Cooperative Grape Harvesting Using Heterogeneous Autonomous Robots

Chris Lytridis, Christos Bazinas, Ioannis Kalathas, George Siavalas, Christos Tsakmakis, Theodoros Spirantis, Efthicia Badeka, Theodore Pachidis and Vassilis G. Kaburlasos *

HUMAN-Lab, International Hellenic University (IHU), 65404 Kavala, Greece; lytridic@cs.ihu.gr (C.L.); chrbazis@cs.ihu.gr (C.B.); kalathas@emt.ihu.gr (I.K.); nesiava@cs.ihu.gr (G.S.); christssakm@gmail.com (C.T.); theospyr1@teiemt.gr (T.S.); evbadek@cs.ihu.gr (E.B.); pated@cs.ihu.gr (T.P.)

* Correspondence: vgkabs@teiemt.gr; Tel.: +30-2510-462-320

Abstract: The development of agricultural robots is an increasingly popular research field aiming at addressing the widespread labor shortages in the farming industry and the ever-increasing food production demands. In many cases, multiple cooperating robots can be deployed in order to reduce task duration, perform an operation not possible with a single robot, or perform an operation more effectively. Building on previous results, this application paper deals with a cooperation strategy that allows two heterogeneous robots to cooperatively carry out grape harvesting, and its implementation is demonstrated. More specifically, the cooperative grape harvesting task involves two heterogeneous robots, where one robot (i.e., the expert) is assigned the grape harvesting task, whereas the second robot (i.e., the helper) is tasked with supporting the harvesting task by carrying the harvested grapes. The proposed cooperative harvesting methodology ensures safe and effective interactions between the robots. Field experiments have been conducted in order firstly to validate the effectiveness of the coordinated navigation algorithm and secondly to demonstrate the proposed cooperative harvesting method. The paper reports on the conclusions drawn from the field experiments, and recommendations for future enhancements are made. The potential of sophisticated as well as explainable decision-making based on logic for enhancing the cooperation of autonomous robots in agricultural applications is discussed in the context of mathematical lattice theory.

Keywords: agricultural robots; cooperation; harvesting; cooperative harvesting; robot operating system; autonomous navigation

1. Introduction

The use of robots in agriculture has been proposed as a solution to various challenges facing the agricultural industry including increasing farming costs, labor shortages, increasing demands for food production, and the need for environmentally friendly farming solutions [1–6]. The potential benefits of employing robots in agriculture have driven research for several years and have resulted in experimental as well as commercially available robotic platforms. The implementation of operational agricultural robots involves an array of supporting technologies including localization (in a greenhouse or outdoor environments), navigation (within crop rows), vision (target plant recognition, e.g., for banana detection [7], pomegranate identification [8], grape cluster detection [9,10] or plant detection for weed control [11]), robotic manipulator control, and power management to name a few. Vision in particular is considered essential for fully autonomous operation. The proposed solutions are usually domain-specific instead of being comprehensive as unifying solutions. Examples of the application of agricultural robots can be found in various crops and for various associated tasks such as corn fertilization [12], weed control in lettuce and broccoli [11], or harvesting strawberries [13]. Automated sensing solutions for viticultural and vinicultural tasks have also been pursued in mapping, monitoring, and
management [14,15], including using robots in typical tasks performed by humans, such as pruning grapevines [16], harvesting [17], spraying [18], monitoring [19,20], maturity estimation [21], and weeding [22].

Cooperative robots in agriculture have been proposed and developed in research settings [23–26]. There are some advantages to deploying multi-robot teams in an agricultural setting. Most significantly, multiple robots can carry out the work in a field in less time than a single robot would, because a larger area can be covered while the robots are working in parallel. Most research studies in this area have focused on tasks such as spraying [27,28] and monitoring [29]. These approaches take advantage of the ability of robot teams to cover more area than when a single robot is used. In this case, it is common that the robot team is comprised of identical robots, with all members having the same capabilities regarding task completion. Therefore, control, coordination, area coverage, localization, and navigation are the main themes explored. A mechanism for area allocation and path coordination is required [30–33]. In other cases, there are tasks that can be undertaken more effectively by using cooperating robots. The task can only be completed by heterogeneous robots, whose skills are complementary. Examples of this can be found when heterogeneous robots are employed, each equipped with software and hardware designed to tackle different goals in the overall task. In many studies, cooperation in such a scenario is implemented as teams of unmanned ground vehicles (UGVs) and unmanned aerial vehicles (UAVs) working together. In this case, UAVs are tasked with mapping and monitoring the work area using cameras, while the ground robots carry out the actual task using manipulators (e.g., in [34]). In [35], a case study is described, where cooperating heterogeneous unmanned vehicles can maximize the information gathered from the field in order to construct crop models which would then be exploited for task planning. Multiple cooperating UGVs have been also used in rice harvesting [36].

Other cooperative approaches involve human–robot cooperation, as is the case in the study presented in [37]. A manned tractor acts as a leader while a robot tractor follows autonomously. There are also examples of cooperation between multiple manipulators in the literature, such as cooperative harvesting of aubergines [38] or coordinated apple harvesting [39]. In the grape harvesting task presented in [17], the proposed robot is designed to perform concurrent grape harvesting using two robotic arms. In [40], the robot tracks a human operator who manually picks fruit in order to assist by carrying the collection crate, but also to create a harvesting map.

