Ref. Ares(2021)2906196 - 30/04/2021

This project is funded by the European Union
G.A No 101016906

Project Number: 101016906
Start Date of Project: 2021/01/01
Duration: 48 months

Type of document D2.1 – V1.0

Requirements, Specifications and Benchmarks

Dissemination level	PUBLIC
Submission Date	2021-04-30
Work Package	WP2
Task	T2:1
Type	Report
Version	1.0
Author	AGRIMESSINA
Approved by	UNIROMA3, UNICLAM

DISCLAIMER:
The sole responsibility for the content of this deliverable rests with the authors. It does not necessarily reflect the opinion of the European Union. Neither the Directorate-General for Communications Networks, Content and Technology, Artificial Intelligence and Digital Industry nor the European Commission are responsible for any use that may be made of the information

This project is funded by the European Union G.A No 101016906

Executive Summary

This deliverable provides an overview of the agronomic activities and requirements of the table grapes vineyard scenario as well as the appropriate performance indices for validating the project outcome.

The vineyard framework is outlined by defining different type of structures, cluster's morphology and phenological scale of table grapes. Harvesting activity is deeply analyzed through cluster detection in different type of structure, quality assessment, cluster collection and placing. Pruning activity is detailed into winter pruning and green pruning with a notion about fertility. Then, canes detection and selection as well as buds selection are described.

Functional specifications for the selected technological solutions are described outlining the main methodologies that will be implemented and hardware that will be employed for networking communication infrastructure, mobile ground platform, dual-arm system, box-exchange mechanism, sensorial equipment, and virtual reality environment.

Finally, agronomic performance indices are defined to effectively assess the quality of the proposed technological solutions.

Table of Content

1 Introduction 5

2 Analysis of the Agronomic Requirements 6

2.1 Table-Grape Cultivation: An Introductory Overview 6

2.2 Harvesting 10

2.2.1 Cluster Detection 10

2.2.2 Quality Assessment for selective harvest 11

2.2.3 Cluster Collection 20

2.2.4 Cluster Placing 20

2.3 Pruning 21

2.3.1 Canes Detection 22

2.3.2 Canes Selection 22

2.3.3 Buds Selection 24

3 Analysis of the Architectural Functional Specifications 25

3.1 Networking Communication Infrastructure 25

3.2 Mobile Ground Platform 26

3.3 Dual-Arm System 27

3.4 Box-Exchange Mechanism 30

3.5 Sensorial Equipment 31

3.5.1 For navigation purposes 31

3.5.2 For agronomic purposes 33

3.5.3 For human-robot interaction 37

3.6 Virtual Reality Environment 44

3.6.1 Virtuality Reality Hardware 44

3.6.2 Virtual Reality Software 46

4 Definition of the Agronomic Performance Indexes 50

4.1 Performance Indexes for Harvesting 50

4.2 Performance Indexes for Pruning 50

Abbreviations and Acronyms

PO	Project Officer
PA	Precision Agriculture
HRI	Human-Robot Interaction
HRC	Human-Robot Collaboration
MRC	Multi-Robot Coordination
ELP	Eichhorn and Lorenz phenological scale
TSS	Total Soluble Solids
TA	Titratable Acidity
CAN	Controller Area Network
SEE	Series Elastic Element
ROS	Robot Operating System
UGV	Unmanned Ground Vehicle
BEM	Box-Exchange Mechanism
DoF	Degree of Freedom
GNSS	Global Navigation Satellite System
IMU	Inertial Measurement Unit
RTK	Real-Time Kinematic positioning
EKF	Extended Kalman Filter
FOV	Field Of View

This project is funded by the European Union G.A No 101016906

1 Introduction

Table grape is today one of the most grown and requested fruit all over the world with around 28 million tons yearly harvested. The intrinsic predisposition for easy moments of consumption along with the natural sweetness make table grapes a versatile fruit that matches the lifestyle of the modern consumer, who is always more demanding in terms of taste and physical characteristics. Research and development are constantly improving the overall quality of fruit meeting high-demand markets and producing, for example, new seedless varieties that are more appealing for a large consumer audience and allow a repetitive positive consumer experience. Ensuring specific physical characteristics of the product and determining the specific degree of ripeness are major challenges for the industry and require skilled labor to carefully handle the fruit.

The global socio-economic situation is facing an increasing shortage of workers across the whole agricultural sector and table grapes is one of the highest demanding in terms of labor. Many Countries like US and Australia have already experienced loss of products for lack of workers in the past seasons. Peruvian farmers faced in season 2020 an unprecedented workers' strike that paralyzed the whole industry for more than two weeks with loss of product and market opportunities.

Precision agriculture in table grapes, one of the less developed agricultural business in terms of automation and robotics, is today both an opportunity and a necessity for the survival of the industry. CANOPIES proposes a novel integrated system where farmworkers can safely collaborate with multi-robot systems for the precision farming of a table grapes vineyard, developing HRI, HRC and MRC methodologies for implementing an effective collaborative paradigm between human workers and multi-robot teams.

As stated by the EU Scientific Foresight Study “Precision agriculture and the future of farming in Europe” PA in the future of the Agri-Food sector will make a significant contribution to food security and safety, promote more sustainable ways of farming, trigger wide societal changes, and develop new skills for workers and farmers.

The three primary technologies that define PA and ultimately the future of farming, i.e., imagery and sensors, robotics and automation, digitalization, and big data, are exploited in CANOPIES as follows:

i. Collection of data in the vineyard with the maximum possible resolution;
ii. Analysis of this data (often by combining them) so as to have a quality analysis of fruit and to decide which clusters are suitable for the harvest;
iii. Agronomic tasks such as harvesting operation and pruning undertaken with great control and precision by using proper equipment.

2 Analysis of the Agronomic Requirements

The European grapes (vitis vinifera) are the most important species of the family vitaceae. Two other commercial species are vitis labrusca (American grapes) and vitis rotundifolia (Muscadine grape). Worldwide grapes are grown for eating, wine, juice, and raisins. European grapes grown in mild subtropical conditions, while both other species do well under temperate conditions.

2.1 Table-Grape Cultivation: An Introductory Overview

Table-grape cultivation aims to fully exploit the environmental resources available in a specific area (sun-light, soil, water) in order to achieve the requested qualitative and quantitative target. Main objectives are:

- Optimize collection of solar energy (efficiency in photosynthetic activity);
- Maximize specific characteristic of each grapes variety;
- Allow mechanization of growing activities to reduce production costs;
- Balance production and homogeneity across the vineyard;
- Promote micro-climatic conditions that allow better control of parasitic diseases;

Table grapes can be grown in a wide variety of soil types. The most important characteristics of the soil are good internal drainage and adequate depth. Grapevines require deep, well-drained soil with a minimum of 0.75 m to 1 m of permeable soil with no impeding layers (shallow bedrock, chemical or physical hardpans) for optimum vine growth.

The type of structure deeply affects all other factors, from leaves and clusters layout to plot fertility, the creation of a specific microclimate and the possibility of mechanization. The most common structures today are Flat-roof Pergola system, Y trellis system, Double-roof system described in the following Table 1.

Name	Flat-roof Pergola system (*Tendone*)	Y trellis system	Double-roof (*Doppio impalco*)
Plant distance	2,50m x 2,50m to 3,00m x 3,00m	2,00m x 2,50m to 2,50m x 3,00m	2,50m x 2,50m to 2,80m x 3,00m
Plant density	1.100 – 1.600	1.300 – 1.600	1.200 – 1.500
Structure height from the ground	2,00m	1,60m	2,00m
Grapes distance from the ground	1,70m – 2,00m	0,80m – 1,50m	1,40m – 1,80m
Min grapes distance from plastic cover	0,30m – 0,50m	0,30m – 0,60m	0,50m – 0,70m
Grapes and canopies area	Canopies and clusters are on two different level, but not structurally divided	Canopies and clusters are in a shared area on different levels	Canopies and clusters are separated by the double structure

Table 1 - Table grapes structure overview

In Figure 1 the main elements of a Flat-roof Pergola system are depicted.

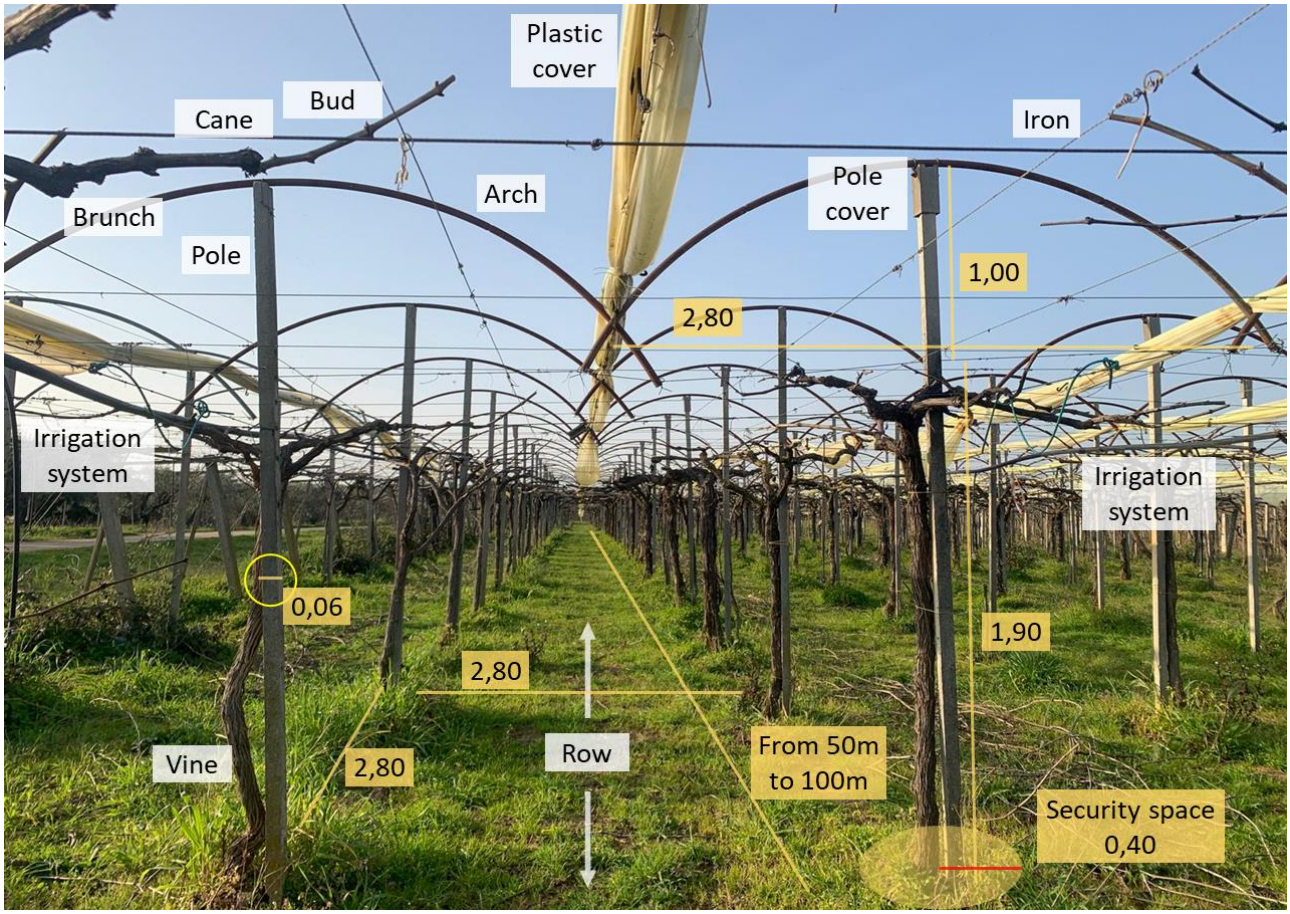

Figure 1 - Vineyard description of Flat-Roof Pergola system.

Table grape fruit develops as clusters (bunches) with each berry attached to the bunch stem (rachis and branches) via a pedicel, which contains vascular bundles (also known as the cap stem). There is much variation between cultivars in stem structure (i.e., length of parts, toughness, and adherence to berries). Impact or shaking of bunches may cause the loss of berries, leaving an unsightly spray of vascular strands (brush). The cluster of table grapes is formed by the berries and the rachis as shown in Figure 2. The stem unites the berry with the rachis that should be strong and wealth enough to avoid loss of berries (shattering). The cluster is connected to the cane thanks to the peduncle.

Number of cluster per cane and number of clusters per plant vary according to the variety, climatic/seasonal condition, vineyard management, soil, plant density and production strategy.

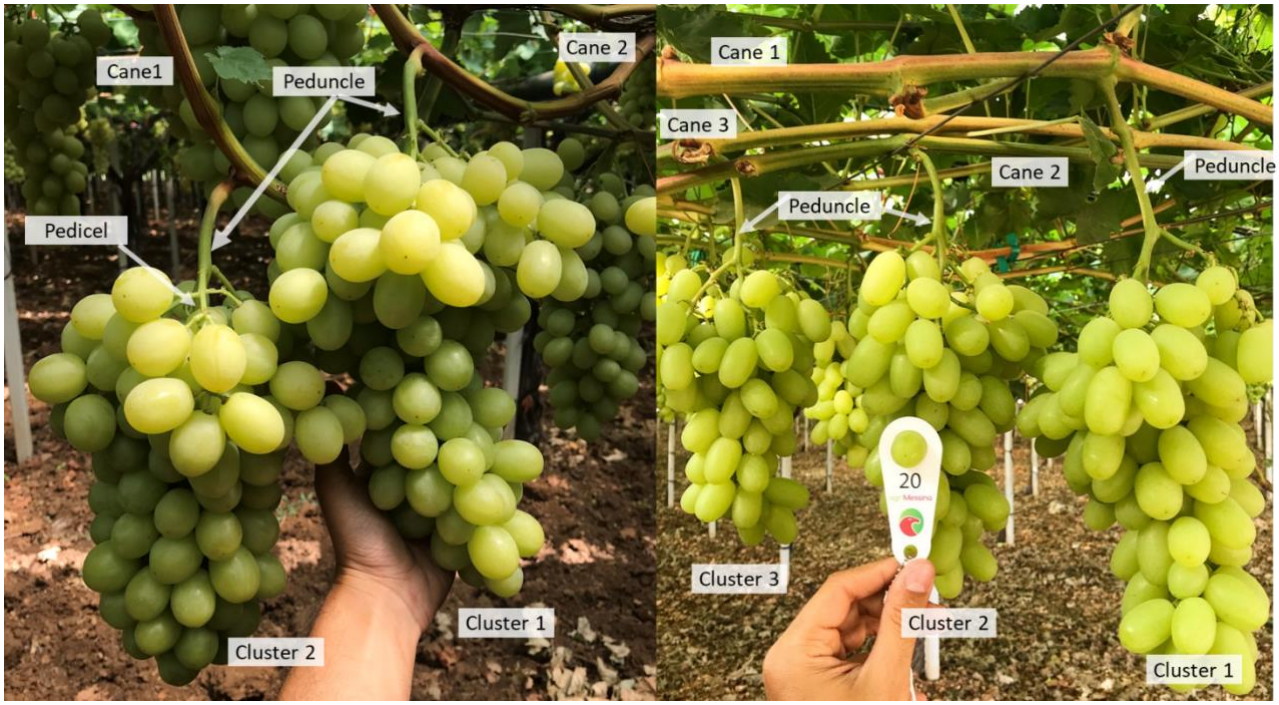

Figure 2 - Cluster description.

