

# Multiuser Virtual Experiences powered by Holoportation Technologies and Multimodal Human-Computer Interaction (HCI)

Mohamad Hjejij  
i2CAT Foundation & Universitat  
Politecnica de Catalunya  
Barcelona, Spain  
Hjejij.mohamad@i2cat.net

Mario Montagud  
i2CAT Foundation & Universitat  
de Valencia  
Valencia, Spain  
mario.montagud@i2cat.net

David Rincon  
Universitat Politecnica de  
Catalunya  
Barcelona, Spain  
David.rincon@upc.edu

## ABSTRACT

This paper elaborates on a set of research objectives to optimize multi-user Social Virtual Reality (VR) experiences. A first objective consists of designing and developing a novel and lightweight Volumetric Video (VV) pipeline to provide 3D realistic user representations (i.e., holograms), increasing the visual resolution and real-time performance compared to state-of-the-art solutions. A second objective consists of designing and implementing novel modular communication architectures, integrating virtualized in-cloud media processing functions with the goal of enhancing the scalability and interoperability in Social VR. A third objective consists of designing and implementing a set of multimodal Human Computer Interaction (HCI) features be integrated in Social VR experiences to enable rich group-based interaction and effective collaboration. All such envisioned contributions will be thoroughly (objectively and subjectively) evaluated in relevant use cases, like culture and tourism, to validate their benefits and shed some light on associated hypotheses.

## CCS CONCEPTS

- Human-centered computing → Human Computer Interaction
- Networks → Network performance analysis

## KEYWORDS

Holographic Communications, Multimodal Interaction, Social Virtual Reality, Volumetric Video

### ACM Reference format:

Mohamad Hjejij, Marion Montagud, and David Rincon. 2024. Multiuser Virtual Experiences powered by Holoportation Technologies and Multimodal Human-Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).  
*WOODSTOCK '18, June, 2018, El Paso, Texas USA*  
© 2018 Copyright held by the owner/author(s). 978-1-4503-0000-0/18/06...\$15.00  
<https://doi.org/10.1145/1234567890>

Computer Interaction (HCI). In *15<sup>th</sup> ACM Multimedia Systems Conferences*

(MMSys'24), April 15-18, 2024, Bari, Italy. ACM, New York, NY, USA, 5 pages.  
<https://doi.org/10.1145/1234567890>

## 1 Introduction

Social Virtual Reality (VR) brings the promise of becoming the next-generation communication, interaction, and collaboration medium, with wide applicability and impact in relevant sectors [1], like education and culture. It allows multiple remote users to join a shared virtual environment for social interaction and even conducting tasks together. Recent studies have preliminarily proven that Social VR can provide enhanced (co-)presence, social meaning and quality of interaction levels compared to traditional 2D videoconferencing solutions, especially if realistic and volumetric user representations are provided in those shared experiences (e.g., [1]). However, existing Social VR platforms and related studies still encounter key limitations and open research questions. On the one hand, existing Volumetric Video (VV) pipelines to provide realistic user representations and multiuser communications encounter relevant limitations in terms of visual resolution, real-time performance, adaptability, and scalability. On the other hand, most Social VR experiences evaluated and demonstrated to date just provide basic interactive features with the VR environment and between the involved users. This PhD thesis aims at providing innovative technological contributions to overcome such limitations, which are structured around three interconnected pillars: (i) **VV pipelines** to enhance the visual resolution, adaptability and performance for real-time human capture; (ii) Adaptive **Communication architectures** assisted with virtualized in-cloud media processing functions (e.g., stream forwarding, mixing, or transcoding) with the goal of enhancing the scalability and interoperability of Social VR and holographic communication services; (iii) **Multimodal Human-Computer Interaction (HCI)** features to enable rich group-based interactions, effective collaboration, and gamified experiences.

