The Latest in Immersive Telepresence to Support Shared Engineering Education

Nicolas Pope
Department of Future Technologies
University of Turku
Turku, Finland
nwpope@utu.fi

Mikko-Ville Apiola
Department of Future Technologies
University of Turku
Turku, Finland
mikko.apiola@utu.fi

Heidi Salmento
Department of Teacher Education
University of Turku
Turku, Finland
heirpo@utu.fi

Najmul Islam
Department of Future Technologies
University of Turku
Turku, Finland
najmul.islam@utu.fi

Marko Lahti
Department of Future Technologies
University of Turku
Turku, Finland
marko.lahti@utu.fi

Erkki Sutinen
Department of Future Technologies
University of Turku
Turku, Finland
erkki.sutinen@utu.fi

***Abstract*—Work in Progress: In this paper we outline our initial findings on the potential of state-of-the-art immersive telepresence to support the practical and collaborative group work element in our remotely taught international degree programmes, specifically a Software Engineering Masters degree. Whilst we adopt widely used distance learning approaches generally, the challenge remains in supporting live synchronous practical sessions that depend heavily, on one hand, on the expertise of the educator and their presence in the shared learning environment, and on the other hand, the students' active engagement within teams partially distributed physically. Any technical progress in improving the "sense of presence" can radically renew our conceptions of (distance) teaching and learning, maybe also in terms of the 21st century skills or even those that we are not yet aware of. The question to address is that if sufficiently high-fidelity, and unintrusive, immersive capture technology is used then how could this be game changing in this area? Previous work using similar immersive technology has failed to achieve the level of quality or scope necessary, however, the past 4 years have seen significant progress in the hardware and algorithms required. To support our degree programmes we have designed and developed a custom live 3D capture system for a higher fidelity immersive experience targeting small groups of 2-6 people collaborating both locally and remotely. Here we will present the system and scope out some of the initial plausible affordances for education through the Conceive, Design, Implement, Operate (CDIO)-model. We focus on the affordances offered by the technology to support the learning of the much needed competences of communication, collaboration, critical thinking and creativity which are very challenging to enhance by conventional, content delivery oriented distance learning approaches.**

***Index Terms*—Sense of presence, distance learning, remote presence, CDIO, C21 competencies**

I. Introduction

As we are all now discovering, there is rapidly increasing pressure to improve support for distance learning. Our own organisation, as with many Universities, is actively expanding our degree programmes to foreign campuses through the use of technology and in doing so we discover the limitations of existing solutions, particularly for group work and collaborative activities. In tandem with our own international expansion, we are also researching, developing and investigating alternative technologies to support our efforts and this paper outlines our progress and plans for the potential use of immersive telepresence. There are two primary strands to our work: 1) research and advance the technology itself due to current limitations with-regards-to groups, and 2) investigate how it can be applied to Science, Technology, Engineering, Mathematics (STEM) education. In this paper we first outline our remote campus context to clarify the underlying assumptions and objectives we face, along with a brief review of previous work from literature on the use of related technologies in similar situations. Then we state the nature of the proposed technology and our progress in creating it, along with our current understanding of what affordances it provides for education. It is believed that the technology can revolutionise distance learning, however, it is an ongoing process to fully appreciate how that will occur in practice. By identifying an initial set of plausible affordances for education we conclude with a plan of action for validating those in the future.

II. The Remote Campus Initiative

The Global South, especially Africa, offers software engineering a whole range of novel and unforeseen challenges, for several reasons. Africa has a vast young population, hungry for learning. The population has successfully adapted novel technologies, like mobile currency, at a pace that exceeds any of the development programmes that the continent is known for: the diffusion of technologies is fast. Few legacy systems or conventional designs would allow the design of culturally native solutions. For the companies in the Global North, the Global South is an attractive business environment for selling and designing products and services and recruiting talent.

