altice labs

HCI boosted by AI: from smart interfaces to immersive cognitive environments

Smart interfaces; Immersive environments; Artificial intelligence; Digital twins; Extended reality; Augmented cognition; Augmented humans

White paper

Version 1.0, September 2020

Introduction

The history of interaction is a convergence path, closing the gap between humans and machines. The first steps were taken by man adapting to machine communication constraints, using binary code and switches to establish a communication channel, but we've been witnessing enormous advances in bringing machines closer to common natural forms of interaction. The human interaction paradox is a huge challenge, not only for its richness and complexity, but also for the level of cognition required to understand it: there is a multiplicity of forms of interpersonal communication, ranging from spoken language to body and facial expressions, and multisensory capacity heavily enriches human interaction, nevertheless introducing extreme complexity, with countless possibilities.

Research in the area of human-computer interaction (HCI) has been progressing aligned with this perspective, exploring the context of application of interaction technologies and enriching the existing multidisciplinary knowledge. This multifaceted approach has allowed us to broaden the technological domain of intervention and evolve into topics such as natural language processing, motion capture and body language recognition, which, complemented by the most recent advances in artificial intelligence such as deep learning and machine learning, potentiated the emergence of smart interfaces.

Smart interfaces and immersive environments

Human interaction is continuous, deeply contextual and inherently multimodal. Verbal communication is complemented with body language and references or even interactions with one's own environment [1].

Natural interfaces aim to transpose this conversational richness into human-computer interactions, simplifying communication and exposing the natural spontaneity of interaction. This semantically rich and multichannel user-centred interaction is characterized by its broad-spectrum, combining voice, image and behaviour [2], whose perception is context-dependent. The need for user control of interaction, by making it user-centric, is crucial for the success of the multimodality of the interaction [3].

Natural interaction and the multimodality of interaction have been enriched by the growing interest of the scientific community and industry in conceptualizing innovative interfaces. The multimodality of interaction assumes a particular role in investigating interaction with intelligent software agents, as described in the study by Norouzi et al. [4]. On the other hand, Silva et al. [5], in a systematic literature review, identified 19 articles that present innovative trends in multimodal interaction. In this paper, the vast majority of identified articles fall into the category of natural interaction, including gestures, voice, vision, smell, and cognition. The evolution of natural interfaces includes the use of tactile surfaces that extend interaction to the environment and enable novel manipulation techniques [6].

On the other hand, advances in affordable devices for immersion and visualization, such as Oculus Rift, HTC Vive, Samsung Gear VR and Google Cardboard head-mounted displays (HMDs), leverage new environments and augmented three-dimensional (3D), virtual reality (VR) content or even mixed [7]. These new emerging environments open new challenges in interaction. Being based on traditional interaction techniques, these environments limit the tracking of users' movements and consequently restrict manipulation in the three-dimensional environment, especially in situations of free interaction in space where the accuracy of spatial interactions is a requirement. These limitations further constrain user performance [7]. Nevertheless, this technology has potential benefits in different fields, from education [7] to smart cities [8], for their collaborative and co-creation capabilities [9]. Mixed reality (MR) generates new types of user experience by enhancing a semantic understanding of the environment and consequently presenting itself as a solution to the complex challenges of interaction design [10].

The user experience provided by these environments is largely due to their immersibility. For Dede et al. [11] immersion is the subjective impression that someone is participating in a comprehensive and realistic experience. Schubert et al. [12] associate immersion with presence as a subjective sensation of being in the virtual place. The authors argue that presence is observable when people interact in the virtual world, claiming that presence is perceived as a result of the active interpretation of a virtual environment, arising when the possibilities for bodily action in the virtual world are mentally represented. Slater et al. [13] introduce the perspective of a first-person body transfer and argue that immersive virtual reality is a powerful tool in the study of body representation and experience.