The annual cycle of the vine starts from the vegetative awake (Italy: March – April) to the entry into winter rest (Italy: November – December) following the specific phenological phases described in Table 2. In the table, ELP refers to Eichhorn and Lorenz phenological scale.

Phenological scale (ELP)	Major stages	Description	Macro phases	Agronomic operation
4	Budburst	Leaf tips visible	Shoot and inflorescence development	Green pruning: suckers and sterile shoots, which do not have inflorescences and are not suitable for the winter pruning, are removed. During this operation basil leaves on both sides of the cluster zone are pulled. Shoots thinning is done during this stage to achieve an optimal distribution, density and leaf area.
12	Inflorescence clear	Shoots about 10 cm long; 5 leaves separated		
17	12 leaves separated	Inflorescence well developed single flower separated		

23	Flowering	17-20 leaves separated: 50% caps off	Flowering	During full flowering vineyard management operations are not recommended due to the high sensibility of the whole micro-system.
27	Setting	Young berries enlarging (>2 mm diam.); bunches at right angles to stem	Berry formation	Shoots positioning is normally done immediately after flowering in order to have an optimal clusters orientation. It improves canopy configuration and reduces the amount of shading created by overlapping shoots. Once the rachis reaches an optimal length, all the clusters are correctly positioned and the leaves that touch and obstruct the grapes are removed. Bunch selection (removal of not-suitable clusters), trimming and thinning are done before bunch closure.
31	Berries pea-size	7 mm diameter		
32	Bunch closure	Berries touching (if bunches are tight)		
35	Veraison	Berries begin to color and enlarge, berry softening begin	Berry ripening	During veraison canopy management is very important: leaf and laterals are removed to open up the interior of the canopy to light and air to promote fruit ripening, reduce disease pressure and increase spray coverage.
38	Harvest	Berries harvest - ripe		
43	Leaf fall	The canes complete their maturation	Senescence	Pruning: This operation affects any parameter

				of the canopy during the growing season

Table 2 - Table grapes phenological phases and related agronomic operations

2.2 Harvesting

After the seasonal cultivation process harvesting is the final operation that completes all efforts made in the vineyards during the whole year. It is performed manually and is carried out only when the fruit reaches the requested maturity level and compliance to market specifications. Grapes are non-climacteric fruit and do not continue to ripen after collection, so harvesting timing is critical. According to the variety and the specific production area, in the Northern Hemisphere (Italy, Spain, Greece, California), it is performed from June until November.

This agronomic operation is composed of different step: cluster detection, quality assessment, cluster collection and cluster placing, described below.

2.2.1 Cluster Detection

All clusters and peduncles need to be recognized from canopies and obstacles. To facilitate this operation and the whole harvesting process, few days before harvest the vineyard is prepared with defoliating process: excessive vegetation is removed to allow bunches get more sunlight and be more visible for workers. The position of clusters changes on three axis x, y, and z according to two factors: the type of structure and the distance between plants.

In a Flat-roof system, shown in Figure 3, clusters are positioned under the canopies surface at the same level, around 180/160 cm. Despite most clusters are well visible, sometimes there are clusters up in the vegetation area (5/10%); iron and canes are at a higher level from the bunches. Wide plant-distance implies clusters to be spread around the plant in a circular area.

Figure 3 - Flat-roof system.

In the Y system, shown in Figure 4, clusters are positioned beside the plant at different levels. Here vegetation and canes are close to clusters and often in a promiscuous zone. Clusters grow at heterogeneous levels (0.7/1.5 m).

Figure 4 - Y trellis system.

In the double-roof system ("*doppio impalco*"), shown in Figure 5, clusters are beside the row of plants as the Y system but at a higher level (1/1.7 cm according to the height of the second roof). Clusters are placed at a lower stage separated from canopies and canes with more consistency in terms of height compared to classic Y structure. Compared to Flat-roof structure, in the double-roof system clusters are closer to the row of plants.

Figure 5 - Double-roof system.

In each type of structure, it is required to remove all shoots and leaves from the clusters area before harvesting operation, preparing the vineyard for the harvesting to have a clear visibility of all clusters and peduncles and facilitate workers.

2.2.2 Quality Assessment for selective harvest

Table grapes are a type of fruit whose commercial quality is closely related to their visual appearance and overall flavor. The right harvest of table grapes is usually determined by different parameters like

skin color, texture softening, titratable acidity, total soluble solids, and sometimes with flavonoid content, and aromatic compounds. Visual attributes of table grapes, such as intensity and uniformity of color, large size of berries, and brightness are the main characteristics that influence consumer choice.

Skin color – From thousands of years color has influenced human decision making; color is associated with many aspects of the decisions we usually take, and many studies have shown that it affects the acceptability, choice, and preference of food.
Grape color is defined during the last phase of fruit growth, when both red and white grapes cultivars acquire their typical pigmentation. Figure 6 reports samples of clusters of table grapes and highlights which colors are appropriate and which are not for them. Red cultivars acquire their characteristic pink, red or violet tints through the presence of higher or lower concentrations of anthocyanins, while white cultivars become translucent and greenish-yellow. Despite the coloring process reflects the grapes maturity stage, it is not strictly related to it and it may occur to find well colored bunches not ripe yet and vice versa. All vineyards are managed to maximize product homogeneity reducing small differences that can be found in the soil, microclimate, and the variability of nature itself. Nevertheless, full homogeneity is very hard to achieve (maybe impossible in nature) so the coloring process is not a perfect homogeneous process. According to the specific variety, weather condition and other minor factors clusters reach the requested color in different moment and different levels so they must be collected at different time by a selective harvesting process. Color is a matter of perception and subjective interpretation, and there is a wide variety of ways to express it. In order to standardize this characteristic, colorimeters and spectrophotometers are used increasingly to express colors numerically. To give clear and simple guideline modern supermarkets provide color acceptability chart to growers trying to help them in carrying out a selective harvesting process. Clusters that are left on the vine for lack of color will be pick up during the second or third harvest round waiting for the achievement of the requested characteristics. Green grapes usually can be harvested 90% in one single round of harvesting while red grapes require more harvesting rounds, i.e., according to the variety from 2 to 5 steps.

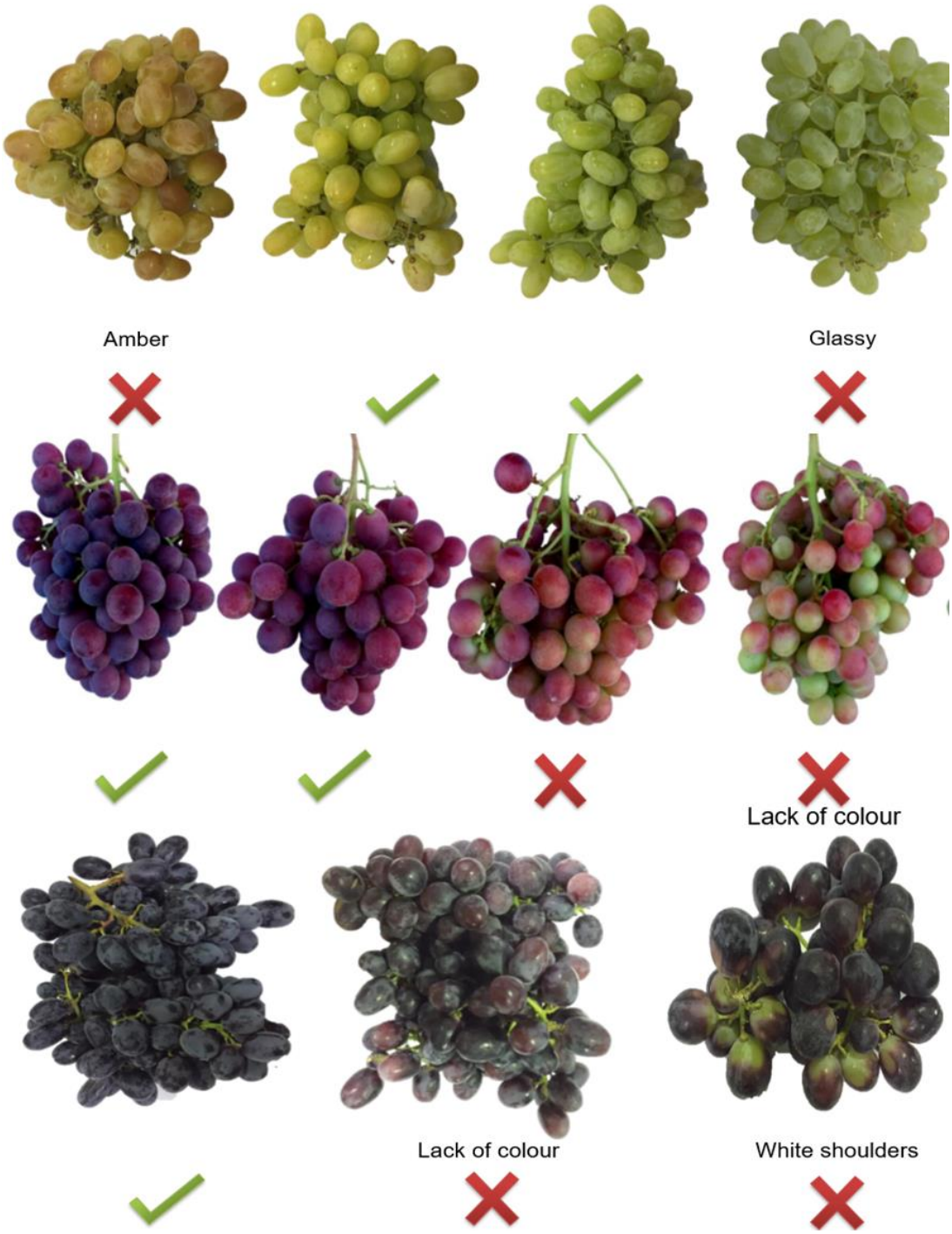

Figure 6 – Color chart.

Sugar level and acidity - The overall flavor is one of the most important characteristics for establishing a continuous consumer preference. Flavor composition has been defined as a complex attribute of quality, in which the mix of sugars, acids, and volatiles plays a primary role (Baldwin, 2002). Sugar and organic acid compositions, which are measured through total soluble solids (TSS) and titratable acidity (TA), are most commonly associated with the taste of fruits like table grapes (Ferguson and

Boyd, 2002; Shiraishi et al., 2010). Differences in the acidity of the table grape at harvest can be due to differences between varieties, environmental conditions, storage time, and other factors (Diakou et al., 1997; Navarro et al., 2001; Liu et al., 2006). Despite the importance of TA to the overall flavor, growers mainly use TSS (sweetness) as an indicator of ripeness and most of the commercial varieties are considered mature when TSS ranges from 15 to 18%. The chart below shows TSS trend for a late red variety that is usually harvested at the end of September. TSS is assessed by spectrometer or refractometer with 3 measurements of random berries per bunch: one at the top of the cluster, where there is usually the highest value, one in the middle and one at the bottom of the cluster, where there is the lowest value.

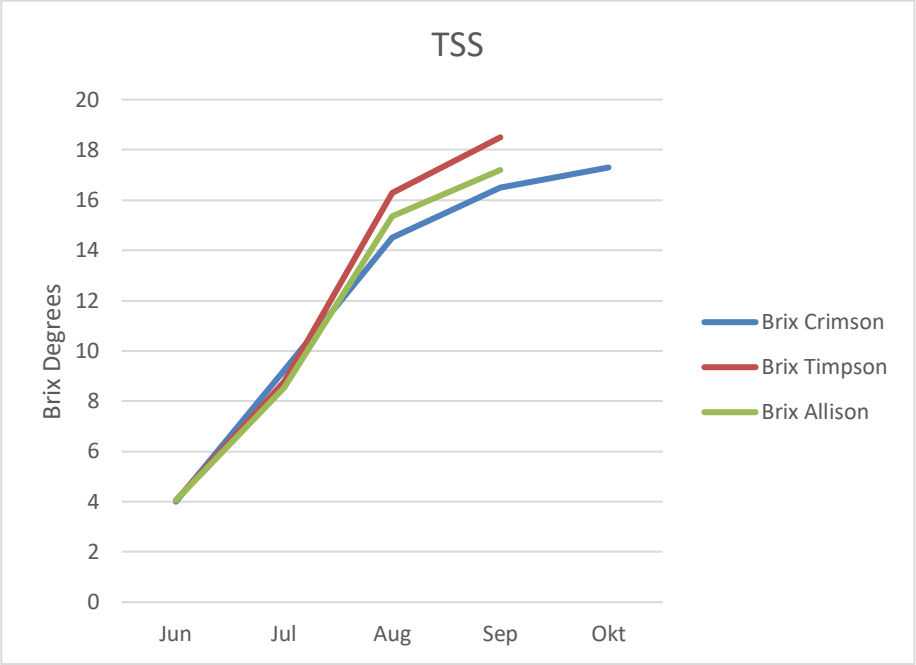

Table 3 - TSS trend.

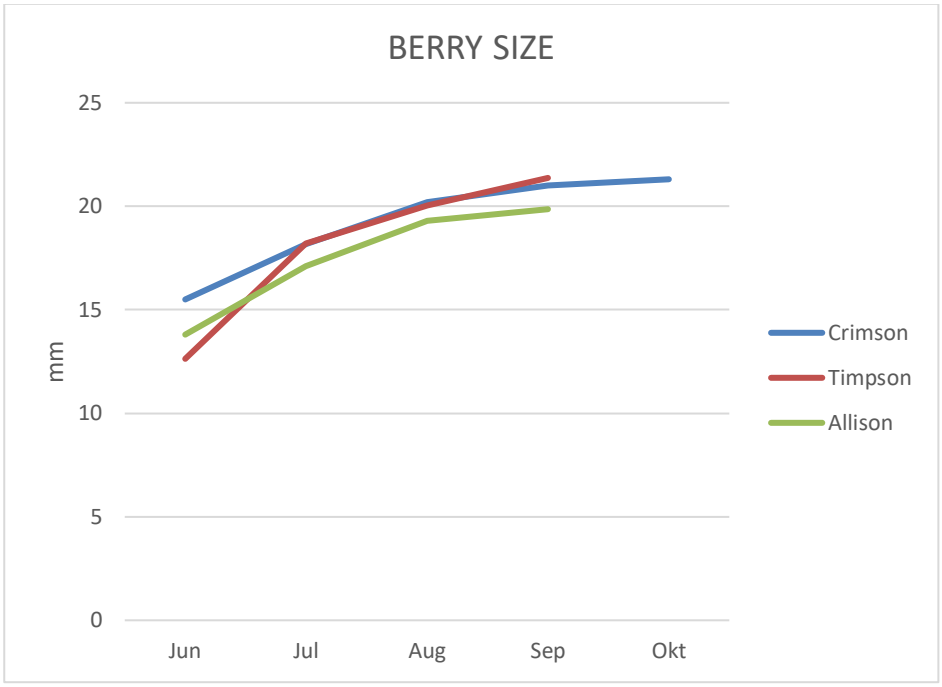

Table 4 - Berry size development.