After mutual discussions since 2015, the challenges and opportunities led the University of Turku (UTU) to agree with the University of Namibia (UNAM) on setting up a remote campus within UNAM's main campus in Windhoek, the capital of Namibia, in 2019. UNAM's leadership indicated the serious demand for software engineers, to be trained in a way that Finland is well-known for: offering a modern and flexible learning environment where future software engineers work with real software companies and learn relevant competencies by designing solutions to real-life problems.

The UTU remote campus, called the Future Tech Lab, is a completely renovated, refurbished and retrofitted learning environment of 200 square meters, consisting of a welcoming zone, a co-learning zone and a co-working zone. The name reminds of its mission: rather than bringing in innovations from the Global North, the Lab is an origin of meaningful designs inspired by the challenges of the South. Besides running a few RDI initiatives and short courses, the campus is now open for applications for its MSc (Tech.) degree program in software engineering, offered by UTU. In fact, the campus offers the first Finnish degree program outside Finland.

The remote campus emphasizes the combination of sustainability and quality. In education, this translates to a learning environment where software engineering can be taught in the base and remote campuses in shared mode. This requirement led us to develop remote presence technologies for learning in a way where learners experience the presence of each other, also when they are working with real-life problems, with stakeholders physically in their own neighborhoods. The remote presence should get significantly beyond the contemporary videoconferencing tools. This, we were reasoning, would also make the remote campus a Future Tech Lab for educational innovations—and inventions—that would provide software engineering education with affordances that would reform the far too traditional ways of learning.

After the first year, we have started to call our remote campus a plug-in campus, to distinguish our setup from the more established satellite campuses. In a nutshell, a plug-in campus is inherently resilient, it can be plugged in an existing host university or other institution so that the host offers basic infrastructure, buildings, and helps the plug-in campus to become an integral part of the local science and innovation ecosystem. As the plug-in campus is swift in its operations, it should be able to carry out all the traditional tasks of a university in an agile way: teaching and learning, research and societal impact. These tasks would require its technological services to adapt to unexpected demands beyond conventional teaching, even as a mobile unit serving the host country's rural areas and their requirements, such as developing health services based on remote presence technologies.

III. BACKGROUND

Teleimmersion is about, in this case, providing an immersive three-dimensional alternative to video. Rather than using immersive virtual environments, real environments are digitally captured live to form a high-fidelity model of a physical space and its occupants. The technology in question is best illustrated by the work done at Microsoft Research on what they term "Holoportation" [1]. Here they use 8 custom depth sensing cameras to capture an individual live in 3D and then transmit and present that individual in their Hololens Augemented Reality head set. More recently both Microsoft and other projects, such as the ongoing VR Together EU Horizon 2020 project [2], have focused on creating portable or consumer friendly versions of this technology designed specifically for capturing individuals to promote social interaction and an improved sense of co-presense [3]. To achieve their portable version, however, they have restricted the resolution and capture scope to individuals only, removing any background or other people. There is presently little work on supporting large groups in a single shared space due to technological limitations. Some work is being done to scale up the technology to groups [4], the problems encountered are data compression, fidelity, cost of equipment and achieving real-time performance. In terms of applications of live 3D capture technology to education, there has been little if any progress since a few early trials using devices such as Microsoft's Kinect [5]–[7]. Whilst there is a plethora or work on immersive virtual learning environments or Collaborative Virtual Environments (CVEs), the same is not true of 3D teleimmersive capture technology [8] and the focus in the literature remains technical in nature with little pedagogical evaluation and a very narrow view of the future potential.

