“Quantifying the phenomenon of immersion in virtual environments”

Authors:
Gutman, Maja – Electrical and Computer Engineering Department, University of California, Los Angeles (UCLA)
Lu, Qiujing – Electrical and Computer Engineering Department, University of California, Los Angeles (UCLA)
Roychowdhury, Vwani – Electrical and Computer Engineering Department, University of California, Los Angeles (UCLA)

A human body plays a profound role in the way one perceives and responds to the world around her and the phenomenon of immersion represents a clear example of paralinguistic body cues that have received little in-depth research so far. With the emergence of new media led by information technology, the impact of such immersion on the individual, and the postmodern society as a whole, has become even more significant. A distinctive feature of new media is the fundamental active role of the user, where media and individuals are not separate but define and create each other, together forming a complex ecosystem. Nowhere is this seamless melding of the individual and media more apparent than in the fields of Virtual Reality (VR) and Brain-Computer Interface (BCI). A growing body of research indicates that a human subject is capable of distributing her immersion across the Virtual Reality Environment (VRE) and that the experience of ‘being present’ or ’immersed’ can be more intense than the corresponding experience in the ‘real’, i.e. physical, world. This phenomenon of immersion inherent to VRE has the potential to reshape an individual more than any other type of media, and VR technology has already shown transformative promise in a number of fields, including medical and psychological treatment, gaming, and education. In spite of such early successes of VRE induced immersion, there is currently no standard or commonly agreed measurement of immersion. The fundamental thesis of this paper is that the effect of VRE and related immersive technologies can be successfully studied only via a trans-disciplinary approach that combines qualitative theoretical models — widely discussed in media studies, phenomenology and psychology — with quantitative data-driven empirical models based on Big Data and modern advances in AI (Artificial Intelligence), Machine Learning (ML), and computational models.

This convergence enables one to borrow and adapt established theories, such as embodied cognition (Merlau-Ponty) along with the body schema concept, and theory of technology as an extension of the body (McLuhan) to develop qualitative models of individuals as prosumers of such interactive media. Moreover, the information technology era provides another prospect that makes the media studies field trans-disciplinary: the promise of quantification.

The *aim of this research paper is to integrate the concept of a human body as a condition of experience with the empirical modeling of immersion by automatically capturing and analyzing the non-verbal signals*, i.e. kinematic features and verbal features (expressions) of users who are experiencing immersion in virtual environments. Preliminary analysis has been conducted on two datasets that consisted of YouTube videos, largely depicting body movements of users interacting with two different types of VR content: (i) survival based / strategy VR games; and (ii) 3D painting VR application. Before developing automatic tools, approximately 150 videos were manually viewed by a human expert, which led to an important qualitative observation: Users' bodies in 3D painting application were creating and intensely occupying the space around them, whereas users in survival based games were mainly escaping from the space. *This body dynamic can be roughly described as a contraction/expansion duality, where two very distinct simulations of situations are at play*. However, can this contextual behavior of the body also be empirically verified, given the aforementioned situations? For this purpose, we are in the preliminary stages of designing a multi-modal AI platform that use both Computer Vision (CV) and Speech Analysis (SA) tools and that can automatically detect, quantify, and analyze bodily expressions associated with immersion.

The pipeline under development comprises of the following modules: (i) Each downloaded video is segmented into clips focusing on each player in an automated manner using both CV and SA tools. (ii) In each image frame of each clip (recall that a standard video has 30 image frames per second of recording) a skeleton comprising of 18 key-points (such as left and right shoulders), and 17 links (such as those representing arms and legs) among these key-points is computed (we use an open-source Deep Neural Network (DNN) based package for extracting these skeletons). (iii) Patterns in the relative positions of these key-points in a frame define certain static poses that are important for quantifying immersion, e.g. sitting down or lying on the floor, or with hands close to one's face. Sequences of movements of key-points over multiple frames define dynamic movement patterns (such as fluid movements of arms in smooth arcs versus movements such as sudden and reflexive reactions, and falling down). (iv) Both the presence and absence of audio signals in a video session provide clues to immersion. To capture these cues, speech processing software packages are being integrated into the pipeline and classifiers for various evocative motifs (such as pleasure, horror etc.) are being developed. (v) Finally, for any given VR recording episode an analytics package that computes different measures that are related to immersion is being developed. A current version of this summary package includes duration and presence of patterns such as composed movements (characterized by coordinated movements of key-points and body parts), body contraction movements (characterized by sudden movements of arms to a defensive position or lifting hands in front of one's face, or falling down), and overall summary statistics such as variation in irregular body movement, average activity level, aural interjections and levels of vocalizations. *Our preliminary results based on this pipeline and an analysis of video clips suggest that there are strong indicators of immersion of subjects in VRE that can be captured and categorized in an automated manner.*

In summary, this paper *lays some of the foundational work for modeling the phenomenon of immersion*, based on theories of embodiment and behavioral modalities of physical sensations experienced in VRE. It focuses on video data in particular, where key body parts are being automatically identified and their movements are being tracked, quantified and visualized. Similarly, for audio processing, various features are being computed, and machine learning algorithms for detecting emotional motifs are being developed. The paper will also demonstrate how biometric signals of immersion can be analysed without the use of biometrical equipment.