Toward Interactive and Distributed Hybrid Extended Reality Experiences

Maurizio Murroni
maurizio.murroni@unica.it
University of Cagliari
Cagliari, Italy

Gabriel Muntean
gabriel.muntean@dcu.ie
Dublin City University
Dublin, Ireland

Mario Montagud
mario.montagud@i2cat.net
i2CAT Foundation and University of Valencia
Barcelona, Spain

Vlad Popescu
vlad.popescu@unitbv.ro
Transilvania University of Brasov
Brasov, Romania

ABSTRACT

Hybrid Extended reAliTy (HEAT) is a project born to pave the way for the next-generation distributed experiences: being realistically immersed (holo-ported) within real captured omni-directional and navigable hyper-realistic 3D spaces, feeling their atmosphere, and sharing these experiences with others, regardless of their location. This paper gives an overview on how HEAT aims at integrating immersive media technologies such as point cloud/holographic imaging, multi-sensorial media, Social VR in a multi-user, feedback-enabled communication system to provide the construction of compelling context-aware and embodied experiences for innovative hybrid XR applications.

KEYWORDS

XR, holography, multi-sensorial.

ACM Reference Format:
Maurizio Murroni, Gabriel Muntean, Mario Montagud, and Vlad Popescu. 2024. Toward Interactive and Distributed Hybrid Extended Reality Experiences. In *Proceedings of the 2024 ACM International Conference on Interactive Media Experiences Workshops (IMXw '24), June 12, 2024, Stockholm, Sweden.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3672406.3672424

1 INTRODUCTION

Multi-user virtual communication and collaboration solutions have gained popularity during the pandemic, as the traditional "in presence" habits of working/learning/entertaining had to be translated to remote/blended ones. Major limitations associated with such remote/blended approaches are mainly related to two aspects. On one hand, there is an inadequacy of how the classic means of audio-video content dissemination (i.e. PCs, laptops, tablets, TV sets) represent real scenarios. On the other hand, the traditional multimedia content itself (i.e., audio and video) is not adequate enough to reflect the real world. 3D scenarios, dynamic volumetric representation of users, stereoscopic 2D, 180°/360° videos, point cloud/holographic imaging have been considered more suitable candidates to reproduce highly realistic experiences. True realistic contents should also include media with multiple sensorial effects (i.e., multi-sensorial media, also called mulsemedia) aimed at increasing the user's experience through the five senses representation (i.e., taste, sight, touch, smell, and hearing), as the real world is perceived. Mulsemedia has been proposed and tested in several scenarios in which it has been proven to enrich the sense of reality with respect to multimedia (i.e., audio-video) [6]. Furthermore, Virtual Reality (VR) headsets have been proposed to enhance users' immersivity levels, allowing them to experience a seamless and rich integration of heterogeneous media formats. Moreover, social interaction between remote users can now be enabled through shared immersive virtual environments, bringing up a new communication medium termed Social VR [9]. Social VR has rapidly attracted a high interest, magnified by the social distancing measures brought by the worldwide pandemic. Social VR allows multiple distributed users getting together in shared virtual environments to socially interact and collaborate. However, the major issue with the current VR approach is sharing of audio-video content only, relying on synthetic avatars for representing users which makes the social cues complicated to be shared. The potentiality of the above technologies is far from being fully exploited, especially due to the lack of a "system perspective" where the technological integration could really make the difference in enabling true immersion and almost-real interaction based on the effective exchange of social cues and proper rendering of user feedback. A step ahead in this sense has been made in [4], where photo-realistic volumetric user representations (i.e., holograms) are used to allow richer identification of the self and others' representations, as well as richer interaction between the users and the VR environment and among themselves. However, full VR experiences have limited and very specific application scenarios (e.g., virtual training, learning, etc). More often, especially in real-time performance scenarios (e.g., live performance, blended learning), real needs are more for hybrid eXtended Reality (XR) applications [5][10], where remote users can experience a real captured environment through immersive XR, while in presence users can visualise and interact with the holograms of remote users integrated in the real environment through holographic rendering. Such interactive hybrid and multi-sensory scenarios have not been fully considered so far and at least three main issues need addressing:

IMXw '24, June 12, 2024, Stockholm, Sweden

ACM ISBN 979-8-4007-1794-9/24/06
https://doi.org/10.1145/3672406.3672424

(1) How can remote users be brought together by capturing and presenting their 3D holograms in a real in presence environment? That is, how can next-generation multimedia and immersive content be effectively processed and delivered over hybrid XR scenarios using low-complexity and low-cost devices?
(2) How can the real in presence environment (audio-visual and multi-sensory) be realistically 3D captured and presented to remote users?
(3) How can in presence and remote users interact? That is, how can remote users' feedback (e.g., social cues, emotions, ratings, etc.) be conveyed to the in-presence users and vice-versa?

The EU Horizon Europe-funded Hybrid Extended reAliTy (HEAT) project was designed to effectively answer the above questions, paving the way for the next-generation distributed experiences. Its aim is to integrate immersive media technologies such as point cloud/holographic imaging, multi-sensorial media, Social XR in a multi-user, feedback-enabled communication system to provide the construction of compelling context-aware and embodied experiences for innovative hybrid XR applications. The resulting system will aim at facilitating the exploitation of agile 3D data acquisition techniques, such as plenoptic point cloud, enhancing interoperability and performance at reduced costs. Examples of HEAT's envisioned mobile multi-sensor volumetric capture setups are shown in Figure 1. As illustrated, HEAT will create an adaptable and modular communication pipeline embodying either encapsulation of different media from classical audio-video to multi-sensorial to holographic video or a combination of them, providing means for efficient encoding, processing, storage, (real-time) streaming and rendering. Different scenarios set in real-world environments are considered for assessment of enhanced XR user experiences: a blended learning, a modern theatre act, a music festival, and an opera show. The pilot actions will follow all GDPR and ethical recommendations.

2 HEAT INNOVATION

HEAT has been conceived around several major pillars to lay the foundation for a new enhanced-media ecosystem in the context of real multi-user social XR experiences: volumetric capture and reconstruction; adaptive delivery and multimodal integration; XR presentation (XR headsets and holographic displays); multi-sensory experience, and interaction techniques.

On the topic of volumetric capture and reconstruction, HEAT will go beyond the state of the art by providing novel technological solutions to enable the agile volumetric (photo-) realistic real-time capture by using a set of off-the-shelf and low-cost sensors such as the one of mobile consumer devices (e.g., smartphones/tablets, 360°/RGB-D cameras), thus overcoming key challenges like varying resolutions of available sensors, their potential limited mobility, and achieving a (semi-) automatic calibration process. Artificial intelligence (AI) techniques will be investigated to enable fast and agile capture in the wild: opportune machine learning (ML) algorithms will be investigated to train capturing systems to reduce complexity and perform near real-time acquisition and processing. Two aspects will be considered: (i) dynamic (real-time) multi-sensor capture and reconstruction of 3D humans (i.e., remote users and live performers), (ii) offline and near-real-time capture and reconstruction of environments or objects. In addition, HEAT will go beyond the state of the art enabling an effective and strategic integration of heterogeneous content modalities and formats, overcoming challenges not only in terms of performance, but also in terms of storytelling and providing increased feelings of realism and immersion beyond 3DoF+ scenarios, at reduced costs [2].

Multi-sensory technology is one of the latest enhancement techniques for XR technologies. HEAT will investigate the production of content with integrated sensorial effects. The current solutions for adding effects into multimedia content mostly consist of manually inputting tags with the desired effects/scents into a video's timeline. Haptic effects are also usually triggered by specific actions manually performed by users with their avatars. Designing (semi-)automatic solutions to enhance multimedia content with sensorial effects will simplify the addition of such effects to both new and existing content and is expected to enhance viewer experience.

