SpECULARIA: Towards Fully Autonomous Robotic Indoor Farming System

Marsela Car Barbara Arbanas Ferreira Jelena Vuletić Matko Orsag

***Abstract*— To support the hypothesis that embracing robotics has the potential to address farming challenges and at the same time replace large and complex farm machinery, this paper proposes designing a farm around a heterogeneous robotic system dubbed SpECULARIA. This multi-robot system has mobile robots deployed to work just like in a warehouse, moving plants grown in containers to make sure every plant receives optimal care in ideal growing conditions. By structuring the work cell environment around a stationary dual arm manipulator, the system can plan and execute procedures to control every plant's growth and hygiene, from seed to harvest. Such a system surpasses current farming robots in scalability and versatility. We showcase compliance control algorithms combined with artificial intelligence which helps build a functional model of the plant. The same approach is used to program different plant treatments. Finally, we benchmark the proposed setup with a classical mobile manipulation approach, to demonstrate its feasibility.**

I. Introduction

Driven by the ever-growing demand, under the constraints of climate and societal changes, modern agriculture evolved into a technology native industry [2]. Comparable to industry, recent key technological advancements in Artificial Intelligence, Internet of Things, and Robotics, expressly found application in growing high-value crops on large-scale farms [3]. In the last decade, however, small and medium-sized manufacturers served as the proving ground of collaborative robotics, the technology that makes it possible to program and deploy robots on small production batches. SpECULARIA is built around this concept, which could have a transformative impact on small and medium-sized family-run farms.

Authors are with Laboratory for Robotics and Intelligent Control Systems, Faculty of Electrical and Computer Engineering, University of Zagreb, 10000 Zagreb, Croatia

To truly embrace robotics in agriculture this paper proposes going beyond the usual approach [4], where a mobile robot equipped with a manipulator is programmed to adapt to the farm environment. In SpECULARIA, the farm is built around a stationary dual robotic manipulator. This provides a structure to the manipulator workspace, enabling robots to execute challenging manipulation tasks such as pruning or pollinating. In turn, they rely on mobile robots deployed to work just like in a warehouse, moving plants grown in containers to make sure every plant receives optimal care in ideal growing conditions. Ideally, the multi-robot system can plan and execute procedures to control every plant's growth and hygiene, from seeds to harvest. In such a setup small aerial robots can be deployed to provide fast daily-run inspections, ensuring all plants are pest and disease free. The main contribution of this paper is to present this novel idea of manipulator-centric robotic agriculture, including its capabilities, limitations, and justification through evaluation against the state of the art mobile manipulator based approaches.

The next section compares our results with the state of the art. Then, we discuss how the hardware design and compliant

Fig. 1: The proposed system, inspired by antic greenhouses of emperor Tiberius (42 BCE–37 CE) [1] goes beyond the current state of the art proposing a heterogeneous multi-robot system comprised of small robots with specific abilities brought to work together. Such a system surpasses current farming robots in scalability and versatility and is a promising solution for family-run organic farms, as shown in the project video at https://youtu.be/k4twazSAuac.

control strategies [5] enable the system to prune plants, pollinate flowers [6], manage irrigation [7], and ultimately pick fruit [8]. An important enabler of the system is the custom-designed deep learning based perception solution, utilizing the diverse sensory apparatus to recognize and track plant status [9]. Section IV disseminates important tools used to train these perception models. In section V we evaluate the concept using our approach to mission planning that combines these technologies into a fully autonomous robotic farming system. We compare the proposed manipulator-centric cooperation approach with standard mobile manipulation.

II. Related work

Just like in [10], [11], we argue that robotics has the potential to address the farming challenges by replacing large complex farm machinery that is custom-built for specific crops with fleets of small versatile autonomous robots. However, when it comes to small and medium-sized, family-owned farms, such an approach is especially slow to catch on, mainly due to high initial costs [12]. A parallel can be drawn between industries where large production lines such as in the automotive industry, have been the major drivers of robotic technology from the start. Equivalent size production lines in agriculture are built around mobile machines and robots equipped with positioning systems, manipulation, and perception tools that enable them to autonomously cultivate large surfaces [11]. Smaller scale robotic systems for greenhouse cultivation also rely on mobile manipulators, i.e. autonomous mobile platforms that bring the multi-DoF manipulators equipped with specific tools to the point of manipulation [13]–[15]. Manipulating fragile objects, such as plants requires a degree of dexterity and adaptability that can be more easily achieved through soft robotics than using classical industrial tools [16]. Building upon recent achievements developed for agricultural applications [17], [18], we have developed a soft robotic gripper for our collaborative manipulator and evaluated it in a greenhouse experimental setup.

Detection in most domains of robotics has experienced a leap in performance and generalization ability since the advent of CNN-based models [19]. Most recent work in agricultural robotics relies on some kind of machine learning based perception. However, the use of deep learning perception methods is associated with the cost of large training dataset generation [20], [21]. In an attempt to reduce the cost of the tedious labeling process, researchers have recently turned to the generation of synthetic datasets for applications in agriculture of various crops and cultures [22]–[24], including a synthetic dataset for the semantic segmentation tasks of sweet peppers [25]. However, for some models trained on synthetic data only, the challenge of overcoming the domain gap can outweigh the benefits of automation [26], [27]. Automated and semi-automated labeling procedures can similarly reduce the time and cost associated with manual annotation while maintaining high levels of accuracy thanks to reducing the domain gap. While some authors use visual markers for automatic annotation [28], higher precision has been achieved using an articulated robot and an RGB-D camera [29]. Thanks to the availability of computational power, this work too heavily relies on deep learning and automation of training related procedures, using transfer learning to fine tune custom object detectors and segmentation models to tailored datasets generated using realistic simulation and collaborative annotation with human operators.

The novel paradigm of robotised greenhouse structured around a static manipulator is shown to be achievable in different plant hygiene and cultivation activities. To fully justify the approach, we conducted a benchmark study against a more conservative solution based on a mobile manipulator. The evaluator compares the costs of optimal plans for multi-robot mission planning and execution in the two scenarios. Some of the best known distributed solutions to the multi-robot mission planning problem are auction- and market-based approaches [30], that usually solve task allocation problems without considering partial ordering between tasks and coupling underlying cooperative missions. Precedence constraints for cross-schedule dependency (XD) problem class was addressed with iterative auctions [31]. Task planning is then usually approached with various optimization-based methods. Exact (optimal) offline solutions [32] exist, but these are computationally expensive and lack reactivity in dynamic environments. Heuristics such as evolutionary computation and other AI optimization methods are a more recent alternative [33]. For mission planning and coordination in our greenhouse setup, we developed a decentralized framework for multi-robot coordination for solving problems of higher complexity (CD class). The framework was used in this study as well. It defines the particular decision-making modules and coordination mechanisms. Tasks are assigned to robots using a greedy function which allows each task assigned to the robot to be the one with the best score for that task and ignores the effects on other task assignments.