## 2 Objectives and Hypotheses

## 2.1 Objectives

The main objective of this PhD thesis is to *design, develop and evaluate innovative technological contributions to effectively enable rich interactive and collaborative experiences in virtual environments through multi-user holoportation and multimodal HCI technologies*. Three Specific Objectives (SO) have been formulated to fulfill this main objective:

- SO1. Design, implementation, and evaluation of a novel real-time and high-resolution VV pipeline and its integration into an existing Social VR platform ([2] [3]).
- SO2. Design, implementation, integration, and evaluation of in-cloud processing components to enhance the scalability and interoperability of Social VR services.
- SO3. Design, implementation, and evaluation of group-based multi-modal HCI features to enable rich interaction and collaboration in group-shared Social VR experiences.

## 2.2 Hypotheses

The research objectives and associated tasks will revolve around substantiating and validating a set of associated Hypotheses (H).

**H1 – Multi-modal interactions with and within VR environments, using state-of-the-art VR controllers:** (H1.1) provide satisfactory usability; (H1.2) can be an effective means to acquire rich and comprehensive information about virtual environments; (H1.3) contribute to meaningful shared and collaborative experiences; (H1.4) can provide added value in different multi-modal VR environments and use cases (e.g., virtual tourism, culture...).

Social VR applications have been applied to various scenarios, yet key limitation still exist when it comes to group-based interaction and collaboration features [4]. This PhD thesis aims to investigate how to effectively combine diverse content formats and multimodal (e.g., VR controllers, hand gestures...) interaction features to enhance interactive and collaborative experiences, with key emphasis in domains like education and culture. Beyond the technical aspects, appropriate methodologies will be adopted to assess the usability, task load index, effectiveness, and potential benefits of these contributions in at least two distinct Social VR scenarios, potentially comparing them to baseline conditions.

**H2 – Realistic and volumetric users' representation in Social VR contribute to enhanced levels of presence, co-presence, and quality of interaction:** Recent studies have provided initial evidence on realistic user's representation formats contributing to enhanced (co-)presence, social meaning and quality of interaction levels in Social VR experiences [1, 6]. However, existing low-cost VV pipelines still encounter remarkable real-time performance, resolution, and adaptability limitations. This PhD thesis aims to contribute with a novel and lightweight pipeline for VV capture, representation, processing, and delivery, providing higher performance and adaptability compared to state-of-the-art solutions in this domain.

**H3 – The use of network-based media processing contributes to increase the scalability, adaptability, and interoperability in Social VR, while still providing satisfactory user experience levels.** This research work will explore how smart networks, cloud

continuum, and orchestration technologies can contribute to enhancing the scalability, adaptability, and interoperability of Social VR services. Diverse in-cloud (virtualized) media processing functions (e.g. selective stream forwarding, mixing and transcoding) will be devised and integrated to analyze their potential implications and benefits. Interestingly, Edge/cloud transcoding features will leverage intrinsic aspects of interactive 3D VV services, including relative distances, viewpoints, positions, and even user roles.

Tailored objective and subjective evaluation methodologies will be adopted to shed relevant light on, and/or substantiate, each of the formulated hypotheses (H1, H2, H3).

## 3 Related Work

### 3.1 Social VR with holographic comms

Recent works have provided initial empiric evidence on remarkable benefits provided by the integration of realistic user representation formats in Social VR, compared to the use of non-realistic representations (i.e., avatars), and even compared to traditional 2D conferencing services (e.g., [1, 3]). Likewise, the study in [4] highlights the wide applicability of Social VR, the potential benefits that can be provided by the addition of realistic user representations therein, and the limited multi-user interaction possibilities in existing Social VR platforms. This PhD thesis aims to build upon recent contributions to enable multi-user Social VR experiences integrating realistic users' representations via VV pipelines, like Point Clouds, captured in real-time via affordable and off-the-shelf RGB-D sensors (e.g., Azure Kinect) (e.g., [1, 2, 3, 5]). Significant efforts will be devoted at proposing novel and lightweight VV pipelines (focusing mainly on capture, reconstruction and encoding steps) to provide superior performance in terms of visual resolution, real-time performance, and adaptability, when compared to state-of-the-art solutions.

