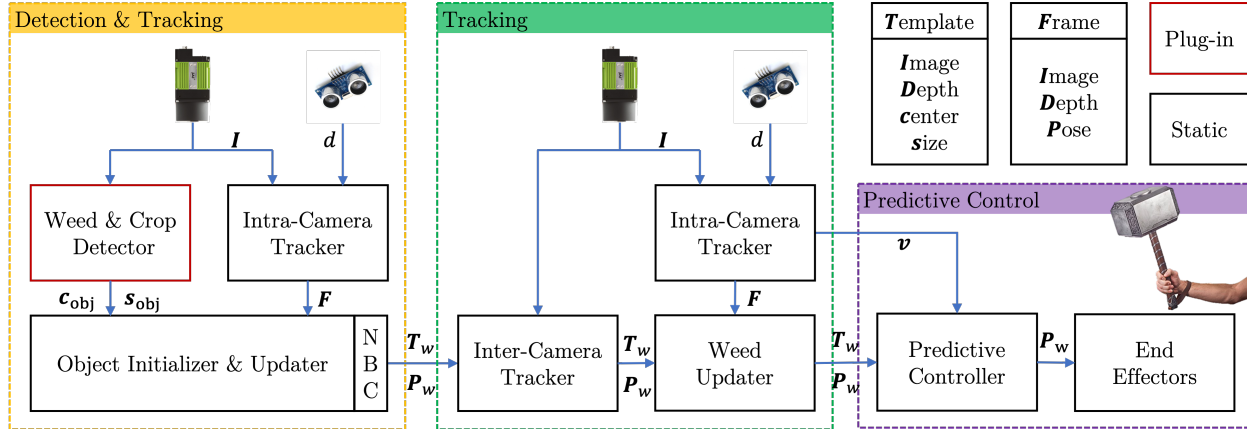


Figure 5: An overview of our proposed weed control system that is composed of weed detection, tracking and predictive control modules. The symbols used in the figure is briefly explained in upper-right corner.

the received objects, propagates their poses up to date in Sec. 4.3. As an object moves out of the FOV of the detection camera, a naive Bayes classifier (NBC) further classifies it as either crop or weed based on the accumulated labels in Sec. 4.4. Once the tracking camera finds a new weed object moving into its FOV, inter-camera tracking performs illumination-robust direct tracking to find its new pose and creates a new template for intra-camera tracking in Sec. 4.2. After repeated intra-camera tracking, updating, and inter-camera tracking, the weeds finally approach the end-effector, where the control algorithm predicts the triggering timing and position of actuation for intervention in Sec. 4.5.

4.1 Crop and Weed Classification

The primary objective of the plant detection and classification module is to enable our proposed weed control system to distinguish the weeds and crops in the real field. We implement state-of-the-art the methods described in (Lottes and Stachniss, 2017) and (Lottes et al., 2018b), where the accuracy and runtime performance for the dataset recorded in Bonn, Germany is presented in Table. 2, where the object-wise metric is used for this evaluation. For both implemented classifiers, the RGB+NIR data acquired by a 4-channel JAI camera in Fig. 6 is used instead of merely using RGB images, due to the fact that NIR information is proven to be especially useful for separating the vegetation from the soil. Besides, both methods provide mainly a pixel-wise semantic segmentation of crops and weeds. The output of our implemented classifier is a set of central locations (center of mass) of the detected weeds and crops with their corresponding areas (bounding boxes) in the image space.

Table 2: Performance of Implemented Classifiers

Approach	runtime [Hz]	recall[%]		precision [%]	
		weed	crop	weed	crop
(Lottes and Stachniss, 2017)	5-8	92	96	98	86
(Lottes et al., 2018b)	5	95	91	73	90

4.2 Intra- and Inter-Camera Tracking

Our proposed intra- and inter-camera tracking modules are essentially direct VO algorithms based on continuous optimization of photometric error over images, which can be generally divided into tracking and mapping procedures. In this section, the mathematical representation of the direct formulation with the costs for intra- and inter-camera tracking is described in Sec. 4.2.1, and the optimization algorithm utilized in this paper is presented in Sec. 4.2.2. Most importantly, our proposed intra- and inter-camera tracking modules are presented in Sec. 4.2.3 and Sec. 4.2.4,



Figure 6: Left: acquired RGB image. Middle: acquired NIR image. Right: crop (green) and weed (red) classification result using (Lottes et al., 2018b)

respectively.

4.2.1 Direct Formulation

A reference frame is defined as a gray-scale reference image $I_r : \Omega \rightarrow \mathbb{R}$ and an inverse depth map $D_r : \Omega \rightarrow \mathbb{R}^+$, where $\Omega \subset \mathbb{R}^2$ is the image domain. A 3D scene point $\mathbf{x} = (x, y, z)^T$ observed from an image is parameterized by its pixel location and inverse depth $d = z^{-1}$ in the reference frame instead of the conventional 3 unknowns. Defining a 3D projective warp function $\pi(\mathbf{x}) = (x/z, y/z)^T$, a pixel $\mathbf{u} = (u, v)^T \in \Omega$ can be back-projected into 3D world as $\mathbf{x} = \pi^{-1}(\mathbf{p}, d) = \mathbf{K}^{-1}\mathbf{p}/d$, where $\mathbf{p} = (u, v, 1)^T$ is the homogeneous coordinate of such pixel and \mathbf{K} is the pre-calibrated camera intrinsic matrix.

A 3D rigid body transformation $\mathbf{G} \in SE(3)$ from the reference frame to frame i can be expressed as follows:

$$\mathbf{G}_{ir} = \begin{bmatrix} \mathbf{R}_{ir} & \mathbf{t}_{ir} \\ \mathbf{0} & 1 \end{bmatrix} \quad (1)$$

where $\mathbf{R}_{ir} \in SO(3)$ and $\mathbf{t}_{ir} \in \mathbb{R}^3$ are the 3D rigid body rotation and translation from reference frame to frame i , respectively. $SE(3)$ and $SO(3)$ are the 3D rigid transformation and rotation group in Euclidean space (Blanco, 2010).

To better explain all costs involved in this paper, the pixel-wise direct error E_k of the k^{th} pixel between reference frame and the i^{th} frame can be generally written as:

$$E_{k,Int} := \sum_{\mathbf{p} \in S_{\mathbf{p}}} w_{\mathbf{p}} \|I_i(\pi(\mathbf{p}')) - I_r(\pi(\mathbf{p}))\|_{\gamma} \quad (2)$$

$$\mathbf{p}' = \mathbf{R}_{ir}\pi^{-1}(\mathbf{p}, D_r(\mathbf{p})) + \mathbf{t}_{ir} \quad (3)$$

where $I(\cdot)$ represents pixel intensity of image I . The set $S_{\mathbf{p}}$ is the set of pixels in a pre-defined local patch in the reference frame, $w_{\mathbf{p}}$ is the weight assigned for each pixel, $\|\cdot\|_{\gamma}$ is the Huber norm, and the subscript $k \in S_{pixel}$ is the index of a sampled pixel selected from a selection algorithm.