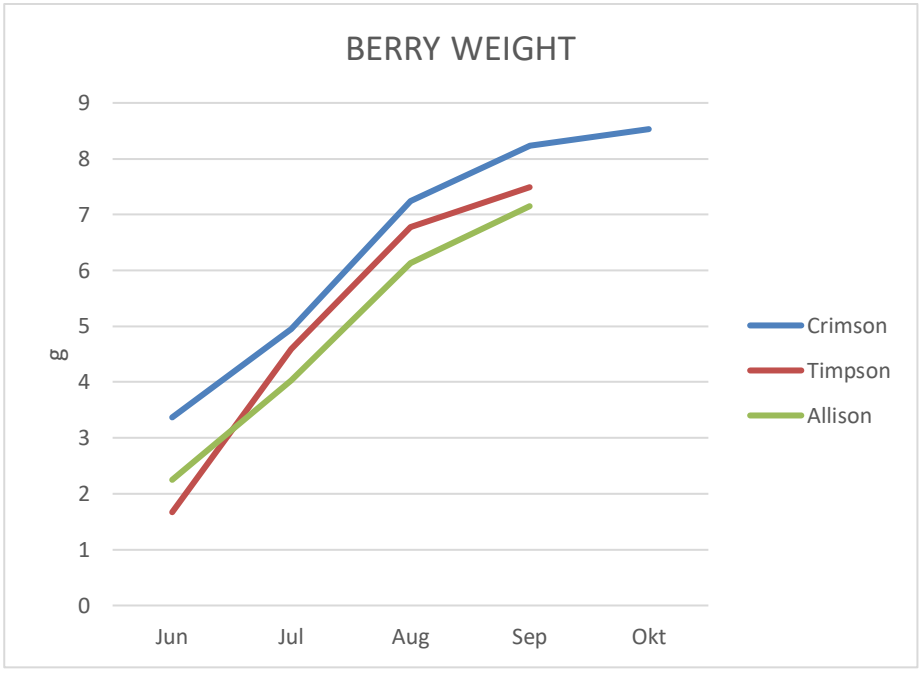

Table 5 - Berry weight development.

Bunch Density - According to specific variety, clusters can be too tight or too loose. An effective way to explain a selection criterion based on looseness and compactness is through the examples provided in Figure 7.

Figure 7 - Looseness and compactness samples.

Berry size – Berry size monitoring, as shown in Figure 8, from the first phenological stages to the maturity phase, can provide indication of grape ripeness crossed with other features, such as berry density. At harvesting time, it is an important parameter that affects the overall quality that consumers expect to receive. It must be considered in terms of cluster of berries (homogeneity) and in terms of size (caliber). For example, clusters with berry size smaller the 15 mm diameter, or with very big and very small berry in the same cluster, should be considered not complaint for the specific customer specifications. The scale of values depends to the specific variety.

Figure 8 - Berry size sample.

Defects - Major and minor defects can occur and affect grapes quality. A supermarket chart with the most common samples of defects is shown in Figure 9 and Figure 10. Some of them can occur at different stages before harvesting and others are more related to post-harvest management.

Figure 9 – Chart of table grape issues (1 of 2).

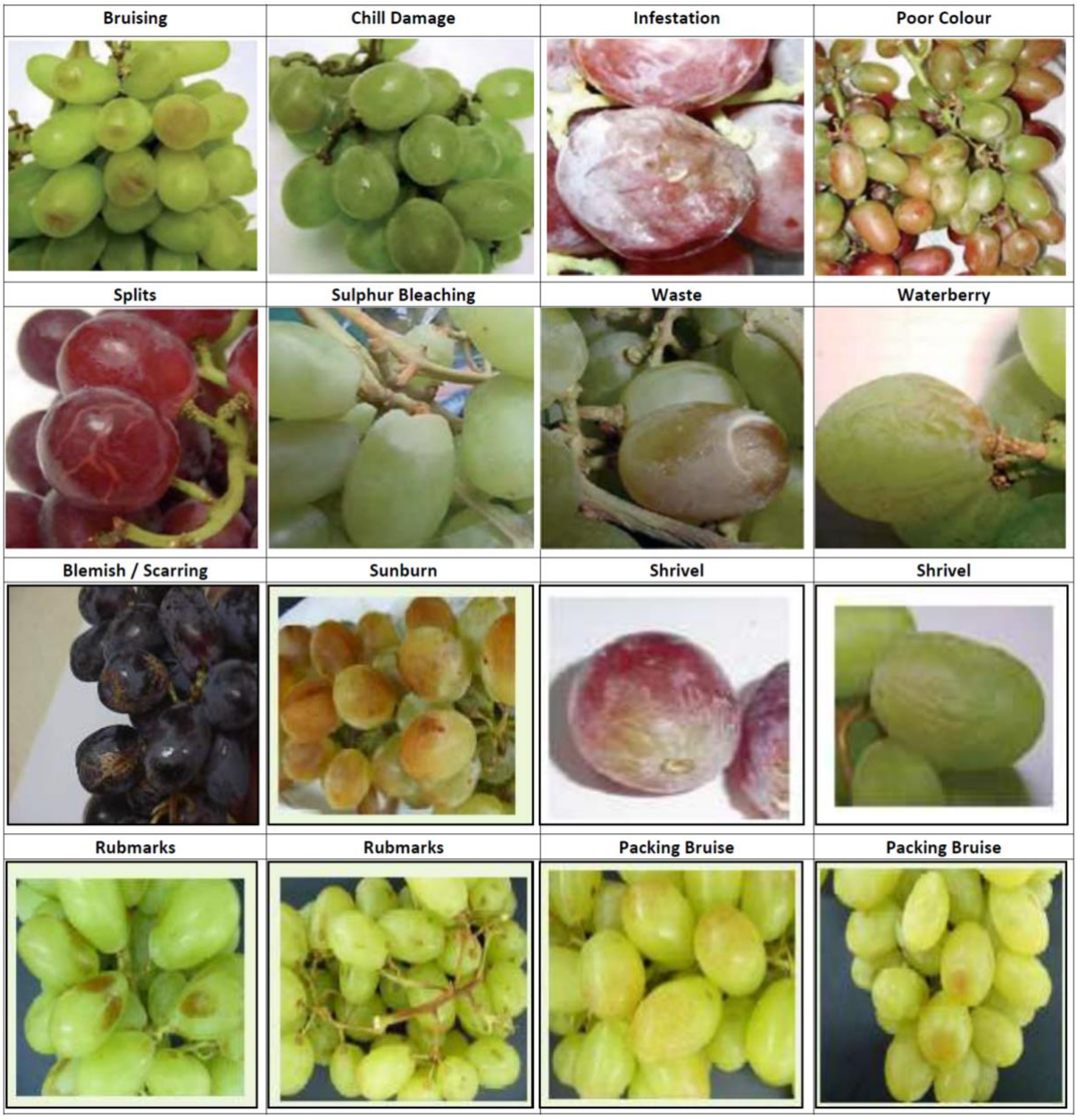

Figure 10 – Chart of table grape issues (2 of 2).

Harvest Definition & Quality Supervision – Before harvesting, a supervisor analyses sugar level and relative color in a representative sample of clusters in the specific vineyard to assess the overall conformity of fruit and then identify the ripe grapes to cut. During harvesting the supervisor gives indications to the workers about the grape to collect empathizing the color as an index to recognize the ripe bunches. During harvesting the supervisor monitors the product conformity giving indications and suggestions to the workers. According to specific customer requirements, market opportunity and weather condition, the supervisor sets the grade of tolerance for waste and issues that can be accepted and explain this to workers. Minor defects can be accepted in some less demanding market while some other markets can be more sensitive to some specific issues compared to others. For this reason, he/she explains to workers the waste-*non-compliant bunches*

that must be left on the vine. Sometimes the defects can be removed manually and so the cluster becomes suitable for the harvesting and must be collected.

Main parameters and quality characteristics of table grapes are summarized in the following Table 6:

Parameters and description	Object	Method
Cluster weight	Client specification	Weighting in the packhouse
Brix° (Soluble solid concentration)	Ripeness and client specification	Refractometer
Titratable Acidity (tartaric acid)	Ripeness and client specification	Titration with sodium hydroxide
Berries per cluster (berries number)	Appearance	Visual analysis
Berry weight	Appearance	Weight scale
Rachis length	Appearance	Meter
Berry length	Appearance	Calliper
Berry diameter	Appearance and client specification	Calliper
Loose Berries (pedicel detachment resistance)	Client specification	Number of detached berries
Decay (number of affected berries)	Appearance and client specification	Visual analysis and removal by the worker
Cracking (number of affected berries)	Appearance and client specification	Visual analysis and removal by the worker
Sun blush (number of affected berries)	Appearance and client specification	Visual analysis and removal by the worker
Browning (number of affected berries)	Appearance and client specification	Visual analysis and removal by the worker
Waterberry (number of affected berries)	Appearance and client specification	Visual analysis and removal by the worker
Compactness (bunch formation and appearance)	Appearance and client specification	Visual analysis according to tolerance scale
Looseness (bunch formation and appearance)	Appearance and client specification	Visual analysis according to tolerance scale
Berries Size Uniformity	Appearance and client specification	Visual analysis
Skin Color (color related to the variety)	Appearance and client specification	Visual analysis
Skin Color Rating (uniformity and intensity)	Appearance and client specification	Visual analysis
Firmness (pressure on the berry)	Eating quality and client specification	By hands

Stem Condition (color and defects)	Appearance and client specification	Visual analysis

Table 6 - Grapes main parameters

2.2.3 Cluster Collection

In this task the worker collects the bunches from the plant walking through the row, harvesting one plant at a time. Each operator works only on one side of the row in couple with a partner who works on the other side of the row. The operator meticulously collects one cluster at a time: after detecting the cluster, the operator assesses the overall quality and possibly cleans any wasted or rotten berry; then, he/she holds with one hand the peduncle close to the cane, leaving enough space to perform the cut with the scissors as shown in Figure 11; finally, he/she uses the second hand to cut perpendicularly the peduncle with the scissor as close as possible to the cane.

Figure 11 - Cluster collection and cutting/holding points.

During this operation, the clusters need to be managed carefully to prevent any damages like scratches or bunch break (rachis break) that would lead to product loss. Non-compliant clusters are left on the vine for a second harvest or for second-class harvest.

2.2.4 Cluster Placing

In this task the operator places the cluster in the box. He/she holds the cluster from the peduncle and releases the bunch in a free area of the box with the steam upwards (steam-up), as reported in Figure 12-left. The clusters are gently placed on one single layer to preserve quality of grapes until the box is full. According to the type of box they are usually filled at average weight of 9/10 kgs. When the box is full the operator starts filling the next box. All boxes are placed close to the plants in the vineyard before the harvesting; usually, one or two boxes per plant are placed according to the overall quantity of fruit. An example of full crates is shown in Figure 12-right.

Figure 12 - Cluster placing.

Full crates are manually stacked on pallets inside the vineyard and later moved outside by tractor with elevator machine. Then pallets are loaded on trucks and carried to the packhouse for the final stages: quality control, pre-cooling, packing, final cooling and loaded to customer.

2.3 Pruning

The pruning technique includes the cuts and other operations, carried out in different periods in order to:

- adapt the plant to the chosen training system;
- make production regular over the years, avoiding production stress and prolonging the efficiency of the vineyard;
- adjust the quantity of product in order to obtain the desired quality;
- keep the development of the aerial part balanced in relation to the root system;
- arrangement of vegetation and production in an appropriate way.

It is important to distinguish winter pruning, which is done in the absence of vegetation (depending on the varieties and production area, from November to January) from green pruning which refers to the operations carried out on the vegetation and ensures the correct balance between vegetative part and fruit. The spatial arrangement of the buds, for winter pruning, considering the future vegetation and production, is carried out according to the training system. The different vegetative expansion, given by the quantity of leaves and shoots, is related to the vigor of the plant. It is very important to produce by achieving an optimal level of vegetative vigor, as this affects some very important qualitative and quantitative parameters (such as bunch and berry sizes).

Depending on the age of the plants, pruning can be classified as:

- *training pruning*, which is carried out on young vines, in the first 1-2 years, which consists in cutting short or long according to the development of the plant;
- *formation pruning*, 2nd year, to give the final shape for the vine, according to the training method;
- *production pruning*, from 3rd year, which represents the management of the vine in the adult phase.

The number of buds per hectare can be:

- Low: 30.000-40.000 buds/hectare;
- Medium: 50.000-60.000 buds/hectare;
- High: 70.000-80.000 buds/hectare;
- Very high: over 80.000 buds/hectare.

Depending on the length of the shoot left for winter pruning, this can be defined:

- Short: 2-3 buds per cane (spurs);
- Medium: 4-7 buds per cane (fruit head);
- Long: 8-10 buds per cane;
- Very long: if more than 10 buds cane are left;
- Mixed: if both spurs and fruiting heads are left on a vine.

Fertility is a genetic characteristic of grapevine cultivars, which means the ability of a certain grape cultivar on formation and production of flower parts, which determine the yield (Martin, 1968; Al-Dujaili, 1980; Lopez-Miranda et al., 2000).

Fertility changes according to:

- Genetic (variety);
- Position of the fruit head (inclination and curvature);
- Position of the buds (bud fruitfulness generally increases from the base to the middle and decreases again toward the distal part);
- Inhibitions that occur on the fruit head.

Many factors could determine a low bud fruitfulness: environmental conditions (light intensity, temperature, and shade), varieties and vineyard management (pruning, irrigation, nutrition, and hormone application). Excess shoot vigor can lead to reduction of bud fertility, which is correlated with the occurrence of bud necrosis (Shikhamany, 1999). It is important to distinguish the potential fertility from the real/actual fertility. The potential fertility (potential bud fruitfulness) is the number of inflorescences primordia in the buds; the real fertility (actual bud fruitfulness) is the number of clusters observed in the field.

2.3.1 Canes Detection

Detection of the main branches and canes of the specific plant. Identify Mother (upper part), and Daughter (lower part) of any internode. Sometimes the operators need to clear the area from other plants canes that occlude the interested plant.

2.3.2 Canes Selection

To put theory into practice, the pruning concepts and methodologies above described can be summarized with a practical and simple rule that most of the time allow to correctly prune plants in full production, 2 years older: ***choose the first good cane***.

Main characteristics to identify ***the first good cane*** are:

- **External** To allow a correct development of the plant structure over the years and increase exposure to sunlight the cane should be on the external side of the branch, considering the main intersection;
- **Minimum diameter** Avoid too small and too big cane diameter
- **Healthy** Not dry and absence of any diseases;
- **Minimum distance** To facilitate the lymphatic flow the cane should be not too close to the internode (usually not closer than 4/5 cm). If distance is more than 20/25 cm, it could be useful to leave a spur that can be used for next year pruning;
- **Minimum length** the cane should be long enough for the requested number of buds (buds selection);
- **Main branches loading balance** Considering the total number of canes requested, increase the number of good canes on the upper part of the plant (Mother). The lower part (Daughter) getting more lymphatic flow is stronger and could take all nutrients weakening the mother. This general rule is valid for main branches, secondary branches, and canes where the number of buds should be higher on the mother side compared to the daughter side. The classification of mother and daughter is exemplified in Figure 13.

With a sequentiality criterion all the remaining ***not first good cane*** must be cut off (non-compliant). For example, if canes target is 5, the operator chooses the first 5 good cane to leave on the plant; all other canes will be identified as non-compliant and so they will be removed.

