POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

EXECUTIVE SUMMARY OF THE THESIS

XR application for remote operations

LAUREA MAGISTRALE IN COMPUTER SCIENCE ENGINEERING - INGEGNERIA INFORMATICA

Author: ANDREA ALESANI

Advisor: PROF. FRANCA GARZOTTO

Co-advisor: TARU HAKANEN

Academic year: 2022-2023

---

1. Introduction

Remote controlling off-highway machinery is a non-trivial task for operators. Many issues arise, in particular regarding situational awareness. It is hard for operators to be aware of distances from camera footage, there is no auditive or haptic feedback about possible collisions or other unpredicted behaviors, and to all of that latency is added. This thesis work marks the beginning of THEIA-XR, a EU funded project that aims to enhance human-machine interaction by implementing extended reality (XR) features to the control systems of this kind of machinery. In particular, this thesis analyzes the human-machine interactions of a particular type of vehicle, namely a reach stacker from Kalmar Cargotec.

Literature review has been conducted on different aspects, namely hardware, software, and user experience related to extended reality or remote-controlled machinery. This work contributes to the literature by providing (i) the design and implementation of an interactive XR application that emulates remote control capabilities for reach stacker operators; (ii) a pilot test to assess the usability of said system; (iii) a redesign of the application's interactions based on the usability test and on the users' feedback.

Figure 1: Kalmar's reach stacker

2. Literature review

Regarding hardware to use for adding XR capabilities, the main choice to make was about using a Head-Mounted Display or a CAVE system. The first can generally be more immersive and allow for more kinds of interactions, while the second allows for collaborative use and does not have to be worn, but has a considerably higher cost [10].

HMD	CAVE
Visually more immersive	Can be collaborative
More interactions with the virtual scene	No cybersickness
Cheaper	No need of wearing

Table 1: Comparison between the strengths of HMD and CAVE systems.

Among HMDs, the devices present different features that can influence the choice of one over the other according to the project's requirements and necessities. They can be either PC-dependent or standalone, some of them support hand-tracking while others only support controllers. Each device also positions differently on the Reality-Virtuality continuum [7], with some that can render virtual objects on top of the real-world view such as the HoloLens 2, some that blend reality and virtuality in the same space (a screen on each eye) thanks to pass-through cameras such as the Varjo XR-3, and finally some devices that are dedicated only to Virtual Reality and are not able to show real-world footage, for example Meta Quest 2 and PlayStation VR 2 only use a black-and-white pass-through for alerting users when they are about to exit the "safe area" or bump into real objects. For this project, Varjo XR-3 was chosen to get the best out of both real-world interactions and virtual interactions. A CAVE version was also tried for the second iteration of the application.

For software, the choice was made to use a game engine, and in particular Unity, in order to have better device compatibility in case the project had to be ported also to other platforms. Unity also has support for Varjo SDK which offers pass-through features, and for Ultraleap SDK, which communicates directly with the Ultraleap sensor located in the XR-3 headset and enables hand-tracking capabilities.

On the user experience side, findings were synthetized into the following categories:

Feedback methods
Visuo-motor feedback is necessary to perform any kind of action, but sometimes it can be not sufficient. Visuo-tactile feedback can be very beneficial when performing high-precision tasks[6]. It might be useful to have physical devices to perform critical and accurate tasks, while other easier tasks may not require tactile feedback to be performed and allow for free hand-tracking.

Agency and ownership
Higher agency might help the user perform tasks faster and more accurately, and a higher level of ownership the user might be more incentivized to not take risks that might lead to accidents, such as collisions between the reach stacker and other objects [11].

User's self evaluation
Even though users might not report the feeling of being more productive, having a high quality image could still improve overall productivity [6].

Professional environment
Small flaws in the system can cause major nuisance if they have to be used on a daily basis[4]. Some barriers of XR, such as adaptation to cybersickness and hand-tracking commands, can be overcome thanks to habits [5]. Since the control system is very complex, it should allow the user to customize it according to their needs [8].