III. Cooperative Multi-Robot Farming

The robotic agriculture concept discussed here relies on the cooperation of specially designed robotic components and human operators. In this section, we discuss the required customization of the hardware components, their limitations, and capabilities.

A. Growing pods organization and distribution

The cornerstone of the system is the team of mobile robots adapted to transport plants around the greenhouse. Inspired by hydroponic farms, but opting for soil rather than water-based nutrient solution that require fixed pipelines, our plants are grown in growth pods. This enables the robots to move the plants around the greenhouse and provide optimal growing conditions factoring in various environmental conditions like sunshine, temperature, air quality, etc. Likewise, it enables the robots to bring the growth pods to the treatment station, where the plants are provided with timely and regular agricultural maintenance procedures.

Fig. 2: Modified Pioneer 3DX and omnidirectional Agilex Scout robots with a lift mechanism. Robots navigate between the growing pods to transport plants to the manipulator workstation. The raised growing pods simplify navigation within a known and static environment. Treating plants on a remote location enables tight packing of the growing pods.

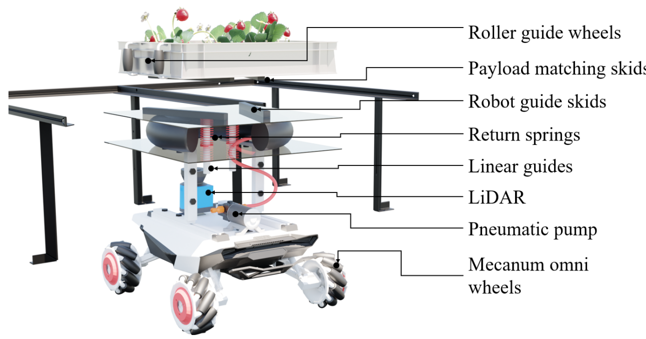

Fig. 3: We built a simple pneumatic mechanism to drive under the plants and lift them. Linear guides and return spring provide rigidity in the structure. We use different mechanical guides on the side and under the growing pods to ensure the plants are properly docked and drive safely on the robot.

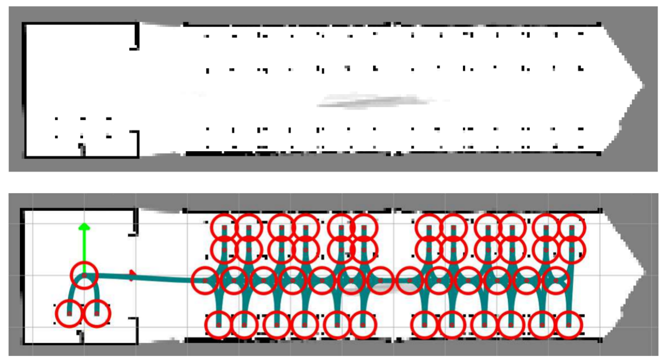

Fig. 4: When the LiDAR is placed as close to the ground as possible, under the plants the greenhouse occupancy grid looks structured and unique, as shown in the upper figure. Once we introduce structure, it is straightforward to treat the planning problem the same as in an industrial warehouse using AGVs. Therefore, we derive roadmaps to help plan and navigate trajectories around the greenhouse.

We take advantage of the fact that the plants are not treated where they grow to pack the growing pods as tight as possible, leaving only a small transport corridor between the plants. The growing pods are raised above the ground, enabling the mobile robots to navigate underneath the plants. This way, the ever-changing environment above the growing pods does not affect mobile robots' localization and navigation.

Promoting reuse and sustainability, we modified several decommissioned Pioneer 3DX robots with a custom lift mechanism that ensures stable transport of growing pods. The modified robot, shown in Fig 2, drives under the growing pod and lifts it slightly to transport it to another location. Since Pioneer robots are vulnerable to a raised center of mass, they can only move small size plants, like lettuce and strawberries. To further reduce the problem of a raised center of mass, once the growing pod reaches the transport corridor, the robot lowers it closer to its center of mass. To transport larger plants like peppers and tomatoes we modified an Agilex Scout mini robot 3. With a footprint comparable to a Pioneer robot, Scout is a sizable improvement with a larger payload and better dynamic capabilities. It also features mecanum omnidirectional wheels, making it easier for the robot to navigate the same environment as the Pioneer. A different custom lift mechanism is designed, based on a pneumatic pump which drives the inflation of a spring-enforced tire inner tube for directed deflation of the mechanism.

Both robots use 2D LiDARs to navigate around the greenhouse. The LiDAR is placed slightly above the wheels, as low as possible to the ground, to ensure only the structured environment beneath the plants is measured. In practice, the robot thus builds a map of legs supporting the growing pods that are fixed to the ground for consistency. Fig 4 shows what the map of the greenhouse looks like from the robot's perspective. To localize in this sparse map environment, we deployed Google Cartographer SLAM software [34]. Taking advantage of a completely structured warehouse-like environment, we opted for the roadmap approach to define permissible routes and checkpoints the robots can take in advance. To ensure these paths are feasible for robots, we followed the path continuity condition defined in [35], combined with appropriate trajectory planning for the selected mobile robot. Using this setup to navigate around the greenhouse proved to work well, however, we modified the vehicles with guide skids to make sure the docking with growing pods works precisely every time. Beneath them, the growing pods are equipped with a matching skid guide system and a set of roller wheels on the side to slide smoothly on the robot as well as into their docking container.

B. *Dual arm manipulation*

Relying on the mobile robots for transport, the plants are treated in an enclosed space that provides structure for the dual arm manipulation system. This way it is easier to perform plant segmentation, and plan plant hygiene procedures without the inconvenience of the ever-changing surroundings. This makes the technology transfer from the laboratory conditions that much easier.

Within SpECULARIA, plant treatment is considered throughout the entire plant life cycle, including activities such as pruning, pollination, rind cleaning, and yield estimation, all the way to fruit picking. Our approach relies on a dual-arm collaborative robotic system upgraded with soft tools and compliant algorithms which enable the manipulation system to interact with frail parts of the plant. Programming soft robots to treat plants beyond simple mechanical compliance requires measuring the interaction with the environment. In [36] we proposed an autoencoder based approach to single out and detect key features of soft components when we record images of their deformation. This allows us to use a camera as a force or tactile sensor to train shape and force estimators. It also enables us to combine other measuring modalities such as flex and force sensors. Measuring the force of interaction between the robot and the plant can be used to safely follow the shape of the plant, for instance during rind cleaning shown in [5].