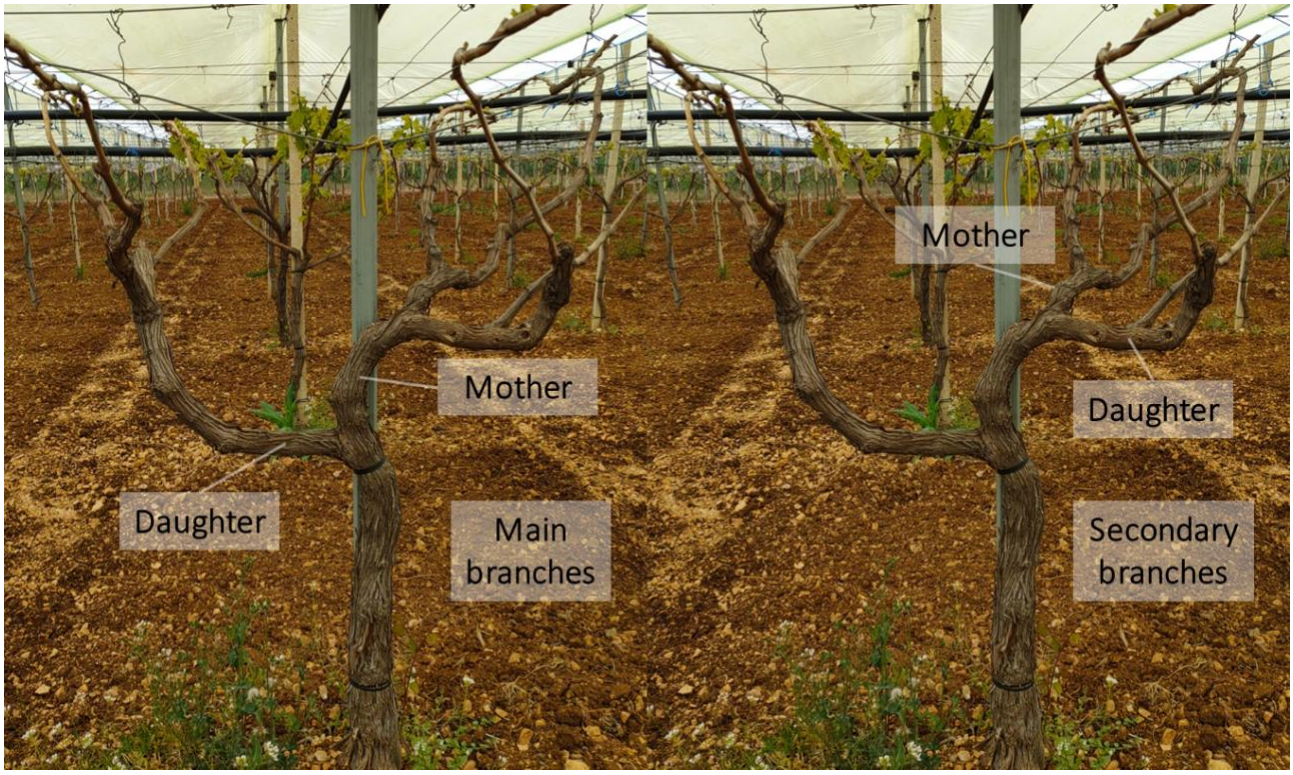

Figure 13 - Mother and Daughter sample.

The cane cut is usually performed with a single hand holding scissors and with the second hand helping to remove obstacles and moving the cane in the right position to execute the cut in the

optimum position and direction: close to the intersection (few mm) and perpendicular to the cane.

Examples of scissors are shown in Figure 14.

Despite big cut to main and secondary branches should be avoided to limit risks of disease and infections, sometimes they are needed and performed with big scissors using both hand of the operator (big cuts require much more strength than normal cut).

Nowadays new electric scissors are also available that allow for faster and easier cutting.

Figure 14 - Pruning tools.

2.3.3 Buds Selection

Seasonally, according to variety, year of production, plant density and production target the grower decide the total number of buds to have in production in the specific vineyard. The total number of buds divided by the total number of plants and the number of canes per plant, gives the number of buds to leave per cane. For example, if total buds target is 50.000, number of plants is 1.400, canes per plant are 5, then buds per cane will be 7. The operator holds with one hand the cane and with the other hand cuts the cane after the 7th bud.

3 Analysis of the Architectural Functional Specifications

In this section, we provide a description of the proposed architecture and related specifications.

The architecture comprises the following components:

- Networking Communication Infrastructure;
- Mobile Ground Platforms;
- Dual-Arm System;
- Box-Exchange Mechanism;
- Sensorial Equipment;
- Virtual Reality Environment.

A detailed description of each component is provided in the following.

3.1 Networking Communication Infrastructure

The Networking Communication Infrastructure relies on a Wireless Network Backbone (WNB) which allows to keep all the robotic components of the system interconnected by providing the possibility to use the ROS architecture in the field.

The selected WNB architecture is based on a set of mesh antennas to create a mesh network on the field, so that UGVs can operate in the field itself. In addition, if the logistics of the experimental site renders it possible, also two long-distance antennas will be deployed to have a control room stored in a remote warehouse connected to the mesh network deployed in the vineyard.

For the sake of easy hardware and software integration, we decided to rely on commercial Wi-Fi solutions based on the 802.11 standard, at either 2.4GHz or 5.0GHz. The WNB consists of the following components:

- UNIFI antennas (Ubiquiti UAP-AC-M-PRO) for creating the mesh network;
- AIRMAX antennas (Ubiquiti LiteBeam 5AC GEN2) for creating the long distance point-to point connection;
- UNIFI Security Gateway (or any other commercial router)

Regarding the mesh network, the UNIFI antennas are wireless connected, i.e., through an uplink-downlink radio connection as specified by the UNIFI Protocol, in order to create a mesh network. In this way, the signal is re-broadcasted through a desired network topology pre-configured from a UNIFI Controller, permitting the connection between all devices in the field.

If the long-distance point-to-point connection is deemed necessary to be implemented and the logistics of the experimental site allows for it, then the first UNIFI antenna can be wired connected to the AIRMAX antenna to be deployed in the field, while the other AIRMAX antenna can be wired connected, i.e., through an Ethernet cable, to any computer in the control room located in the warehouse and can be orientated to the second antenna placed in the field to maximize communication bandwidth.

Concerning the specifics of the selected antennas, the Ubiquiti UNIFI UAP-AC-M-PRO is an access point wireless dual-band 802.11AC 3x3 MIMO equipped with an integrated omni-directional antenna, ideal to cover ample external zones. In addition, the AIRMAX LiteBeam 5AC GEN2 is an ultra-lightweight airMAX ac CPE device, that can provide 23 dBi of gain for long-distance connectivity, with throughput 450+ Mbps, and which uses a directional antenna pattern for improved noise immunity.

Regarding the router, this is required for handling the interconnection over the network by providing basic network services such as DHCP. In particular, two different configurations are possible according to whether the network is going to include or not the point-to-point long distance connection. If so, then the router will be placed inside the control room, i.e., an area dedicated to the project to be defined within the facilities of the vineyard, otherwise it will be placed within the field attached to the first mesh antenna.

3.2 Mobile Ground Platform

An initial search for appropriate mobile platforms that would cover the research requirements has been done before the beginning of this project. As described in the research proposal, the consortium has then identified the Alitrak DCT350P, shown in Figure 15, as a capable solution.

Alitrak's platform is an electrically powered mobile base which uses tracks for its mobility. The platform is market-ready and is already used in domains with similar operational requirements as the vineyards where we will deploy and validate all proposed methods. An example of validated operating domains is construction. Therefore, the platform is proven to operate under dry conditions with a lot of heat and dust.

Figure 15 – Alitrak DCT350P mobile base.

The DCT350P has a track design allowing mobility in uneven terrains while being stable by design and not allowing for the platform all the equipment to tip over. The tracked design provides agility to maneuver in the narrow space between the vine tracks, i.e., using skid steering the mobile platform can perform 360 degrees of rotation while remaining on the same spot. Moreover, the platform has the capacity to allow for up to 700kgs of payload, more than enough for all required equipment and mechanisms developed in this project. The platform also comes with multiple holes on its main frame, thus allowing for multiple mounting considerations for all required equipment. In addition, safety features are provided with the DCT350P, such as an emergency stop installed in the back of the platform, and a wireless emergency stop installed on its wireless remote control. Moreover, a "heartbeat" feature between the wireless controller and the platform is implemented by Alitrak, where once the platform is losing connection with the controller it automatically switches off. With the help of Alitrak, the consortium is also considering the addition of extra wireless emergency stops. Finally, using the included power source (batteries) DTI is able to design and implement an electrical power distribution system (currently designed at 24V and 240W) that covers the power requirements of all sensors and computers for navigational purposes. The mobile platform has space for extra batteries and electrical components that will cover the needs for the dual-arm system on the farming robot and Box-Exchange mechanism on both farming and logistic robots.

For the control and navigation of the mobile platform an appropriate compact computer will be used. The consortium has identified the Intel NUC 11 Pro Mini PC NUC11TNKv7. The computer contains the latest generation CPU technology, i.e., Intel® Core™ i7-1185G7 Processor (12M Cache, up to 4.80 GHz) which will be able to cover all the processing requirements of the sensor drivers and developed algorithms. We are going to equip the computer with 64GB of RAM which is paramount for the multiprocessing requirements of our platform. Furthermore, the computer will equip the NUC with a 1TB drive that will allow the storage of the operating system and developed software, while also providing space for recording ROS bags, i.e., timestamped logs from all sensors through the ROS middleware. The NUCs also come equipped with a variety of inputs and outputs, such as high bandwidth USB, Thunderbolt ports, Ethernet and WiFi6 networking capabilities. The included I/O will guarantee high bandwidth and low latency connectivity with all the sensors simultaneously. Finally, the power requirements of the computer, i.e., allowed input voltage in the range of 12V-24V, allow an easy integration using the designed power distribution system.

3.3 Dual-Arm System

The dual-arm system provides manipulation capabilities to the overall integrated system. These skills are essential to carry out the harvesting and pruning operations described in Section 2. The design and development of the dual-arm system will be mostly in charge of the partner PAL which has consolidated expertise in the field.

In particular, in CANOPIES, PAL wants to further advance its manipulator platform and develop a product in a new niche that will create opportunities in worldwide markets in the agricultural segment, especially in agronomic manipulation where market ready solutions are scarce. New opportunities will be also enhanced by the occasion to create a version of the upper-body of their robot that is standalone, independent by the base, that can be mounted on several mobile bases.

This covers alterations in the mechanical design of the current robot TIAGo++ (see Figure 16) in order to increase its capabilities. The starting specifications of the TIAGo++ robot are the following:

- Base footprint: Ø 0.54 m;
- Weight: ~ 100 kg;
- Torso lift stroke: 0.35 m;
- 7 DoFs each arm;
- Arm payload (without end-effector): 3 kg;
- Arm reach: 0.87 m;
- Pan-tilt head with a RGB-D camera;
- Optional Nvidia Jetson TX2;

Where the RGB-D camera is a specific type of depth-sensing device that works together with a RGB sensor camera to augment the conventional image with depth information (related with the distance to the sensor) in a per-pixel basis. This may provide a significant contribution to solve or simplify many challenging tasks, such as object detection, scene parsing, pose estimation, visual tracking, semantic segmentation, shape analysis, image-based rendering, and 3D reconstruction.

Figure 16 - TIAGO++ robot.

The new design will include an additional DoF in the torso with a minimum range of 300°, for rotating about itself to enlarge even more its workspace, and torque sensors or Series Elastic Element (SEE) and brakes in each joint of the arm to allow it to be compliant and increase its safety rating when operating along with humans. This design will prevent the exert of any high forces against a person or obstacles, which is a very important asset in the collaborative application. The arms will be also equipped with force sensors mounted at the end-tip for measuring the contact forces in the interaction with the environment. These will be mounted between the robot custom end-effector and the tool whose dynamics will be identified and compensated. The arm communication bus will be migrated from CAN to EtherCAT or a similar one allowing a higher bandwidth. Furthermore, all the sensors included in the arm as well as the new communication bus will pave the way to implement a broader set of controllers for controlling the force applied by the robot as admittance and impedance control. A camera is needed on the end-effector to have information about the cut and the grasping as detailed in the following Section 3.5.2. During the design phase, an analysis will be carried to check the feasibility of integrating either a RGB or a RGB-D camera. The dual-arm system will include speakers and a microphone and a light on the body of the robot for lighting the dark spot or when the sun light is not good enough. Then, the flat tray on the back and the part on the head will allow to add more sensors needed for the other functionalities. The head covers and the current RGB-D camera may be removed and replaced by other sensors depending on the functional requirements.

The agronomically adapted dual-arm design will be accompanied by the design of custom grippers that will allow the execution of the harvesting and pruning tasks. The current design includes the development or integration of a robust gripper (important for the pruning) and in the next deliverables we will describe whether we are going to integrate electric scissors and a net gripper in

the gripper (used for harvesting) or whether we will go for other solutions and other tools or towards the integration of third party end-effectors with the required tools attachable as add-ons.

Here some specifications of the agronomic dual system are resumed:

1) Upper body with its own battery and computer;
2) Prismatic joint in the torso with stroke of 0.35 m;
3) Rotational joint in the torso with a minimum range of 300°;
4) 2x7 DoF arms with torque sensors or SEE and brakes and 6-axis F/T sensors in the end-tip (before the end-effector);
5) Migration from CAN to EtherCAT to have higher control bandwidth;
6) New end-effector with under-actuated fingers and different custom add-ons to attach tools required by the agricultural operations;
7) Pan-tilt mechanism on top of the upper body to allocate sensors with a payload around 1 kg;
8) Cameras at the end-tip of the arms;
9) Speakers and microphone for ASR and TTS;
10) Extra lights on the body to illuminate dark spots;
11) ROS API.

Figure 17 shows a first draft of how the dual-arm system will look like when mounted on the mobile base. More details on the specifications and design of the dual-arm system will be discussed in the following deliverables due to the design pre-study currently being ongoing.