The intensity-based direct formulation, in essence, compares pixels based on the brightness constancy assumption and treat the lighting change affected pixels as outliers. However, the illumination cannot be explicitly controlled between cameras, thus the illumination-robust direct image alignment methods are required for our system. The global affine model in (Klose et al., 2013) (Engel et al., 2015) provides one possible solution that is capable of compensating the additive and multiplicative global lighting or exposure changes. The pixel-wise energy function is described as:

$$E_{k,GAff} := \sum_{\mathbf{p} \in S_{\mathbf{p}}} w_{\mathbf{p}} \|I_i(\pi(\mathbf{p}')) - \beta_i - \frac{e^{\alpha_i}}{e^{\alpha_r}} I_r(\pi(\mathbf{p})) - \beta_r\|_{\gamma} \quad (4)$$

where $\alpha_{i,r}$ and $\beta_{i,r}$ are global illumination affine model parameters, which are jointly optimized at each iteration. Combined with Huber norm, this *affine* method can work well in an environment without substantial local illumination changes.

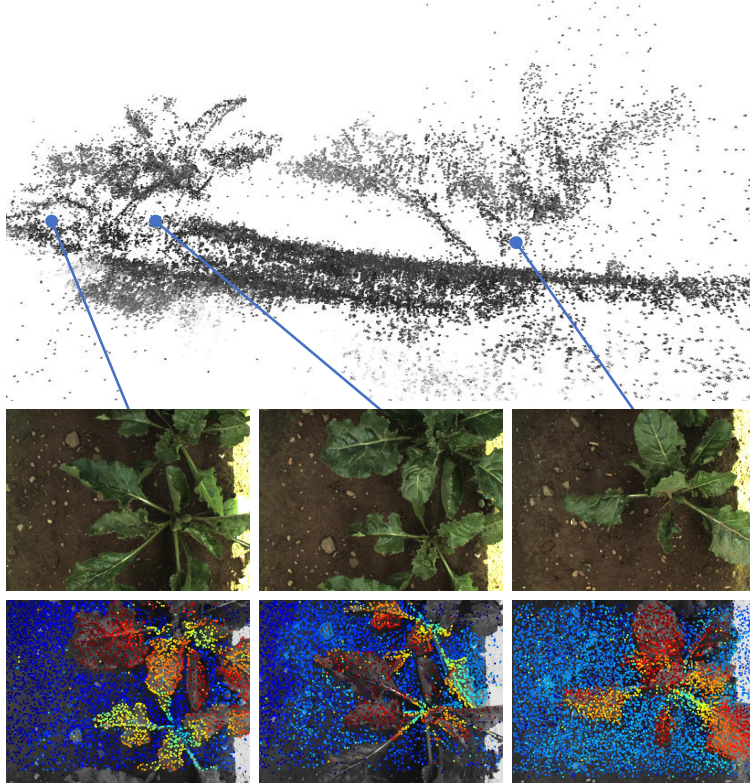


Figure 7: An example of reconstructed inverse depth map and a 3D pointcloud of plants and ground surface from our proposed intra-camera tracking algorithm.

4.2.2 Windowed Optimization

A sliding window optimization framework using the Gauss-Newton algorithm, described in (Engel et al., 2018), is utilized to achieve real-time motion estimation and 3D structure mapping. The optimization problem, in Eqn. 2, is furthered reduced to solve a nonlinear least-square minimization problem on Lie-manifolds. The corresponding Lie Group component $\xi \in \mathfrak{se}(3)$ is introduced to represent the 6-DoF camera pose, where this element can be mapped to $G \in SE(3)$ through the exponential mapping as:

$$G = \exp_{\mathfrak{se}(3)}(\xi) \quad (5)$$

and the update rule in Lie Manifold can be performed through logarithm and exponential mapping as:

$$\xi_{ik} = \log_{\mathfrak{se}(3)}(\exp_{\mathfrak{se}(3)}(\xi_{ij}) \cdot \exp_{\mathfrak{se}(3)}(\xi_{jk})) \quad (6)$$

For monocular VO algorithms, there are typically two phases performing the optimization: tracking phase and reconstruction phase. In the tracking phase, the inter-frame camera pose G_{ir} is estimated given the depth map D_r as a prior. In the reconstruction phase, both the depth map D_r and the camera pose G_{ir} are jointly optimized to improve the overall performance. To boost the tracking robustness over large camera motion, a minimization over an image pyramid, named coarse-to-fine approach in (Engel et al., 2014), is utilized to achieve a good trade-off between precision and speed.

4.2.3 Intra-Camera Tracking

The primary objective of our proposed intra-camera tracking algorithm is continuously tracking the 3D positions of detected crops and/or weeds until retraced in the next camera or being treated. Instead of tracking individual plants, we

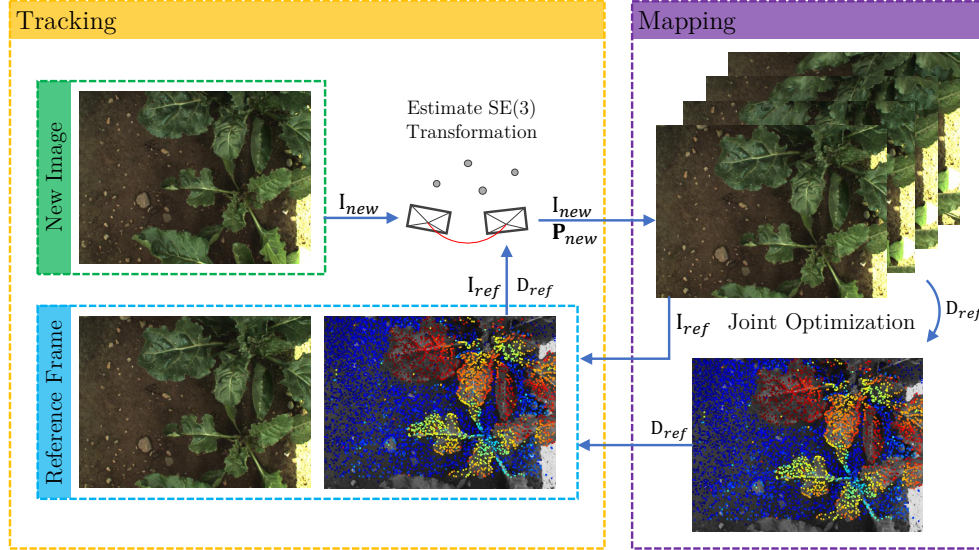


Figure 8: The overview of our proposed intra-camera tracking and mapping algorithm.

formulate our multi-object tracking as a VO algorithm that performs camera tracking and ground object mapping by exploiting the domain knowledge that all tracked objects are stationary on the ground in Fig. 7. Unlike conventional multi-object tracking algorithms receiving classification results, extracting the image template, then propagating its pose and tracking, our proposed VO approach recovers the 3D scene structure before obtaining object information, then formulating each template as a combination of trimmed image and inverse depth map for later tracking upon arrival of classification results. As a result, our proposed intra-camera tracking strategy guarantees a constant-time operation in spite of the change of the amount of tracked objects.

To improve the accuracy and robustness of weed estimates, the intra-camera tracking is developed based on direct structure and motion formulation, which does not use features at any stage of the algorithm. Specifically, the intensity-based photometric error in Eqn. 2 is minimized during optimization since the appearance changes due to illumination variation is negligible under artificial lighting condition, thus satisfying brightness constancy assumption. The algorithm consists of two major components: camera tracking and scene structure mapping in Fig. 8, where the map is represented as inverse depth map as (Civera et al., 2008).