There are only a few autonomous grape harvesting robotic systems reported in the literature (e.g. [17,41]) and, to the authors’ knowledge, there are no cooperative grape harvesting solutions. This paper proposes such a heterogeneous robotic system with the focus being on the cooperative execution of the harvesting task. The rationale behind this approach is the fact that viticultural robots have size limitations due to the narrow width of a vineyard row. Therefore, a single robot can only carry a limited number of grapes. By employing one or more helper robots dedicated to carrying grapes, this limitation is minimized. Extending the work presented in [42], which dealt with the coordinated navigation of two robots inside a vineyard, the present paper introduces and demonstrates a cooperative multi-robot harvesting system as a solution to the aforementioned grape capacity problem. The concepts presented in [42] are further refined in order to integrate the actual harvesting operation. The harvesting scenario that has been implemented for the purposes of this paper involves two heterogeneous robots navigating inside the rows of a grapevine in a leader–follower formation. The robot team consists of an expert robot responsible for harvesting and a helper robot responsible for carrying. At certain predefined positions, the robots stop, and the expert robot performs the harvesting, while the helper robot carries the harvested grapes. The paper describes the two robotic platforms and the additional equipment that were used in order to fulfill the requirements of their distinct roles in the cooperative harvesting scenario. In addition, the various software subsystems (sensing, navigation, communication) and the related algorithms necessary for the implementation of the cooperative robot system are examined. The desired functionality of the
robots has been validated in field experiments in an actual vineyard. The paper describes the field experiments, and the observations that have been made are discussed to support the validation of the system and the evaluation of the cooperative approach.

A theoretical consideration of this work regards decision-making based rigorously on logic. Note that decision-making in machine learning, including agricultural applications, is typically pursued through classic modeling techniques by optimizing an “objective function” using training data. Recall that classic modeling is traced back to work by Newton and Gauss, whereby a parametric model is fit to measured data in the Euclidean space $\mathbb{R}^n$ for optimal parameter estimation. However, objective functions cannot accommodate either semantics or logic. An interesting approach has been proposed in the literature for decision-making by an inclusion measure function in the context of the “lattice computing (LC) paradigm” including (lattice-ordered) data semantics as well logic for explainable decision-making [43], as discussed below.

The present paper is structured as follows: In Section 2, the hardware and software used in the field experiments are presented. In Section 3, the field experiments are described and the experimental results are presented. Section 4 discusses the experimental results, and in Section 5, conclusions and recommendations for further extensions of the work are made.

2. Materials and Methods

This section describes the robotic platforms and the various subsystems and algorithms that contribute toward the cooperative grape harvesting task. While all the subsystems are necessary for effective cooperative operation, emphasis is placed on the cooperative strategy proposed in this paper, the effectiveness (successful execution) and efficiency (task duration) of which will be examined in the field experiments.

2.1. Robotic Platforms

The robotic platforms chosen for the implementation of the harvesting experiments were the RB-Vogui and RB-Eken wheeled robots by Robotnik [44] because they offer complete solutions for indoor and outdoor operations and include sensing, navigation, and communication capabilities. The robots are designed such that they can also be customized to fit the needs of the desired task. In terms of software, since they are ROS-based systems, their software can be customized by integrating custom ROS nodes and packages. Both robots operate under ROS Melodic installed on a Linux-based computer. They are equipped with various sensors including 2D and 3D LiDAR sensors, GPS, and depth cameras. GPS accuracy was improved by pairing the robots with one Real-Time Kinematics (RTK) base for each robot. Additionally, the robots were equipped with temperature and humidity sensors in order to monitor environmental conditions as well as the temperature inside the robots.

Each robot was assigned a different role in the harvesting task. On one hand, the RB-Vogui robot (the expert robot) was assigned the actual harvesting role; for this, it was fitted with a Kinova Gen3 arm carrying a depth camera for grape recognition and a custom cutting tool as the end effector for cutting the grapes. Moreover, it has an Nvidia Jetson AGX Orin to handle the computationally intensive machine vision operations. To meet the power demands of the additional hardware, a 24 V battery was added and placed inside a custom steel box located at the back of the robot, which also houses the Orin computer, the electronics for the environmental monitoring sensors, and a network hub which connects the arm and the Orin computer with the robot. On top of this box, two removable grape cluster baskets were placed. On the other hand, the RB-Eken robot (the helper robot) acted as the helper in the harvesting task by collecting and storing the harvested grapes, since it has a much larger load capacity (300 kg) than the RB-Vogui robot. It features a Universal Robots UR10e arm (12.5 kg of load capacity) with an OnRobot RG2 gripper as the end effector and a large storage basket for storing grapes.
2.2. Software Architecture

The software developed for the various functions carried out by the robots was arranged as ROS packages in the robot operating system installed on the robots. Each package has its own functionality and handles a particular task. These custom packages work in conjunction with the pre-installed packages which ensure the robots’ basic functionality such as sensing, motor control, power management, manipulator drivers, etc. Figure 1 shows the custom packages operating on the robots.

More specifically, the following packages were implemented:

- **Controller package.** The controller package is responsible for the coordination of all packages. It gathers information published from all packages and transmits it to the base station. It receives commands from the base station and initiates the task. It is also responsible for the communications between the robots.

- **Sensors package.** This is the package that collects the readings received from the custom Arduino-based data collection hardware that was specifically developed for these robots and includes temperature and humidity sensors.

- **Arm package.** This package includes methods that determine the manipulator’s movements and predefined poses. Depending on the arm type (Kinova Gen3 or UR10e) the arm package also includes methods to control movement sequences for harvesting and basket handling.

- **Navigation package.** The navigation package contains methods related to the robot’s localization, planning, and movement, including methods for positioning robots relative to the vineyard row and relative to each other. It also publishes the current position and orientation of the robot to the system.

- **Measurements package.** This package is responsible for gathering and synchronizing sensor measurements and statuses from all devices on the robotic system, and then publishing it to the system.

- **Gripper package.** This package is used to control the gripper and the cutting tool.

- **Marker package.** The marker package operates on the RB-Eken robot and is used to detect the Aruco marker fixed at the back of the RB-Vogui robot. When a marker is detected, the package publishes marker pose information to the system, to be used in other packages.