IV. TELEIMMERSION PLATFORM

For any immersive remote presence technology, it must achieve a certain threshold of fidelity or quality to enable a sense-of-presence and to compete with traditional video technologies. Our prototype setup consists of 6 stereo pairs of cameras, each pair connected to a desktop computer with a GPU (NVIDIA GeForce RTX 2080) for the depth processing. The cameras are arranged around the edges of the room or group to passively capture the scene without intrusion. At a maximum resolution of 1920x1080 pixels and with a depth range of 1m to 12m, we estimate the depth in the scene of each pixel using Semi-Global-Matching [9] combined with recent enhancements to improve edge sharpness [10], non-frontal surfaces [11] and unreliable textureless areas. Our desired performance would be 25 frames per second (fps) and this is currently being achieved only at 1280x720 pixels due to a need for ongoing optimisation work. However, the resolution of the depth image is less crucial than with the colour feed. In addition to achieving a high enough resolution for the technology to compete with standard video conferencing, it is important to minimise specific types of errors in the depth estimates which produce the greatest distortion. These errors include edge sharpness, excessive smoothing or distortion of faces in particular, and holes. Keeping the frame-rate above 20 fps is also crucial if this is to remain a low-latency live experience. One of the challenges encountered regardless of technology used is a flickering or wobbling effect in the video feed due to noise and error in the depth estimates that changes with each frame. To improve upon this, temporal smoothing is applied using hardware optical flow estimates, along with the intelligent utilisation of past depth estimates when generating the next estimate, but further work is required before this distracting problem can be considered resolved.

Once the depth data is captured from each camera, the camera views are merged together using camera pose information gained through a calibration step to achieve a high-quality fusion with reduced noise distortion and errors in the different individual depth estimates from each camera. Following fusion the data is compressed by converting the depth data to 10-bit Yuv 4:2:0 colour format and tiling all the depth sources into one or more 4K video frame(s) that is then encoded with lossy compression using hardware (NVENC HEVC) for real-time performance. These videos can then be transmitted to multiple remote locations. After transmision the multiple video feeds for colour and depth are decoded, compression artefacts are filtered and then arbitrary 3D views can be rendered locally using a straightforward point-cloud or mesh rendering combined with reprojection to the colour video feeds to add texture. The colour feeds should be at a higher resolution, perhaps 4K but at least 2K in our case, depending on bandwidth, to achieve sharpness. Since the viewer is able to move to arbitrary locations they are in effect zooming the original colour video which makes any missing or blurred details more visible than with a fixed camera position. Overall the rendering can occur well in excess of 90fps, allowing smooth motion using a VR headset, and can be performed by any number of users. Additionally, virtual objects, including live screen captures placed on walls or tables, can be mixed into the final scene. Face detection, ArUco marker detection and other kinds of analysis are or can be performed to supplement the model, allowing, for example, screen captures to be placed on markers or allowing virtual cameras to track and move with individuals in the scene such that a user can view through the eyes of the remote individual. Further examples of educational affordances, or kinds of interaction to support learning, are to be discovered as a central aim of this ongoing work.

V. AFFORDANCES FOR ENGINEERING EDUCATION

Table I lists examples of affordances of the next generation teleimmersion technology. Next we consider the educational implications and prospects through the Conceive, Design, Implement, Operate (CDIO) -model [12] that integrates the critical aspects of engineers' competences for tackling real-world challenges in engineering education. For each aspect, we raise key affordances that reform education in the given aspect; however, each affordance can usually be mapped to others aspects as well.