In the camera tracking phase, the full 6-DoF camera pose is recovered through whole image alignment against estimated scene model: given the new image I_i and latest frame $\{I_r, D_r\}$, the relative pose $\xi_{ir} \in \mathfrak{se}(3)$ is estimated by minimizing Eqn. 2 using image pyramid (Engel et al., 2015). As a non-keyframe VO algorithm, the incoming images are always tracked to the latest frame that includes an estimated inverse depth and grey-scale image. Then, the tracked image is fed into the mapping layer for structure estimation and pose refinement, in order to formulate a new frame to be tracked.

The scene structure mapping layer, in essence, is a joint optimization algorithm over successive tracked frames with inverse depth parameterization. As the new tracked image is available, all frames in a predefined sliding window are passed into the optimization layer, where both the scene model and the previously tracked camera poses are jointly estimated: given a reference frame $\{I_r, D_r\}$ and a set of tracked Images $\{I_i : 0 < i - r < s\}$ with their poses $\{\xi_{ir} \in \mathfrak{se}(3) : 0 < i - r < s\}$, both these camera poses ξ_{ir} and the scene structure D_r are refined by joint minimization of Eqn. 2, where s represents predefined window size. The estimated inverse depth map is propagated to the latest tracked image to formulate a new frame for camera tracking.

Many direct VO algorithms prefer to use all pixels with significant gradient, named semi-dense methods. Incorporating as many pixels as possible into optimization is proven to improve the robustness and accuracy of pose estimates, however, also compromising the real-time performance. In this work, a sparse point selection strategy (Engel et al.,

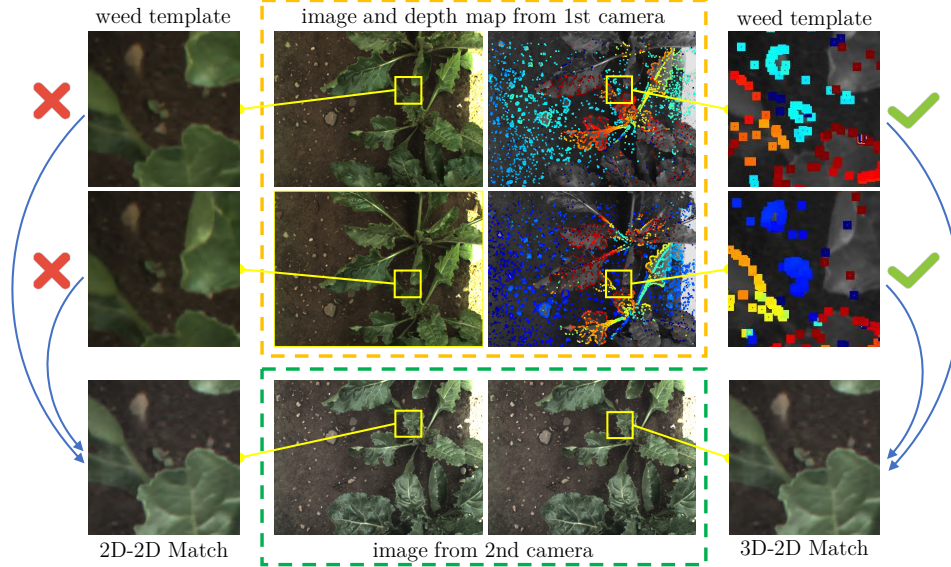


Figure 9: The comparison of 2D-2D image matching and 3D-2D direct template matching methods for inter-camera tracking, where appearance variety induced by the viewpoint changes could make the conventional image matching completely fail.

2018) is utilized to select well-distributed pixels spreading all over the image regions with sufficient gradients, which is proven to provide high-precision estimation in real time. The 3D structure of the point candidates are initialized using the propagated inverse depth maps or pre-defined ground model through bilinear interpolation, then being tracked and optimized in camera tracking and environmental mapping phases.

It should be noted that our proposed intra-camera tracking is essentially a monocular VO algorithm that suffers from scale ambiguity, which is useless for real-world control applications that require absolute scale. In this work, the filtered average ground depths from sonar and the integrated displacement information from wheel odometry are fused with the camera pose, and the scene structure estimates from tracking layer for absolute scale recovery.

4.2.4 Inter-Camera Tracking

The inter-camera tracking is designed to retrace the 2D positions of weeds using the information from images acquired from another camera, which forces us to take the illumination difference between cameras into consideration. To compensate the appearance changes of weeds due to lighting variation, we can either plug global or local illumination-robust costs into our optimization framework described in Eqn. 2. As stated in (Park et al., 2017), the global illumination-robust costs are capable of capturing whole image additive and/or multiplicative offsets, but failing to distinguish local lighting changes; the local ones, in contrast, can compensate both global and local illumination changes in the image, however, presenting significantly smaller convergence basin. For our weed control unit, both the global and regional illumination differences can be observed between images from different cameras due to the different lighting arrangements.

Taking advantage that only the 2D positions of weeds in image space is of interest other than the estimation of camera poses, we extract the small frame template of each weed combined with a global illumination-invariant cost to perform local image alignment using Eqn. 4, which alleviates the convergence issue by making local illumination consistency assumption. Given the image I_i in the current camera and the weed frame template $\{I_r, D_r\}$ from the previous camera, the relative pose $\xi_{ir} \in \mathfrak{se}(3)$ is estimated by minimizing Eqn. 4 using the image pyramid approach (Engel et al., 2015). Then, the weed center and its template boundary are transformed into the current frame using the pose estimate, which is used to generate a new template for intra-camera tracking. It should be noted that the retrieval of weeds object is

achieved by using 3D-2D direct template-based matching instead of using conventional 2D-2D image correspondence since the change of viewpoint could induce significant changes of appearance of objects especially for the ones with high depth variance as shown in Fig. 9.

4.3 Object initialization and update

There are two primary objectives of the object updater: (1) the object updater propagates the 3D positions of tracked objects using subscribed camera pose estimates and keep updating them when received new poses and depth information from intra-camera tracker; (2) The object updater is responsible for passing objects across cameras. For example, as one object in camera-1 moves into the visible area of camera-2, the object updater of camera-1 passes the object template and its predictive pose into the inter-camera tracker of camera-2. If successfully tracked, the object information is removed from camera-1 and sent to object updater of camera-2. Otherwise, the object updater keeps tracking this object until being retraced.

To initialize the items in object updater with the unpredictable delays, a time-stamp indexed frame buffer is preserved in the memory to ease the later 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 recovered inverse depth map. When a classification result arrives at the object initialization layer, the corresponding frame is extracted from the buffer given its time-stamp. The templates of the detected plants, consisting of a grey-scale image and inverse depth map, are extracted with the centroid depth estimated using bilinear interpolation. Then their poses are propagated using inter-frame poses in the buffer. When the object has transformed to the latest camera frame, 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.