In an attempt to generalize over various types of crops and to enable executing so many different agricultural procedures, we developed a two-finger soft robotic gripper SofIA (Soft Finger AI-Enabled Hand) presented in [37]. To increase the capabilities of the original design, we enforced the SofIA with hinges designed as passive serial joint supports that allow the SofIA to perform equally well in both horizontal and vertical approaches. This enabled picking both heavier tomato fruit as well as delicate ripe strawberries. Performing equally well in all approach directions provided more degrees of freedom when planning agricultural procedures on the plant.

Treating plants away from the location where they grow has a significant drawback. Moving the plants around the greenhouse makes it impossible to construct an effective pipe system for watering plants on the spot. Instead, the proposed dual-arm manipulation system has to water the plants when the mobile robot brings them to be treated. Here we make use of the multi-modal sensory setup described in the following section, which enables us to measure everything there is to know about the plant's needs. Using this information, the dual arm manipulation system can distribute water and nutrient in an approach personalized for each plant, which can be dubbed true precision farming.

C. Surveillance and high level control

The farming system as a whole is monitored by a UAV. An eye-in-hand camera mounted on an aerial manipulator enables close-up recording of all the plant parts and the soil, while keeping the UAV at a safe distance. A straightforward solution to fly above with a camera mounted on a gimbal would only provide canopy recordings, and the propellers with their prop wash might damage the plants or induce other problems such as poorly timed pollination or spread of potential diseases due to limited UAV maneuvering space. Depending on the detected plant states, UAV-provided measurements are an important autonomously generated input to the mission planning system. Pollination and yield control activities are planned based on the detected flowers, timely harvesting is based on detected fruit mass, and watering is based on plants' water stress level estimation. Similar to the developed NDVI-based water stress estimation, future

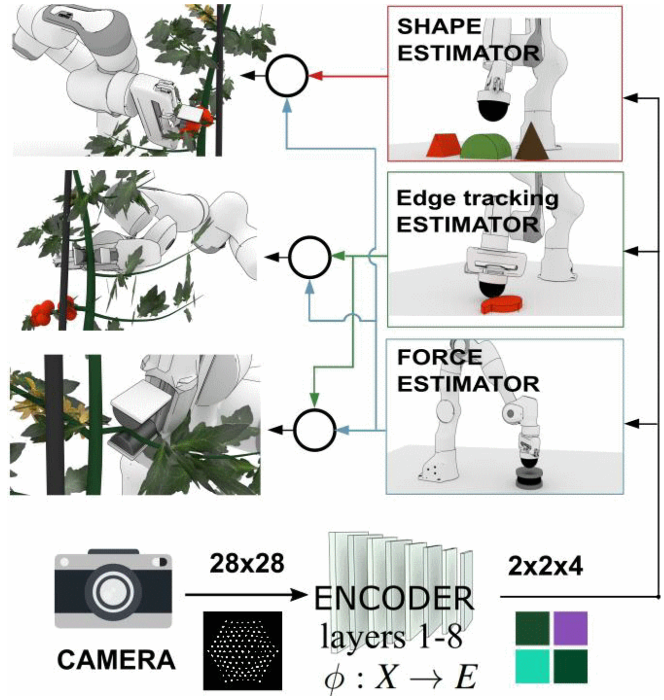

Fig. 5: Using a latent space between the encoder and the decoder of a simple convolutional neural network enables us to use a camera as a force or tactile sensor. We use the features defined this way to train shape and force estimators. It also enables us to combine other measuring modalities such as flex and force sensors and ultimately train the robot to execute compliant manipulation.

work includes modelling of other plant states based on the collected multi-spectral information.

The other, more classical input provider is the human operator. The envisioned paradigm assumes human intervention will still be inevitable, either in cooperation with or in place of robots. Once the mobile robots bring the plants to the workstation, this area can be observed as a standard industrial collaborative workcell.The compliant approach ensures safe interaction with the manipulator and the learn by demonstration approach is following ISO 15066 similar to the standard industrial collaborative workcells. Ensuring the safety of mobile robot navigation and collision avoidance is a task of the navigation system. This is made easier and safer since the plants are packed as tight as possible leaving no room for humans to walk around, and thus no need to share the environment with the robots.

IV. Collaborative Artificial Intelligence

Collaborative robotics is an emerging field that involves the use of robots to work alongside humans. Inspired by its success in robotics, SpECULARIA extends the collaborative approach into the field of AI. We envision a system where the farmer and the robot work alongside each other to generate datasets. These are then used to train detection, segmentation or classification models for specific tasks, such as identification of location, size, and health of individual plants and plant parts.

Automating so many plant-treating procedures drives us to develop versatile and robust perception solutions. The

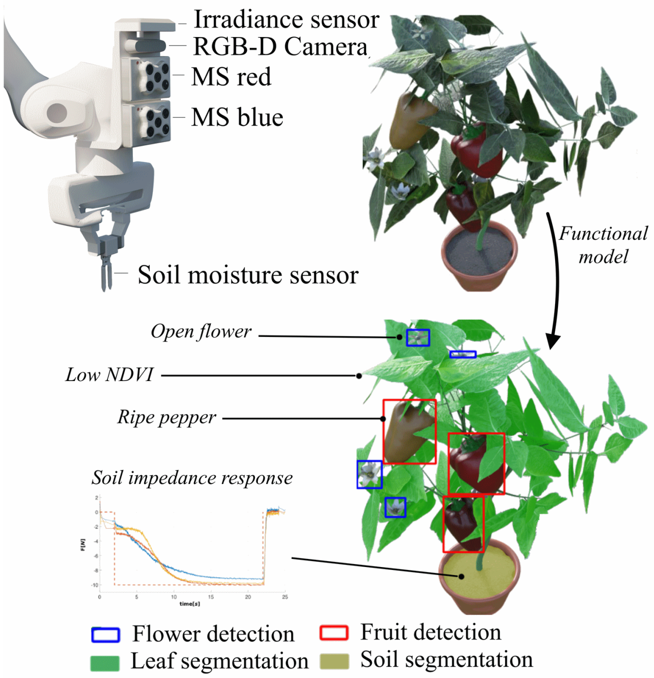

Fig. 6: A multi-modal sensory setup in an eye-in-hand configuration can produce a functional 3D model of a plant, suitable for manipulation planning. The multi-spectral camera system RedEdge-MX includes two MS cameras, *blue* and *red*, each capturing five distinct spectral bands, and an irradiance light sensor. It is combined with classical soil moisture sensors and compliant control to feel the plant. Storing information from different modalities we can draw conclusions about plant health and growth cycle.

robot vision module developed within the SpECULARIA project relies on a combination of state-of-the-art approaches, namely multi-modal sensory apparatus and CNN-based inference, in an attempt to generate a functional 3D model of a plant suitable for manipulation planning.