User feedback conveying and rendering is currently one of the critical points of virtual and mixed-mode reality applications, being available only in limited scenarios and with the use of additional hardware such as haptic gloves or vests. In HEAT, together with the industry partners specialised in holographic and multi-sensory technology and with the partners involved as end users, we will develop feasible means of conveying and rendering meaningful user feedback (e.g., social cues, emotions, ratings…) to all involved users in both virtual and mixed modes. Biometric data acquisition (e.g., eyegaze, heartbeat) and AI and ML techniques will be investigated to automatically produce dynamic and accurate feedback. As to Adaptive delivery and multimodal integration, volumetric video brings remarkable challenges in terms of real-time processing and adaptive delivery compared to traditional 2D video, mostly due to the much larger amount of captured data and the recent emergence of such formats. HEAT plans to advance the state-of-the-art in this topic by: (i) devising and adopting novel semantic-, position- and viewport-aware strategies for content encoding and segmentation, exploiting AI-based techniques, like autoencoders, as well as denoising and completion methods; and (ii) exploiting the capabilities of network virtualization and Cloud Continuum (e.g., Edge Computing), to maximise scalability and interoperability, by offloading heavy processing functions from the clients to virtualized in-cloud media processing functions, as preliminarily done in [2].

All the above-mentioned pillars have an impact on the perceived user Quality of Experience (QoE) [3], both at remote and in presence users. Therefore the HEAT project will also focus on better evaluating the user QoE for hybrid XR applications. In particular, noteworthy is the evaluation of the effect of multisensorial content on Quality of Interaction (QoI) [7]. Furthermore, the latest developments in terms of AI algorithms and its use for XR media consumption will be investigated as well as further directions on how to confer intelligence to XR, the so-called Affective Computing, or Artificial Emotional Intelligence [8].

3 ENVISIONED HEAT PLATFORM

To enable the hybrid interactive and immersive multi-sensory XR experiences envisioned in HEAT, innovative media components and

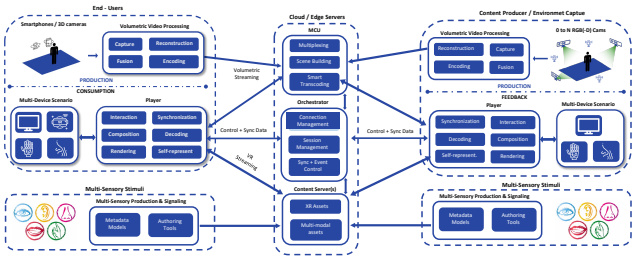

Figure 1: HEAT high-level system architecture

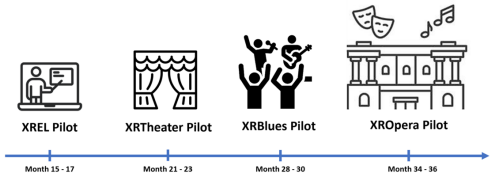

Figure 2: Pilot development timeline

modules need to be developed, evolved, and integrated, re-defining current-day end-to-end media workflows. This involves the specification and development of a modular, elastic and standards-compliant end-to-end platform, integrating:

- Advanced agile heterogeneous multi-sensor capture setups for volumetric media (Point Clouds, NeRF, etc.);
- Authoring tools and well-defined metadata models for the integration of multi-sensory stimuli (haptic, scent);
- Processing and coding for delivering reliable user feedback;
- Low-latency, adaptive and viewport-aware delivery strategies, potentially assisted by cloud computing;
- Orchestration components for signalling sessions and contents, dynamic session management, scene composition, and rich media synchronisation;
- Interaction Engines, including communication and synchronisation features, with interfaces between users, contents, applications and specific output devices for rendering user feedback;
- Rendering capabilities, including heterogeneous consumption devices with their required hardware and software components;
- Strategies to adapt the presentation of contents, based on the available resources (consumption devices, bandwidth, etc.), preferences, interaction patterns, and session and context status, assisted with advanced content- and context-aware QoE models.