In a recent study, Kang [14] highlights the positive impact of assisted natural interaction in virtual reality environments, increasing the level of empathy, efficiency and emotional effect. In this paper, the author evaluates various technologies that allow three-dimensional body interaction, allowing users to control the contents of the environment in a similar way to the real world by partially or totally moving their body. LaViola [10] refers that the investigation of interactions in 3D space has received considerable attention from the community, presenting several strategies to improve the accuracy of recognition of three-dimensional gestures, in particular broad sets of gestures, which can be associated with their context. The contextualization of interaction in three-dimensional environments leads to emerging concepts presented by Lee et al. [15]. Ubiquitous virtual reality (U-VR) is an emerging paradigm that assures the user applications according to their interaction context [16]. To ensure the pervasiveness of virtual reality environments, this paradigm is based on three pillars: collaborative mediated reality, wearable mediate reality, and context-aware mediated reality. Similarly, the concept of pervasive augmented reality is introduced by Grubert et al. [17] as "***a continuous, ubiquitous and universal augmented interface of information in the physical world***". Practical applications of this kind of applications are commonly presented as "***context-aware augmented reality***" [18], emphasizing the conjugation of augmented reality (AR), intelligent virtual agents, and the Internet of things (IoT) analyzed in the systematic review by Norouzi et al. [19]

In a nutshell, the convergence we see from research in the domain of human-computer interaction goes towards Engelbart's view of machines as enablers of human intelligence [1]: smart interfaces are not intended to automate tasks, but to improve human decision making by interacting with the real world through virtual interactions.

Exploratory research

Figure 1 – Engelbart computer mouse prototype (1968) [25]

Mouse and graphical user interface (GUI) paradigm deeply shaped the way we have been interacting with computers for decades (**Figure 1**), so in this world dominated by mice, keyboards and joysticks for games, all based on similar triggering principles for interaction, the release of Microsoft Kinect in 2010 was a truly disruptive innovation.

Our research in immersive 3D environments has made evidence of its remarkable potential in collaborative activities, but the classical GUI is quite limitative for user embodiment (avatar). Natural interaction is required for improved immersivity, ease of use and pervasiveness, and motion recognition provided by Kinect and its integration with virtual worlds in Online Gym project allowed the creation of a low-cost multi-user virtual gym prototype [20]. Tests conducted with real users shown the ease of interaction for the elderly and the improved socialization features, but also highlighted some significant shortcomings, namely in the recognition of fine gestures, the accuracy of recognition, depth and concealment.

Throughout recent years, advances in VR and AR technology catalyzed the emergence of a series of devices based on new acquisition paradigms, including computer vision, electromyography and gyroscopes. To address previously identified problems, project InMERSE explored the fusion and contextualization of interaction in immersive environments: fine gestures using Myo and LeapMotion devices with greater precision (**Figure 2**), complementing Kinect motion capture, plus context synchronization between immersive display devices, namely (by then state of the art) Oculus Rift and Google Glass. Furthermore, to achieve the technological objectives of the project, a multi-device framework was created allowing to abstract motion recognition and its application semantics, thus also mitigating technological obsolescence in an area where devices are coming and going at a fast pace. The software library for natural interaction was made available in open-source [21] and allowed the development of two demonstrators: a digital signage prototype and an interactive, immersive multi-user installation [22]. Although being a big advance relatively to most of the limitations found in previous projects, this approach is not yet a solution for widespread use cases because of the complexity and cost of the required mix of the devices.

More recently, a small revolution has been happening in the software side, with the introduction of new toolkits for gestures recognition based on state of the art machine learning/deep learning techniques. ARaNI project explored these emerging technologies in AR/MR scenarios.