A. Structural plant model

The sensory setup of SpECULARIA robotic system provides multi-spectral raw data recorded in the indoor farmhouse workspace. The data is processed through the developed perception system and the final result is a structural plant model with all relevant functional parts needed for robotic manipulation planning, such as the one shown in Fig. 6. The model should contain concise information for planning and execution of manipulation activities. For example, harvesting relies on the detection of target fruit, and on the segmentation of leaves to plan a harmless grasp. These modules are developed as MobileNet V2 SSD architecture for detection, and DeepLab V3 semantic segmentation model, as a compromise between the accuracy and the inference speed of these simpler models, to satisfy the requirements of real-time robot control. Manual pollination and yield control rely on flower detection. Watering is planned depending on relevant vegetation indices estimation, and the execution is safe thanks to the compliant control algorithm relying on soil surface segmentation. The multi-modal semantic information of the scene is incorporated into the 3D functional model for each plant instance, allowing planning and execution of treatment procedures, and recording development over time.

B. Multi modal feel and appearance sensing

The perception in compliant manipulation relies on visual feedback but goes beyond the state of the art by combining multiple sensing modalities. A set of multi-spectral cameras is introduced into the perception hardware setup in addition to an RGB-D camera, since decision-making for automation in the agricultural domain often relies on different vegetation indices such as the normalized difference vegetation index (NDVI).

The developed sensory setup shown in Fig. 6, in which multi-spectral cameras are calibrated with respect to an RGB-D camera, was presented as the Multispectral-Depth (MS-D) system [38]. This system consists of an Intel RealSense RGB-D camera and a RedEdge-MX dual camera system installed on a custom-designed mount, as shown in Fig. 6 on the left. The calibration of all 10 spectral cameras of this MS-D is developed using a robotic manipulator that enables it to later be detached and used manually or with another manipulator [38]. More importantly for the SpECULARIA paradigm, it also enables accurate registration of close-range multi-spectral images, extending multi-spectral sensing into close-range sensing, as opposed to state-of-the-art remote multi-spectral sensing approaches, such as aerial imaging of fields using UAVs.

Measuring a plant's appearance can provide invaluable insight into its health, but it is only the first step a farmer would take to learn more about the plant. After observing it, the farmer usually approaches it to feel the plants and draws conclusions based on touch. Relying on compliant control algorithms, robots can interact with plants with no a priori knowledge of their shape [5]. Attaching moisture sensors as end effectors allows us to dig deep into the soil around the plant and measure water conditions near the roots. Compliant control algorithms ensure there is no harm done to the roots of the plant. Cleaning, drying out, and recalibrating sensors before each attempt enables us to achieve consistent results and measure across multiple plants or even multiple positions around a single plant. Cleaning is an important step that stops us from spreading potential pests or diseases across plants, which can be automated or performed in collaboration with farmers. All of these procedures would be hard, if not impossible to program on a moving mobile manipulator platform in a constantly changing environment.

The same approach to deriving unknown mathematical models based on the autoencoder shown in III, is used to fuse feel and appearance measurements. We train using the latent space between the encoder and decoder part, which holds important information about the unmodeled features in the dataset. By decoding the multispectral images into matrices and combining them with impedance and moisture sensor measurements, we can train a model to recognize the functional model of the plant.

C. Dataset generation

One of the key challenges in developing robotic visual perception systems in unstructured (non-industrial) environments, such as agriculture, is the need for large, high-quality training datasets. These datasets are critical for training CNN models to recognize and respond to different plants, growth stages, and environmental conditions. Collecting real-world data can be costly and time-consuming, which has led researchers to explore the use of synthetic data generation techniques. This is especially relevant for the SpECULARIA use-case, where many different plant parts need to be detected, segmented, and localized. The synthetic approach can, however, suffer from a domain gap, a challenge that reduces the capabilities of the CNN models trained on synthetic data to generalize to real-world data.

Inspired by the learn-by-demonstration approach in collaborative robotics, we have developed a collaborative dataset generation and labeling procedure capable of overcoming the domain gap issues arising in synthetic data generation. Similar to the learn-by-demonstration concept, in our semi-automated labeling procedure a human operator teaches the robot to generate a set of labeled examples on a single sample. This collaborative AI approach finally generates a functional plant model, carrying both semantic and geometric information about the plant state. Combining these two approaches enables us to build reliable, feature-rich datasets.

1) Synthetic data generation: Thanks to the assumption of complete control over the working conditions in the indoor robotic manipulation setup such as the one envisioned within SpECULARIA, we can create a realistic simulation environment that enables us to train the network and deploy it more easily for domain adaptation. SpECULARIA synthetic generation pipeline is developed on top of Blender, ROS, and Gazebo [39]. An RGB-D camera is simulated on top of Blender scene rendering capabilities and used as a sensor in the Gazebo robotic simulation. Then, a simulated robot with an eye-in-hand camera can be used to generate an image dataset of an object from a predefined set of viewpoints.

In the plant model generation pipeline, as shown in Figure 7, simulated 3D models are generated using real objects. First, we generate a single realistic model of an object. For example, a realistic pepper model is generated by applying photogrammetry onto a set of photographs of a real pepper. Models of leaves and flowers can be extracted from the photographs as planar objects. These objects are then provided with a third dimension using 3D mesh manipulations. Other parts such as the stem and plant pot can be modeled manually, and enriched with realistic textures for better domain adaptation.

These simulated models are then used in the procedural generation to create an entirely realistic and diverse synthetic dataset by employing smart augmentation techniques. These include a pseudo-random choice of plant parts, their visual, positional with morphological parameters, and background images [8]. During the simulation rendering process, auto-generated labels are produced for object detection, localization, and semantic segmentation. We filter the pixel-perfect labels to more closely resemble the level of detail of manually labeled data [9].