Technological possibilities and use cases
Among other aspects, it can be noted that teleoperation VR interfaces are usually found as two implementations: (i) egocentric where the user perceives the world from the robot's point of view, and (ii) robocentric where the user moves freely in the virtual environment as an observer [9].

3. Design and Implementation

In order to build an application that best suited human needs and best reflected Human-Centered Design, the design and implementation cycle was divided into a first iteration, a pilot test, and a redesign. Future work includes conducting a second test to assess the improvements of the second iteration, and from there a virtuous cycle of redesign and user tests can be conducted to better refine the user experience.

To develop and run the application on the Vario

XR-3, a high-end PC was required. The main specifications of the PC that was used are the following:

- OS: Windows 10 Pro
- CPU: Intel Core i9-10900KF @ 3.70GHz, 10 Cores, 20 Logical Cores
- RAM: 64GB @ 3200MHz
- GPU: NVIDIA GeForce RTX 3090 48GB VRAM

The application works so that it can be run only on a normal monitor showing a digital twin of an harbor which a reach stacker and a container to grab, and at any moment the user can wear the Varjo XR-3 HMD to enter the virtual scene in first person, while still being able to see the real monitor and the physical devices used to control the vehicle thanks to Mixed Reality features.

Figure 2: The user can use the application through a monitor and wear the HMD at any time. Thanks to Mixed Reality capabilities, the user can still see the monitor, as well as other physical devices such as the keyboard and the joystick.

3.1. First iteration

The application features were divided into two sections, called input and output. The former describes the affordances that the user has in order to provide input to the system, while the latter describes how the system informs the user about what is happening in the virtual scene (and thus in the real world). Regarding input, all the available commands to control real-world objects are bound to physical devices, while other commands, such as the ones to move the user around the scene, are available on virtual menus available with hand-tracking.

Physical controls use different devices to control different parts of the vehicle, the driving system is delegated to a Logitech Extreme 3D joystick, while the control of the boom and spreader is done using the keyboard (this changed in the second iteration). In this first prototype, the reach stacker can only translate forward/backward and left/right, and it can not rotate. The simplification was made for the sake of time, but it turned out to be to leave out too many critical aspects of driving a reach stacker during the pilot test.

Hand-tracking interactions are made possible thanks to two virtual menus, a “rotation menu” and a “movement menu”. The rotation menu consists in a sphere that can be grabbed and moved around a central cube in order to rotate the virtual scene, while the movement menu consists in a series of buttons and sliders to move around the virtual scene (only translation, as the rotation is managed by the previous menu). The menus are accessible by grabbing a cube to which they are bound, and moving it away from its anchor which is next to the left hand’s palm. Sliders are used in the virtual menus to allow the user to move freely inside the virtual scene. A two-dimensional slider is used to control movement on the ground plane, while a one-dimensional slider is used to control the height from the ground. The sliders were initially mapped to reflect the movement of the virtual world, because initially the user had a pass-through vision of the real room and the virtual scene was only occupying the center of a room.

Figure 3: The physical devices used for the first iteration include a Varjo XR-3 HMD, a keyboard and a Logitech Extreme 3D joystick.

The mapping was then changed when a virtual harbor scene was added and it completely occluded the real room. At that point, the slider was mapped to reflect the user movement inside the virtual scene.

There is a third slider, which is used to change the scale of the scene, that does not behave the same way as the other two sliders. While the first two sliders reset their position to the center after being released (they are mapped to the velocity of the movement, rather than to the position), the third one stays in the position where it was left (the slider value is mapped to the actual scale, not to the speed at which the scale changes). It was initially thought that the different behavior or similar menus would be confusing for users, but no users complained about it and they all seemed to find it intuitive.

One button on the movement menu enables a "driver mode", which teleports the user inside the operator's cabin and binds the user position to the reach stacker's position. This way, when the vehicle moves, the user moves together with it the same way it happens when the operator is sitting in a real reach stacker. Activating the driver mode would also disable the visibility of the monitor, which would have otherwise occluded the view.