4.4 Biased naive Bayesian classifier

To protect the valuable plants from being eliminated by the robot, an incremental naive Bayesian classifier with biased probabilistic model is utilized to filter out falsely detected weeds (which are actually crops) from the detection system. Considering a conditional probability model $P(C_i|\mathbf{l}_i)$, with sequentially received classified label $\mathbf{l}_i = l_{i,1}, l_{i,2}, \dots, l_{i,n}$ from the detection system and C_i as the output label from the Bayesian classification, the formula can be written as follows:

$$P(C_i|\mathbf{l}_i) = \frac{P(\mathbf{l}_i|C_i)P(C_i)}{P(\mathbf{l}_i)} \propto P(\mathbf{l}_i|C_i)P(C_i) \quad (7)$$

where $P(C_i)$ is the prior of an object belonging to weed or plant, and $P(C_i|\mathbf{l}_i)$ is conditional probability that is pre-defined with a probability model in Table. 3.

Table 3: Proposed Biased Probability Model

$P(\mathbf{l}_i, C_i)$	$\mathbf{l}_i = weed$	$\mathbf{l}_i = crop$
$C_i = weed$	0.8	0.2
$C_i = crop$	0.1	0.9

It should be noted that the probability model is biased because we adopt a multiple-time weed removal strategy at different growth stages of value crops, where one particular weed rarely survives under such treatments. For this weed control scenario, destroying a valuable plant is always considered to be a much more serious problem than just failing to remove a weed from the application point of view. The value listed in the table was set through small-scale trails using recorded in-field dataset, where the probability model that achieved (1) no crop was classified as a weed and (2) most weeds are classified as weeds were chosen for our application. For other weeding strategies, the user should reset the values of the probability model to ensure performance.

We follow a standard incremental implementation of this filter described in (Murphy et al., 2006), 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 value crops are deleted from the object updater and are not tracked anymore.

4.5 Predictive Control

At a high-level, a motion-model-based predictive controller is designed to predict the trigger timing and horizontal position, thus allowing for an operation-while-driving strategy. To better explain our high-level control design, all mathematical equations are derived concerning the first row of the stamping tool, and the same formulation can be obtained on all rows of stamping tool and the single-row sprayer in the same way.

Considering an extrinsic calibrated camera-tool system, we already have the homogeneous transformation matrix from the last camera frame to the first row of the stamping tool \mathbf{T}_{r1}^{cam} . The i^{th} object position $\mathbf{P}_{cam,i}$ in camera frame and the estimated velocity \mathbf{v}_{cam} from intra-camera tracker can be transformed to the coordinate system of the stamping tools first-row, such 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:

$$\mathbf{P}'_{r1,i} = \mathbf{T}(\mathbf{R}_{r1}^{cam} \mathbf{v}_{cam} t_{pred} + \mathbf{R}_{r1}^{cam} \mathbf{v}_{cam} t_{delay}) \mathbf{T}_{r1}^{cam} \mathbf{P}_{cam,i} \quad (8)$$

where \mathbf{R}_{r1}^{cam} is the rotation matrix extracted from \mathbf{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 $\mathbf{P}'_{r1,i} = [x'_{r1,i}, y'_{r1,i}, z'_{r1,i}, 1]^T$ 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.

5 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. In this section, the experimental setup including robot platform and the relevant information about the datasets are described in Sec. 5.1. The quantitative experimental evaluations of our proposed multi-camera tracking system in Sec. 5.2 and the inter- and intra-row weeding in Sec. 5.4 and Sec. 5.3 are presented regarding execution accuracy, robustness, and run-time property to support our claim. In the end, the generalization capability of our proposed weed control system is briefly discussed in Sec. 5.5.

5.1 Field Robot and Datasets

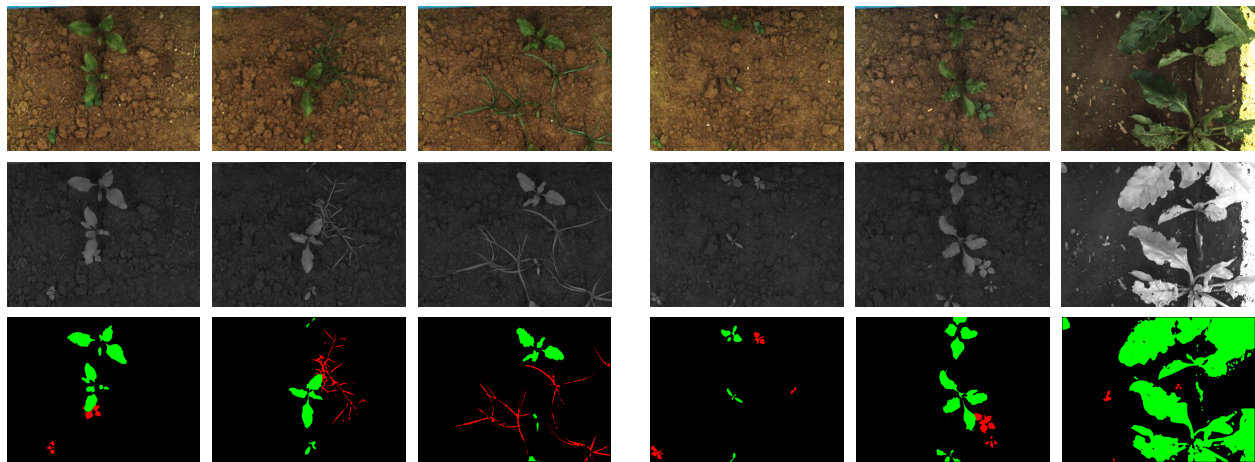
The BoniRob is a multi-propose field robot designed by BOSCH DeepField Robotics, which provides an application bay to install various tools for specific tasks. In recent years, a variety of agriculture applications have been successfully developed and validated using this robot, such as selective spraying,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

In this paper, we mainly evaluate our proposed system on sugar beet fields in Renningen (Germany), Ancona (Italy), and Zurich (Switzerland) in Table. 4. In Renningen, we used faked weed targets (leaves) to quantitatively evaluate the performance of multi-camera tracking and inter-row weeding in different terrain condition, speed and various simulated classification delays. In Ancona and Zurich, the intra-row weeding is evaluated in a real field with typical weed species in Fig. 10. The two datasets vary between 2-leaf and 4-leaf growth stages in Fig. 10, which are the main stages of weeding for farmers. It should be noted that the sugar beets in Zurich dataset are at least four times

larger compared to the ones in the Ancona dataset, which can be regarded as a harder dataset due to its increased the environmental complexity.

Table 4: Dataset Information

	Renningen	Ancona	Zurich
# sequences	12	8	5
# images	21,591	7,499	4,218
approx. crop size	-	0.5 - 100 cm^2	100 - 900 cm^2
approx. weed size	0.5 - 3.0 cm^2	0.5 - 60.0 cm^2	0.5 - 60.0 cm^2
driving speed	0.05 - 0.5 m/s	0.1 - 0.3 m/s	0.1 - 0.2 m/s
classifier	Vegetation Detection	(Lottes and Stachniss, 2017)	(Lottes and Stachniss, 2017)
runtime	$\sim 8 Hz$	3 - 5 Hz	3 - 5 Hz



(a) typical weeds in dataset

(b) different growth stages of crop in dataset

Figure 10: The overview of the dataset used in this paper: (a) three typical weeds frequently observed in the datasets, where the RGB, NIR, and classification mask of pigweed (Left), crabgrass (Middle) and quackgrass (Right) are presented; (b) three typical grow stage of crop in the datasets recorded in Ancona (Left and Middle) and Zurich (Right).

