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 D8.4 – V1.0

VR Multi-Human-Multi-Robot Interaction Component

Dissemination level	PUBLIC
Submission Date	2023-03-15
Work Package	WP8
Task	T8:4
Type	Report
Version	1.0
Author	PALEBLUE
Approved by	UNIROMA3, UNICLAM, UNIROMA1

Executive Summary

The 3D virtual environment is enabled with networking multi-user functionality, allowing several VR operators for humans to join. Full body modelling for humans is used for humans, alongside with physical modelling, to provide full-body movement and interactions. Robots are allowed into the same networking environment and are spawned in the real-time scene as physical and graphical entities. The previously developed physical interaction models between humans and robots are combined and adapted in a networking scenario, thus enabling multi-human-multi-user interactions by the means of kinematic physics modelling.

Table of Content

Abbreviations and Acronyms

ROS	Robot Operating System
URDF	Unified Robot Description Format
VR	Virtual reality
WP	Work package

1 Introduction

For the CANOPIES project, we created a multi-robot multi-human simulation environment which can be used to train, develop, and test robotic actions in collaboration with humans without the need of a physical environment. Among the benefits of this simulation environment, there are:

- Work with ripe grapes all year round;
- Perform experiments with a few or even many robots and humans together;
- Change the environmental conditions at will;
- Perform repeatable experiments;
- Train neural networks with many variations of the same situation.

In Deliverable [D2.4] we described the simulation environment, including support structure, vines, and grapes. Deliverable [D3.4] presents the basic components of the robot, mainly focusing on the mobile base, while [D4.3] is about the simulation of the robot's sensors. Deliverable [D5.6] describes how the robots can interact with the environment. Deliverable [D6.4] describes the interaction between robots and humans, while [D7.3] describes the interactions between different robots in the same scene.

Finally, this document presents the setup where we have multiple humans in the scene, enabling multi-human-multi-user interactions as per the requirements. It extends the setup described in [D7.3] with the ability to have multiple humans in the environment, while the interaction between humans and robots described in Deliverable [D6.4] also works for the inter-human interaction. The benefit of this functionality advancement is that it allows to have several humans in the scene, thus opening up the possibility for a wide array of simulation scenarios. This enables the training of the robots in environments with multiple people, simulating scenarios with multiple people and have live interaction between different people during simulation. Lastly, the simulation can also be used to train people how to work with robots in the field.

Chapter 2 revisits the simulation setup and how it differs from the single-human setup. Chapter 3 highlights the various ways people can be included in the scene. Then Chapter 4 describes the various ways people can interact with each other.

2 Setup design

In [D7.3] we introduced the possibility to include multiple robots into the simulation. In this document we describe the extension of having multiple humans, resulting in a multi-robot, multi-human simulation.

Like for the multi-robot setup, we take the real-world situation as the goal for the multi-human, multi-robot virtual reality setup. Humans and robots should be able to interact with each other in the simulated environment in the same way as in reality.

2.1 Real world setup

The primary way of interaction for humans is by using eyesight: seeing what is going on in the simulation is key. Humans will primarily respond to visual cues so the visual representation of the situation should be optimal. Next to that, with multiple humans in the scene, humans will communicate with each other verbally about the situation and their (intended) actions.

Robots in the scene will have to deal with multiple humans in their environment, possibly giving them (different) signals at the same time as using voice or gestures.

2.2 Simulation setup

For each humans controlled with a VR device, a local environment simulates all the relevant items. The rendering of the vineyard with all other robots and other humans is most relevant. The result of this is visible in the VR headset. The other robots visible for the human will be the remote robots as described in [D7.3]. This is mostly the visible appearance of the robot itself. Humans from other