- **Task package.** This is the package that is responsible for initiating and managing the task received from the base station. Depending on the robot (RB-Vogui or RB-Eken) it executes the appropriate methods that perform the task and coordinates the robot’s actions. The package continuously publishes the status of the task to the system.

- **Vision package.** This package contains the methods concerning the machine vision operations required by the RB-Vogui robot. The package loads the machine vision models and, when requested by the system, grapes and stem recognition are performed. The package publishes the location of the detected objects.
2.3. Base Station

The base station consists of a laptop computer and a wireless access point to which the robots and the laptop are connected. Through the purpose-built base station’s software, the operator is able to select the desired task (in this case harvesting), plan the desired path, select where the robot is to perform harvest, and define various task parameters such as the percentage of grapes to be harvested in the case of green harvesting. Additionally, the operator can oversee the progress of the task by viewing the robots’ location in the vineyard, monitoring the various measurements collected by the robots’ sensors, and observing their camera feeds. In addition, the base station’s software allows for saving the data collected by the robots as well as their progress logs in a database for later examination.

In order to plan the path that the robots are to follow during the harvesting task, the user must first select the map of the target vineyard. Vineyard maps are produced in advance using images taken from an aerial drone, deployed to conduct a comprehensive aerial survey of a vineyard. These high-resolution images are then processed in such a way that the maps contain location information and features such as the vineyard rows and other obstacles. This processing is carried out in Agisoft Metashape, a software application specifically designed for processing aerial imagery. The software generates three map types: an orthomosaic, a high-resolution composite image of the vineyard, and a binary image. The binary image, in particular, is generated by extracting and classifying the Digital Point Cloud (DPC). This classification results in the creation of a Dense Elevation Model (DEM) comprising various classes of objects and ground points. By subtracting the DEM of ground points from the overall DEM, a DEM of differences (DDM) is obtained. This DDM serves to filter out the crop heights while excluding ground-related information. As a result, a binary DDM is produced, which is then used in the generation of the navigation path, in conjunction with the orthomosaic image. More specifically, the generation of the navigation path consists of three steps: (1) the preprocessing step, which provides the field images in an appropriate format, (2) the crop row detection algorithm which uses the Hough transform to detect the angle of the crop lines, and (3) the navigation mapping process to define the final route for the robots [45]. In the process of developing the navigation system for in-row waypoints, users have the flexibility to choose the desired distance between waypoints and save this configuration for future use. Additionally, the waypoints for turns between rows are determined by identifying the peaks of the rows, which are then used as circular points for navigation calculations. The user is able to select different areas of the vineyard in which the robots are to operate, and the corresponding waypoints for the entire mission (from an initial location to an end location) are generated.

The generated waypoints are characterized by whether they are locations where harvesting is to be performed or not. The user can select at which waypoints the robots are to stop and perform harvesting. This enables the robots to focus harvesting in specific locations (for example around vineyard trunks) where a larger concentration of grapes is expected to occur. It also enables green harvesting, usually taking place before veraison, where only a percentage of grapes need to be harvested.

Finally, the base station’s computer hosts a server based on the MQTT messaging protocol, which handles communications with the robots. Through the MQTT protocol, the base station can send commands to the robots, receive status reports from the robots, and forward messages sent from one robot to the other. The robots are equipped with the mqtt_bridge ROS package which is responsible for converting MQTT messages to ROS messages and vice versa. This allows two-way communication between the robots and the base station and between each other.

2.4. Localization and Navigation
In order to perform navigation, the robots need to first localize themselves in the environment. To achieve this, the robots possess a suitably transformed local copy of the aforementioned vineyard map, indicating the location of the vineyard rows. Localization is achieved using the adaptive Monte Carlo localization (AMCL) method [46]. More specifically, the AMCL method is a probabilistic algorithm that uses sensor readings to produce an estimate of the robot’s pose in 2D. The possible poses of the robots are represented as a distribution of particles. During the operation of the robot, the laser scans received from the LiDAR sensors are continuously compared against the known map, and a particle filter is applied in order to determine the likelihood of each particle being in the actual state (pose) of the robot. The particles with higher likelihoods of representing the actual pose of the robot are used in subsequent iterations of the algorithm to produce new particle distributions. The method is adaptive because the size of the sample sets varies using the KLD (Kullback–Leibler Distance) sampling method. The probabilistic nature of the method is suitable for robot localization in the vineyard since discrepancies between LiDAR measurements and the preloaded map are to be expected; as opposed to the preloaded map, vineyard rows do not have smooth surfaces due to varying foliage density. Therefore, starting from a known initial state (position and orientation), the pose of the robot needs to be estimated and continuously updated.

Having established the location and orientation of the robot in the vineyard, the robot needs to move from one location to the next. The map coordinates of the waypoints as estimated by AMCL are then used by the Timed Elastic Band (TEB) motion planning [47]. The algorithm generates an initial trajectory connecting the current location to the next waypoint. While the robot is moving, the trajectory is optimized in real-time based on multi-objective optimization. The objectives are the duration of motion and the distance from obstacles and are constrained by robot velocity and acceleration factors. The behavior of the algorithm can be altered by providing different weights to the optimization objectives and also by setting parameters such as the minimum allowed distance to the obstacles. The ability to adjust this particular parameter is important since a robot can be required to approach the vine at a small distance in order to perform the harvesting.