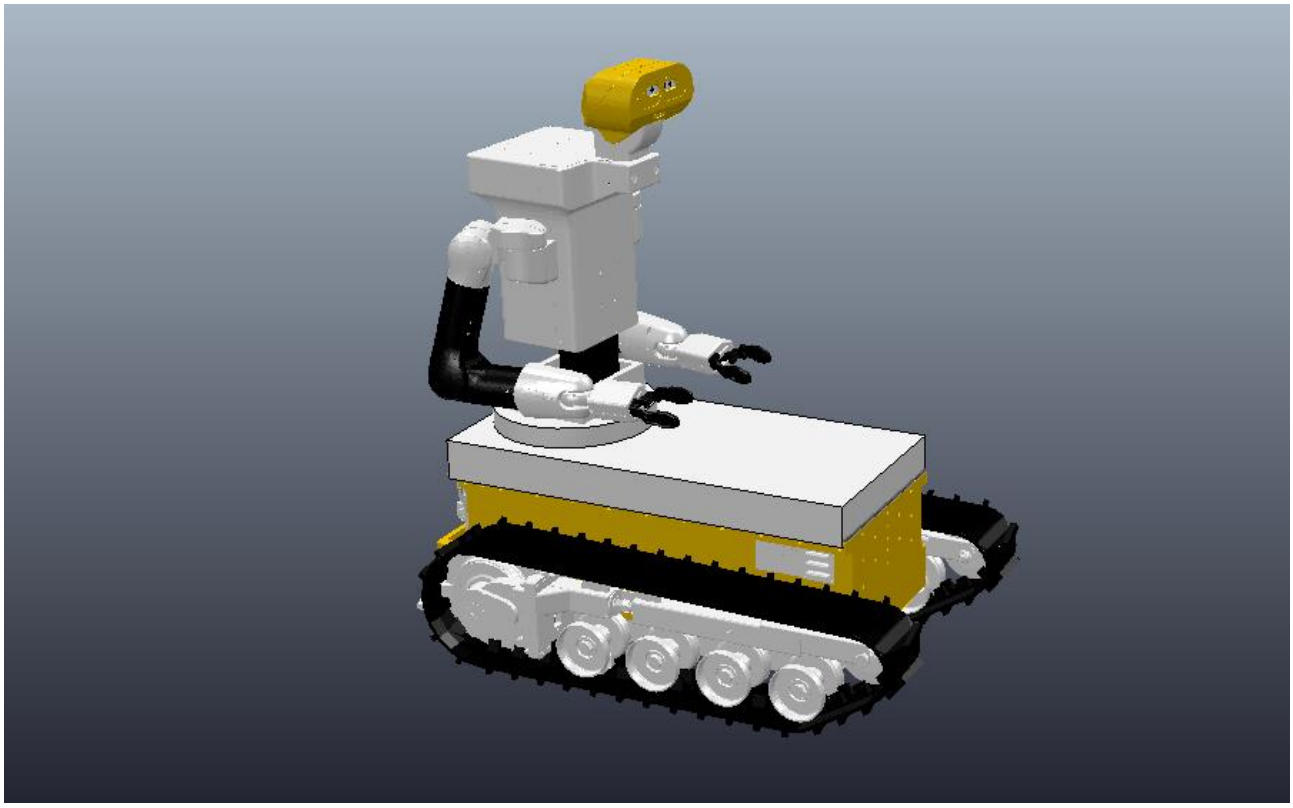

Figure 17 - Draft of the dual-arm system mounted on the mobile base.

3.4 Box-Exchange Mechanism

The Box-Exchange-Mechanism (BEM) is the mechanism that enables to *i)* transport boxes filled with grapes from a farming robot to a logistic robot and *ii)* provide empty boxes back to the farming robot, thus servicing and allowing the continuous operation of the farming robots in the field. Multi-robot coordination strategies will be implemented to synchronize and position farming and logistic robots.

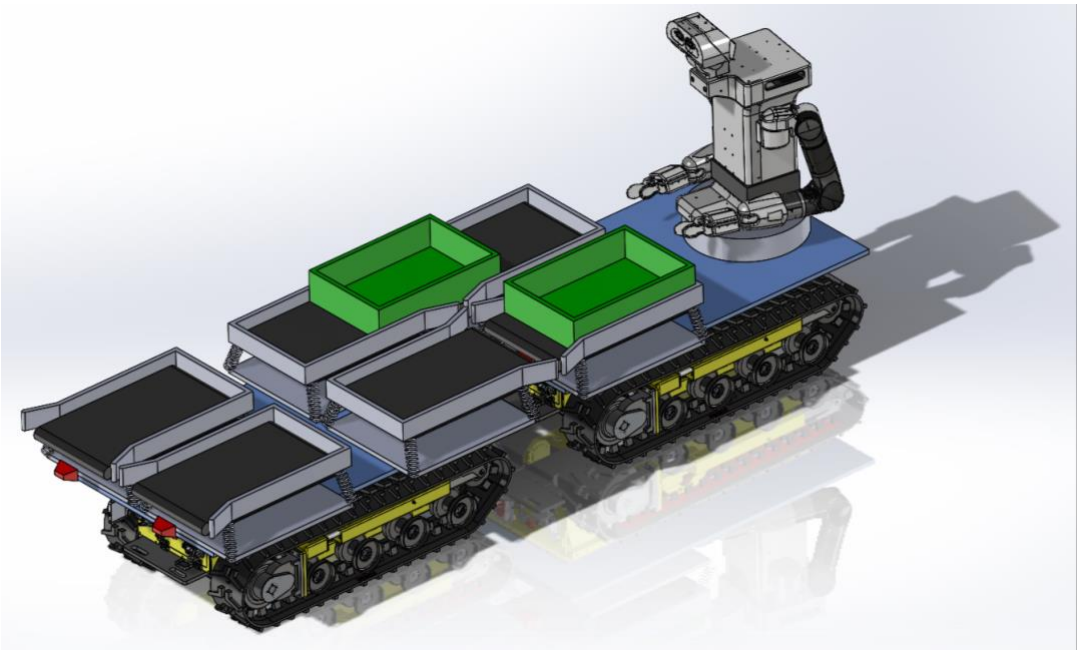

Figure 18 - Concept for the BEM.

The initial concept for a BEM, shown in Figure 18, includes the use of two mechanisms, one implemented on the logistic robot and one on the farming robot. In both mechanisms, there is an extended use of conveyor belts that fit the specifications of the boxes used in the experimental process, i.e., size and weight.

For the logistic robot, two conveyors covering most of the top plate of DCT350P will be used. Therefore, there are considerations for overextending the footprint of the mechanism and therefore using the mounting positions already available by the mobile platform. One of the conveyor belts will collect the boxes from the farming robot while another will transfer empty boxes back to the farming robot. This mechanism should be able to carry more than 4 boxes, thus being able to service at least two farming robots.

For the farming robot, two smaller conveyor belts will be used. It is currently considered that the footprint of the mechanism will be half of the one used in the logistic robot, while respecting the specification of the experimental task and boxes.

Some preliminary considerations about the specifications include:

- The transport of boxes will potentially be assisted with metal rollers and rubber bumpers that will be mounted on the mobile platform.

- An independent power supply and management system will be required in order to achieve the maximum autonomy, thus allowing for more experimental and validation time.
- BEM control hardware must be robust for the test conditions and compatible with connection with the rest of the system, thus allowing for easy and repeatable control.
- The BEM will potentially be supported by sensors that will verify the state of the boxes, i.e., position, fullness, etc.

Partner DTI is also working on several other configurations that will either enhance the current concept.

3.5 Sensorial Equipment

3.5.1 For navigation purposes

In order to effectively navigate the vineyard, sensors are required to enable localizing each robot, measuring attitude, position and velocity, with respect to a global reference frame, and mapping the vineyard itself. For this purpose, the sensorial equipment for navigation relies on three main components:

1. Septentrio GPS- Real-Time Kinematic positioning (RTK) system;
2. SBG Ellipse-E Inertial Measurement Unit (IMU);
3. Ouster SO1-64 Lidars.

Each component is detailed in the following.

GPS-RTK system - The GPS-RTK positioning system provides accurate location information. This capability facilitates the development of safe and autonomous guidance algorithms for the unmanned vehicles. Moreover, it allows to collect geo-referenced data, such as images or 3D laser cloud-points. The selected GPS-RTK positioning system exploits the availability of the WNB infrastructure for transmitting GPS corrections over the field.

The GPS-RTK positioning system consists of a base station, Septentrio AsteRx-U Global Navigation Satellite System (GNSS) receiver, which is common for all robots and the following elements for each robot:

- 2 GPS antennas, Septentrio PolaNt-x MF;
- 1 GPS modules Septentrio AsterRx-m3 Pro+.

The resulting system architecture is described in the following. The base station is connected to the WNB to transmit GPS L1/L2/L5 and GLONASS L1/L2/L3 signal corrections. After receiving these corrections, the GPS module mounted on-board each robot applies them to an initial localization provided by the GPS antenna, generating a more accurate localization. The latter, as explained in the following, is further corrected using information from the IMU sensor.

Concerning the specifics of the selected elements of the GPS-RTK system, the AsteRx-U base station is a multi-channel, multi-frequency GNSS receiver, which is suited for applications that require precise heading information in addition to position. The following satellite signals are simultaneously tracked:

GP, GLONASS, Galileo (optionally activated), BeiDou (optionally activated), SBAS, IRNSS and QZSS. Furthermore, it guarantees centimetre-level (RTK) and dm-level (PPP) position accuracy.

The PolaNt-x MF antenna incorporates low-noise amplifiers and allows multi-frequency GPS, GLONASS, Galileo, BeiDou, QZSS, NavIC, Compass and L-band signal reception.

The AsterRx-m3 Pro+ GPS module is a triple-frequency multi-constellation GNSS receiver which provides, in dual antenna mode, heading and pitch or heading and roll information along with reliable and accurate positioning. Moreover, high update rate of 100 Hz guarantees low latency.

This combination of technologies allows reaching the following specifics:

- Accuracy: < 1 cm + 1 ppm
- Initialization time: < 10 s
- Operating range: < 40 km
- Latency: < 20 s

Finally, Septentrio products are fully integrated in the ROS middleware providing packages[1] that are easily configurable and make the integration of the GPS-RTK system straightforward.

IMU sensor - The IMU sensor provides the system with a dead reckoning capability and integrates the RTK-GPS system to achieve a more accurate location information. More specifically, each robot is equipped with an IMU sensor which is connected to the GPS module and an Extended Kalman Filter (EKF) fuses the data coming from the IMU sensor and the GPS module. Moreover, by exploiting the dual antenna mode of the GPS module, the heading information is also provided by the IMU sensor.

The SBG Ellipse-E IMU is a lightweight sensor (50 grams) which guarantees accuracy of 0.05° for roll and pitch (using RTK), 1 cm RTK/PPK + 1 ppm for position, and 0.03 m/s for velocity. Moreover, it natively provides the connection to an external GNSS module and implements the real-time EKF data fusion algorithm.

Finally, also the IMU sensor is fully integrated in ROS middleware for which the package sbg_driver[2] is available, drastically easing the system software integration.

Lidar sensors - The lidar sensor provides 3D point clouds of the surrounding environment which are fundamental for mapping it as well as for safely navigating.

Each robot is equipped with two Lidar sensors mounted on the mobile base to ensure 360° coverage of the environment.
Ouster SO1-64 is a mid-range lidar sensor with vertical resolution of 64 channels and horizontal resolution equal to 512, 1024, or 2048, allowing to elaborate 1,310,720 points per second. It guarantees an operating range of 120 m and vertical field of view of 45°, with precision ±0.7 – 5 cm.

[1] http://wiki.ros.org/action/fullsearch/septentrio_gnss_driver
[2] http://wiki.ros.org/sbg_driver

As for the previous navigation sensors, Ouster SO1-64 provides a ROS driver[3] which facilitates the software integration.

3.5.2 For agronomic purposes

General analysis for target detection and localization: For the target detection and localization tasks there are two different scenarios with clearly different features, namely: *i)* detection of grape clusters and the peduncle for harvesting; and *ii)* detection of canes and buds for pruning. Both scenarios require that the detections are achieved with performance rates fast enough to enable their utilization within a control loop, while providing robust estimation through tracking. Also, both harvesting and pruning will be performed in the same environments, so there are common elements (specific distribution and dimensions of the vineyard) that will impact sensor performance and usefulness.

Detection of grape clusters and peduncles is performed during harvesting operations. There are three main sensor technologies available at the market that could be used in the dual-arm configuration, namely Lidar, active RGB-D vision sensor and stereo vision (or passive stereo vision) sensor accounting for restrictions and costs (payload, monetary, computation, etc.). For this concrete task, the most convenient sensor would be active RGB-D vision sensor (considered the default option in most literature), deployed preferably in one of the arms used to track the clusters as mentioned in the description of the dual-arm system in Section 3.3. This would allow tracking clusters with enough accuracy to harvest them without risking damaging them (as table fruit, quality requirements are much more stringent compared to wine grape). Notice that correct cluster harvesting requires robust detection and positioning of the cluster, the peduncle, and the cane as explained in Section 2.2 and as shown in Figure 11. This implies that, depending on the final positioning of the sensor, a pan/tilt system could be required to reduce occlusions produced by the same arm/cluster involved in the operation.

Although the discussed active RGB-D technology is the standard approach to fruit detection according to state-of-the-art literature, there are risks associate with uncontrolled outdoor illumination, thus, at least one secondary sensor is required. This requirement can be solved through deployment of a secondary vision sensor in another arm or on top of the dual-arm robotic configuration, as mentioned in Section 3.3. Deployment of a stereo vision sensor should be also the preferred, as it will completely remove the risk associate with utilization of IR technologies completely. If the stereo vision sensor on top of the dual-arm configuration is considered to solve this problem, the resolution of this sensor should be assessed to guarantee accuracy in the cluster detection. In this scenario, the pan-tilt mechanism of the Tiago robot, described in Section 3.3, will be exploited so that perception task and robotic arm task does not interfere with each other.

With respect to the cane detection tasks, as these are to be performed in the context of pruning operations, vines will not present foliage. This means that the number of possible occlusion-producing elements will be greatly diminished, as the canes and buds will be directly exposed. Robustness is critical, as instead of damaging a cluster, the risks associated to pruning operations is

[3] https://github.com/CPFL/ouster

potentially damaging the vine plants. Classical skeleton extraction will not be able to provide all the relevant information required for pruning policies, so additional perception and modelling capabilities are required, with parametrical modelling being the preferred method in literature.

Given the default uncontrolled illumination found in this scenario (no foliage present), utilization of artificial illumination on board on the robotic system should be strongly considered and studied, as active RGB-D technology available at feasible costs has been consistently reported to be incapable of delivering robust detection when exposed to open sunlight. Still, even using artificial illumination, deployment of at least a stereo vision system is needed. The positioning of this sensor should be one that guarantees accurate detection of the canes and buds to prune. A high-resolution stereo rig deployed on top of the dual-arm configuration can satisfy this restriction, as long as the combination of sensor resolution and computational power available provides an estimation good enough to model the canes in such fashion that the pruning point can be correctly assessed. Alternatively, an additional stereo vision sensor with lower resolution requirements could be deployed in an arm, as mentioned above and in Section 3.3, to enable short distance detection and modelling of the canes/vineyards, and close tracking of operations with respect to the cutting point.

Lidar technologies will also be considered in both detection tasks. We will consider the deployment of a Lidar sensor in an arm, close to the end effector. Current designs and specifications make this an unlikely option. Still, the consortium has not fully precluded an upgrade including market-ready solid-state LiDAR solutions, which could lead to viable solutions. Alternatively, we will consider the deployment of the Lidar sensor close to the robot torso.

Additionally, the inclusion of an active RGB-D vision-based system (as the Intel Realsense and the Mynt Eye P proposed in the following) or certain types of passive stereo vision sensors (Zed 2 camera), will require deployment of CUDA compatible hardware platforms to enable utilization of state-of-the-art deep learning (instance/panoptic segmentation) and high-performance stereo vision.

Sensor specifications:

In the following, possible commercial solutions for vision-based and lidar sensors are reported.

Vision-based sensing:

- Intel Realsense D435i (active IR RGB-D sensor with IMU)
- Intel Realsense D415 (active IR RGB-D sensor)
- Intel Realsense D455 (active IR RGB-D sensor)
- Mynt Eye P (active IR RGB-D sensor)
- Stereolabs Zed 2 (passive RGB-D/classical stereo vision sensor)

The following Table 7 reports the specifics of the above vision sensors.