These components will be integrated within an end-to-end platform, comprising key parts: Production and Preparation, Orchestration and Distribution, and Consumption and Interaction. A high-level system architecture of the HEAT platform, with its key components and their communication interfaces, is sketched in Figure 2 . Its conception departs from key outcomes of the EU H2020 VR-Together project (coordinated by i2CAT): HoloMIT, a worldwide pioneering platform to enable real-time multi-party real-time holographic communications between remote users [1]. However, its evolved architecture, components and interfaces will be more accurately specified and refined during the project lifetime, based on the selected use cases and derived requirements, and ongoing research findings and associated developments.

As a whole, the HEAT platform will be able to accommodate the multimedia modalities, distribution technologies, interaction modalities, multi-sensory output devices, Internet of Things (IoT) sensors and actuators, and the wide set of heterogeneous consumption devices (e.g., XR headsets, holographic displays, smart wearables etc.) considered in the project. However, a key aspect of the HEAT platform, and of each of its innovative components, is that it will be designed based on standard-compliance and modularity. On the one hand, the platform and its components will be adaptable to interoperate with both state-of-the-art and next-generation broadband workflows. On the other hand, they will be extensible to accommodate additional and/or emerging features not addressed in the project, but that are relevant in the media consumption landscape (e.g., adaptive storytelling based on the audience's feedback, delivery workflows, etc.).

4 APPLICATION SCENARIOS

HEAT will consider four distinct pilot cases as examples for the deployment of the developed technologies in real scenarios. The pilots will be used to validate the applicability and effectiveness of the proposed system through test assessment that will aim at estimating the perceived user's QoE, as well as readiness and acceptability of HEAT contributions. The pilots are listed in ascending order based on their estimated complexity. The timeline for the pilots is depicted in Figure 2.

4.1 XREL Pilot

A lecture, that is a classroom/laboratory environment including a teacher and laboratory equipment will be captured, using volumetric cameras, microphones, and multi-sensory acquisition devices. The whole captured environment (i.e., the live lecture) will be dynamically processed and delivered to the remote students via XR headsets and multi-sensory actuators (i.e., olfaction dispensers, haptic gloves/vests). The remote students will be captured at their site using available devices (e.g., 360°/plenoptic/RGB-D cameras/stereo microphones from smartphones or dedicated cameras), being their hologram tele-ported into the classroom that will be equipped with holographic displays to allow teachers and remote students to socially interact, sharing visual social cues with them.

4.2 XRTheatre Pilot

In a cultural performance of modern theatre, the environment (i.e., scenography, stage and performers) will be captured with volumetric cameras, microphones and multi-sensory acquisition devices. Remote audiences will experience the environment together with the in presence audience, with the use of XR headsets and multi-sensory actuators (i.e., olfaction dispensers, haptic gloves/vests), like being seated in a specific position upon their request. The remote audience will be captured at their site, using available devices

(e.g., 360°/plenoptic cameras/stereo microphones from smartphones or dedicated cameras), being their hologram tele-ported in the theatre via 3D holographic displays and/or XR headsets.

4.3 XRBlues Pilot

The performance of a blues/rock band in a club will be captured using volumetric cameras and multi-sensory acquisition devices. Sound will be dynamically captured from the stage mixer and synchronised with the visual part. Remote audience will experience the environment together with the in presence audience, with the use of XR headsets and multi-sensory actuators (e.g., olfaction dispensers, haptic gloves/vests). The remote audience will be captured at their site, using available devices (e.g., 360°/plenoptic cameras/stereo microphones from smartphones or dedicated cameras), and their hologram tele-ported in the club via 3D holographic displays.