TABLE I
EXAMPLES OF AFFORDANCES OF TELEIMMERSION

	Affordance	**Example**
1.	Shared interaction with real objects	Use of a real whiteboard, not a digital board. Or bringing some scientific apparatus to a classroom
2.	Observation of subtle details of objects	Particularly relevant for recognising peoples faces
3.	High-fidelity body language and (facial) expression	Subtle body language and expression can indicate the emotional state or level of comprehension of a team mate
4.	Peripheral vision	Monitoring the activities of other group members in the background, allowing a greater recognition of what is going on
5.	Arbitrary and shifting focus of attention by each individual participant	Can choose which speaker or group member to look at for cues on whether they are understanding you or that task.
6.	Depth cues for improved focus and attention	Improved sense of space.
7.	Isolation and extraction of scene elements	Cut out objects or people to remove distractions, or place them in a different setting. Augmenting with virtual elements
8.	Physically implausible points of view	From another persons perspective, viewing from behind their eyes.
9.	Unrestricted or informal positioning in the physical environment	Can sit on the floor or play in the corner and still be visible and able to participate
10.	Freeform groupings within the single common space	Fluid, informal, short-lived discussions between individuals.
11.	Feedback from touch (tactile?)	Shaking hands, patting on the shoulder and small social interactions of this kind
12.	Externalised memory for third-person unrestricted reflection and reapraisal	Pause or playback allows details previously missed to be revisited, including those beyond the original field of view
13.	”Extrasensory perception” or supplemented perception	Real-time digital analysis and reporting of body-language and events occurring outside of the users present field-of-view. ”What is going on behind your back?”
14.	Supporting or manipulating the ”sense of place”	The learning activity or meeting with remote individuals can occur in a familiar space or be deliberately located in a strange space
15.	Personalised spatial organisation of supplementary data	Additional information such as notes or screens can be placed and located within a 3D space, using the users natural ability to navigate such spaces

A. Conceive

Conceiving a real-world challenge translates to an educational challenge of enhancing students' minds, senses and bodies to grasp and comprehend the problem at hand.

For example, Affordance #1 (see Table I) allows shared interaction with real objects, and enables, for example, the use of a real whiteboard, not a digital board, and easy use of any educational equipment, such as scientific apparatus or robotics toolkits in the engineering laboratory or classroom. This has clear educational benefits as the use of equipment is not limited like in a VR-, or video-based environment.

Affordances #5, #6, #7, and #8 widen the possibilities for self-expression of individuals in the learning environment, enabling creativity, and also overcoming limitations in traditional environments, such as the possibilities to see through visual obstacles. This also relates to individuals' sense of intimacy, when learners from different parts of the globe experience being ”physically” closer to one another. Very unique new questions also arise. For example, how can out-of-body experiences and viewing through a ceiling or wall help change our appreciation of what is happening around us? Can teachers get a better perspective of the class in this way? Can students take on the perspective of the teacher, and if so, what educational potential may this have? This would also enable situations, where individual students' screens, eg. progressing program codes can be projected to the walls of the environment for others to see.

B. Design

Design refers to the ways that engineering students learn to convert their conception of the challenge into the design process, and finally, the pilot. Pedagogically, the challenge is how to integrate the design exercise to the real-world problem and its owners throughout the process.

Affordances #2, #3, and #4 together provide previously unseen possibilities for the transfer of emotional information via human mirror cell system, which enables capturing and processing of emotional information, and related growth of emotional intelligence, as opposed to a traditional video-, or VR-based system. Facial expressions and body language give important information for the teacher, as well as the students. The environment also provides a unique possibility for collecting related educational data. Learning of groupwork dynamics is essential in design and engineering processes. Co-design can be made with stakeholders located where they are. The system allows free or flexible formation of groups, or n-to-m relations between groups and their participants, which helps to prepare for a chaotic working life.

C. Implement

Pedagogically, implementing a real-world system requires learning of traditional engineering assets, but applying them in an agile way where learners and teachers are continuously aware of the process.

Affordances #3 and #11 facilitate multi-sensory on-the-fly learning analytics for making learning more focused and relevant and, thus, promote a shared awareness of the learning process and the interventions that it might call for.

D. Operate

Operating a real-world engineering system requires that an engineer has learned to observe and control a complex system where the users usually represent diverse geographical settings and cultural backgrounds.

Affordances #2 and #3 allows for observing other operators' actions and attitudes in an intercontextual, glocal learning environment.

Affordance #6 improves focus which is critical for observations and empirical learning.

Affordances #9, #10, #11, and #12 allow “multi-channel” interaction processes in the engineering and operational environment. Parallel discussions and conversations in smaller groups are made possible, with aspects such as touch, handshakes, and related sense of presence, which makes the interaction authentic. The environment allows revising and revisiting what has happened before. This enables, for example, playing back a previous conversation or demonstration, which might be useful in many learning situations.