	Realsense D435i	Realsense D415	Realsense D455	Mynt Eye P	Stereolabs Zed 2

Video Resolution	1920 × 1080	1920 × 1080	1280 × 720	2569x720	up to 4416x1242
Depth Resolution	1280 × 720	1280 × 720	1280 × 720	1280 × 720	Equal to video resolution
RGB Frame Rate	30 fps at maximum resolution	30 fps at maximum resolution	30 fps at maximum resolution	60 fps at maximum resolution	15 fps at resolution 4416x1242, 60 fps at resolution 2560x720
Depth Frame Rate	30 fps at maximum resolution	30 fps at maximum resolution	30 fps at maximum resolution	60 fps at maximum resolution	Equal to RGB fps
IMU	X				
IR	X	X	X	X	
FOV	H:87° V:58° D:95°	H:65° V:40° D:72°	D:95° H:86° V:40°	D:86° H:75° V:40°	D:120° x H:110° x V:70°
Range	0.3 to 3 m	0.5 to 3 m	0.6 to 6m	0.2 to 4.2m	0.2 to 20m
Weight	72 gr	72 gr	387 gr	184 gr	124 gr
Dimension	92x25x25 mm	99x23x20 mm	124x26x29 mm	124x33.3x32.5 mm	175 x 30 x 33 mm
Support ROS Melodic	X	X	X	X	X

Table 7 – Specifics of the possible vision sensors.

The appropriate cameras will be selected by carrying out laboratory trials as well as through extensive trials in the Aprilia vineyard to validate the two sensors in realistic outdoor conditions.

Lidar range finder sensing:

- 3D LiDAR on the shoulder or close to the torso (as a backup for safety reasons).

In summary, a vision-based sensor (active RGB-D or stereo camera) is to be deployed in the main operating arm, close to the end effector. A secondary vision sensor based on a different technology should be ideally deployed in the second arm. Should this placement not be possible, this sensor can be placed on top of the dual-arm configuration or in the base platform, considering resolution, orientation decoupling from the platform (through gimbal/joints), and possible occlusions and effects of illumination. Lidar technology can be used to provide support sensing capabilities, and help with data fusion procedures, joining non-overlapping sensor spaces. It will also be used for safety reasons, in case that the stereo vision sensors have erratic detections of the brunches or the canes.

The exact configurations will be defined after some of the hypotheses exposed in the analysis have been validated in an early experimental phase.

General Analysis for Quality Assessment: As explained in Section 2.2.2, quality of the grapes for the choice of which bunches to pick at a certain time is related to a variety of different physical and aesthetic characteristics which are summarized in the following for the sake of completeness:

- Physical: it is mainly related to the ripeness of the fruit. The main physical parameter is the soluble sugar content in the berries (°Brix).
- Aesthetic: all the berries of a grape have to be in perfect conditions, i.e., without signs of illnesses, rot, pests, or similar problems. In addition, it is important that during the process of the berries are not scarred, pierced, or in any way damaged.

The main sensors to determine the quality of the bunches on the field are vision sensors, since other physical quantity sensors (e.g., portable spectrometers) are not usable on a mobile platform without complex actuations or are excessively expensive. In addition, all the aesthetic characteristics can be detected by RGB sensors alone, since, by definition, are visible characteristics. That said, the number of sensors and their characteristics depend on the specific condition of the environment where the robot operates: the vineyard has specific dimensions, and this constrains the focus length and the field of view of any visual sensor to be used. In particular, early experiments in the Corsira field with the available hardware (MAPIR camera, JIR NIR camera, Realsense L515 RGB D sensor) showed that it will be beneficial to have large FOV devices in order to be able to see most of the grape plant when the robot is close to them.

While it is possible to assess the external quality of the berries by RGB cameras alone, to reduce risks and to increase accuracy, it is advisable to also use depth sensor, to being able to reconstruct the 3D shape of the bunches, and thus the dimension, orientation and other geometric characteristics can be easily measured. In addition, the robot has to interact with the vine plant to gain the best perspective of the bunches and, in the end, to pick them, so some kind of 3D sensor is needed in any case. With this in mind, we observed in the field that active stereo-based sensors, such as the L515 and the Microsoft Kinect are not able to give any useful information due to the intense brightness of the environment. For this reason, the 3D sensors that are considered in this specification are stereo cameras. To achieve good short-medium range combination, more than one sensor will be part of the final setup.

Assessing the °Brix using only passive visual sensors is much more challenging, and it is an active research question. We hypothesize that it is possible to gain some information using some form of multi-spectral sensor, since the physical response of the soluble solids to light is something that is actively used in different agricultural contexts. At the same time, most of the research in these contexts is performed in the lab using hyper-spectral sensors with controlled light. Using this lab equipment, it has been shown that it is possible to test the °Brix up to a 1% error, however it is not easy to know the accuracy of a similar sensor under the direct sun in the vineyard field. In addition, hyperspectral sensors are very expensive, and other solutions have to be explored. Multi-spectral cameras use less frequencies and have a much wider bandwidth compared to hyperspectral sensors,

so we will need to perform experiments in the field to see if the combination of the RGB and multi-spectral sensors is able to keep the °Brix estimate near the 1% error. Even if this will not be the case, the common practice among the workers in table grapes vineyards is that one worker estimates the quantity of sugar of some bunches in the field using a portable spectrometer and then gives visual instructions to the other workers in order for them to pick the right bunches. Since the workers select the bunches to pick only on the visual aspect, we are confident that RGB information is enough to decide which bunches to pick, even if we are not able to measure the exact sugar content.

Sensor Specification:

Stereo Cameras: the same cameras reported for the target detection and localization (and summarized in Table 7) will be considered.

Other Cameras:

The following devices will be specified after an initial experimental phase will be completed. In particular, the multi-spectral device will be defined by experimentation on spectral response during the first harvesting season. Monocular and omnidirectional cameras will be added if the experimentation will show some advantage in doing it.

- Multi-spectral camera (e.g. Silios CMS-V);
- Omnidirectional RGB camera.

Other sensors:

The following sensor could be useful for initial guesses of 3D point cloud reconstruction. However, as for the target detection and localization, we will consider if it is possible to use one of the Lidars mounted for navigation purposed also to accomplish this task before considering the addition.

- 3D lidar or similar sensor

3.5.3 For human-robot interaction

General Analysis for Safety: According to the world's first specifications of safety requirements for collaborative robot applications (ISO/TS 15066[4]), collaborative operations between humans and robots may include one or more of the safeguarding methods, which are depicted in Figure 19 and which are application dependent.

[4] ISO/TS 15066:2016 Robots and robotic devices - Collaborative robots

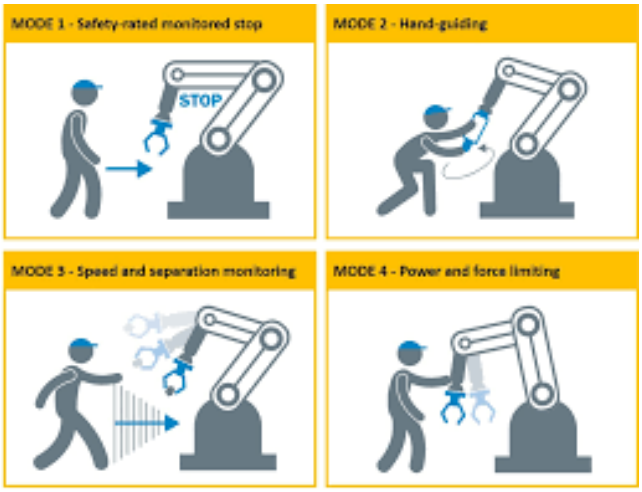

Figure 19- The four forms of collaboration identified by the robot safety standards ISO 10218-1/2 and by the technical specification ISO/TS 15066 (image from work[5]).

In Safety-rated monitored stop mode, the robot is stopped during the interaction when the operator is in the collaboration space. This is the simplest type of collaboration since the robot and the operator can work together, but not at the same time. In Hand-guiding mode, the safety of the human–robot collaboration is assured by the fact that the robot is guided manually and moves at reduced speed. In Speed and separation monitoring mode, speed and motion of the robot are supervised and adjusted based on the position and speed of the operator in the workspace space. In Power and force limiting mode, physical contact between the robot system and the human operator is controlled by limiting the contact force to values such as damages or injuries do not happen.

In Figure 20, a typical configuration of human and robot during non-physical interaction in the case of the CANOPIES project is reported. Obviously, as mentioned before the first requirement to be met in such a scenario is the human safety, and, in this perspective, risk evaluation is a crucial aspect of any safety design process since if a risk/potential hazard is not assessed, then it cannot be controlled and accounted for[6]. However, it is straightforward that this process requires adequate sensors in order to have the robot sense and then, assess the level of safety in real-time during the interaction.

In the case of the CANOPIES project, sensors to be adopted for human-robot-safe interaction consist of both 3D Lidars (also used for navigation purposes as in Section 3.5.1) and RGB-D sensors mounted as discussed in the above and shown in Figure 20, where the minimum set of them mounted on the farming robot are displayed; additional RGB-D sensors, if required, will be mounted either on the robot upper-body or on the mobile base in order to increase the field of view at short-medium distances. Concerning the logistic robots, 3D lidars and RGB-D sensors will be used as well.

[5] Emanuele Magrini, et al. Human-robot coexistence and interaction in open industrial cells, Robotics and Computer-Integrated Manufacturing, Volume 61, 2020.
[6] ISO 12100:2010 Safety of machinery - General principles for design - Risk assessment and risk reduction

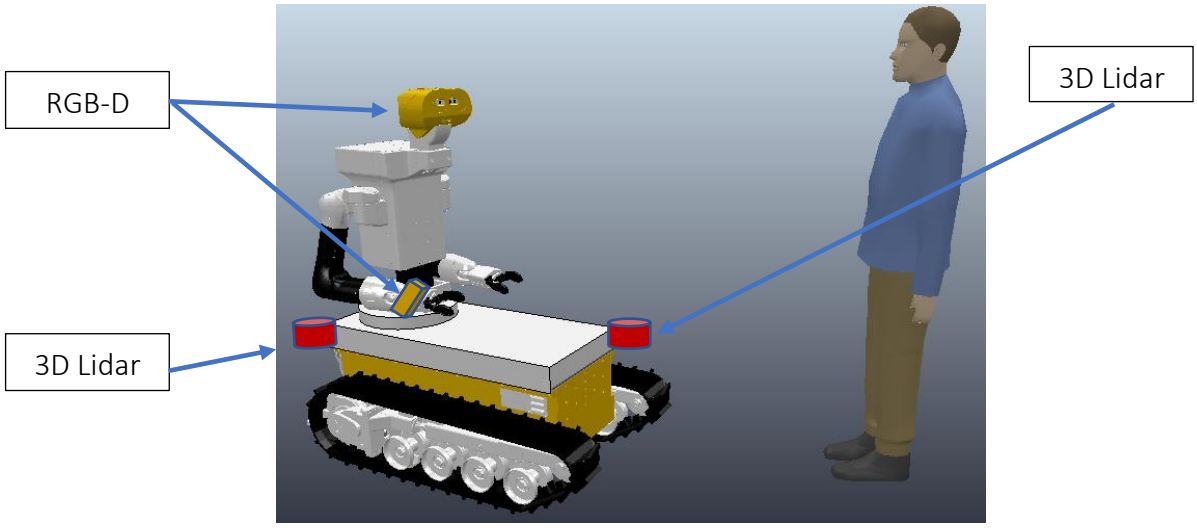

Figure 20 - Typical configuration of human and robot during the non-physical interaction (the robot is not in its final design).

Concerning RGB-D sensors, the same cameras identified for target detection and localization (and summarized in Table 7) will be considered for human safety. RGB-D sensors will be exploited together with OpenPose[7] that is a real-time multi-person system to jointly detect human body, hand, facial, and foot key-points on single images; in particular, it allows to retrieve a representation of the human body as a kinematic tree which is particularly useful for human-robot safety purposes.

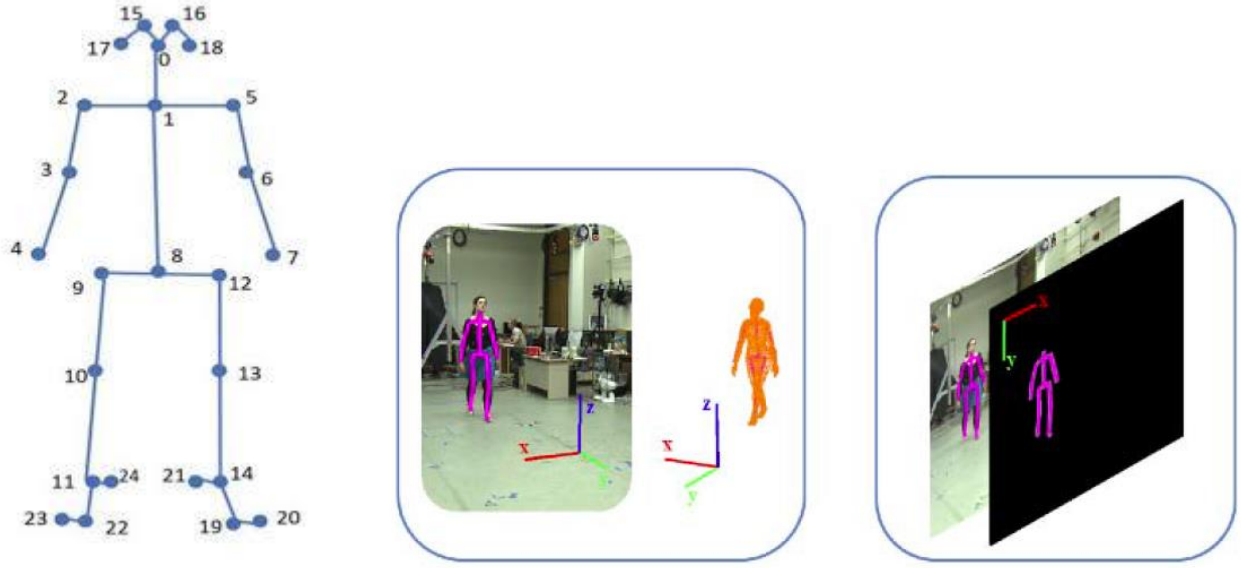

Figure 21 - Kinematic tree obtained from a RGB-D sensors and OpenPose.

Concerning 3D Lidar sensors in Figure 20, Ouster SO1-64 lidars were selected and purchased as already mentioned in Section 3.5.1. The basic idea is to use these sensors not only for navigation purposes but also to monitor the human presence at long distances, so as to complement RGB-D sensors, which mainly work at short-medium distances; in this way, anticipatory robot behaviors

[7] https://github.com/CMU-Perceptual-Computing-Lab/openpose

might be undertaken, and the RGB-D sensors configuration might be adjusted in order to have the human within their field of view.

In the case of physical interaction like in the scenario depicted in **Error! Reference source not found.**, safety needs to take into consideration power and forces limiting requirements, as in ISO/TS 15066[8], by considering both transient and static contacts, as well as voluntary and accidental ones. During the physical contact, the mobile base will be kept still; in detail the base will be stopped whenever the distance between the human and the robot falls below a given distance and this will be achieved thanks to the 3D Lidars and RGB-D sensors mentioned above.