While AMCL and TEB motion planning operate on location in the map coordinate frame, the paths generated by the base station are lists of waypoints expressed in terms of GPS (latitude/longitude) coordinates. Therefore, for the robots to be able to plan and execute their motion, the received list of GPS waypoints is first converted to a list of local map coordinates at the beginning of the task. This is possible since the robots are also equipped with GPS devices whose signal is corrected using RTK bases, as mentioned in the previous section, and so the robots’ position in both coordinate systems (local map and GPS) is known at any time. The conversion of GPS latitude/longitude coordinates to map coordinates involves a series of steps. First, the GPS coordinates of the initial location of the robot and the desired GPS coordinates of a waypoint (goal location) have to be converted to the UTM (universal transverse Mercator) coordinate system. This involves the calculation of the meridional arc and then the eastings and northings, which are the distances in meters from the central meridian and the equator, respectively. Then, the difference between the initial and goal coordinates yields the map coordinates of the goal location according to Equation (1):

\[
\begin{align*}
    x_{\text{map}} &= x_{\text{utm}} - x_{\text{utm}_0} \\
    y_{\text{map}} &= y_{\text{utm}} - y_{\text{utm}_0},
\end{align*}
\]

where \((x_{\text{utm}_0}, y_{\text{utm}_0})\) and \((x_{\text{utm}}, y_{\text{utm}})\) are the UTM coordinates of the starting location and the goal location, respectively, and \((x_{\text{map}}, y_{\text{map}})\) are the resulting map coordinates of the goal location. These coordinates can then be utilized by the local planning TEB algorithm, and the robots are able to sequentially visit each of the waypoints of the prescribed path.

2.5. Coordinated Navigation
In addition to their ability to move toward prescribed locations, the robots are able to navigate in a coordinated manner inside the vineyard. The preliminary coordination algorithm was presented in [42]. In the leader–follower architecture adopted here, the expert robot is visiting the waypoints produced by the software of the base station and received at the beginning of the mission. At each waypoint, the expert robot transmits a command to the helper robot to move in such a way so that the helper robot is always a waypoint behind. In other words, the expert robot instructs the helper robot to always occupy the expert robot’s previous position. The expert robot waits until the helper robot reaches that position and is notified when this happens. In the case where the waypoint reached by the expert robot is designated as a harvesting position, the helper robot is instructed to prepare to support harvesting. Preparing for harvesting support means that the helper robot is required to first roughly approach the expert robot and then fine-tune its position so that it rests at a preset distance of 40 cm and also assumes the same orientation as the expert robot, using visual cues as is described later in Section 2.6. At this distance, the UR10e manipulator on the helper robot can reach the expert with ease and is able to assume any poses necessary to successfully carry out cooperative harvesting. The navigation algorithms for the coordination of the two robots are summarized in the flowcharts of Figure 2 below.

![Flowcharts](image)

**Figure 2.** The algorithms for the coordinated motions of the (a) expert robot (Robot A) and (b) the helper robot (Robot B).
2.6. Vision

Vision is used in two specific tasks, which are necessary for cooperative harvesting, namely grape and stem detection, and positioning the robots relative to each other in preparation for harvesting.

2.6.1. Grape and Stem Detection

The vision ROS package resides in the Jetson Orin platform, which is external to the robot’s main computer. It is therefore launched remotely at startup, but it is part of the same ROS network as the robot. The vision package is responsible for storing the vision pre-trained models and performs grape and stem recognition and localization via a 3D camera upon request from the main system. For this research, the You Only Look Once (YOLO) v7 E6E model was chosen since it is well-suited for object detection.

To train the model, a dataset that comprised 10,000 photos was collected. These photos represented various objects, such as grapes, trunk, and ripening phases of the vineyard. Two methods were used to capture the images: (a) a DJI Mavic Mini Drone, which flew between the rows of the vineyard, and (b) a handheld camera operated by an individual walking through the vineyard rows. To optimize the annotation process, the software “Label Studio” was used because it allows image annotation by multiple users working in parallel. The annotation process involved the participation of 70 contributors, who collectively annotated 3500 images using rectangular boxes. The resulting dataset consisted of eight predefined classes that the model had to recognize: Grape, Leaves, Irrigation System, Branches, Grass, Pillar, Stem, and finally the Trunk. However, for the harvesting task, only two classes, namely Grape and Stem, are considered. For the training process, the YOLO v7 E6E model was employed. The images collected from the various sources were resized to 1280 × 724 pixels and training was conducted using four Type A100 GPU cards. The dataset was divided into a training subset comprising 75% of the data, and the testing subset consisted of the remaining 25%. The training process spanned 2500 epochs, and the duration was approximately 10 days. Figure 3 shows the resulting precision for all classes after training.

![Figure 3](image.png)

Figure 3. Precision in object recognition for all classes in the dataset.

The evaluation results demonstrate the model’s effectiveness in accurately identifying all object classes. The results of Figure 3 show that the model is not very effective regarding the stem class which exhibits the lowest performance. This is due to the fact that the stem was clearly visible in significantly fewer images in the dataset compared to objects of other classes. Figure 4 shows an example of object recognition using the robot’s camera in the field.
It should be noted that the current vision module is limited to grape cluster recognition and does not determine grape maturity. This means that, at this stage, the system does not select mature grapes for picking, and the robots operate under the assumption that they are deployed at an appropriate harvesting date.

2.6.2. Robot Relative Positioning

Vision is also utilized for the relative positioning of the robots when the expert robot reaches a harvesting position. This particular function of vision is handled by the marker ROS package. When the expert robot is at a harvesting position, the helper robot is required to position itself relative to the expert robot with greater accuracy than that possible by the estimation of the robot’s position using the AMCL method. For this reason, a vision-based approach is required. The marker node continuously searches for an Aruco marker in the helper robot’s camera stream, and upon detection, it returns the marker’s pose. Therefore, when the marker placed on the back of the RB-Vogui robot is detected, the position of the expert robot relative to the helper robot can be determined with accuracy. This position can then be used to adjust the helper robot’s position behind the expert robot. Figure 5 shows the view from the helper robot’s camera and the detected pose of the marker.