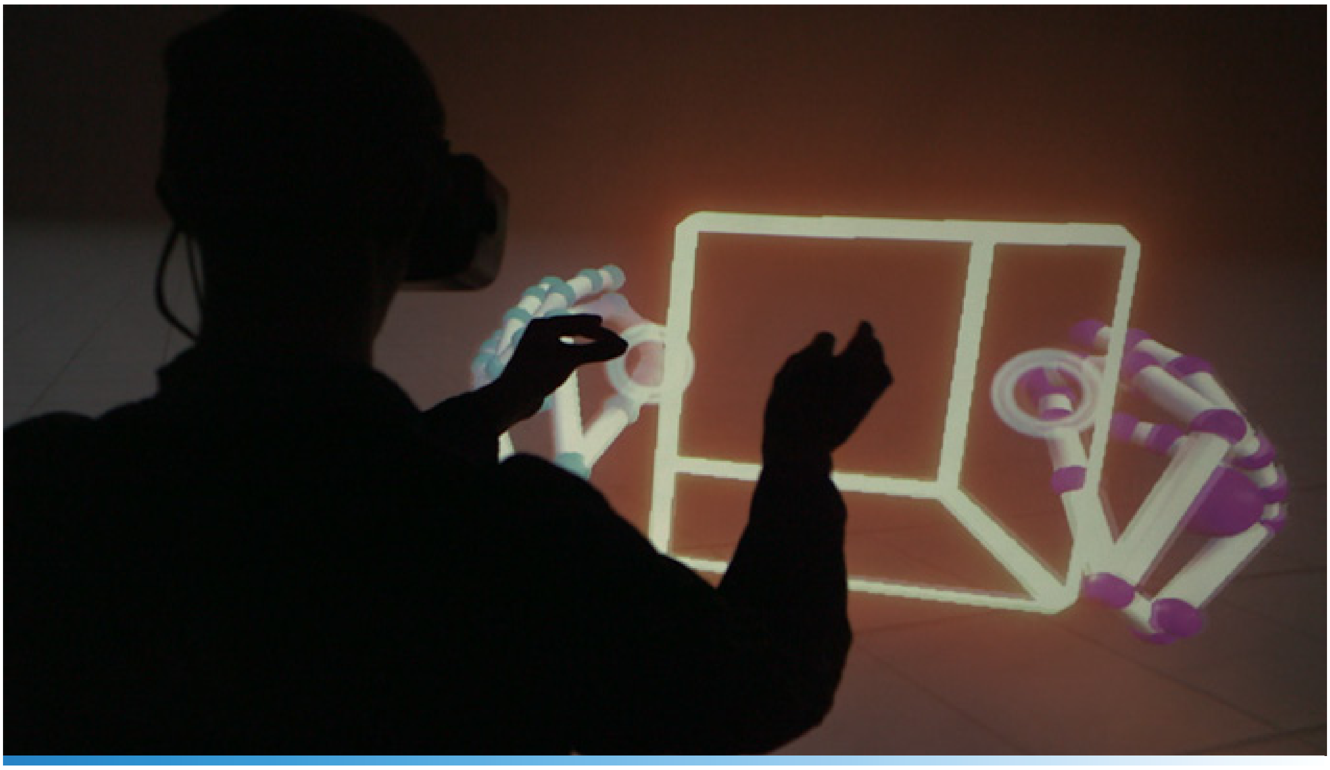

Figure 2 – LeapMotion tracking hands VR demo [26]

The SmartMirror prototype took advantage of the latest AI techniques for motion recognition using very low-cost devices and modest processing capabilities. The use of the OpenPose framework [23], allowed the creation of a multimodal "how to make a tie knot" demonstration without any special interaction device besides a simple cheap webcam (**Figure 3, top**). Gestures aside, user can also interact in plain voice, to guide the prototype flow while having his hands busy.

In another prototype, we explored Microsoft HoloLens mixed reality headset applied to a human body 3D model exploration demonstrator (**Figure 3, bottom**). In addition to using the device capabilities for multimodal interaction using gestures and voice, we also addressed several aspects of the collaborative interaction within the MR environment, allowing multiple users to interact with the virtual 3D model simultaneously (via HoloLens or a web interface, easily extendable to other devices).

In few years, we have definitely walked away from the limited and somewhat clumsy natural interaction in 3D virtual worlds requiring special devices. The convergence of several rapidly expanding technological areas is creating the opportunity for smooth shared eXtended Reality (XR, AR/VR/MR) collaborative interactive scenarios that we may antecipate being soon widely available, enriched with even more interesting capabilities.