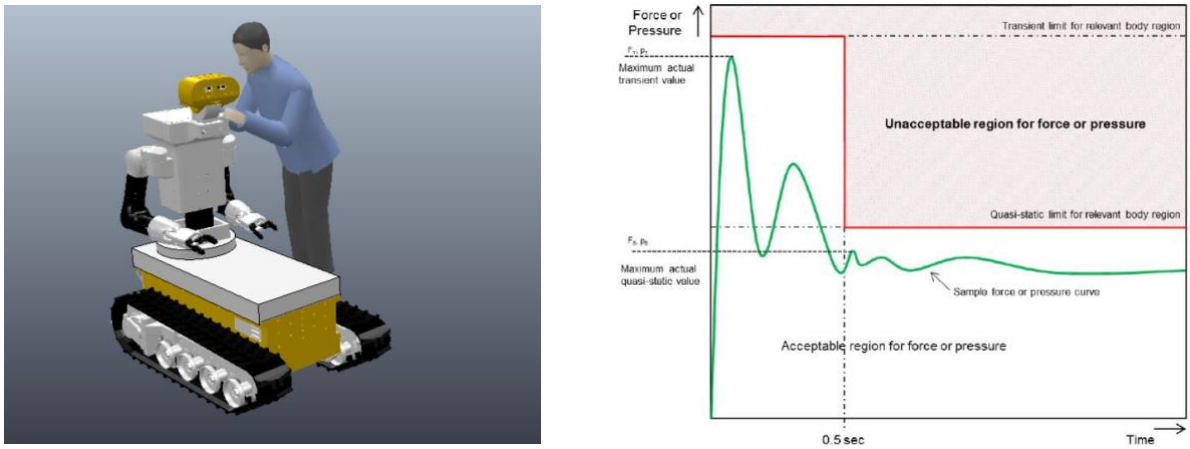

Figure 22 Left. Human and robot during physical interaction (the robot is not in its final design). Right. Power and force limiting plots as in ISO/TS 15066:2016.

In order to achieve a safe physical human-robot interaction, joint torque sensors of the farming robot upper-body, as detailed in Section 3.3, will be available to sense the human contact and limit the interaction forces. These sensors will be exploited together with vision sensors in order to estimate the contact point and correctly estimate the contact forces and enforce the right robot behavior.

General Analysis for Human-in-the-loop: For human-in-the-loop task planning and learning from demonstration, KTH will first validate the developed methodologies on the infrastructure available at the KTH labs. Next, the developed methodologies will be integrated on the farming and logistic robots described in Sections 3.2 and 3.3. In the following, the hardware that will be used for the preliminary validation at KTH is described.

For the online human-in-the-loop replanning tasks, two mobile manipulators with different sensorial equipment are considered. For simpler scenarios, a configuration including a 5-DoF Interbotix WidowX 250 robot manipulator (providing positional feedback, voltage levels, load, and compliance feedback), and a Nexus holonomic base, is chosen. For more complex missions, a robot design with enhanced sensor feedback is, nonetheless, essential. This configuration includes a HEBI Robotics robot manipulator with two versions, one with 6-DoF (model A-2085-06) and another one with 7-DoF

[8] ISO/TS 15066:2016 Robots and robotic devices — Collaborative robots

(model A-2099-07), and a HEBI Robotics holonomic base (model A-2121-01). In contrast to the Interbotix system, both these manipulators can be controlled in position, velocity, and torque, and, more importantly, provide torque feedback and internal IMU accelerations and angular velocities in all the joints.

For learning from demonstrations tasks, Rokoko smartsuit, a motion capture system with 19 IMU sensors is considered to collect human limbs and whole-body movements. For a validation in simulation, an Oculus VR system with a headset and two trackers is chosen for demonstrating skills in a virtual reality environment. The configuration of robotic systems for reproducing the skills include ABB YuMi robots (two 7-DoF arms) and two Franka Emika Panda arms (7 DoF each) with a dual-arm setup. All robots are fix-based, position controlled and equipped with two-finger grippers. ABB YuMi has a payload limit of 0.5kg thus is suitable for manipulating lightweight objects and a closer collaboration with humans. In contrast to that, Franka Emika arms can be used for grasping and transporting up to 3kg fruits.

General Analysis for Communication: In order to set a truly collaborative framework for humans and robots explicit communication is a key element. By explicit communication we refer to all the actions that are executed with the specific intention of achieving a goal that involves the other party. Speech act theory, which is at the basis of agent communication languages, distinguishes several categories of speech acts, including those that are more relevant to our application for example *directives,* which roughly correspond to commands and *expressives* that are focused on the exchange of information.

The explicit communication between robot and humans will take place through a variety of channels, including spoken language, gesture, sound. Moreover, although in the preliminary discussions held in preparation for the proposal it was agreed that the user should not be overwhelmed with additional devices, we plan to explore solutions based on tools such as pointing devices or Augmented Reality.

Sensors:

In order to enable the use of spoken language a *microphone* is needed to obtain a quality sound signal for speech processing. The device should provide a good compromise between quality of signal and burden for the user.

We do not foresee the use of specific cameras or devices to capture human gestures, but we will rely on the equipment that will be specifically acquired for the implicit communication and safety of the human.

Actuators:

Besides *speakers* that will need to provide utterances in natural language clearly understandable by the user, sounds accompanying maneuvering are also to be investigated.

Lighting emitters are also foreseen to combine visual and sound information to replicate and reinforce specifically, but not restricted to, alert signals.

Pointing devices could be used both by the user and by the robot in order to disambiguate references to specific elements of the environments (e.g. a grape bunch). The key feature of a pointing device would be visibility in daylight, while conforming to safety standards.

Augmented Reality settings will also be considered, whose main purposes will be to share references to elements of the environment and to create awareness on the internal state of the robot.

General Analysis for Body Prediction/Planning in Agreement: Human-robot interaction is required in several agronomic operations and in some of them, the robot has to predict the human body motion and use it for planning the robot body motion in agreement with the human or for modifying the robot navigation for safety reasons.

For all the operations where the human and robots are in close proximity, the robot has to detect the human motion. This will be used for modifying the robot navigation and also for knowing the human distance when the robot it is doing a close agronomic operation with the human.

For the rest of the agronomic operations, the human body prediction is required in the following tasks:

- Human removes obstacles and canopies;
- Human helps in quality control/supervision;
- Human harvest “hard to catch” clusters;
- Human helps positioning clusters in the box for better layout;
- Human removes rotten / waste berries;
- Human places “hard to catch clusters”.

In all these tasks, the robot can be close to the human, and has to anticipate where the human arms or body will be. Specifically, the robot needs to know the position and orientation of the human joints, head, and human end effectors (the hands). This information has to be updated frequently (in less than 1s) in order that the robot can modify its motion plan (of the platform and the arms), to do the task in agreement with the robot and anticipate risky operations for safety reasons.

The robot will make the body and arms motion plans using the information of the human body prediction. The motion plans will be shared-plans, that means that the robot will use the human body prediction to create a motion plan that achieves the goal and in a safety manner. These shared plans will be done in agreement with the human, in order that both (the human and the robot) understand the plan that will be executed. In case that the human disagrees with the robot plan, the human will propose alternatives and the plan will be modified.

Sensor specifications:

Stereo cameras: same cameras as for target detection and localization (summarized in Table 7).

Lidar range finder sensing:

- Ouster SO1-64 lidars (as defined in Section 3.5.1).
- 3D LiDAR on the shoulder or close to the torso (as a backup for safety reasons)

A first proposal of the sensors that will be used and their location can be seen in Figure 23 and Figure 24, in which operations of picking a cluster and placing it in the box are considered, respectively.

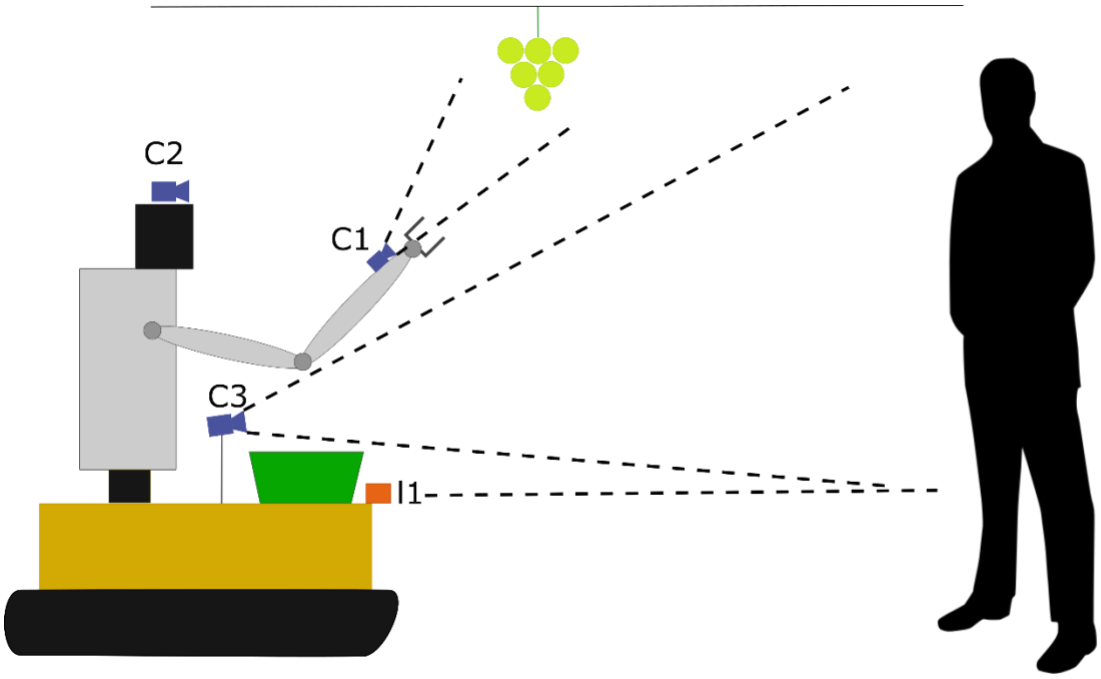

Figure 23 - The robot and the human are in the harvesting area. Cameras (c1,c2,c3) and lidar (l1) are shown.

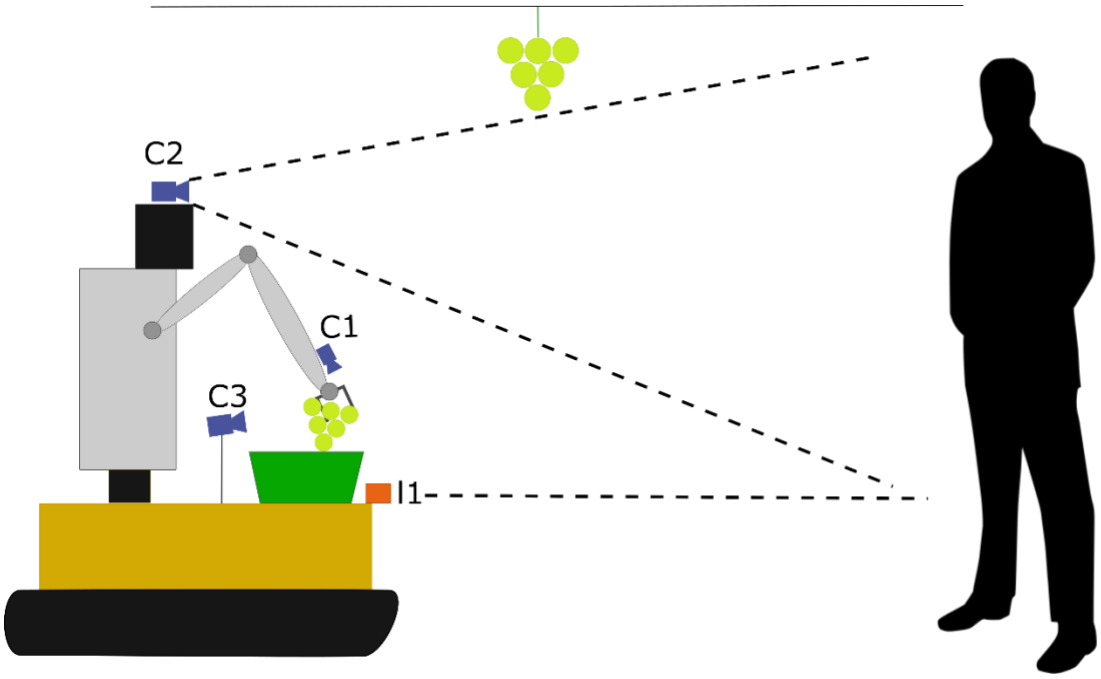

Figure 24 - The robot is placing a cluster in a box and human in helping in the operation. Cameras (c1,c2,c3) and lidar (l1) are shown.

In the two figures it is indicated the location of the sensors. The positioning of the sensors has not been yet definitely decided, and it will be decided once we can analyze all possible human operations

and understand the placement of the sensors that minimize the partial occlusions due to the human and robot arms.

To summarize, from a preliminary analysis the sensors that will be used are:

- Lidar L1: Located in from of the platform for detecting the legs of the humans;
- Stereo camera C1: Located in the arm of the robot for precise detection of the clusters, canes and peduncles;
- Stereo camera C2: Located at the head of the robot for precise detection of the clusters, canes and human body;
- Stereo camera C3: Located at the platform, in front of the robot torso, for detection of clusters, canes and the human body.

A safety Lidar on the shoulder or on close to the torso will be included in case that the stereo vision sensors have erratic detections of the human body, the bunches, or the canes.

3.6 Virtual Reality Environment

In this section, specifications concerning the virtual reality environment are briefly described. In particular, first the virtual reality hardware specifications are given, then the virtual reality software specifications are described.

3.6.1 Virtuality Reality Hardware

The hardware setup for the Virtual Reality needs to cover different aspects of the real-time modeling and simulation process, and thus is composed of a number of different elements.

The high-level set of elements can be listed as the following:

1. PC for running 3D simulation: modeling the virtual environment and the actors, simulating the physics, the sensors, and the motors;
2. PC for rendering graphics: visualizing the environment and actors;
3. Hardware for bridging 3D simulation and ROS: used for robot participants to sense the environment and interact with it;
4. Virtual Reality headsets: used by human participants to sense the environment and interact with it;
5. Networking server: used for synchronization of multi-participant sessions;
6. Networking components.

Figure 25 shows a simplified architectural integration diagram.

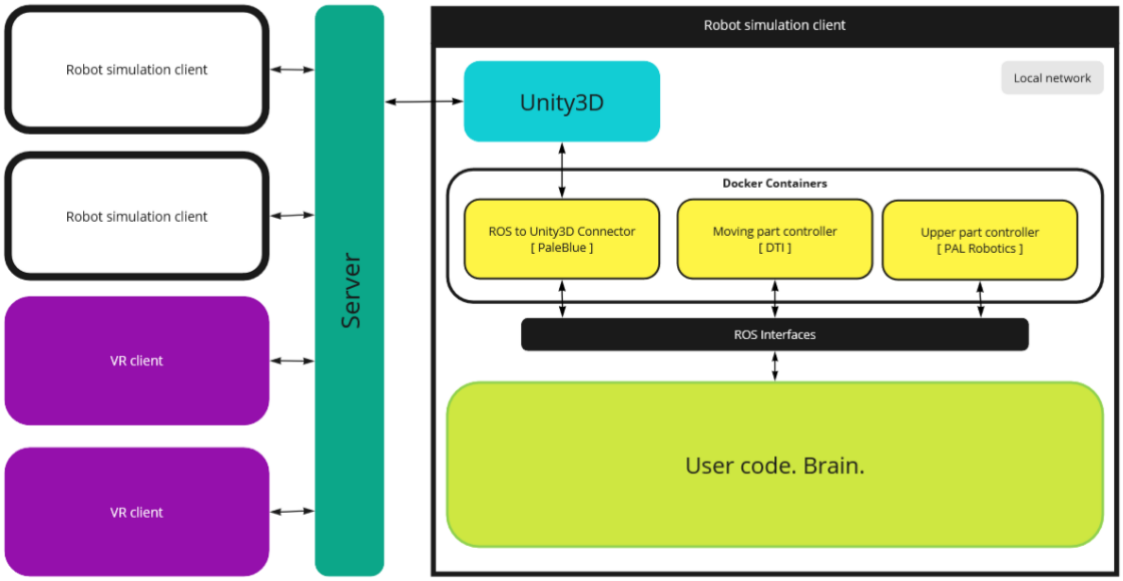

Figure 25 – VR Integration Architecture.