2.7. Cooperative Harvesting

When the robots have stopped and are properly positioned in a waypoint where harvesting is to be carried out, the expert robot initiates the harvesting process by setting its Kinova Gen3 manipulator to an observation pose, preset at an appropriate height where
grape clusters are visible, and requests grape recognition from the vision package. If a grape is detected, then the vision package returns the 3D location of the detected grape. The manipulator then approaches the grape, at a distance where the stem is more visible, and stem recognition is requested from the vision package. When the stem is detected and its location and orientation are calculated, the manipulator moves its end effector to the stem, performs a cut with the cutting tool, and moves in order to deposit the cut grape in one of the two temporary collection baskets on the expert robot. It proceeds by assuming the observation pose again to detect more grapes. This process is repeated until a temporary basket reaches its capacity. When this happens, the expert robot instructs the helper robot to pick up the full temporary basket and empty it into its own, larger basket. Using a similar process as the one used for the accurate relative positioning of the two robots prior to harvesting, the helper robot uses the position of the Aruco marker at the back of the expert robot in order to determine the location of the two temporary baskets on the expert robot, given that the baskets are at fixed and known positions relative to the marker. The location of each basket is initially determined in relation to the camera’s reference frame. For this information to be effectively utilized for the planning of the UR10e arm’s movement and particularly to establish the goal location of its end effector, the marker’s position needs to be expressed in relation to the robot’s reference frame. To achieve this, the position of the center of the marker with respect to the camera \((x_{CAM}, y_{CAM}, z_{CAM})\) is determined as the position \((x_R, y_R, z_R)\) of the center of the marker with respect to the robot’s base using Equation (2):

\[
\begin{bmatrix}
    x_R \\
    y_R \\
    z_R
\end{bmatrix} = \begin{bmatrix}
    \cos \theta & -\sin \theta & 0 & l_x \cos \theta - l_y \sin \theta \\
    \sin \theta & \cos \theta & 0 & l_y \cos \theta + l_z \sin \theta \\
    0 & 0 & 1 & l_z
\end{bmatrix} \begin{bmatrix}
    x_{CAM} \\
    y_{CAM} \\
    z_{CAM}
\end{bmatrix}, \tag{2}
\]

where \(l_x, l_y,\) and \(l_z\) are the distances of the camera from the robot’s base and \(\theta\) is the marker’s yaw. Considering the basket handle and the end effector’s fingers, the end effector’s roll and pitch are set to constant values as required. However, in order to compensate for possible misalignment between the robots, the yaw of the end effector takes into account the orientation of the marker parallel to the ground plane, as seen in Equation (1). This ensures that the end effector approaches the basket handle at right angles even if the robots are not perfectly aligned along their longitudinal axes. With this information, the helper robot can accurately and safely approach, grasp, and move the correct temporary basket.

During the time that the helper robot handles the temporary basket, the expert robot is free to proceed with cutting another grape if one is detected, but it cannot deposit it to the other temporary basket until the helper robot signals that the basket area is clear. This reduces the time that the expert robot is idle and allows parallel but safe operation. The number of grapes that each basket can hold can be determined in advance, based on an estimation of the average size of the grapes of the target grape variety. This value would depend on the target grape variety. The harvesting process continues until the expert robot’s manipulator assumes the observation pose and does not detect any grapes. At this time, if there are grapes inside a temporary basket, even if the basket is not full, they are picked up by the helper robot, and when this process is completed, the robots retract their manipulators and resume their coordinated navigation toward the next waypoint. The full cooperative harvesting algorithm for both robots is illustrated in Figure 6.
3. Experimental Results

To assess the methods described in the previous sections, the two robots were deployed in a vineyard for field experiments. More specifically, the selected vineyard was in the privately owned vineyard Ktima Pavlidis located in the region of Drama, in northern Greece. The vineyard was previously mapped using an aerial drone, and a detailed map that included GPS coordinates was produced. A section of the map was isolated, reflecting the section of the vineyard in which the field experiments were to be conducted. This map was then loaded into the base station’s software as well as the two robots. Using this information, a path was planned as shown in Figure 7. The path included waypoints guiding the robots from an initial position to harvesting locations inside a vineyard row. Three harvesting locations were selected for this path.
Figure 7. The path planned by the base station’s software. The red circle with the yellow filling is the initial position of the expert robot. The yellow circles are waypoints and the light blue points are waypoints where harvesting is to be performed.

At the beginning of the experiment, the robots were positioned at an initial location, and adequate time was allowed (approximately 10 min) until an accurate enough GPS location was retrieved. During this initialization time, the vision model was loaded. After that, the experiment was initiated. The robots were observed to follow the intended path by visiting the prescribed waypoints sequentially, with the intended coordinated navigation behavior. At each waypoint, the expert robot waited for the helper robot to reach the expert robot’s previous waypoint. When this happened, the expert robot proceeded to the next waypoint. As anticipated, the paths followed by both robots between any two waypoints were not straight but were executed as the result of the planning process occurring under the TEB path planning algorithm described above. Figure 8 shows the two robots within a vineyard row while coordinating their motions.

Figure 8. The helper robot following the expert robot while the latter sequentially reaches the preset waypoints inside a vineyard row.

At each harvesting location, the expert robot stopped and waited for the helper robot to initially approach and then position itself with precision in relation to the expert robot, prior to the actual harvesting. According to the relative positioning scheme originally proposed in [42], the helper robot initially approached the expert robot, and then, after de-
etecting the Aruco marker located on the expert robot, corrected its positioning by adjusting the distance from the expert robot. Figure 9 shows the two robots correctly positioned prior to harvesting.