VI. CONCLUDING REMARKS

Based upon our own challenges to expand software engineering education within a distributed learning environment shared by two physical spaces, one in the Global North and the other in the Global South, we started a project to design and develop an immersive telepresence platform for meeting up the requirements.

As the description of our current prototype shows, the affordances of the current solution substantially exceed those of conventional videoconferencing or other related technologies that, while technically opening new opportunities, pedagogically have reduced the ways that a learning community can function in a distributed setting.

The affordances of the developed immersive telepresence pilot, when analyzed in the framework of the CDIO model of engineering education, indicates that our solution will not only facilitate a shared learning experience but fundamentally reform it.

Limitations such as cost and bulk of equipment are being put aside to allow us a chance at exploring the possibilities of the technology prior to such systems being made affordable and portable. We will be further developing the technology in the context of the plug-in campus in Namibia over the coming few years, putting both the technology and our proposed affordances for education under a spotlight, as well as investigating what presently unimaginable applications there may be. Through direct engagement with educators we are ideating potential affordances before using co-design and Design Science Research to elaborate on the most promising applications. We will subsequently conduct studies involving both University and school students in Finland and Namibia using well developed scales to measure social, spatial and cognitive presence as well as learning outcomes for various group activities and contents in order to evaluate the different uses of this technology.

REFERENCES

[1] S. Orts-Escolano et al., “Holoportation: Virtual 3D teleportation in real-time,” UIST 2016 - Proc. 29th Annu. Symp. User Interface Softw. Technol., pp. 741–754, 2016.
[2] “Social VR like never seen before,” VRTogether. [Online]. Available: https://vrtogether.eu/. [Accessed: 11-Apr-2020].
[3] S. N. B. Gunkel, H. Stokking, T. De Koninck, and O. Niamut, “Every-day Photo-Realistic Social Vr: Communicate and Collaborate With an Enhanced Co-Presence and Immersion”, 2019.
[4] P. Stotko, S. Krumpen, M. Weinmann, and R. Klein, “Efficient 3D reconstruction and streaming for group-scale multi-client live telepresence”, Proc. - 2019 IEEE Int. Symp. Mix. Augment. Reality, ISMAR 2019, pp. 19–25, 2019.
[5] G. Kurillo and R. Bajcsy, “3D teleimmersion for collaboration and interaction of geographically distributed users”, Virtual Real., vol. 17, no. 1, pp. 29–43, 2013.

[6] C. Fifelski, A. Brinkmann, S. M. Ortmann, M. Isken, and A. Hein, "Multi depth camera system for 3d data recording for training and education of nurses", Proc. - 2018 Int. Conf. Comput. Sci. Comput. Intell. CSCI 2018, pp. 679–684, 2018.
[7] D. Economou, I. Doumanis, L. Argyriou, and N. Georgalas, "User experience evaluation of human representation in collaborative virtual environments", Pers. Ubiquitous Comput., vol. 21, no. 6, pp. 989–1001, 2017.
[8] B. Ens et al., "Revisiting collaboration through mixed reality: The evolution of groupware", Int. J. Hum. Comput. Stud., vol. 131, no. February, pp. 81–98, 2019.
[9] H. Hirschmüller, "Stereo processing by Semiglobal Matching and Mutual Information," IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 2, pp. 328–341, 2008.
[10] K. R. Kim and C. S. Kim, "Adaptive smoothness constraints for efficient stereo matching using texture and edge information," Proc. - Int. Conf. Image Process. ICIP, vol. 2016–Augus, pp. 3429–3433, 2016.
[11] [1] D. Scharstein, T. Taniai, and S. N. Sinha, "Semi-global stereo matching with surface orientation priors," in Proceedings - 2017 International Conference on 3D Vision, 3DV 2017, 2018.
[12] E. F. Crawley, "MIT CDIO Report #1 The CDIO Syllabus: A Statement of Goals for Undergraduate Engineering Education." Massachusetts Institute of Technology. 2001.