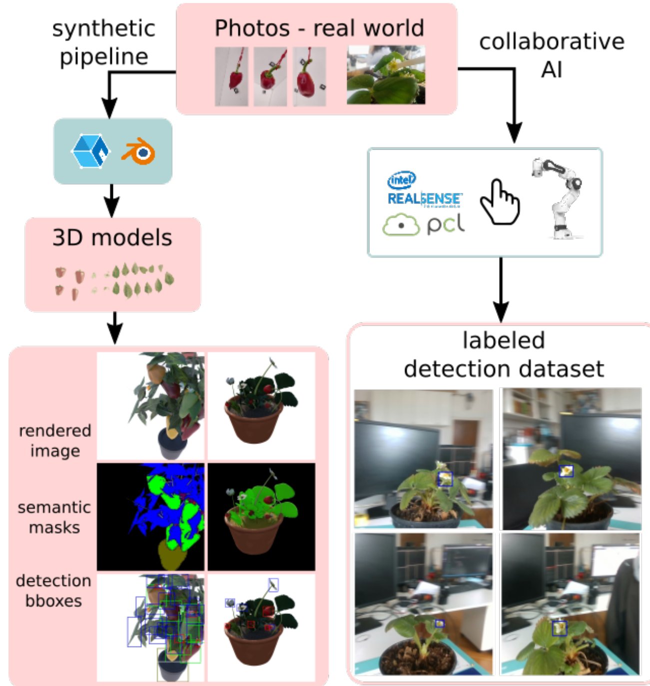

Fig. 7: Pipelines developed for training dataset generation in SpECULARIA activities. The synthetic approach relies on Meshroom for photogrammetric 3D model generation. Blender is used as the backbone of the rendering and labeling procedure. Examples of synthetic data for deep learning model training used and publically released under SpECULARIA project shown as results of the synthetic pipeline. Semi-automated dataset labeling on real world images is designed as a collaborative procedure, where a human operator and cobot equipped with an RGB-D camera cooperate. Examples of the dataset generated through a semi-automated labeling pipeline, along with automatically generated labels is shown in as the result of the collaborative pipeline.

2) Semi-automated annotation: To overcome the limitations of synthetic data, a semi-automated annotation method was developed in [6], where a human operator teaches a robotic system to generate a set of labeled samples by manually labeling a single sample. The setup consists of an RGB-D camera mounted in an eye-in-hand configuration as in the MS-D system, along with a set of real-world examples of the object to be detected. The proof of concept was shown in the strawberry flower detection use case, where the robot records the plant with visible flowers as detection targets from a set of predefined poses along a pre-computed trajectory. The result of this process is a database of unlabeled samples containing an RGB image, a depth image, and the global pose of the camera thanks to the calibrated eye-in-hand RGB-D camera.

The collaboration relies on a human operator who annotates the targeted object (in this case, the flower) by drawing a minimum bounding rectangle around it in a single RGB image. The 3D position of the labeled object is obtained by applying the bounding rectangle to the organized point cloud, and the corresponding 3D position is located in all the other recorded samples. Projecting the identified 3D points into their respective RGB images, segments the object of interest and provides a potentially large set of automatically annotated samples of an object recorded from different perspectives relying on a single manual annotation by a human user. One of the main benefits of this approach is relying on the same distribution of input data during the training and deployment phases. Models trained on higher-quality images tend to fail in real applications due to a significant domain gap. Exploiting the structured and constant environment around the dual-arm robotic manipulation system, we use the same sensory setup for both training and validation, resulting in a robust perception module regardless of hardware limitations.

V. Robot Cooperation VS Mobile Manipulation

To evaluate the SpECULARIA system, we envision an opposing robotic setup having a single, more complex (and more expensive) mobile manipulator, as a representative of the state of the art robotised greenhouse approach. The systems are compared in the same greenhouse, which consists of several rows of growth tables with bell pepper plants. Special care was taken to ensure that this layout is fair, in the sense that both systems could reach and handle all plants. For the multi-robot system, two workstations are provided on the sides of the manipulator, where two mobile robots can operate. The UGVs bring the plants to the workstations of the static robot manipulator arm. The underlying assumption is that the manipulator can only treat the front half of the plant. To reach behind the plant, the manipulator requires cooperation from the UGV to rotate the plant. The opposing system, which relies on the mobile manipulator, must therefore go around to reach the other side of the plant. Consequently, the proposed greenhouse is spread out so that the mobile manipulator can reach both sides of the plants.

To enable the heterogeneous robot teams to work together, we deploy the developed mission planning and coordination procedure for both types of systems. Mission planning is a complex problem that is often summarized with two distinct sub-problems, namely the task allocation problem (the question of *who does what?*) and the task scheduling problem (the question of *how to arrange the tasks in time?*).

The tasks we model in this paper fall into the class of problems XD[ST-SR-TA] defined in the taxonomy in [40]. These task types require execution by a single robot (SR, single-robot tasks), and robots are allowed to execute only one task at a time (ST, single-task robots). The task allocation and scheduling procedure considers both current and future tasks (TA, time-extended assignment). In terms of complexity, these tasks involve cross-schedule dependencies (XD), where the constraints of one robot depend on the plans of other robots.

A. Problem definition and modeling

The mission planning procedure considers a problem where a team of heterogeneous robots $R = \{1, \dots, m\}$ is available to perform a collection of simple single agent tasks (*actions*) $A = \{1, \dots, n\}$. A solution to the problem described is a set of time-related actions (*schedule*) for all robots. Formally, the schedule s_i for each robot $i \in R$ is defined as $s_i = \{(a, a^s, a^f)\ \forall a \in S_i\}$, where S_i is the set of actions assigned to robot i, and $a^s(a^f)$ are the start (finish) times of action a.

Each action $a \in A$ is associated with a couple $(d_a(i), c_a(i)), \forall i \in R$, where $d_a(i)$ represents the duration of action a when executed by robot i and $c_a(i)$ represents its cost. Each robot estimates the duration and cost of a future action based on the current state of the system and its capabilities.

All solutions to the problem must satisfy the *precedence constraints* stated in the mission specification: If action $a \in A$ must finish before action $b \in A$ begins, a constraint is generated as $prec(a, b)$. This constraint enforces $a^f < b^s$, where a^f and b^s specify the times when action a finishes and b begins.

In solving this problem, we model task planning as Multi-Depot Vehicle Routing Problem (MDVRP) with heterogeneous fleet and precedence constraints. Essentially, MDVRP is a routing problem for vehicles with limited payloads to pick up or deliver items at different locations. The goal is to find routes for vehicles with the lowest cost without exceeding the vehicle capacity. Given the many similarities between task planning and MDVRP, by modeling the task planning problem as MDVRP, we can apply many optimization techniques already available for this model to our problem. The concept of a depot in MDVRP is directly associated with the initial robot state, and the vehicle in MDVRP represents the robot itself. The concept of customer and customer demand relates to actions in task planning and the cost of each action, respectively. Consequently, routes as solutions to VRP problems represent ordered actions in the final robot schedules in the task planning model.

B. Solution approach

We employ multi-objective optimization using a distributed genetic algorithm with mimetism and knowledge sharing to solve the specified problem. This approach produces near-optimal solutions offering time scalability properties. The method we use is inspired by the Coalition-Based Metaheuristic (CBM) algorithm [41], with a specific implementation to meet our problem requirements.