4.4 XROpera Pilot

In a cultural performance of opera, the opera scenography and performers will be captured with volumetric cameras, microphones and multi-sensory acquisition devices. Remote audiences will experience the environment together with the in-presence audience, with the use of XR headsets and multi-sensory actuators (i.e., olfaction dispensers, haptic gloves/vests), like being seated in a specific position upon their request. The remote audience will be captured at their site, using available devices (e.g., 360°/plenoptic cameras/stereo microphones from smartphones or dedicated RGB-D cameras), being their hologram tele-ported in the theatre and rendered via 3D holographic displays. In addition, multi-sensory feedback (e.g., haptic, heart bit) from remote users will be acquired and the possibility of dynamically rendering to the performers via wearable devices will be investigated.

5 CONCLUSIONS

In this paper, the ambition of the Hybrid Extended reAliTy (HEAT) project has been presented. HEAT aims at integrating immersive media technologies such as point cloud/holographic imaging, multisensorial media, Social VR in a multi-user, feedback- enabled communication system to provide the construction of compelling context-aware and embodied experiences for innovative hybrid XR applications. The project will provide well designed and fully tested scenarios in real-world environments for enhanced XR experiences: a blended learning, a modern theatre act, a music festival and an opera show.

ACKNOWLEDGMENTS

This study was carried out within the "HEAT – Hybrid Extended reAliTy" Project GA 101135637 funded by the EU Horizon Europe Framework Programme (HORIZON).

This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

REFERENCES

[1] 20022. *HoloMIT*. Retrieved April 2, 2024 from https://i2cat.net/holomit-the-3d-holoconferencing-solution-for-a-sustainable-future/

[2] Gianluca Cernigliaro, Marc Martos Cabré, Mario Montagud, Amir Ansari, and Sergi Fernández. 2020. PC-MCU: point cloud multipoint control unit for multi-user holoconferencing systems. 47–53. https://doi.org/10.1145/3386290.3396936

[3] Henrique Debarba, Mario Montagud, Sylvain Chagué, Javier Herrero, Ignacio Lacosta, Sergi Fernández, and Caecilia Charbonnier. 2022. Content format and quality of experience in virtual reality. *Multimedia Tools and Applications* (03 2022). https://doi.org/10.1007/s11042-022-12176-9

[4] Sergi Fernández, Mario Climent, Gianluca Cernigliaro, and David Rivera. 2021. Toward Hyper-Realistic and Interactive Social VR Experiences in Live TV Scenarios. *IEEE Transactions on Broadcasting* PP (11 2021), 1–20. https://doi.org/10.1109/TBC.2021.3123499

[5] Charles Hughes, Christopher Stapleton, Darin Hughes, and Eileen Smith. 2005. Mixed Reality in Education, Entertainment, and Training. *IEEE computer graphics and applications* 25 (11 2005), 24–30. https://doi.org/10.1109/MCG.2005.139

[6] Lana Jalal, Matteo Anedda, Vlad Popescu, and Maurizio Murroni. 2018. QoE Assessment for IoT-Based Multi Sensorial Media Broadcasting. *IEEE Transactions on Broadcasting* 64, 2 (2018), 552–560. https://doi.org/10.1109/TBC.2018.2823914

[7] Lana Jalal, Roberto Puddu, Maria Martini, Vlad Popescu, and Maurizio Murroni. 2023. A QoE Model for Mulsemedia TV in a Smart Home Environment. *IEEE Transactions on Broadcasting* 69, 1 (2023), 179–190. https://doi.org/10.1109/TBC.2022.3217676

[8] Jina Lee, Saeyeon Lee, and Dongcheol Kim. 2022. Problems and Issues of Emotional Computing and Artificial Emotional Intelligence: A Review of Rosalind W. Picard's thesis "Affective Computing: Challenges". 360–364. https://doi.org/10.1109/BCD54882.2022.9900683

[9] Li J. Cernigliaro G. et al. Montagud, M. 2022. Towards socialVR: evaluating a novel technology for watching videos together. *Virtual Reality* 26 (2022), 1593–1613. https://doi.org/10.1007/s10055-022-00651-5

[10] Maximilian Speicher, Brian D. Hall, and Michael Nebeling. 2019. What is Mixed Reality?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3290605.3300767

Received 28 April 2024; revised 14 May 2024; accepted 14 May 2024