Instead of manually labeling all possible weeds from original images, we merely label the weeds and sugar beets from the detected objects from the classification layer for ground truth evaluation. Obviously, this ground truth labeling method is biased by the implemented classifier. However, our proposed weed control system merely aims at reducing the number of false positives from the classifier, as well as improving the treatment performance, rather than boosting the weed and crop detection. Considering that counting of undetectable weeds or crops cannot support our claims, we choose to label all detected objects and present our evaluation results up to the classification precision.

5.2 Multi-Camera Tracking

The proposed multi-camera tracking performance is extensively evaluated regarding accuracy, robustness, and runtime performance in a real field using the Renningen Dataset. We compare our proposed intra- and inter-camera tracking algorithm with our previous work (Wu et al., 2019) using EKF-based trackers in real field conditions under various crop growth stages, terrain surfaces, speeds, and classification delays. The ground truth centroid positions of the plants are provided by a simple vegetation detection classifier, which holds a 1.82 pixel repeatability.

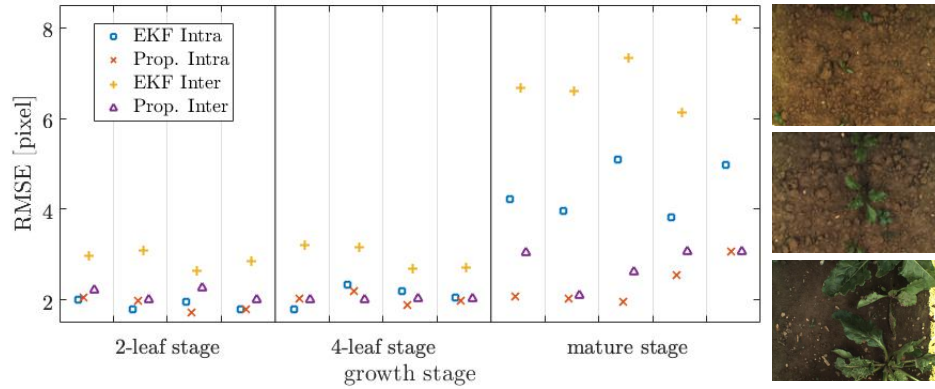


Figure 11: Tracking accuracy analysis on effect of different crop growth stages using Ancona and Zurich Dataset. Left: the intra- and inter-camera tracking RMSE comparison between EKF-based approach and our proposed optimization-based approach are versus test sequences with different crop growth stages. Right: Example images acquired in four typical real field condition (Top-to-Bottom): flat surface, rough surface, flat surface with plants, and rough surface with plants.

5.2.1 Performance under different growth stages

In this paper, one of our major contributions is that our system can deliver high-precision 3D position tracking of objects in high depth variance environments compared with our previously proposed weed control systems. To support our claim, the tracking precision under different crop growth stages are studied in term of the tracking accuracy. In this evaluation, the Ancona and Zurich dataset with the trained weed/crop classifier are used to examine tracking performance, where the vehicle speed for this test ranges from $0.1m/s$ to $0.2m/s$. The rooted mean square error (RMSE) of object center in image space is chosen as the evaluation metric to evaluate the intra- and inter-camera tracking accuracy quantitatively. We treat our previous work using the EKF-based multi-camera tracker as state-of-the-art method and compare it with our proposed approach to validate our claims.

In Fig. 11, we can generalize three major observations. (1) Our proposed system shows comparable intra-camera tracking accuracy at 2-leaf and 4-leaf crop growth stages, which is attributed to the fact that the plants appeared in the image are not high enough to present large appearance variance due to viewpoint changes during tracking. (2) Having ruled out the adverse effect of the appearance changes due to viewpoint variation at 2-leaf and 4-leaf stages, the inter-camera tracking precision of our proposed system still presents an improvement over EKF-based method, which is due to the involvement of illumination-robust cost. (3) At the crop mature stage, the crop is high enough to make significant appearance changes during tracking. Our proposed system presents a significantly higher accuracy in both intra- and inter-camera tracking phases compared with the EKF method, which is the joint effect of both illumination-invariant cost and the robust 3D-2D matching algorithm.

5.2.2 Performance under different terrain conditions

The experiments presented here are designed to evaluate the intra- and inter-camera tracking accuracy of our proposed system quantitatively in various field conditions. The Renningen dataset is chosen for this test, where a $0.2m/s$ vehicle speed and a simple vegetation detection algorithm with a $0.138s$ constant classification delay are observed. It should be noted that the Renningen dataset uses small leaves as fake targets and there is no big plants in the acquired images to produce structure-induced appearance changes. The rooted mean square error (RMSE) of object center in image space is again chosen as the evaluation metric to evaluate the intra- and inter-camera tracking accuracy quantitatively. Our previous work is treated as state-of-the-art method to compare.

In Fig. 12, both intra- and inter-camera tracking accuracy are plotted versus different terrain conditions, where we

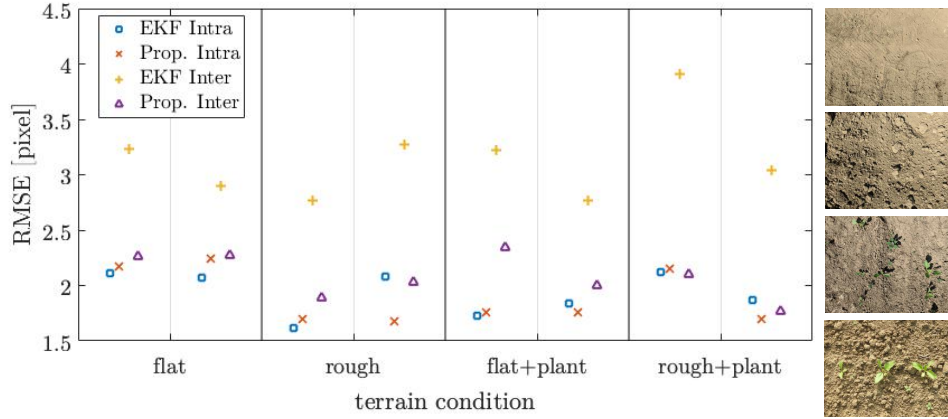


Figure 12: Tracking accuracy analysis on effect of terrain conditions using Renningen Dataset. Left: the intra- and inter-camera tracking RMSE comparison between EKF-based approach and our proposed optimization-based approach versus test sequences with different terrain conditions. Right: Example images acquired in four typical real field condition (Top-to-Bottom): flat surface, rough surface, flat surface with plants, and rough surface with plants.

can find three significant observations. (1) Both methods present good intra-camera tracking accuracy with an RMSE at around 2 pixels, which means our proposed framework is capable of tracking multiple objects with a precision as good as our previous work. (2) Our proposed illumination-robust inter-camera tracking algorithm improves the object position estimation across cameras compared to our previous work. (3) Analyzing the effect of the terrain roughness and the density of plants on both tracking frameworks, we can observe that the overall tracking precision is correlated with the number of edge pixels with significant gradients, where both the rough terrain and the plants can provide such edges for motion estimation and template matching. In contrast, the small height variation caused by the terrain roughness plays a minor role in tracking with the down-looking camera setup.