The algorithm is similar to Genetic Algorithm (GA) in that the solutions in both are represented as chromosomes that are modified by genetic operators during optimization. An important difference is that in CBM the choice of the operator to apply is not completely stochastic, but relies on stored information about past actions and their potential to obtain solutions more likely to lead to better results. Moreover, CBM is a distributed algorithm that runs on multiple agent nodes (i.e., robots). The robots can share the acquired knowledge and the suboptimal solutions they have found. Therefore, each robot not only learns from its own experience but also exhibits mimetic behavior.

Inspired by an evolutionary process, solutions expressed as chromosomes contain genetic material (genotype) that defines each solution. For our problem, this refers to the assignment of actions to different robots and their order within the schedule. Each chromosome is associated with a phenotype that evaluates the genotype and generates schedules for task sequences based on the temporal properties of the tasks (task duration, time of transitioning between tasks).

The algorithm maintains a population of solutions to which it applies various genetic operators. The individuals in the population are evaluated using a double-rank strategy. Specifically, we use the makespan of the schedule and the total cumulative cost as criteria. Makespan is the elapsed time between the start and finish of a sequence of tasks in a group of robots. In the first part of the evaluation, we use a Pareto ranking procedure that assigns ranks to all solutions based on the non-dominance property (i.e., a solution with a lower rank is superior to solutions with higher ranks regarding all objectives). Therefore, the solutions are stratified into multiple ranks based on their ability to meet the optimization objectives. The second part of the evaluation is the density function, which determines the solution's similarity to other individuals in the population. Finally, rank and density scores are combined in the fitness of a solution *fitness* $\in \mathbb{R}$.

We chose genetic crossover and mutation as search guiding operators and adapted them to the specifics of our problem. The implemented crossover operator is a version of the Best-Cost Route Crossover (BCRC). For mutation, we use the intra-depot and inter-depot swapping operators, as well as single customer rerouting. The details of the individual operators can be found in [42].

C. Simulation setup

To evaluate the proposed approach, we set up the SpECULARIA system in the Gazebo simulator. The static greenhouse map created in Blender is known to all robots for navigation purposes. We assume ideal robot localization, since this is realistically achievable in such a determinate environment. The system including mission planning is based on ROS, and we use the OMPL library for path planning.

The greenhouse layout we used for the simulation is shown in Fig. 8. The structure consists of eight tables, each with four bell pepper growth pods. The stationary manipulator workstation is located in the center of the structure. For this layout, we ran several simulations with a random number of peppers per plant. First, we evaluated the performance of the mission planner for the two configurations, a single mobile manipulator versus a stationary manipulator assisted by two UGVs, on a fruit-picking scenario, using the a priori set-up ground truth numbers.

TABLE I: Simulation setups for the use case. For each setup, a random number of fruit per bell pepper plant was generated for 20 of 40 available plants.

simulation setup	1	2	3	4	5
total pepper number	61	80	89	113	118
average pepper number per plant	3.05	4.0	4.5	5.7	6.0

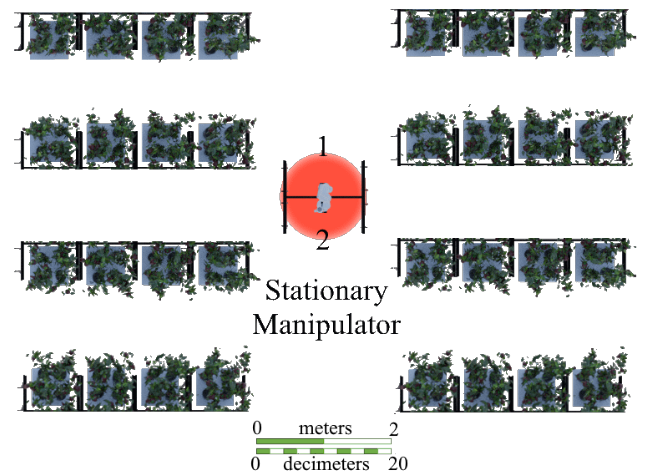

Fig. 8: Showing the layout of the proposed greenhouse structure consisting of eight tables, each with four bell pepper growth pods. The stationary manipulator workstation is located in the center of the structure.

To evaluate the planner, we randomly generated 20 of the 40 available plants to bear fruit. For these plants, a random number of peppers between 1 and 5 was generated from a uniform distribution for the left and right sides of the plant. The simulation setups for this use case in terms of the number of peppers are listed in Table I, where we report the total number of fruits in the entire greenhouse and the average number of peppers per plant.

The duration of pepper fruit harvesting was estimated at $10s$. The energy cost of the mission for a Pioneer 3-DX robot used to transport plants is estimated at around $30W$. In the case of the mobile manipulator, we propose a more robust solution of a Clearpath Robotics Husky A200 with a drive power of $400W$. The maximum speeds for both robots are set to $0.5m/s$, considering the sensitive payloads they carry, namely the heavy and expensive equipment, or plants in the other case. Given the duration of each task, calculated using the path length determined by OMPL on a static map and the speed of the robot, we calculate the energy consumption in kJ by multiplying the duration by the power requirement. This is a simplified form of a cost function and serves the purpose of testing the planning system. For a more accurate model, one should take into account the robots' battery capacity.

D. Evaluation results

The comparative results of the two system configurations are shown in Fig. 9. The static configuration uses a stationary manipulator assisted by two UGVs, and the mobile configuration assumes a robotic arm aboard a larger UGV. Based on the results, we can see that the mobile configuration of the system has a faster execution of the mission (lower makespan). This is to be expected since the plants are processed directly and the travel time of the tasks associated with each plant operation is lower. However, this speed comes at an overall higher energy cost. For the static manipulator configuration, the total cumulative cost is on average 8.4 times lower than the setup with a more robust UGV unit.

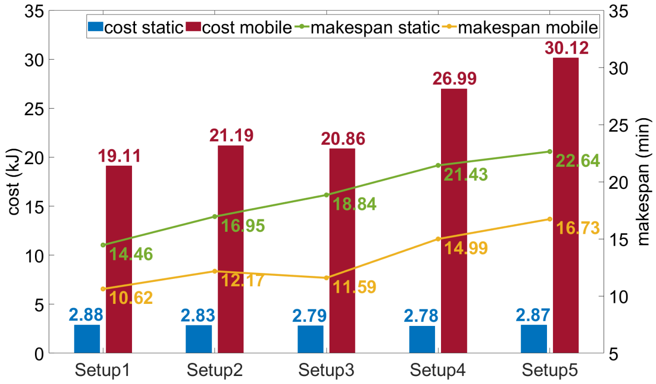

Fig. 9: Total cumulative costs and makespans of generated schedules on five different simulation setups for the two robotic system configurations. Based on the duration of each task, we calculate the energy cost in kJ by multiplying the duration and the power requirement of each configuration. The static configuration uses a stationary manipulator assisted by two UGVs and the mobile configuration assumes a robotic arm onboard a larger UGV.