### 3.2 Scalability and Interoperability:

The study in [7] provides an in-depth review of immersive VV services. On the one hand, that study reflects on the stringent bandwidth, processing, and delay requirements in such services. On the other hand, that study highlights that the adoption of network-assisted delivery and processing functions (e.g., transcoding, format conversion, remote rendering) can contribute to optimizing the performance, interoperability, and costs in such domain.

In addition, two recent works have provided relevant evidence on how network-assisted processing functions can enhance the scalability and interoperability in immersive video services. First, the work in [8] proposed an in-cloud Multipoint Control Unit (MCU) for Point Cloud streams and integrated it into the end-to-end holo-portation platform from [3, 4]. Apart from mixing streams, that MCU implements novel optimization strategies, like scene fusion, transcoding, and adaptive delivery, with the goal of providing single optimized streams to each target client. In [8], it is shown that the addition of the MCU, compared to only using a Selective Forwarding Unit (SFU) [4], allows for increasing the

number of users per session, while lowering the processing and bandwidth requirements on the client side, and without significantly increasing the end-to-end delays. Second, the work in [9] proposed a new architecture for eXtended Reality (XR) leveraging in-cloud strategies for remote rendering and cloud-edge synchronization, which showed to provide enhancements in terms of scalability and interaction performance.

This PhD thesis builds upon the insights from [8] and aims at leveraging the potential of smart networks and cloud continuum technologies to increase the scalability and interoperability of Social VR services, by devising novel modular Edge-assisted communication architectures.

### 3.3 Multi-Interaction in Social VR

Gamification serves as a crucial strategy and method for driving behavioral change, engagement, and motivation in video games and VR services, contributing to increased user participation and social interaction while minimizing attrition. This approach is gaining momentum across diverse industries, including medical, training and therapy, among others. Supported by multimodal HCI techniques, users can interact with virtual environments using speech, gestures, and gaze, creating authentic immersive experiences. Gamification and multimodal HCI in VR have the potential to enhance the user performance, motivation, and engagement. However, existing Social VR platforms still provide limited support for multi-modal interaction features, especially when it comes to multi-user interaction [4]. This PhD thesis aims to investigate how to leverage recent advances in the field of multi-modal HCI, which are mostly focused on single-person human-to-environment interaction [10, 11, 12], to enable effective group-shared interactive and collaborative Social VR experiences in relevant use cases, like education and culture.

## 4 Ongoing and Completed Work

This section reports on ongoing and completed work, and some preliminary results, for two of the three technological pillars of the thesis: VV pipelines, and integration of multimodal interaction features in Social VR experiences. These technological contributions have been integrated into the end-to-end Social VR

platform from [2, 3], resulting in an evolved architecture sketched in Figure 1.

### 4.1 VV Pipeline for Holoportation using RGB-D format

State-of-the-art VV pipelines for **holoportation** commonly rely on Time Varying Meshes (TVM) [1] or Point Cloud [2, 3, 5, 8] formats, which still face key challenges in terms of visual resolution, adaptability, and real-time performance. This PhD thesis is exploring how to exploit the potential of affordable off-the-shelf sensors, like Azure Kinect, to strategically process the capture video and depth (i.e., RGB-D) streams to provide VV representations, while leveraging the capabilities of standard 2D video codecs, in terms of GPU processing and parallelization. The achieved results so far allow for processing and delivering holographic representations with over 300K voxels per frame and 30 frames per second (fps), compared to earlier performance limited to 80K voxels and 15fps [2], which results in higher photo-realism (see Figure 2).

### 4.2 Metrics Measurement and Monitoring

The PhD thesis aims to assess the effectiveness, performance and reliability of newly devices technological advancements in multiuser Social VR services. To be able to determine the benefits and limits, a modular sub-system is being developed and integrated to measure, report and visualize, in near real-time, a wide set of cross-layer Quality of Service (QoS), performance, resource usage and activity metrics, from different key components and modules of the end-to-end Social VR platform, like captureurs, encoders, decoders, renderers, and stream processing. Examples of such metrics include: delays, fps, bandwidth, CPU (%) / GPU (%) / RAM (MB) usage, etc.