5.2.3 Performance under different classification delays

In this paper, one of our major claims is that our system can work with a much higher range of classification delays without compromising overall tracking accuracy and efficiency compared to 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. In this evaluation, the Renningen Dataset is chosen, and a simple vegetation detection algorithm with a fake classification delay generator is used 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. The rooted mean square error (RMSE) of the 2D centroid position of the retraced object is chosen as the evaluation metric to evaluate delayed initialization accuracy of our proposed system quantitatively, and the central comparison system is set to be our previous work - the EKF-based multi-camera tracking algorithm.

In Fig. 13 (Top), our proposed system shows similar performance in phase (a), but improved initialization precision in phase (b) and (c) compared with the EKF method. Combining with the phase description, we can find that the tracking accuracy limits our proposed initialization precision. Having received detection results from the delayed classifier, the initialization is equivalent to the intra-camera tracking if the object appears in detection camera as phase (a), or is equivalent to inter-camera tracking if it appears in tracking camera as phase (c), or in-between as phase (b). Since the illumination-robust cost has improved the inter-camera tracking accuracy of our proposed system, the delayed initialization also shows better precision in phase (b) and (c).

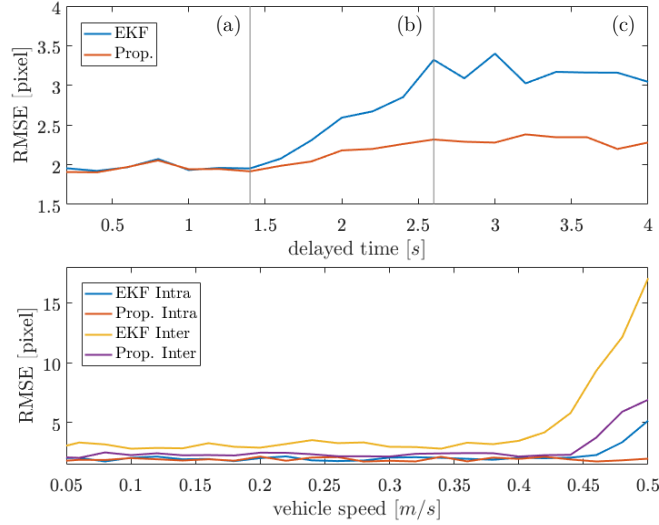


Figure 13: Top: the RMSE [pixel] versus various simulated classification delays [s] in the delayed initialization phase are plotted, which can be generally divided into 3 phases based on the current locations of objects when detected: (a) all detected objects are still in the detection camera, (b) detected objects appears in both detection and next tracking cameras, and (c) all detected objects are approached to tracking camera. Bottom: the RMSE of both intra- and inter-camera tracking versus vehicle speeds are plotted to study the effect of vehicle speed on our proposed system.

5.2.4 Performance under different vehicle speed

To examine the system capabilities, the tracking accuracy, both within or across cameras, are evaluated in different vehicle velocities using the Renningen Dataset. A simple vegetation detection algorithm with a $0.138s$ average classification delay is chosen. It should be noted that the maximum allowed speed could only reach up to $0.3m/s$ for mechanical design reasons, for this reason, the results for higher velocities are simulated by sampling recorded images to study the limitation of our proposed system. In Fig. 13 (Bottom), the intra- and inter-camera tracking error of both methods increases exponentially after a certain speed, which is attributed to the reduced overlap area between sequential images.

5.3 Evaluation of Weed Control

Table 5: Treatment Rates

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

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 with an average $0.138s$ classification delay is utilized to generate classification results.

To evaluate 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



Figure 14: Left: example picture of field robot performing weed control evaluation. Middle: sample pictures of leaves after mechanical weed removal. Right: sample pictures after selective spraying.



Figure 15: Left: an in-row selective spraying test is conducted in real-world field in Ancona (Italy), where the marks after spraying can be clearly seen in the picture. Right: the performance of in-row mechanical weeding is presented in simulated test environment in Bonn (Germany), where the faked targets made of leaves attached on modeling clay is used to visualize the execution results.

performed in short paths to control the test condition in an outdoor environment better. For short-path tests, we use $5 - 10/m^2$ targets and repeat such test for 10 times for each test speed.

To evaluate selective spraying, we set up a webcam to monitor targets after spraying, taking the advantage of the wet area being 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. 14, and the quantitative experimental results are provided in Table 5. The successful treatment rates regarding different vehicle speeds and terrain roughness are presented in the table, where the leaves with average $10mm$ radius are chosen as targets. We don't have an experimental evaluation of mechanical weeding with a speed higher than $0.2m/s$ because driving the vehicle higher than this would introduce a huge tangential force on the stamper, which might lead to the malfunction of the stamping tool. From the table, we can observe that the successful treatment rate is almost invariant with the speed in both flat and rough field ground.

Table 6: Weed Classification and Treatment Rate

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

5.4 Evaluation of Intra-Row Weed Control

The purpose of the in-row weed control evaluation is to analyze the whole system performance, and we choose the final treatment rates as the measure of treatment precision. The overall performance evaluation, named in-row weed control, is performed in the sugar beet fields and simulated environment. The reason why we introduce the extra experiment in a simulated environment is that 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 after some trials. 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. 15. 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 summarized in Table 6. The real weeds serve as targets in 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 a good precision.

5.5 Discussion about Generalization

Having proven that our proposed weed control system can successfully detect, track, and execute the weeds in the real-world sugar-beets field, one of the remaining questions is how the generalization capability of our proposed framework on other crops is. From the theoretical point of view, our proposed multi-camera tracking system could track any detected objects fixed on the ground if the terrain presents enough textures. Our proposed predictive control module is invariant with geometric properties and appearance of objects since the predictive controller treats the executed objects as shapeless points. The only component that is highly affected by crop and weed species is the weed and crop detection module. However, our proposed framework holds the ability to plug-in any weed and crop detectors as needed. As a result, our proposed weed control framework can easily generalize to different crop and weed species by incorporating appropriate weed and crop detectors. Besides, the latest work (Bosilj et al., 2019) shows that transfer learning between different crop types can significantly reduce the training time, which shed light on fast the crop and weed generalization.

6 Conclusions

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 caused by the plant detection algorithms, and a 3D multi-camera multi-object 3D tracking algorithm is developed to provide high-precision tracking results across cameras. To boost the robustness of our proposed system, a 3D mapping layer is introduced to enable the 3D-2D template matching, which compensates the appearance changes due to viewpoint variation during operation, while an illumination-robust cost is incorporated to rule out the appearance changes due to lighting difference. A biased naive Bayesian filter is designed to remove the false positives from the detector in the complex field terrain. To adopt an operation-while-driving strategy, both low- and high-level control strategies are deliberately designed for fast-actuation high-precision weed removal. The tracking and control performance of the proposed system is extensively evaluated in different terrain conditions and crop growth stages 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.

Acknowledgments

This work has partly been supported by the European Commission under the grant number H2020-ICT-644227-FLOURISH. We furthermore thank the people from BOSCH Corporate Research and DeepField Robotics for their support during joint experiments, and particularly Tillmann Falck and Elisa Rothacker-Feder.