More specifically in the figure, the Unity3D block contains both the PC for running 3D simulation and the PC for rendering graphics. It interfaces with robot controllers and the robot logic through ROS. Moreover, it interfaces with other networking clients through the server PC. These networking clients can be VR clients (used by human participants), or robot simulation clients (used for robot participants).

The hardware selection will be performed based on the following considerations:

- PC for running 3D simulation. Investigation and choice will be made between headless (no user interface), and headful (with user interface). This PC will run an instance of Unity 3D for simulation. Another consideration will be made for combining elements (1) and (2), that is performing simulation and rendering on the same PC.
- PC for rendering graphics. This PC will also run an instance of Unity 3D, but for rendering. The minimum configuration is considered to be at Intel Core i5, 8 GB RAM and a graphics card of Nvidia GTX 2070 or better.
- Hardware for bridging 3D simulation and ROS. This most likely will be a process running on Windows or Linux platform for translating messages between ROS and VR networks.
- Virtual Reality headsets. Investigation will be made between desktop VR (needs a PC for rendering to run) and standalone VR (rendering runs inside the headset on Android platform and does not need connection to a PC for rendering). The main candidate headsets at this moment are Oculus Rift S and Oculus Quest 2.
- Networking server. The prime candidate for network synchronization is Photon Server, and it can be run either in the cloud infrastructure of Photon, or on-premises.

Figure 26 shows a human participant using desktop VR: VR headset and controllers, where the headset is connected by a cable to the PC running the real-time simulation and 3D rendering.

Figure 26 – Human participant using desktop VR.

3.6.2 Virtual Reality Software

The simulator software should be a representation of the real-life system (robots, environment, people). The simulator should replicate the real-life system and its contents, so that testing and validation of the system can take place in the simulator.

The considerations for designing the software are as follows:

- multiple robots should be able to participate;
- multiple people should be able to join;
- the environment should be adaptable to the necessary environmental conditions for testing.

Creating the simulator software will involve working in a close relation between the consortium participants with regular iterations on the software for ROS, robot controllers, and simulation (Virtual Reality) parts.

The first step is the identification of the environment where the robots and the humans will collaborate. A basic environment for simulation is shown in Figure 27, while Figure 28 and Figure 29 show the mobile platform in the simulated environment and controlled through ROS and the TIAGo++ robot in the simulated scene, respectively. This environment will be iterated from simple to complex,

with set dimensions, and will be based on vines structures with few aleatory parameters. This simulation will be based on the specifications in Section 2.

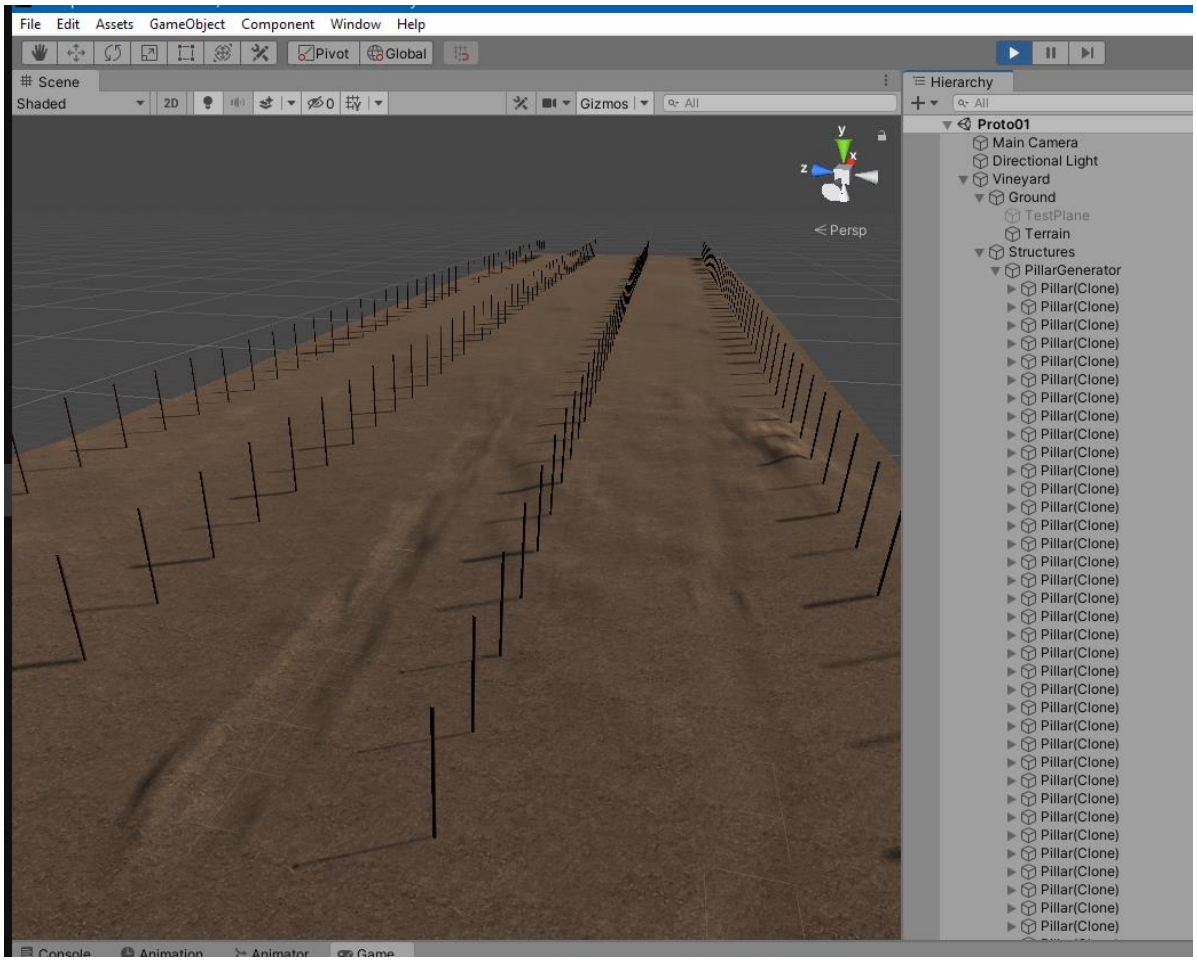

Figure 27 – Basic environment for simulation, with trees and actors to be added later.

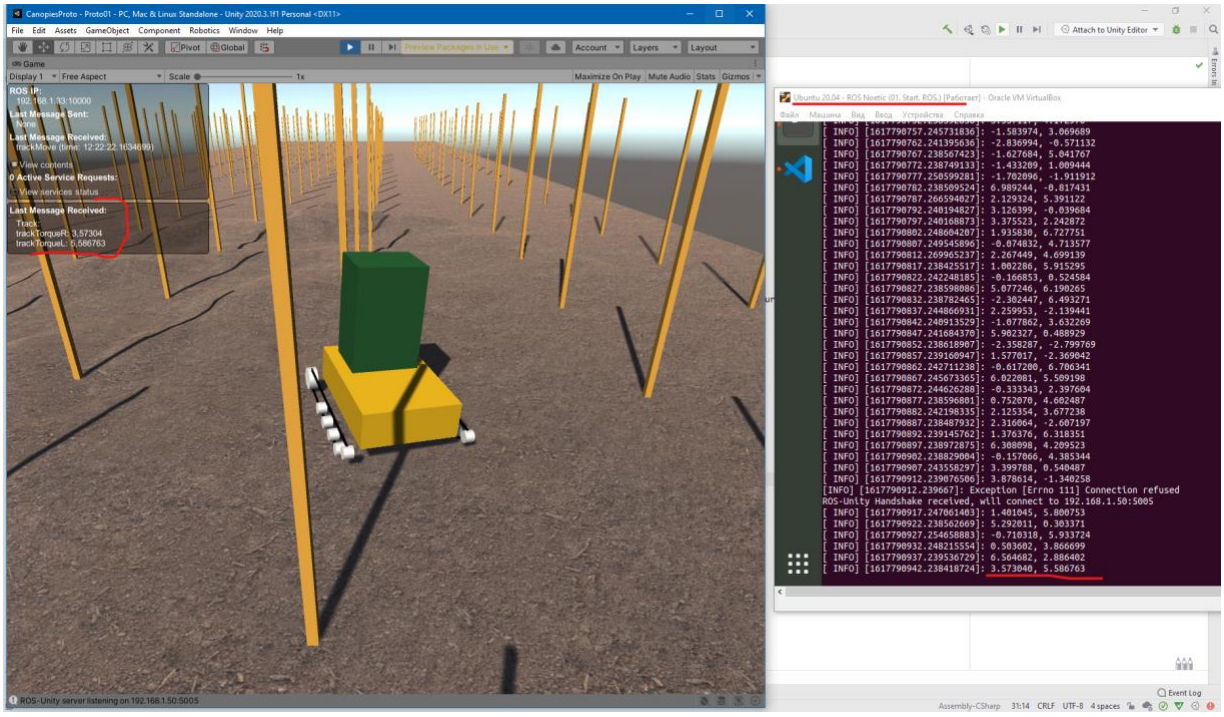

Figure 28 – The robot digital twin is placed in a basic simulated environment and controlled from ROS.

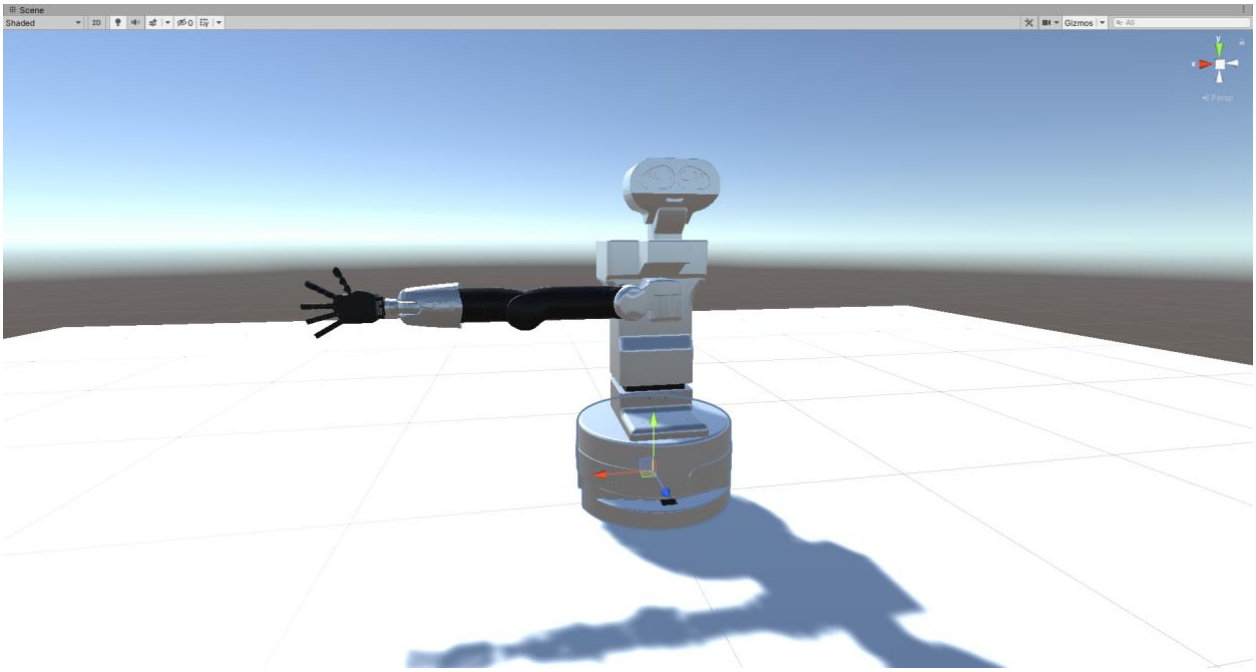

Figure 29 – Simulating joint movements of the TIAGo++ robot in the simulator.

Close collaboration with the consortium members will allow us to update the simulation and get closer to the real behaviors, with reference in the tests done on demo plots. Iterative development with continuous feedback is the key to creating a product that as close as possible to real life behavior. Using the simulator, different actors will be able to test behaviors, logic, configurations, algorithms, and work routines prior to performing real-life tests in the field.

4 Definition of the Agronomic Performance Indexes

Here we provide a description of the general performance indexes from an agronomic standpoint that should be carefully evaluated to properly assess the potential benefits associated with the adoption of innovative technological solutions. These performance indexes are derived according to the two main agronomic interventions that are more relevant within the context of table-grape vineyards maintenance, i.e., harvesting and pruning.

4.1 Performance Indexes for Harvesting

1. Percentage of successful detection of grape clusters for harvesting (# of clusters detected / tot # of clusters). This indicates the effectiveness of clusters detection.
2. Percentage of compliant clusters not harvested (# of compliant clusters not harvested / tot # of compliant clusters). This indicates how many good clusters are mistakenly left on the vine.
3. Percentage of brix-non-compliant clusters harvested (# of brix-non-compliant clusters harvested / tot # of clusters harvested). This indicates compliance to target brix specification.
4. Percentage of color-non-compliant clusters harvested (# of color-non-compliant clusters harvested / tot # of clusters harvested). This indicates compliance to target color specification.
5. Percentage of waste-non-compliant clusters harvested (# of waste-non-compliant clusters harvested / tot # of clusters harvested). This indicates compliance to target waste specification.
6. Percentage of clusters harvested & damaged by the robot (# of clusters harvested damaged / tot # of clusters harvested). This indicates how many clusters are damaged by the robot.
7. Percentage of clusters correctly placed in the box (# of clusters correctly placed in the box / tot # of clusters harvested). Indicates proper box filling.

4.2 Performance Indexes for Pruning

1. Percentage of successful detection of branches for pruning (# of branches detected per plant / tot # of branches per plant). This indicates the effectiveness of branches detection. (for example, plant1 10/12 branches detected, plant2 11/11, plant3 8/10: 88%)
2. Percentage of cut successfully executed (# of successful cut / tot # of cut). This indicates if the cut is executed correctly compared to the agronomic requirements.
3. Percentage of plant with non-compliant branches left (# of plants with non-compliant branches left / tot # of plants). This indicates how many plants were pruned incorrectly leaving some canes that should been removed.
4. Percentage of plant with compliant branches erroneously cut (# of plants with compliant branches cut/ tot # of plants). This indicates how many plants were pruned incorrectly cutting some canes that should been left.
5. Percentage of damaged plants (# of damaged plants / tot # of plants). This indicates how many plants are damaged by the robot during the pruning operation.