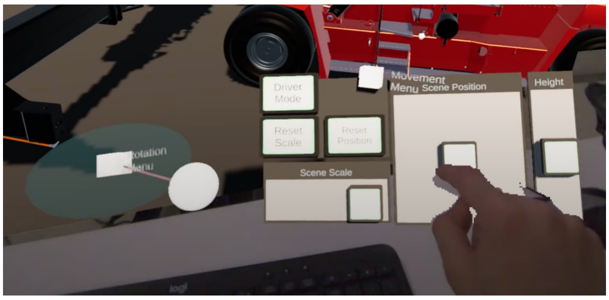

Figure 4: Virtual interactions include a rotation menu to rotate the virtual scene around the user, and a movement menu with sliders to change the scale of the virtual scene and allow free movement inside the scene.

Regarding the output, the design was made so that multiple cameras were implemented to render on different devices. An XR Camera is used to show footage on the Varjo XR-3, and 3 "traditional" cameras are rendered to the monitor. The main one is a perspective camera that is placed right in front of the operator's cabin and points towards the front of the vehicle. The other two cameras are orthographic cameras and they show a top and a side view of the reach stacker's spreader, to better align the twistlocks to the container's casting corners. The use of orthographic cameras (which do not exist in the real world) reduced the number of cameras needed, as a single camera is sufficient instead of having a camera per each of the container's corners.

Distance lines were implemented between the twistlocks and their respective casting corners to help the user align them. Text indicating the distance (in meters) between each twistlock and its casting corner is shown on top of the line.

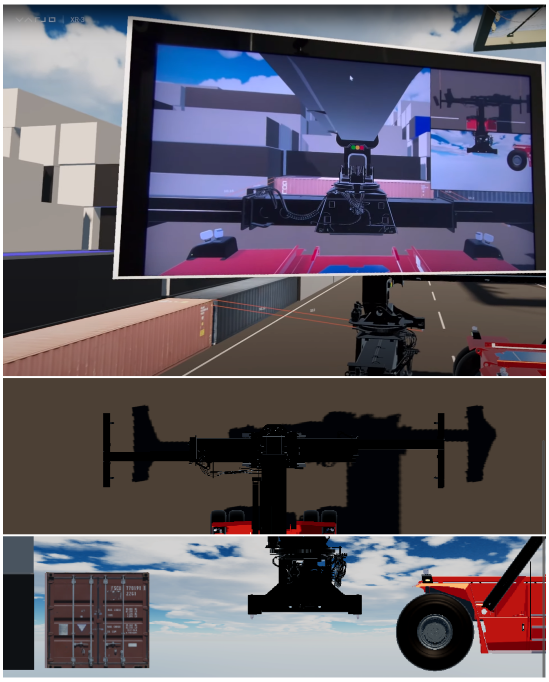

Figure 5: The real monitor is still visible even after wearing the HMD thanks to its Mixed Reality capabilities. On the monitor, three cameras are rendered. A first perspective camera is shown as the main view, while two other orthographic cameras can be noted in the top left of the screen. Bigger images of the orthographic cameras are visible in the second and third images.

The container to grab is color-coded so that the user is aware when it is close enough to the spreader to be picked up and when it has effectively been picked up. Sound was implemented as an alerting system for emergencies. A beeping mechanism notifies the user when a person or a physical object is getting closer to the vehicle. The beep sound is sent spatially, so its direction

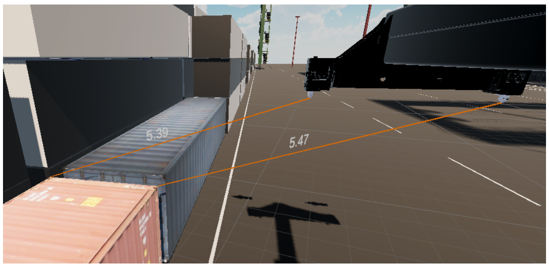

Figure 6: Distance lines are drawn between each twist lock and its relative casting corner. A text label on top of each line shows the distance in meters.

will reflect the direction of the emergency, and sound for people has a higher pitch than sound for objects.