![Figure 9. Helper robot positioned behind the expert robot, ready to support harvesting.](image)

Regarding the cooperation between the two robots during harvesting, the process was performed as planned by the cooperative harvesting algorithm illustrated in Figure 6. During the experiments, it was observed that the preset grape cluster capacity of the temporary baskets affected the synchronization of the two robots. For the field experiments, the capacity of the temporary baskets was set to two grape clusters. As the proposed algorithm dictates, after reaching the capacity of either one of the baskets, the expert robot sent a message to the helper robot to empty the temporary basket and proceeded with cutting the next grape. However, the time required for the helper robot to empty the basket was significantly longer than the time required for the expert robot to harvest a grape cluster. This resulted in the expert robot remaining idle for a considerable time while waiting for the helper robot to complete its operation and for the basket area to be safe for the expert robot’s arm to approach. This idle time will depend on the preset capacity of the temporary basket as well as the speed of the helper robot’s arm. Figure 10 shows the helper robot picking up and emptying the temporary basket.

![Figure 10. The helper robot (a) picks up the temporary basket and (b) empties the temporary basket into its own storage basket.](image)

As shown in Figure 10, the grape clusters are first deposited into the temporary basket by the expert robot and then deposited into the storage basket. The fact that the harvested grapes are dropped to the baskets from a certain height has the risk of damaging the grape clusters. In order to minimize any possible damage, the following steps were taken in advance during the laboratory tests: In the case of approaching and depositing in the temporary basket, the lowest possible point of a grape cluster safely clearing the basket
while the arm is moving toward the center of the basket was estimated, and also the distance between the cluster and the basket at the time of gripper release was minimized. In the case of the storage basket, the temporary basket was set to rotate at a low speed, and in addition, the distance between the grape and the storage basket was also set to be as short as possible while avoiding any possible collisions. During the field experiments, and more specifically during grape deposition, some individual grapes were separated from the grape clusters, but no damage to the individual grapes was observed. This is acceptable for viniculture, where individual grape integrity is crucial, as opposed to the preservation of grape cluster integrity.

In terms of the actual harvesting of the grapes, it was observed that the preset observation pose was appropriate, set at the approximate height where grape clusters are present. The vision module recognized the grape clusters unobscured by leaves reliably and, using the depth information of the camera, was able to direct the end effector of the robotic arm to approach each grape cluster. All detected grape clusters were eligible for harvesting. However, the stem recognition, which takes place after the initial approach of the arm to the grape cluster, was not as reliable. This is due to the fact that for many of the recognized grape clusters, the stem was either fully or partially occluded by the individual grapes or by leaves. Occlusions of this kind hindered object recognition and thus the location information for stem approach and cutting was unavailable to the arm. When stem location information was missing, the arm was directed to assume the observation pose and search for the next grape cluster and repeat the process. A failure to detect the stem adds a delay of approximately 10 s in total to the process. This time includes a grape cluster arm approach, a five-second timeout for unsuccessful stem detection, and arm retraction. Another reason for unreliable stem recognition was image overexposure when the grape cluster was directly exposed to the sun, especially in cases where the visible stem was very short. Figure 11 shows the robotic arm cutting a recognized stem.

![Figure 11. Expert robot’s robotic arm cutting a grape cluster.](image)

During the field experiments, some task quantities were recorded in order to evaluate the efficiency of the current implementation. Table 1 shows the measured average duration of the various tasks and sub-tasks that comprise the cooperative harvesting task and the total time required for the entire cooperative harvesting cycle.
Table 1. Duration of the sub-tasks in the cooperative harvesting cycle.

<table>
<thead>
<tr>
<th>Task</th>
<th>Sub-Task</th>
<th>Duration (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initialization</td>
<td>GPS localization</td>
<td>10:00</td>
</tr>
<tr>
<td></td>
<td>Loading vision model</td>
<td>0:15</td>
</tr>
<tr>
<td>Harvesting</td>
<td>Successful grape detection</td>
<td>0:01</td>
</tr>
<tr>
<td></td>
<td>Arm movement to the grape cluster</td>
<td>0:02</td>
</tr>
<tr>
<td></td>
<td>Successful stem detection</td>
<td>0:01</td>
</tr>
<tr>
<td></td>
<td>Arm movement to cut stem and cutting</td>
<td>0:05</td>
</tr>
<tr>
<td></td>
<td>Arm movement to deposit grape cluster</td>
<td>0:16</td>
</tr>
<tr>
<td>Cooperation</td>
<td>Arm movement to grab and move the temporary basket</td>
<td>0:23</td>
</tr>
<tr>
<td></td>
<td>Depositing grapes to the storage basket</td>
<td>0:09</td>
</tr>
<tr>
<td></td>
<td>Arm movement to return to initial position</td>
<td>14</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1:11</td>
</tr>
</tbody>
</table>

It should be noted that the values stated above are approximations and are subject to change in the following iterations of the development process. For example, the duration of the grape cluster approach and cutting varies depending on the location of the detected cluster. The objective of the field experiments was primarily to verify the functionality of the prototypes and to demonstrate the feasibility of the cooperative harvesting process. For this reason, some of the task durations are intentionally longer. More specifically, the speed of both manipulators was lowered for safety reasons, especially while approaching and handling the baskets. The same is also true for robot speeds during navigation. It is clear that the efficiency of cooperative harvesting is not optimal and can be significantly improved by a) increasing the speed of manipulators and b) improving the synchronization of the two robots while the temporary basket is being emptied, as described above.

4. Discussion

The field experiments described in the previous section were aimed at testing the operation of the proposed cooperative harvesting system. The experiments were conducted successfully and the prototypes performed well in a real-world agricultural setting. The main aspects of the robots’ functionality were (a) navigation, (b) harvesting, and (c) cooperation. In this section, each of these aspects is discussed. The potential of explainable decision-making based on logic for enhancing the cooperation of autonomous robots in agricultural applications is also discussed.