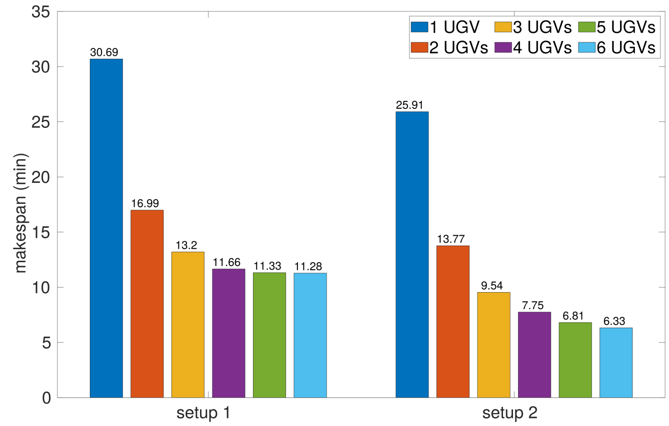

Fig. 10: Comparison of makespans of a robotic system consisting of a stationary manipulator assisted by a varying number of UGVs $\in [1, \dots, 6]$. The duration of the operation for each bell pepper fruit was estimated to be $10s$ and $5s$ in Setup 1 and Setup 2, respectively.

Fig. 11: Showing the dual arm manipulation system picking various fruit, pollinating flowers, and pruning the plant. We are showcasing the results of our first experiment, growing 60 sweet peppers, 60 hot peppers and 40 strawberries, demonstrating the feasibility of the proposed system.

To reduce the discrepancy of makespans for these two setups, one solution is to support the stationary manipulator with more transport UGVs. Such a setup reduces the potential idle time of the manipulator while waiting for the plants to be brought for treatment, as multiple robots supply each workstation slot. In Fig. 10, we can observe how the number of transport UGVs affects the total makespan of the mission. Some plant treatments might take less time to complete. For instance, spraying fruit against pesticides might be faster than picking fruit. To test the effect of the duration of the treatment we simulated two scenarios with time estimates around $10s$ and $5s$ in Setup 1 and Setup 2, respectively. From the results, we can observe how increasing the number of UGVs rapidly decreases the mission makespan until a certain saturation point where adding more UGVs to the system barely affects the mission duration. Using these two scenarios, we can see that the impact of adding more UGVs to the system is more pronounced for missions with shorter operation durations for each fruit. In Fig. 10, we can observe a more distant saturation point on the right plot with $5s$ operations, as opposed to the left graph with $10s$ operations.

Overall, the use of a stationary manipulator with supporting UGVs has many advantages, both economically and in terms of the long-term deployment potential of the system. Since larger UGVs capable of carrying a robotic arm consume more energy, batteries would need to be recharged more frequently, resulting in a lot of idle time during which the robots are not performing useful work. In addition, these robots tend to be more expensive than smaller UGVs, so the system with multiple smaller UGVs still makes more economic sense.

VI. Conclusion

Collaborative robots can be highly effective in alleviating the labor costs of organic agriculture, even in small and medium-size production scales. State-of-the-art micro sensing solutions are used to build multi-modal functional models of plants for timely and autonomous monitoring, planning and execution of agrotechnical processes, as shown in multiple experiments conducted within SpECULARIA project (Fig 11). With robotic perception built from deep and machine learning models, such as joint feature extraction and multi-modal information fusion in autoencoder neural network architecture, these complex systems can now be run on simple and affordable hardware, bringing state-of-the-art robotics and AI research results closer to the market. Another important takeaway is to consider new paradigms, such as replacing the expensive highly specialized robot with a heterogeneous team of simpler robots, resolving the same tasks as efficiently thanks to development of novel planning solutions.

VII. Acknowledgement

This work has been financed by Croatian Science Foundation under the project Specularia UIP-2017-05-4042. The work of doctoral student Jelena Vuletic has been supported in part by the "Young researchers' career development project–training of doctoral students" of the Croatian Science Foundation funded by the European Union from the European Social Fund.

References

[1] J. C. Loudon, *An Encyclopædia of Gardening*. Longman, Rees, Orme, Brown and Green, 1826.

[2] I. Charania and X. Li, “Smart farming: Agriculture’s shift from a labor intensive to technology native industry,” *Internet of Things*, vol. 9, p. 100142, 2020.

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

[4] C. Lehnert, I. Sa, C. McCool, B. Upcroft, and T. Perez, “Sweet pepper pose detection and grasping for automated crop harvesting,” in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 2428–2434.

[5] M. Polic, M. Car, F. Petric, and M. Orsag, “Compliant plant exploration for agricultural procedures with a collaborative robot,” *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2768–2774, 2021.

[6] J. Vuletić, M. Polić, and M. Orsag, “Robotic strawberry flower treatment based on deep-learning vision,” in *International Workshop on Human-Friendly Robotics*. Springer, Cham, 2023, pp. 189–204.

[7] M. Polic, M. Car, J. Tabak, and M. Orsag, “Robotic irrigation water management: Estimating soil moisture content by feel and appearance,” in *2022 19th International Conference on Ubiquitous Robots (UR)*. IEEE, 2022, pp. 16–22.

[8] M. Polic, J. Tabak, and M. Orsag, “Pepper to fall: a perception method for sweet pepper robotic harvesting,” *Intelligent Service Robotics*, vol. 15, no. 2, pp. 193–201, 2022.

[9] J. Vuletić, M. Polić, and M. Orsag, “Procedural generation of synthetic dataset for robotic applications in sweet pepper cultivation,” in *2022 International Conference on Smart Systems and Technologies (SST)*. IEEE, 2022, pp. 309–314.

[10] D. Ball, P. Ross, A. English, P. Milani, D. Richards, A. Bate, B. Upcroft, G. Wyeth, and P. Corke, “Farm workers of the future: Vision-based robotics for broad-acre agriculture,” *IEEE Robotics & Automation Magazine*, vol. 24, no. 3, pp. 97–107, 2017.

[11] A. Pretto, S. Aravecchia, W. Burgard, N. Chebrolu, C. Dornhege, T. Falck, F. Fleckenstein, A. Fontenla, M. Imperoli, R. Khanna *et al.*, “Building an aerial–ground robotics system for precision farming: an adaptable solution,” *IEEE Robotics & Automation Magazine*, vol. 28, no. 3, pp. 29–49, 2020.

[12] S. Patnaik, S. Sen, and M. S. Mahmoud, *Smart Village Technology: Concepts and Developments*. Springer, 2020.

