VR and AI: The Interface for Human and Non-Human Agents

Lukasz Mirocha

School of Creative Media, City University of Hong Kong

lukasz.mirocha@my.cityu.edu.hk

Abstract

The paper analyses how real-time 3D graphics, VR and AI, allow users to create a new type of interface/environment for human and non-human agents' collaboration and learning. It argues that spatial media (VR, AR) should be considered as real-time software and media interfaces, rather than multi-media projections. The study is informed by software and platform studies, critical theory and media studies perspectives.

The Ultimate Display as an Interface

In 1965, Ivan Sutherland suggested that "the ultimate display would be a room within which the computer can control the existence of matter." [1] Nearly a decade later, in one of the Star Trek: The Animated Series' episodes, a peculiar technology appeared – the holodeck. Today, we could describe it as an ultimate VR environment designed for work and entertainment. Lately, thanks to the combination of latest developments in AI, computer graphics and VR – we seem to be closer to turn these 20th century dreams into reality.

Berry and Dieter write that in the last decade, "computation [has become] experimental, spatial and materialized in its implementation, embedded within the environment and […] even within the body."[2] Following Grau, we could ask about affordances and limitations of immersive and real-time software media, and objectify them "through knowledge and critique of the image production methods." [3] Analysing a new software and media ecology for the creation of virtual or hybrid environments that open new dimensions in human-machine interaction can help us to understand not only the conditions behind these phenomena but also their wider cultural impact.

Simulated Reality – NVidia's Applications

At Siggraph 2017, NVidia showed the Isaac Robot that had been trained in a virtually created world to play dominos with human players. Isaac Robot was firstly trained in Isaac Sim, a virtual training environment. [4] The environment was based on a modified version of a game engine – Unreal. Isaac Sim offers fully integrated and high-fidelity visuals and physics simulation. Thanks to a set of AI algorithms for deep reinforced learning in a virtual environment, a virtual robot can iterate and learn much faster than in a real world. The same rationale lies behind NVidia Drive Sim. [5] It is a virtual training environment that utilizes high fidelity visuals and physics to simulate real-world driving in different weather lightning and traffic conditions. The photorealistic data streams generated by the software are compatible with the same sensors and chips that are used in physical autonomous cars currently tested by the company. Ultimately, a physical testing car can be firstly trained in Drive Sim and then use its knowledge in a real-life situation.

At a certain level of generalization, we can then conclude that the rationale behind NVidia's experiments is to have two instances of agents (robots and cars), the virtual one that is trained through reinforced learning techniques, and the physical one, that consists of a physical "body" and makes use of data gathered by its virtual counterpart to perform tasks in a non-virtual environment. By "body" we mean the same array of sensors and actuators for perception, navigation and manipulation both in the virtual and in the physical world. NVidia's applications of real-time graphics and AI is a continuation of research conducted by other companies and researchers, e.g. Xerox Research (pre-training of computer vision algorithms for autonomous vehicles in a virtual environment created by Unity 3D engine) [6]; or Princeton, Darmstadt University and Intel (real-time recognition of objects, such as road signs, people, and cars, by a machine learning system in a modified video game environment (GTA V). [7]

VR as an Approximate and Simplified World Simulation and Interface

NVidia's achievements prove that today we can use sophisticated software/hardware ecologies to create virtual environments. Conceptually, these environments take advantage of the principles of "approximation to" and "simplification of" when simulating selected properties of the physical world - visuals, audio, physics etc. [8] As a result, at a very basic level, these environments function as streams of data and algorithms processed in real-time by efficient hardware platforms. These data streams can be converted into output material (cues) that can be delivered at the same time both to human agents (e.g. as a dimensional 3D space visualized and interacted with as a VR experience), and to non-human agents. The simplification of the physical world to several data streams (for instance video feed, radar, and proximity sensor) that would not be suitable for human agents, is sufficient for the robot to operate in a physical environment. We can observe a comparable situation when the roles are reversed. Human agents can operate in a VR environment that offers an approximate simulation of physical world by stimulating human senses with simplified cues (visual, auditory, and haptics).

If we were to assess the conceptual status of virtual environments used in the examples presented above, we could follow Galloway's idea of interfaces as "processes" and "zones of activity." [9] Immersive CGI-based environments could be considered, not as multi-sensory projections, but rather as interactive, real-time interfaces. Grau observed that technological developments, like VR, bring us closer to "images as dynamic virtual spaces." [10] In fact, the key characteristic behind VR is that it is a real-time (dynamic) and multi-sensual (multi-cues) medium, where, thanks to a projection of convincing stimuli, an *immersant* (human or non-human) can feel a sense of presence inside a virtual space. Bolter and Grusin explicitly say "the responsive character of the environment, gives VR its sense of presence." [11]

The presented examples show that VR environments are in fact zones of activity that simulate ontologies, create horizons of possibility – defined by affordances of systems that can deliver specific visual, auditory, haptic and data cues to the agents involved. The unique design affordances and constraints implemented in VR environments shape their status as cultural software that today mediates people's interaction with media and other people. Soon, as the Isaac example shows, they will also mediate human-non-human interaction and communication. Therefore, if we consider VR environments as media interfaces, we are getting access to yet another perspective for analysing different models of representing and accessing digital information in today's media ecology, populated both by human/non-human agents.

References

1. Ivan E. Sutherland, "The ultimate display," *Proceedings of the IFIP Congress* (1965), 506-508.
2. David Berry and Michael Dieter, eds. *Postdigital Aesthetics: Art, Computation and Design* (Basingstoke: Palgrave Macmillan, 2015), 3.
3. Oliver Grau, *Virtual art: From Illusion to Immersion* (Cambridge, Mass: MIT Press, 2007), 202.
4. Voices of VR website, accessed August 15, 2018, https://bit.ly/2A7N44w
[5] Nvidia website, accessed August 15, 2018, https://bit.ly/2pJv8r9
[6] Adrien Gaidon et al., *Virtual Worlds as Proxy for Multi-Object Tracking Analysis* (2016).
[7] Princeton University website, accessed August 12, 2018. https://bit.ly/2vbr1YO
[8] Jason Gregory, *Game Engine Architecture* (Boca Raton: Taylor & Francis, CRC Press, 2018), 9.

[9] Alexander Galloway, *The Interface Effect* (Cambridge, UK: Polity, 2012), VII, 36.
[10] Oliver Grau, *VirtualArt: From Illusion to Immersion* (Cambridge, Mass: MIT Press, 2007), 345.
[11] Jay David Bolter, and Richard Grusin, *Remediation: Understanding New Media* (Cambridge, Mass.: MIT Press, 2003), 16.

Biography

Lukasz Mirocha is a PhD Candidate at SCM, CityU. He is interested in media aesthetics, design (particularly VR, AR, MR) and software studies.