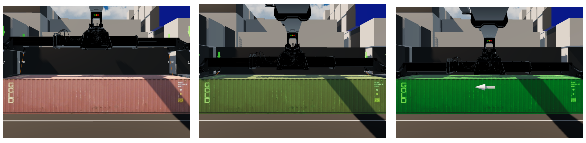

Figure 7: Container is color-coded. If the spreader is close enough to the container to grab it, it becomes slightly green. If the container is grabbed, it becomes green.

3.2. Pilot test

A pilot test was conducted on March 16th, 2023. It involved four employees from Kalmar Cargotec, who are not reach stacker operators but have a lot of experience with its control system, as they are part of the research team for the vehicle. The evaluation was carried out with a System Usability Scale (SUS) test [2], followed by user interviews. The test was designed so that users had a general introduction and tutorial, and then each of them had a very quick hand-tracking tutorial before starting the task, which consisted in driving the vehicle to approach and pick-up the target container. Users could then choose to put the container on a target area inside the target, or to make free use of the application features, for example repositioning to different sides of the vehicle to find better views. The first iteration of the project scored 59 points in the SUS test, which is considered to be marginal (even though not statistically relevant due to the low number of testers) [3][1]. User interviews highlighted a few critical problems, and a handful of possible improvements and suggestions.

3.3. Redesign

The pilot test represented the basis on which to start the redesign process. One of the most important comments about the first iteration was that the simulation was too simplified and was hiding one of the hardest parts of picking up a container, which consists in aligning each twist lock to its casting corner. A physical simulation of the vehicle was then added to the system, allowing the user to steer and adding a new level of complexity to the system. Moreover, the spreader was updated to be able to rotate on a plane parallel to the ground. To make up for the new complexity, and following the users' suggestion that distance lines were not helpful enough for the alignment task, visual hints were revisited. A new arrow guides the user to make the reach stacker point towards the center of the container. When this is achieved, the user is guaranteed to be able to pick-up the container without the need of rotating the vehicle anymore. The main arrow disappears and four new arrows appear, one per each twist lock, that point towards the horizontal direction towards the casting corner and become green when the twist lock is correctly aligned. A new "activate screen" button was added, to address the issue signaled by the users that there were cases in which the real monitor was needed even if the driver mode was active. A few other features were added, such as the possibility to make the container transparent to see behind it, and a cast time before grabbing or releasing the container to avoid unintentional slips. The application was also adapted to run on a CAVE

Figure 8: The second iteration introduced a Logitech G29 steering wheel to drive the reach stacker. The Extreme 3D joystick was mapped to control the boom and spreader.

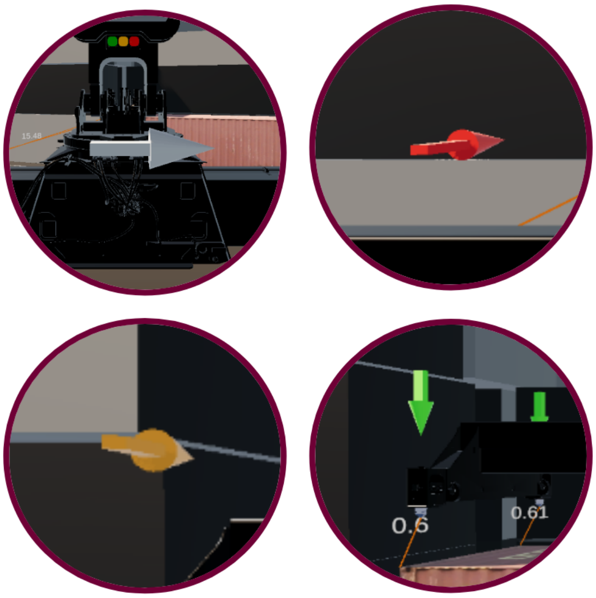

Figure 9: New visual hints were added. The white arrow is visible until the vehicle points towards the center of the container. At that point, four small arrows appear on top of each twist-lock to guide the user in aligning the spreader to the container.

system (without using an HMD).