[13] R. Barth, J. Hemming, and E. J. van Henten, “Design of an eye-in-hand sensing and servo control framework for harvesting robotics in dense vegetation,” *Biosystems Engineering*, vol. 146, pp. 71–84, 2016.

[14] C. Lehnert, A. English, C. McCool, A. W. Tow, and T. Perez, “Autonomous sweet pepper harvesting for protected cropping systems,” *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 872–879, 2017.

[15] Y. Xiong, Y. Ge, L. Grimstad, and P. J. From, “An autonomous strawberry-harvesting robot: Design, development, integration, and field evaluation,” *Journal of Field Robotics*, vol. 37, no. 2, pp. 202–224, 2020.

[16] S. Abondance, C. B. Teeple, and R. J. Wood, “A dexterous soft robotic hand for delicate in-hand manipulation,” *IEEE Robotics and Automation Letters*, vol. 5, 2020.

[17] E. Navas, R. Fernandez, D. Sepulveda, M. Armada, and P. Gonzalez-De-Santos, “Soft gripper for robotic harvesting in precision agriculture applications,” *2021 IEEE International Conference on Autonomous Robot Systems and Competitions, ICARSC 2021*, 2021.

[18] A. Gunderman, J. Collins, A. Myers, R. Threlfall, and Y. Chen, “Tendon-driven soft robotic gripper for blackberry harvesting,” *IEEE Robotics and Automation Letters*, vol. 7, pp. 2652–2659, 4 2022.

[19] L. Fu, F. Gao, J. Wu, R. Li, M. Karkee, and Q. Zhang, “Application of consumer rgb-d cameras for fruit detection and localization in field: A critical review,” *Computers and Electronics in Agriculture*, vol. 177, p. 105687, 2020.

[20] S. Hinterstoisser, O. Pauly, H. Heibel, M. Martina, and M. Bokeloh, “An annotation saved is an annotation earned: Using fully synthetic training for object detection,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019, pp. 0–0.

[21] S. Khan, B. Phan, R. Salay, and K. Czarnecki, “Procsy: Procedural synthetic dataset generation towards influence factor studies of semantic segmentation networks.” in *CVPR Workshops*, 2019, pp. 88–96.

[22] M. Di Cicco, C. Potena, G. Grisetti, and A. Pretto, “Automatic model based dataset generation for fast and accurate crop and weeds detection,” in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 5188–5195.

[23] J. Olatunji, G. Redding, C. Rowe, and A. East, “Reconstruction of kiwifruit fruit geometry using a cgan trained on a synthetic dataset,” *Computers and Electronics in Agriculture*, vol. 177, p. 105699, 2020.

[24] K. Zhang, Q. Wu, and Y. Chen, “Detecting soybean leaf disease from synthetic image using multi-feature fusion faster r-cnn,” *Computers and Electronics in Agriculture*, vol. 183, p. 106064, 2021.

[25] R. Barth, J. IJsselmuiden, J. Hemming, and E. J. Van Henten, “Data synthesis methods for semantic segmentation in agriculture: A capsicum annuum dataset,” *Computers and electronics in agriculture*, vol. 144, pp. 284–296, 2018.

[26] C. Doersch and A. Zisserman, “Sim2real transfer learning for 3d human pose estimation: motion to the rescue,” in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Curran Associates, Inc., 2019.

[27] A. Pashevich, R. Strudel, I. Kalevatykh, I. Laptev, and C. Schmid, “Learning to augment synthetic images for sim2real policy transfer,” in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 2651–2657.

[28] T. Kiyokawa, K. Tomochika, J. Takamatsu, and T. Ogasawara, “Fully automated annotation with noise-masked visual markers for deep-learning-based object detection,” *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1972–1977, 2019.

[29] V. Ilin, I. Kalinov, P. A. Karpyshev, and D. Tsetserukou, “Deepscanner: a robotic system for automated 2d object dataset collection with annotations,” *ArXiv*, vol. abs/2108.02555, 2021.

[30] M. Otte, M. Kuhlman, and D. Sofge, “Multi-robot task allocation with auctions in harsh communication environments,” in *2017 International Symposium on Multi-Robot and Multi-Agent Systems (MRS)*, 2017, pp. 32–39.

[31] H. Mitiche, D. Boughaci, and M. Gini, “Iterated local search for time-extended multi-robot task allocation with spatio-temporal and capacity constraints,” *Journal of Intelligent Systems*, vol. 28, no. 2, pp. 347–360, 2019.

[32] G. A. Korsah, “Exploring Bounded Optimal Coordination for Heterogeneous Teams with Cross-Schedule Dependencies,” 1 2011.

[33] J. G. Martin, J. R. D. Frejo, R. A. García, and E. F. Camacho, “Multi-robot task allocation problem with multiple nonlinear criteria using branch and bound and genetic algorithms,” *Intelligent Service Robotics*, vol. 14, no. 5, pp. 707–727, 2021.

[34] W. Hess, D. Kohler, H. Rapp, and D. Andor, “Real-time loop closure in 2d lidar slam,” in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 1271–1278.

[35] M. Kokot, D. Miklić, and T. Petrović, “Path continuity for multi-wheeled agvs,” *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7437–7444, 2021.

[36] M. Polic, I. Krajacic, N. Lepora, and M. Orsag, “Convolutional autoencoder for feature extraction in tactile sensing,” *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 3671–3678, 2019.

[37] D. Stuhne, J. Tabak, M. Polić, and M. Orsag, “Design and prototyping of soft finger ai-enabled hand (sofia),” in *2022 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*. IEEE, 2022, pp. 1581–1586.

[38] J. Vuletić, M. Car, and M. Orsag, “Close-range multispectral imaging with multispectral-depth (ms-d) system,” *Biosystems Engineering*, vol. 231, pp. 178–194, 2023.

[39] A. Ivanovic, M. Polic, J. Tabak, and M. Orsag, “Render-in-the-loop aerial robotics simulator: Case study on yield estimation in indoor agriculture,” in *2022 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, 2022, pp. 787–793.

[40] G. A. Korsah, A. Stentz, and M. B. Dias, “A comprehensive taxonomy for multi-robot task allocation,” *The International Journal of Robotics Research*, vol. 32, no. 12, pp. 1495–1512, 2013.

[41] D. Meignan, A. Koukam, and J.-C. Créput, “Coalition-based metaheuristic: A self-adaptive metaheuristic using reinforcement learning and mimetism,” *Journal of Heuristics*, vol. 16, pp. 859–879, 12 2010.

[42] B. A. Ferreira, T. Petrović, M. Orsag, J. R. M. de Dios, and S. Bogdan, “Distributed allocation and scheduling of tasks with cross-schedule dependencies for heterogeneous multi-robot teams,” https://arxiv.org/abs/2109.03089, 2021.