4.1. Navigation

At an individual robot level, navigation is performed using established methods for localization and planning [46,47]. These methods are used to allow the robots to localize themselves within a pre-mapped vineyard and navigate to consecutive desired locations accurately and safely by planning safe routes between locations. At the robot group level, the proposed coordinated navigation algorithm was successfully utilized to guide the helper robot to follow the expert robot to the desired locations. Inter-robot communication was used in order to coordinate the robots’ movements so that they are performed in stages (i.e., only one robot is in motion at any time) and also allow exchange of location information. More specifically, WiFi communications were used for the exchange of information between the robots [48], with each robot being controlled by a separate ROS master. The data exchanged by the robots were limited to simple leader-to-follower commands and locations, so there was no need for a more sophisticated communications system with a larger bandwidth. By executing the robot motions in stages, it was ensured that the trajectory planned for each robot movement remained constant and could not be affected in real-time by the other robot’s movement while the two robots were in proximity [42]. In addition, the algorithm provided a mechanism that allowed correct positioning
of the robots prior to harvesting, in order to ensure that manipulation operations using the robotic arms were executed successfully. This was accomplished by a ROS node running on the helper robot that uses the camera to detect Aruco markers and calculate the pose and distance of the expert (leader) robot, and thus determine the desired helper robot position relative to the expert.

4.2. Harvesting

The actual harvesting operation involved a combination of machine vision and robotic arm control on the part of the expert robot. On the machine vision side, a model was trained so that the expert robot’s camera could be used to identify a number of objects present in the vineyard. For the harvesting task discussed in this paper, grape cluster and stem classes of the trained model were used. The model was trained after first assembling a dataset of 10,000 vineyard images. When the expert robot’s camera (located on the robotic arm) receives an image, either grape cluster or stem recognition is performed on that image. Because the camera also provides depth information, the location of the cluster or the stem in 3D space can be calculated. This location is then utilized to dictate arm motion and the desired end effector’s eventual position. Potential complications that can occur in the field are two: (a) difficulties in the object recognition phase due to occlusions or environmental conditions and (b) recognition of objects in locations that the robotic arm cannot reach. Recognition problems due to environmental conditions mainly relate to overexposure due to direct sunlight on the grape cluster. Depending on the orientation of the sun compared to the position of the two robots at the time of harvest, this problem could potentially be reduced if the helper robot can provide a shading mechanism using its manipulator while the expert robot is at the stage of grape recognition. Such a functionality would enhance the benefits of using a robot team.

Regarding the movements of the robotic arm, these were a combination of predefined poses, Cartesian planning, and goal-based planning, all performed using the MoveIt trajectory planning ROS library [49]. Goal-based planning for both Kinova and UR10e arms was achieved using the RRT-Connect planning algorithm. The sequence of motions comprising the harvesting task aimed at performing the desired function while taking into account obstacles that were not included in the internal description of the robot. For example, after cutting the grape, the arm was set to move back in order to create some distance between the cutting tool and the plant itself, since the robot was near the plant (at approximately 0.5 m). The cutting tool then moves directly upwards until a desired height is reached, so that a rotation places the cut grape on top of the temporary basket. This sequence of motion was designed such that unpredictable plans produced by the MoveIt library and which would damage the vine are avoided. Such unpredictable behavior was observed in earlier laboratory experiments when the target was near the maximum distance the arm could reach. Accuracy of motion was observed to be nominal for the Kinova arm and therefore adequate for the grape-picking task.

4.3. Cooperation

As mentioned earlier, the two robots have separate ROS masters. The complexity of the robots due to the large number of nodes being active on each, renders impractical the use of a common ROS master using separate namespaces [48]. At the same time, since the robot team only consists of two robots, the use of a ROS Multi-master package is not necessary. As is the case with coordinated navigation described earlier, the two robots only exchange limited information in a master–slave manner, where the master (the expert robot) sends action messages (commands) to the slave (the helper robot), and the slave responds when its action is complete.

Communications are achieved via a WiFi network established by the base station’s access point. The robots do not exchange messages with each other directly, but instead, the messages are first sent to the MQTT server running on the base station and are then forwarded to the recipient robot. For this reason, the messages have a recipient field. More
specifically, the transmitting robot creates a ROS message which is then converted to an MQTT message through an mqtt_bridge ROS node installed on both robots. The mqtt server receives the message and immediately broadcasts it. Both robots receive the MQTT message. The message is then converted into a ROS message, but only the designated recipient acts on the received information.

In terms of physical interactions between the robots, it was imperative that the grasping of the temporary baskets by the helper robot was as accurate as possible. This is the reason why visual cues were utilized, through the use of a camera detecting the pose of a large Aruco marker, whose location on the expert robot in relation to the temporary baskets is well-defined. A ROS node constantly detecting the given marker and returning the pose of the detected marker was developed. The helper robot is idle until it receives a command from the expert robot. When the expert robot sends a command to the helper robot to empty one of the temporary baskets, the helper robot samples the data published by the aforementioned ROS node, in order to acquire the latest information on the location of the marker, and consequently determine the location of the baskets. This is achieved through a series of transformations that calculate the pose of the basket handles in relation to the helper robot’s reference frame.

The proposed cooperative harvesting strategy that was evaluated in the field experiments offers the capability of partitioning the cooperative task into two sub-tasks, harvesting and carrying, each allocated to a different robot with different capabilities. This allows the deployment of two heterogeneous robots specialized in a different portion of the overall task. The expert robot is responsible for harvesting, and therefore it is equipped with an arm with a specialized cutting end effector (in line with other proposed solutions for grape cutting such as in [50]), a depth camera, and additional equipment to augment the computational capabilities of the robot, which is necessary for the machine vision operations. It temporarily stores the harvested grape clusters in its temporary baskets. On the other hand, the helper robot can be significantly less complex as it only requires a robotic arm and a transportation basket to support the harvesting operation. The helper robot needs only to collect the grapes harvested by the expert robot by manipulating the expert robot’s temporary baskets. The advantage of this architecture is that it is possible that the system can be scaled such that multiple helper robots can be deployed. Waiting times during cooperative harvesting such as those described in the previous section can be reduced by adjusting the temporary baskets’ capacity and the operational speed of the arms.