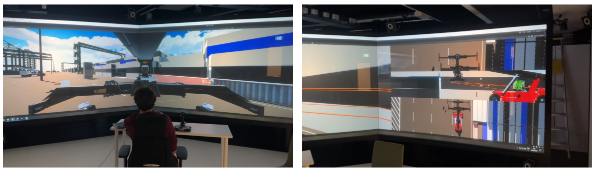

Figure 10: The application running on the “Powerwall" wall-sized display. The main camera footage is shown on 70% of the screen (left), and other available cameras are rendered in the remaining 30% of the screen (right).

4. Conclusions and future perspectives

Several conclusions can be drawn from this thesis work, both from the technical side and from the user experience side. Addressing the research question, XR Camera helped a lot for some tasks (i.e., driving) but less for others (i.e., aligning containers to casting corners). Mixed Reality allowed to have both 1st person view and “traditional” cameras.
Thanks to *First iteration -> pilot test -> redesign* a lot of user's issues were addressed even before finishing this thesis work.
A second user test is needed to assess the usability of the system.
It would be interesting to study how interactions change at different stages of the Reality-Virtuality continuum.
A new engine dedicated to XR would need to focus on having different reference frames and separate physics engines.

5. Acknowledgements

I want to thank VTT for offering me the opportunity of working on the amazing world of extended reality with the most skilled and fun team I could ever find. In particular, I want to thank Taru Hakanen for the trust she put in me since the first video call we had when I was still in Paris. I want to thank Kaj Helin for guiding me through the planning and design of my work. Jaakko 'Jaska' Hautamäki, Jaakko Karjalainen, Petri Tikka, you all made me feel like at home. Special thanks to Andrea Peña for the LaTeX tips and for listening to me venting about my accommodation and life events throughout these months.
Special thanks to Franca Garzotto for taking on the challenge of supervising a student working on peculiar vehicles from far far away.

This work is part of the THEIA-XR project, which has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101092861. Any information of results reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains.

References

[1] Aaron Bangor, Philip Kortum, and James Miller. Determining what individual sus scores mean: Adding an adjective rating scale. *J. Usability Studies*, 4:114–123, 5 2009.

[2] John Brooke. *SUS—A Quick and Dirty Usability Scale.* Taylor and Francis, 1996.

[3] John Brooke. Sus: A retrospective. *J. Usability Studies*, 8:29–40, 2 2013.

[4] Kjell Brunnström, Elijs Dima, Tahir

Qureshi, Mathias Johanson, Mattias Andersson, and Mårten Sjöström. Latency impact on quality of experience in a virtual reality simulator for remote control of machines. *Signal Processing: Image Communication*, 89:116005, 8 2020.

[5] Natalia Dużmańska, Paweł Strojny, and Agnieszka Strojny. Can simulator sickness be avoided? a review on temporal aspects of simulator sickness. *Frontiers in Psychology*, 9, 8 2018.

[6] Mie C S Egeberg, Stine L R Lind, Sule Serubugo, Denisa Skantarova, and Martin Kraus. Extending the human body in virtual reality: Effect of sensory feedback on agency and ownership of virtual wings. In *ACM International Conference Proceeding Series*. Association for Computing Machinery, 8 2016.

[7] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. Augmented reality: a class of displays on the reality-virtuality continuum. pages 282–292, 12 1995.

[8] Laurent A Nguyen, Maria Bualat, Laurence J Edwards, Lorenzo Flueckiger, Charles Neveu, Kurt Schwehr, Michael D Wagner, and Eric Zbinden. Virtual reality interfaces for visualization and control of remote vehicles, 2001.

[9] Eric Rosen, David Whitney, Elizabeth K Phillips, Daniel Ullman, Elizabeth Phillips, and Stefanie Tellex. Testing robot teleoperation using a virtual reality interface with ros reality, 2018.

[10] SkyReal. The vr cave, halfway between reality and virtuality.

[11] Rutger Vlek, Jan Philip van Acken, Evine Beursken, Linsey Roijendijk, and Pim Haselager. *BCI and a User's Judgment of Agency*, volume 12, pages 193–202. Springer Science and Business Media B.V., 2014.