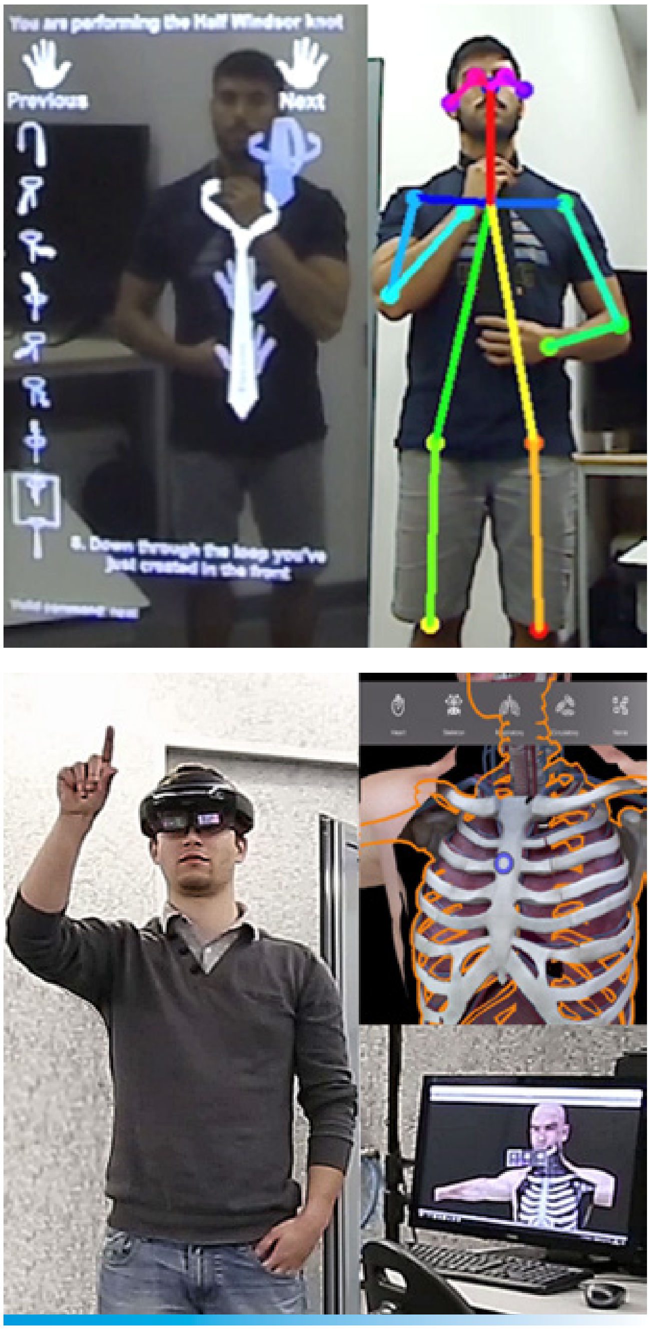

Figure 3 – AraNI prototypes

Raising up augmented humans: a roadmap through seamless AI for HCI

Exploratory research carried on by Altice Labs in collaboration with University de Trás-os-Montes e Alto Douro yielded a set of outcomes that allow us to anticipate a roadmap for the potential of smart interfaces and some of the building blocks for their combination with extended reality environments.

Strictly at the interaction level, the use of AI is already happening, with the integration of deep learning techniques associated with image processing and gesture recognition, such as the powerful OpenPose library explored in the ARaNI project. This open-source real-time multi-person system allows very complex use cases, by jointly detecting human body, hand, facial and foot key points, up to a total of 135 key points on single images. Besides, it's based on TensorFlow [24], an end-to-end open-source platform for machine learning, originally developed by researchers and engineers working on the Google Brain team, which became a popular de facto standard with a comprehensive and flexible ecosystem of tools, libraries and community resources.

The potential for using AI techniques in interaction goes far beyond their application in computer vision. We can foresee AI facilitating interaction, with the ability to predict user actions and adapting to their habits, based on a comprehensive set of personalization aspects, usage information, and also the context of the interaction, using human-inspired techniques such as neural networks embedded into the interactive systems. In this way, it will be possible to make tasks simpler and consequently improving the performance and quality of the interaction, opening new horizons for assisting complex tasks, more prone to human error, but which require human cognition in the decision process.

This role of interaction facilitator reverses a more conservative view of technology as an obstacle for general access to knowledge and the information society development, becoming, in fact, a catalyst for technology's ability to scale up, due to the broadening potential of personalization, fitting each and every user, and as such facilitating education and contributing to a more inclusive (information) society. At this turning point, technology takes on its responsibility to act as a bridge to overcome communication and social interaction hurdles.