References

- Alismail, H., Browning, B., and Lucey, S. (2016). Direct visual odometry using bit-planes. *arXiv preprint arXiv:1604.00990*.
- Alismail, H., Kaess, M., Browning, B., and Lucey, S. (2017). Direct visual odometry in low light using binary descriptors. *IEEE Robotics and Automation Letters*, 2(2):444–451.
- Åstrand, B. and Baerveldt, A.-J. (2002). An agricultural mobile robot with vision-based perception for mechanical weed control. *Autonomous robots*, 13(1):21–35.
- Bawden, O., Kulk, J., Russell, R., McCool, C., English, A., Dayoub, F., Lehnert, C., and Perez, T. (2017). Robot for weed species plant-specific management. *Journal of Field Robotics*, 34(6):1179–1199.
- Bechar, A. and Vigneault, C. (2017). Agricultural robots for field operations. part 2: Operations and systems. *Biosystems Engineering*, 153:110–128.
- Blanco, J.-L. (2010). A tutorial on se (3) transformation parameterizations and on-manifold optimization. *University of Malaga, Tech. Rep*, 3.
- Blasco, J., Aleixos, N., Roger, J., Rabatel, G., and Molto, E. (2002). Ae—automation and emerging technologies: Robotic weed control using machine vision. *Biosystems Engineering*, 83(2):149–157.
- Bloesch, M., Omari, S., Hutter, M., and Siegwart, R. (2015). Robust visual inertial odometry using a direct ekf-based approach. In *Proceedings of International Conference on Intelligent Robots and Systems (IROS)*, pages 298–304. IEEE.
- Blue River Technology (2018). Blue river technology. <http://smartmachines.bluerivertechnology.com/>.
- Bosilj, P., Aptoula, E., Duckett, T., and Cielniak, G. (2019). Transfer learning between crop types for semantic segmentation of crops versus weeds in precision agriculture. *Journal of Field Robotics*.
- Bradski, G. and Kaehler, A. (2000). Opencv. *Dr. Dobb's journal of software tools*, 3.
- Chang, C.-L. and Lin, K.-M. (2018). Smart agricultural machine with a computer vision-based weeding and variable-rate irrigation scheme. *Robotics*, 7(3):38.
- Chikowo, R., Faloya, V., Petit, S., and Munier-Jolain, N. (2009). Integrated weed management systems allow reduced reliance on herbicides and long-term weed control. *Agriculture, ecosystems & environment*, 132(3-4):237–242.
- Civera, J., Davison, A. J., and Montiel, J. M. (2008). Inverse depth parametrization for monocular slam. *IEEE transactions on robotics*, 24(5):932–945.
- Crivellaro, A. and Lepetit, V. (2014). Robust 3d tracking with descriptor fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3414–3421.
- Dai, A., Nießner, M., Zollhöfer, M., Izadi, S., and Theobalt, C. (2017). Bundlefusion: Real-time globally consistent 3d reconstruction using on-the-fly surface reintegration. *ACM Transactions on Graphics (TOG)*, 36(4):76a.

- Datta, A. and Knezevic, S. Z. (2013). Flaming as an alternative weed control method for conventional and organic agronomic crop production systems: a review. In *Advances in agronomy*, volume 118, pages 399–428. Elsevier.
- Engel, J., Koltun, V., and Cremers, D. (2018). Direct sparse odometry. *IEEE transactions on pattern analysis and machine intelligence*, 40(3):611–625.
- Engel, J., Schöps, T., and Cremers, D. (2014). Lsd-slam: Large-scale direct monocular slam. In *European Conference on Computer Vision*, pages 834–849. Springer.
- Engel, J., Stückler, J., and Cremers, D. (2015). Large-scale direct slam with stereo cameras. In *Proceedings of International Conference on Intelligent Robots and Systems (IROS)*, pages 1935–1942. IEEE.
- Engel, J., Sturm, J., and Cremers, D. (2013). Semi-dense visual odometry for a monocular camera. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 1449–1456.
- Fennimore, S. A., Slaughter, D. C., Siemens, M. C., Leon, R. G., and Saber, M. N. (2016). Technology for automation of weed control in specialty crops. *Weed Technology*, 30(4):823–837.
- Furgale, P., Rehder, J., and Siegwart, R. (2013). Unified temporal and spatial calibration for multi-sensor systems. In *Proceedings of International Conference on Intelligent Robots and Systems (IROS)*, pages 1280–1286. IEEE.
- Gobor, Z. (2013). Mechatronic system for mechanical weed control of the intra-row area in row crops. *KI-Künstliche Intelligenz*, 27(4):379–383.
- Gonçalves, T. and Comport, A. I. (2011). Real-time direct tracking of color images in the presence of illumination variation. In *Proceedings of International Conference on Robotics and Automation (ICRA)*, pages 4417–4422. IEEE.
- Greene, W. N., Ok, K., Lommel, P., and Roy, N. (2016). Multi-level mapping: Real-time dense monocular slam. In *Proceedings of International Conference on Robotics and Automation (ICRA)*, pages 833–840. IEEE.
- Griepentrog, H.-W., Nørremark, M., Nielsen, J., and Ibarra, J. S. (2007). Autonomous inter-row hoeing using gps-based side-shift control.
- Haug, S., Biber, P., Michaels, A., and Ostermann, J. (2014). Plant stem detection and position estimation using machine vision. In *Proceedings of the International Workshop on Recent Advances in Agricultural Robotics (RAAR)*.
- Hillocks, R. J. (2012). Farming with fewer pesticides: Eu pesticide review and resulting challenges for uk agriculture. *Crop Protection*, 31(1):85–93.
- Klein, G. and Murray, D. (2007). Parallel tracking and mapping for small ar workspaces. In *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*, pages 225–234. IEEE.
- Klose, S., Heise, P., and Knoll, A. (2013). Efficient compositional approaches for real-time robust direct visual odometry from rgb-d data. In *Proceedings of International Conference on Intelligent Robots and Systems (IROS)*, pages 1100–1106. IEEE.
- Kraemer, F., Schaefer, A., Eitel, A., Vertens, J., and Burgard, W. (2017). From plants to landmarks: Time-invariant plant localization that uses deep pose regression in agricultural fields. *arXiv preprint arXiv:1709.04751*.
- Kunz, C., Weber, J. F., Peteinatos, G. G., Sökefeld, M., and Gerhards, R. (2018). Camera steered mechanical weed control in sugar beet, maize and soybean. *Precision Agriculture*, 19(4):708–720.
- Lamm, R. D., Slaughter, D. C., and Giles, D. K. (2002). Precision weed control system for cotton. *Transactions of the ASAE*, 45(1):231.
- Langsenkamp, F., Sellmann, F., Kohlbrecher, M., Kielhorn, A., Strothmann, W., Michaels, A., Ruckelshausen, A., and Trautz, D. (2014). Tube stamp for mechanical intra-row individual plant weed control. In *18th World Congress of CIGR, CIGR2014*, pages 16–19.