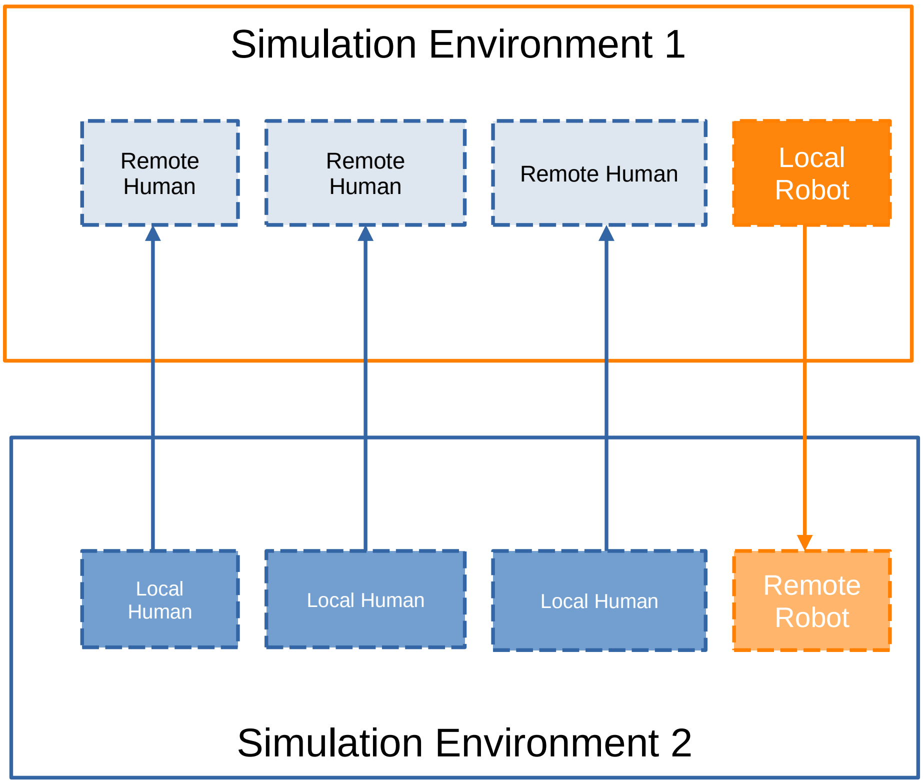

Figure 1: Projecting local robots and humans on remote simulation environments

simulation environments visible in the scene are also a reflection of the pose of each human derived from the tracking information from the VR headset and hand controllers.

For example, in Figure 1, the local environment of the robot is Simulation environment 1, while the local environment for the humans is Simulation environment 2. The network synchronisation will ensure that the state of the robot will be visible as a Remote Robot in Simulation Environment, while each human will have a Remote twin in Simulation Environment 1 which is synchronized through the network.

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

3 Adding humans

It is possible to introduce different types of humans in the scene:

1. *Static humans* which do not move.
2. *Animated humans* which can move using pre-recorded animations.
3. *Controlled humans* who are controlled directly by a user using a virtual reality device.

All these types of humans can be combined in the same environment if necessary.

Static and animated humans can be included in an environment using the dynamic scene configuration. The position, orientation and appearance can be specified for each human. The distinction between static and animated humans is that the latter have an animator controller which can play various pre-recorded animations, as detailed in the following.

3.1.1 Static humans

Static humans can be placed as obstacles in the environment which should be recognized by the robots. An unlimited number of humans can be placed in the scene using either dynamic scene configuration or ROS service calls.

Figure 2 Static avatars in the environment

The dynamic scene configuration can be used to configure a configuration with various humans in the scene. Their place, orientation and appearance can be set as can be seen in Figure 2.

3.1.2 Animated humans

Animated avatars are like static avatars, but have animations attached to them which can be played. These animations can be started and stopped using ROS commands (see Figure 3).

Figure 3 Character with a 'wave' animation

With ROS commands, it is possible to play a different animation for each animated human in the scene. It is possible to select an animation from a set of animations built-in to the simulation. Animations can be added, removed or updated in the Unity3D development environment. It is desired for consortium members to add, remove or change avatars and animations in the simulation environment to facilitate training and testing the robots behaviour. To enable this in a convenient way, a separate repository is available for relevant partners which contains the avatars and animations used. This repository can then be used to generate specific builds of the simulation environment which use these avatar and animations. The generation of these build can be triggered of those partners such that they can easily update the avatars and animations for their specific needs.