This process of interaction transformation is not complete without the mutual involvement of all actors in the course of action, and literature makes evidence of the role of immersion as a stimulus for human involvement. Virtual reality presents an opportunity for reinforcing this immersion, with the new real-world representation potential brought by the IoT sensorization.

Digital twins are a particular example of this merger, being accurate dynamic virtual representations of the physical entities, which are being applied in multiple areas, notably Industry 4.0. In these scenarios, virtual replicas allow to improve operations, increase efficiency or even predict problems. Furthermore, whole lifecycles can be adressed as early as in the design phase, and simulation of new functionalities or brand new products, before a physical prototype is built. Lessons learned in the virtual environment can then be applied in a real environment, already knowing its consequences (**Figure 4**).

Digital twins concept also means an evolution of virtual reality, bridging over the traditional real/virtual segregation, which offers total immersion on the virtual side but involves a clear boundary to the real environment. In this transformation, virtual reality is combined seamlessly with the real world, often giving way to augmented reality or even mixed reality realistic fusion, with enrichment coming from information formerly scattered among several computer applications or in purely virtual environments. The concept of smart everything anywhere, as a new dimension of the Internet of things, is a two-way portal between real and virtual that enables this new augmenting reality.

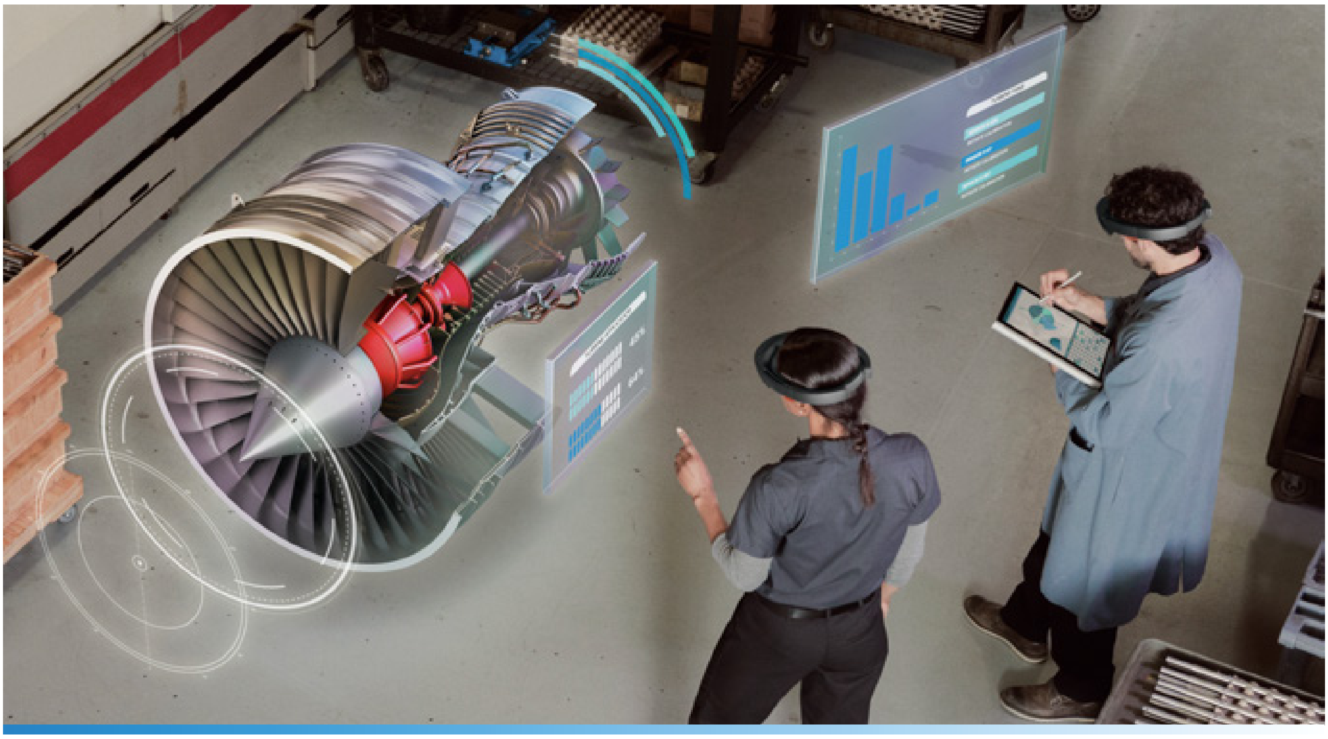

Figure 4 – XR in Industry 4.0 context [27]