- Lee, W. S., Slaughter, D., and Giles, D. (1999). Robotic weed control system for tomatoes. *Precision Agriculture*, 1(1):95–113.
- Lottes, P., Behley, J., Chebrolu, N., Milioto, A., and Stachniss, C. (2018a). Joint stem detection and crop-weed classification for plant-specific treatment in precision farming. *arXiv preprint arXiv:1806.03413*.
- Lottes, P., Behley, J., Milioto, A., and Stachniss, C. (2018b). Fully convolutional networks with sequential information for robust crop and weed detection in precision farming. *arXiv preprint arXiv:1806.03412*.
- Lottes, P., Hoferlin, M., Sander, S., Müter, M., Schulze, P., and Stachniss, L. C. (2016). An effective classification system for separating sugar beets and weeds for precision farming applications. In *Proceedings of International Conference on Robotics and Automation (ICRA)*, pages 5157–5163. IEEE.
- Lottes, P. and Stachniss, C. (2017). Semi-supervised online visual crop and weed classification in precision farming exploiting plant arrangement. In *Proceedings of International Conference on Intelligent Robots and Systems (IROS)*, pages 5155–5161. IEEE.
- McCool, C., Perez, T., and Upcroft, B. (2017). Mixtures of lightweight deep convolutional neural networks: applied to agricultural robotics. *IEEE Robotics and Automation Letters*, 2(3):1344–1351.
- Meilland, M., Comport, A., Rives, P., and Méditerranée, I. S. A. (2011). Real-time dense visual tracking under large lighting variations. In *British Machine Vision Conference, University of Dundee*, volume 29.
- Michaels, A., Haug, S., and Albert, A. (2015). Vision-based high-speed manipulation for robotic ultra-precise weed control. In *Proceedings of International Conference on Intelligent Robots and Systems (IROS)*, pages 5498–5505. IEEE.
- Midtby, H. S., Mathiassen, S. K., Andersson, K. J., and Jørgensen, R. N. (2011). Performance evaluation of a crop/weed discriminating microsprayer. *Computers and electronics in agriculture*, 77(1):35–40.
- Mur-Artal, R. and Tardós, J. D. (2017). Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE Transactions on Robotics*, 33(5):1255–1262.
- Murphy, K. P. et al. (2006). Naive bayes classifiers. *University of British Columbia*, 18.
- Naio (2018). Naio. <https://www.naio-technologies.com/en/agricultural-equipment/weeding-robot-oz/>.
- Newcombe, R. A., Lovegrove, S. J., and Davison, A. J. (2011). Dtam: Dense tracking and mapping in real-time. In *Proceedings of International Conference on Computer Vision (ICCV)*, pages 2320–2327. IEEE.
- Nieuwenhuizen, A., Hofstee, J., and Van Henten, E. (2010). Performance evaluation of an automated detection and control system for volunteer potatoes in sugar beet fields. *Biosystems Engineering*, 107(1):46–53.
- Nørremark, M., Griepentrog, H. W., Nielsen, J., and Sjøgaard, H. T. (2008). The development and assessment of the accuracy of an autonomous gps-based system for intra-row mechanical weed control in row crops. *Biosystems engineering*, 101(4):396–410.
- Pannacci, E., Lattanzi, B., and Tei, F. (2017). Non-chemical weed management strategies in minor crops: A review. *Crop protection*, 96:44–58.
- Park, S., Schöps, T., and Pollefeys, M. (2017). Illumination change robustness in direct visual slam. In *Proceedings of International Conference on Robotics and Automation (ICRA)*, pages 4523–4530. IEEE.
- Pizzoli, M., Forster, C., and Scaramuzza, D. (2014). Remode: Probabilistic, monocular dense reconstruction in real time. In *Proceedings of International Conference on Robotics and Automation (ICRA)*, pages 2609–2616. IEEE.
- Reiser, D., Martín-López, J. M., Memic, E., Vázquez-Arellano, M., Brandner, S., and Griepentrog, H. W. (2017). 3d imaging with a sonar sensor and an automated 3-axes frame for selective spraying in controlled conditions. *Journal of Imaging*, 3(1):9.

- Sa, I., Chen, Z., Popović, M., Khanna, R., Liebisch, F., Nieto, J., and Siegwart, R. (2018). weednet: Dense semantic weed classification using multispectral images and mav for smart farming. *IEEE Robotics and Automation Letters*, 3(1):588–595.
- Siltanen, S., Hakkarainen, M., and Honkamaa, P. (2007). Automatic marker field calibration. In *Proceedings of the Virtual Reality International Conference (VRIC)*, pages 261–267.
- Slaughter, D., Giles, D., and Downey, D. (2008). Autonomous robotic weed control systems: A review. *Computers and electronics in agriculture*, 61(1):63–78.
- Søgaard, H. and Lund, I. (2005). Investigation of the accuracy of a machine vision based robotic micro-spray system. *Precision Agriculture*, 5:613–620.
- Søgaard, H. T. and Lund, I. (2007). Application accuracy of a machine vision-controlled robotic micro-dosing system. *Biosystems Engineering*, 96(3):315–322.
- Steward, B. L., Tian, L. F., and Tang, L. (2002). Distance-based control system for machine vision-based selective spraying. *Transactions of the ASAE*, 45(5):1255.
- Strasdat, H., Montiel, J., and Davison, A. J. (2010). Scale drift-aware large scale monocular slam. *Robotics: Science and Systems VI*, 2.
- Stühmer, J., Gumhold, S., and Cremers, D. (2010). Real-time dense geometry from a handheld camera. In *Joint Pattern Recognition Symposium*, pages 11–20. Springer.
- Tillett, N., Hague, T., Grundy, A., and Dedousis, A. (2008). Mechanical within-row weed control for transplanted crops using computer vision. *Biosystems Engineering*, 99(2):171–178.
- Tillett, N., Hague, T., and Miles, S. (2002). Inter-row vision guidance for mechanical weed control in sugar beet. *Computers and electronics in agriculture*, 33(3):163–177.
- Underwood, J. P., Calleija, M., Taylor, Z., Hung, C., Nieto, J., Fitch, R., and Sukkarieh, S. (2015). Real-time target detection and steerable spray for vegetable crops. In *Proceedings of the International Conference on Robotics and Automation: Robotics in Agriculture Workshop, Seattle, WA, USA*, pages 26–30.
- Urdal, F., Utstumo, T., Vatne, J. K., Ellingsen, S. A. Å., and Gravidahl, J. T. (2014). Design and control of precision drop-on-demand herbicide application in agricultural robotics. In *Proceedings of International Conference on Control Automation Robotics & Vision (ICARCV)*, pages 1689–1694. IEEE.
- Van der Weide, R., Bleeker, P., Achten, V., Lotz, L., Fogelberg, F., and Melander, B. (2008). Innovation in mechanical weed control in crop rows. *Weed research*, 48(3):215–224.
- Vigneault, C. and Benoît, D. L. (2001). Electrical weed control: theory and applications. In *Physical control methods in plant protection*, pages 174–188. Springer.
- Weis, M., Gutjahr, C., Ayala, V. R., Gerhards, R., Ritter, C., and Schölderle, F. (2008). Precision farming for weed management: techniques. *Gesunde Pflanzen*, 60(4):171–181.
- Wu, X., Aravecchia, S., and Pradalier, C. (2019). Design and implementation of computer vision based in-row weeding system. In *Proceedings of International Conference on Robotics and Automation (ICRA)*. IEEE.
- Wu, X. and Pradalier, C. (2019). Illumination robust monocular direct visual odometry for outdoor environment mapping. In *Proceedings of International Conference on Robotics and Automation (ICRA)*. IEEE.
- Xiong, Y., Ge, Y., Liang, Y., and Blackmore, S. (2017). Development of a prototype robot and fast path-planning algorithm for static laser weeding. *Computers and Electronics in Agriculture*, 142:494–503.
- Young, S. L. (2018). Beyond precision weed control: A model for true integration. *Weed Technology*, 32(1):7–10.