The animations are usually pre-recorded motion-capture animation of gestures to control the robots stores as part of an FBX file. To train the robots recognizing these gestures, a lot of training data needs to be collected with different human appearances and variations in the execution of the gestures. An example video of such animations in the simulation environment can be found at the following link: Canopies_Animated_Human.m4v (https://tinyurl.com/canopies-vr-human-animation).

Using this setup, it is possible to train, develop and test neural networks using any combination of appearance, gesture execution, the pose relative to the robots, lighting etc. with 3D capture data in a repeatable manner.

3.1.3 Controlled humans

Human characters in the scene can also be controlled using virtual reality devices. These devices track the head and hands of real users and can project these on the characters in the simulated environment. The character pose is derived from the tracked pose of the user using inverse kinematics for the upper body and procedural animations for the legs. See also [D6.4] – *VR Human-Robot Interaction*.

4 Shared networking environment

In this chapter we describe the various methods of interaction between multiple humans and between robots and multiple humans. We distinguish the following types of interaction:

- Visual interaction
- Physical interaction
- Audio interaction.

4.1 Visual interaction

Using the networked setup, all human and robotic clients will be able to see each other and their movement in real-time. This enables different kinds of interaction.

4.1.1 Human pose synchronization

The poses of each human are each determined on one client computer. We call this the local human. This is the VR client for controlled humans, but for static and animated humans, this can also be a robotic client. When the dynamic scene configuration of a specific client includes static or animated humans, all other clients will see these humans when this client connects to the simulation network.

The human seen on all other clients will be called the remote human and it is a reflection of the pose of the local human. The pose of the local human is continuously synchronized across the network.

4.1.2 Human-Human

Static and animated humans serve as scene objects for the perception by the robots. Human users cannot control them, neither can they view the scene through the eyes of these virtual human objects. The human users controlling humans in the scene however will be able to see all the other virtual humans, including static and animated humans. This includes hand gestures, including finger movements. Facial expressions are not included as they are not in scope of CANOPIES project.

4.1.3 Human-Robot

Controlled humans will also be able to see other robots, including the poses of the mobile base, the dual arms. The interaction of the robots with the environment will also be visible. For instance, when a robot cuts a bunch of grapes and moves it toward the box, it will be visible that the grape bunch is held by the robot with the gripper and the grape bunch will no longer be present on the grapevine.

Furthermore, boxes used by the robot to collect the harvested grape bunches will be exchanged between harvesting and logistic robots using the box exchange mechanism (see [D7.3]), but could also be picked up by humans.

Finally, the robots are equipped with LED light strips which are used to indicate their intentions like movements to the humans in the environment (see Figure 4). These light strips can be controlled from ROS changing the colour in (segments of) the light strips or make (parts of) the light strip blink. This can be used to make clear that the robot is going to stop, start to move or make a turn to the indicated side.