4.4. Logic-Driven Decision-Making Based on Data Semantics

On one hand, typical machine-learning methodologies carry out, either explicitly or implicitly, numerical feature extraction followed by “number crunching” data processing in the space $\mathbb{R}^n$, ignoring data semantics altogether. On the other hand, the LC (lattice computing) paradigm considers semantics represented by the partial order of the data; furthermore, data processing is pursued by the lattice-meet (‘’$\cap$”) and/or join (‘’$\cup$”) operations; therefore, data semantics is retained throughout data processing as explained next.

Given a mathematical lattice $(L, \sqsubseteq)$, an inclusion measure is a function $\sigma: L \times L \rightarrow [0, 1]$, which satisfies, by definition, conditions (C1) $u \sqsubseteq w \Leftrightarrow \sigma(u, w) = 1$ and (C2) $u \sqsubseteq w \Rightarrow \sigma(x, u) \leq \sigma(x, w)$. An inclusion measure supports at least two different modes of reasoning, namely Generalized Modus Ponens and Reasoning by Analogy. The following two equations define two different inclusion measures, respectively.

$$\sigma(x, u) = \nu(u)\nu(x-u),$$

$$\sigma(x, u) = \nu(x\cap u)\nu(x),$$

where $\nu: L \rightarrow R^+_0$ is a parametric positive valuation function, which by definition satisfies both $\nu(x) + \nu(y) = \nu(x\cap y) + \nu(x\cup y)$ and $x \sqsubseteq y \Rightarrow \nu(x) \leq \nu(y)$.

Previous applications in agriculture as well as elsewhere, regarding either classification or regression, have employed an inclusion measure in a lattice of a single type of data...
such as either real numbers or tree data structures or distributions or ontologies [21,51]. The aforementioned data are expected to be significant in agricultural applications because they represent the real world more accurately; e.g., a distribution of measurements/estimates represents “all order data statistics”, a tree data structure may represent more accurately a plant, etc. Hence, original data semantics is involved in the computations without distorting the data by arbitrarily transforming them to real numbers. In particular, in lattice computing (LC), histograms are treated as histograms, tree data structures are treated as tree data structures, etc. all along during data processing. In addition, note that the good results reported in various publications have been attributed to the parametric optimization of the underlying positive valuation function of an inclusion measure function.

An advantage of particular interest of an inclusion measure is its rigorous extension to hierarchical data structures which emerge as the Cartesian product of disparate mathematical lattices. Hence, it becomes possible to fuse disparate types of data semantics toward sophisticated decision-making. An additional level in a Cartesian product of mathematical lattices regards intervals; hence, (information) granules emerge, represented by intervals, resulting in explainability in LC, as it has already been demonstrated by granule-based rules induced in numerous publications.

The primary objective of this work has been the development of a physical system prototype of heterogeneous autonomous robots for cooperative grape harvesting as explained above. The application of LC algorithms for sophisticated decision-making in agriculture, also beyond viniculture, as outlined in this subsection, will be elaborated in future works, as it has been demonstrated in [51].

5. Conclusions

This paper has described field experiments that were carried out in order to assess the functionality of two cooperating agricultural robots in a grape harvesting scenario. The field experiments have shown that the proposed robotic systems and the associated algorithms that were developed have been operational and effective in a real-world setting and have allowed the robots to perform the harvesting task successfully. The cooperative aspect of the experiments has demonstrated the benefit of using heterogeneous robots, where the complementary capabilities of each robot offer a more complete solution to the harvesting task.

Further enhancements of the proposed cooperative harvesting implementation are currently underway and include an additional process in which the helper robot returns to a specific location to empty its storage basket and returns to the location of the expert robot to resume the support of harvesting. Another enhancement under development is to improve the current machine vision models in terms of object recognition accuracy and to include additional machine vision capabilities, such as the automatic determination of the grapes’ maturity so that the robots harvest only the mature grapes. The integration of this grape maturity estimation capability using vision is currently ongoing and is described in [52]. Apart from improving the existing framework for harvesting, the authors’ goal is to proceed with the implementation and testing of other agricultural tasks such as leafing, tying, and spraying, to be carried out by an individual or cooperating robots, in order to provide complete robotic solutions for the viniculture industry.

Future work will transfer inclusion measure-based decision-making from the lab to the field toward unified, explainable decision-making in cooperative autonomous agricultural robots, potentially also involving humans.

**Author Contributions**: Conceptualization, C.L., T.P. and V.G.K.; methodology, C.L.; software, C.L., C.B., G.S., C.T. and T.S.; investigation, I.K. and E.B.; resources, I.K. and E.B.; data curation, C.B. and E.B.; writing—original draft preparation, C.L., G.S. and T.P.; writing—review and editing, C.L., V.G.K., C.B. and T.P.; visualization, C.T. and T.S.; supervision, V.G.K. and T.P.; funding acquisition, V.G.K. All authors have read and agreed to the published version of the manuscript.
Funding: We acknowledge support of this work by the project “Technology for Skillful Viniculture (SVtech)” (MIS 5046047), which is implemented under the Action “Reinforcement of the Research and Innovation Infrastructure” funded by the Operational Program “Competitiveness, Entrepreneurship and Innovation” (NSRF 2014–2020) and co-financed by Greece and the European Union (European Regional Development Fund).

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank the Ktima Pavlidis winery (Drama, Greece) for providing their facilities and support during data collection and the field experiments with the robots.

Conflicts of Interest: The authors declare no conflict of interest.

References


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.