Looking ahead, we may expect a convergence of interaction technologies that have been following autonomous research and development streams, towards an integration in a unified interface. The materialization of this vision of pervasive, personalized and customizable mixed reality environments is strongly conditioned by the computing capabilities of mobile and wearable devices. On the other hand, moving the processing to high availability cloud architectures is conditioned by latency, with a major impact on the user experience. The deployment of 5G and technologies such as edge and fog computing, ensuring low latency of the network infrastructure and all the needed processing capacity, will definitely accelerate the introduction and contribute to market widespread of interesting and powerful mixed reality solutions.

Ultimately, technological conditions for experiencing rich mixed reality in pervasive environments will be met, and we will witness symbiotic man-machine cooperation, with high computational power augmenting human cognition and facilitating/enabling people to accomplish a wide spectrum of tasks. Engelbart's vision of technology as a catalyst for human capacity can be materialized in this hybrid interaction environment, with the imperceptibility between real and virtual allowing the genesis of Augmented Humans who, acting naturally and endowed with technology, are empowered in their full cognitive potential and decision-making power.

References

[1] J. Landay, “Smart Interfaces for Human-Centered AI,” *The Stanford Institute for Human-Centered Artificial Intelligence*, 2019. [Online]. Available: https://hai.stanford.edu/news/smart-interfaces-human-centered-ai.

[2] W. Liu, “Natural user interface- next mainstream product user interface,” *2010 IEEE 11th Int. Conf. Comput. Ind. Des. Concept. Des. 1*, vol. 1, pp. 203–205, 2010.

[3] D. Roscher, M. Blumendorf, and S. Albayrak, “A meta user interface to control multimodal interaction in smart environments,” in *Proceedings of the 14th international conference on Intelligent user interfaces (IUI ’09)*, 2009, pp. 481–482.

[4] N. Norouzi et al., “A Systematic Survey of 15 Years of User Studies Published in the Intelligent Virtual Agents Conference,” in *Proceedings of the 18th International Conference on Intelligent Virtual Agents (IVA ’18)*, 2018, pp. 17--22.

[5] T. Silva, P. Almeida, J. Abreu, and E. Oliveira, “Interaction paradigms on iTV: A survey towards the future of television,” *IMCIC 2018 - 9th Int. Multi-Conference Complexity, Informatics Cybern. Proc.*, vol. 2, no. Imcic, pp. 18–23, 2018.

[6] D. Mendes, F. M. Caputo, A. Giachetti, A. Ferreira, and J. A. Jorge, “A Survey on 3D Virtual Object Manipulation: From the Desktop to Immersive Virtual Environments,” *Comput. Graph. Forum*, vol. 38, pp. 21–45, 2019.

[7] R. R. Divekar et al., “Interaction challenges in AI equipped environments built to teach foreign languages through dialogue and task-completion,” *DIS 2018 - Proc. 2018 Des. Interact. Syst. Conf.*, pp. 597–610, 2018.

[8] L. Morgado et al., “Cities in Citizens’ Hands,” *Procedia Comput. Sci.*, vol. 67, no. September, pp. 430–438, 2015.

[9] T. Kohler, J. Fueller, K. Matzler, and D. Stieger, “CO-creation in virtual worlds: The design of the user experience,” *MIS Q. Manag. Inf. Syst.*, vol. 35, no. 3, pp. 773–788, 2011.

[10] L. J. Joseph J., “Context aware 3D gesture recognition for games and virtual reality,” in *ACM SIGGRAPH 2015 Courses (SIGGRAPH ’15)*, 2015, pp. 10:1--10:61.

[11] C. Dede, “Immersive interfaces for engagement and learning,” *Science (80-.).*, vol. 323, no. 5910, pp. 66–69, 2009.

[12] T. Schubert, F. Friedmann, and H. Regenbrecht, “Embodied Presence in Virtual Environments,” in *Visual Representations and Interpretations*, 1999, pp. 269--278.