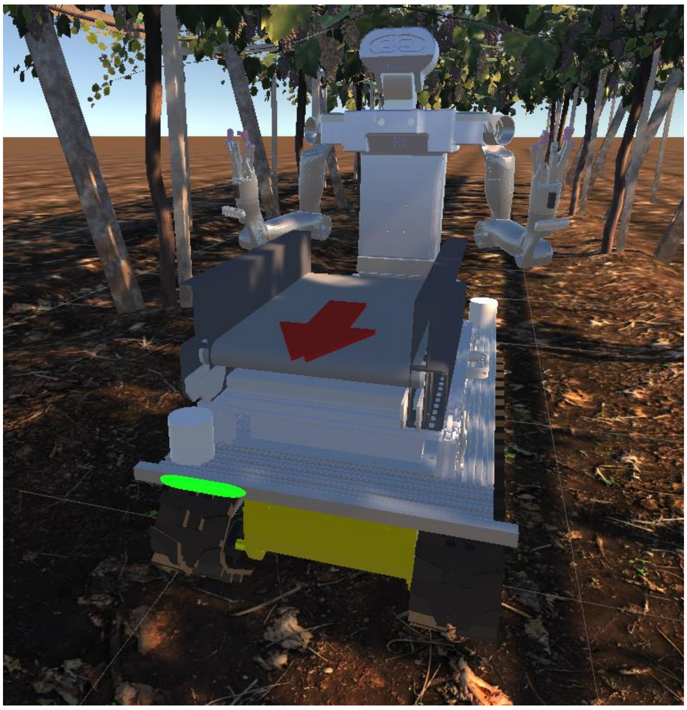

Figure 4 Robot with light signals

4.1.4 Robot-Human

Robots will be able to perceive humans in the scene using available optical sensors as the lidars and depth cameras. The logistics robot only has lidars by default but can optionally have a depth camera configured using an updated URDF (see [D5.6] – *VR Mobile-Robot*). The harvesting robot has two lidars and three depth cameras by default: one inside the head and one at the end effector of each arm. Optionally, an additional depth camera can be configured like on the logistics robot.

The simulated lidars will show the presence of humans in the scene like their real counterparts do. This means that the global image of the person is visible but detailed movements like finger movements will be hard to perceive. The lidars can detect the presence of the humans as obstacles and use that information in their path planning software.

The depth cameras have the same resolution of their real counterparts and will be able to perceive the pose, gestured and movements of the humans (see Figure 5). Machine learning can then be used to process the images from the simulated cameras to possible gestural commands which can be acted upon.

Figure 5 Robot and human in the 3D scene

4.2 Physics interaction

The physics interaction between humans and robots have been described in [D6.4]: humans are able to interact with robots by grabbing and pushing, while robots are able to collide with humans. To these types of interactions, this deliverable adds limited physics interaction between controlled humans. Only interactions between the hands of the humans is simulated. This enables the most important interactions while keeping the simulation simple, as that is important for maintaining good real-time performance of the simulation.

As the animations are mainly used to train the neural networks of the robots, no physics interaction is simulated with static and animated humans. This means that collisions will not be simulated: human hands and robot parts will be able to move through those humans.

4.3 Audio interaction

Another important interaction type is auditory: humans can use their voice for communication and robots will be able to use speakers to emit sounds and a microphone to register sounds in the environment.

4.3.1 Human-Human

Voice interaction between humans in the environment is possible: people can talk to each other like in normal life. The volume of this communication is adjusted for distance: when another human is standing close to a controlled human it will sound louder compared to a human standing further away.

4.3.2 Human-Robot

Robots have microphone which can be used from ROS. The microphone simulation uses the ROS *audio_common* package

4.3.3 Robot-Human

Robots can play sounds from ROS which can be heard by controlled humans in VR and by spectators using normal, non-VR, screen devices.

5 Conclusion

Building upon the components we created in work packages 2 -7, it was easy to extend the simulation with multiple humans. Single humans could already participate in the simulation as described in Deliverable [D6.4]. Network synchronisation of the human state including voice and physics interaction was already needed for Human-Robot interaction. Having the option to include multiple robots created a scalable solution, which could also be used to include multiple humans.

The main focus was therefore on the interaction between different types of humans. Humans controlled with VR devices can see each other and can interact by voice and physically with their hands, like they can with robots. With the inclusion of static and animated humans, we made it easier to execute test scenarios with multiple humans in the scene without the need for multiple VR headset.

6 References

Id	Title	Version	Date
[D2.1]	Requirements, Specifications and Benchmarks	1.0	2021-04-30
[D2.2]	Specifications and KPIs for the two CANOPIES robot prototypes	1.0	2021-06-30
[D2.3]	Dimensioning of the Real-World Scenario for Final Demo	1.0	2021-10-29
[D2.4]	VR Farming Environment Specification	1.0	2021-10-29
[D3.4]	Simulated Farming Environment and Basic Robotic Components	1.1	2023-01-16
[D4.3]	VR Agronomic-Oriented Perception Component	1.1	2023-01-16
[D5.6]	VR Mobile-Robot Component	1.0	2022-03-31
[D6.4]	VR Human-Robot Interaction	1.1	2023-01-16
[D7.3]	VR Multi-Robot Component	1.0	2022-10-31