[13] M. Slater, B. Spanlang, M. V. Sanchez-Vives, and O. Blanke, “First person experience of body transfer in virtual reality,” *PLoS One*, vol. 5, no. 5, 2010.

[14] J. Kang, “Effect of Interaction Based on Augmented Context in Immersive Virtual Reality Environment,” ***Wirel. Pers. Commun.***, vol. 98, no. 2, p. pp 1931–194, 2018.

[15] Y. Lee, S. Oh, C. Shin, and N. Woo, “Recent trends in ubiquitous virtual reality,” ***Proc. - Int. Symp. Ubiquitous Virtual Reality, ISUVR 2008***, no. May 2014, pp. 33–36, 2008.

[16] S. Jang, S. Kim, and W. Woo, “When VR meets UbiComp,” ***Virtual Real.***, pp. 1–2, 2005.

[17] J. Grubert, T. Langlotz, S. Zollmann, and H. Regenbrecht, “Towards Pervasive Augmented Reality: Context-Awareness in Augmented Reality,” ***IEEE Trans. Vis. Comput. Graph.***, vol. 23, no. 6, pp. 1706–1724, 2017.

[18] A. H. Basori, A. M. A.-G. Al-Sharif, A. O. F. AL-Oufi, A. O. Almagrabi, and O. M. Barukab, “Intelligence Context Aware Mobile Navigation using Augmented Reality Technology,” ***J. Inf. Syst. Eng. Bus. Intell.***, vol. 4, no. 1, p. 65, 2018.

[19] N. Norouzi, G. Bruder, B. Belna, S. Mutter, D. Turgut, and G. Welch, “A Systematic Review of the Convergence of Augmented Reality, Intelligent Virtual Agents, and the Internet of Things,” in ***Artificial Intelligence in IoT. Transactions on Computational Science and Computational Intelligence***, 2019, pp. 1–24.

[20] F. Cassola, L. Morgado, F. de Carvalho, H. Paredes, B. Fonseca, and P. Martins, “Online-Gym: A 3D Virtual Gymnasium Using Kinect Interaction,” ***Procedia Technol.***, vol. 13, no. December, pp. 130–138, 2014.

[21] L. Fernandes et al., “InMERSE Framework,” 2015. [Online]. Available: https://bitbucket.org/Apidcloud/inmerse-framework/src/master/.

[22] L. M. Alves Fernandes et al., “Exploring educational immersive videogames: an empirical study with a 3D multimodal interaction prototype,” ***Behav. Inf. Technol.***, vol. 35, no. 11, pp. 907–918, 2016.

[23] G. Hidalgo, Z. Cao, T. Simon, S.-E. Wei, H. Joo, and Y. Sheikh, “OpenPose: Real-time multi-person keypoint detection library for body, face, hands, and foot estimation,” 2019. [Online]. Available: https://github.com/CMU-Perceptual-Computing-Lab/openpose.

[24] “TensorFlow,” 2019. [Online]. Available: https://www.tensorflow.org/.

[25] Wikimedia Commons, “Wikimedia Commons,” 9 December 2018. [Online]. Available: https://commons.wikimedia.org/wiki/File:SRI_Computer_Mouse.jpg.

[26] Ultraleap Ltd, “LEAP MOTION” 2019. [Online]. Available: https://www.leapmotion.com/press/.

[27] Microsoft, “Mixed Reality,” 2019. [Online]. Available: https://news.microsoft.com/presskits/windows-mixed-reality/.

Authors

Hugo Paredes

Assistant Professor with Habilitation at UTAD, Portugal

Assistant Coordinator of CSIG at INESC TEC, Portugal

hparedes@utad.pt

https://www.linkedin.com/in/hparedes/

Fausto de Carvalho

Senior Technology Consultant

Altice Labs, Portugal

cfausto@alticelabs.com

https://www.linkedin.com/in/fausto-de-carvalho/

Contacts

Address
Rua Eng. José Ferreira Pinto Basto
3810 - 106 Aveiro (PORTUGAL)

Phone
+351 234 403 200

Media
contact@alticelabs.com
www.alticelabs.com