Such sub-system is being implemented by using Prometheus and Grafana, by adopting a Publish-Update-Transfer (PUT) strategy, and transmission via a PUSH gateway to Prometheus for storage. Grafana dashboards provide a user-friendly interface for real-time querying and visualization of these metrics, enhancing monitoring capabilities (see Figure 3).

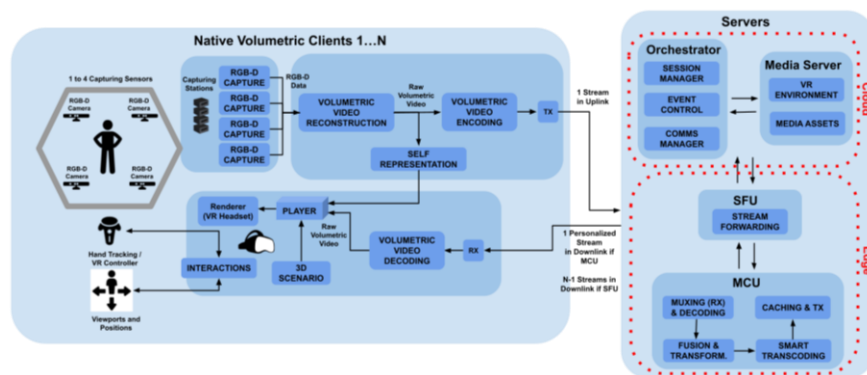


Figure 1: Evolved and Extended System Architecture of the proposed Social VR platform





Figure 2: User Representation using the devised VV pipeline



Figure 3: Example of Grafana dashboard for metrics' monitoring

In next stages, metrics will be also measured and monitored for in-cloud media processing modules to allow for a comprehensive evaluation of Social VR experiences.

### 4.3 Multiuser Social VR Experiences:

The PhD thesis has contributed to the development of two multi-user Social VR games. The first one, **UndergroundVR**, immerses users into an escape room scenario set within the virtual reconstruction of a real museum and city locations. This game requires the collaboration between participants for deciphering hints, escaping, and exploring city locations through virtual underground tunnels. It incorporates diverse multimodal HCI feature, including teleportation, object manipulation, touch interactions, lock deactivation, and content visualization. The second game, **SoulsVR**, follows a similar structure but places a strong emphasis on storytelling. Users embark on a journey through 360° video environments to uncover clues and historical artifacts, culminating in completing a love letter amidst a sea storm. Both games are designed to facilitate cohesive group-shared experiences through multimodal HCI feature and have the capacity to integrate holoportation technology [2,3], thus serving as perfect environment and use case for evaluating the technological contributions from the PhD thesis. In addition, the PhD thesis has

also contributed to a third multi-modal Social VR experience, which allows teleporting couples to a highly realistic reconstruction of a museum for a shared visit and exploration. This VR experience integrates a set of Points of Interest (PoI) and multi-modal interaction features to get rich information about the environment being visited. Interestingly, this experience also allows adding master / follower roles for each of two users to assess the impact of this strategy on the overall user experience and engagement.

Both objective and subjective testing will be conducted for these experiences, aligned with the research objectives and hypotheses of the PhD thesis.

## 5 Conclusions and Future Work

Social VR holds the potential to revolutionize communication and collaboration, offering immersive multiuser experiences in shared virtual environments. This paper has outlined ongoing research objectives to advance the state-of-the-art regarding three main research challenges in the field of Social VR: VV pipelines for increased realism and performance; adaptive and scalable communication architectures; and efficient adoption of multimodal HCI for rich shared experiences. The ongoing PhD thesis addresses specific hypotheses, aligned with these objectives and open research questions. Each of the planned technological contributions will be objectively and subjectively evaluated in diverse Social VR use cases, such as culture, education, and tourism.

Future work will be targeted at completing the associated research, development and evaluation tasks toward a successful completion of the PhD thesis, and creation of scientific and dissemination impact.

## ACKNOWLEDGMENTS

This work is supervised by Prof. Mario Montagud from i2CAT Foundation & University of Valencia (Spain) and Prof. David Rincón from Polytechnic University of Catalonia (Spain). This work received funding from the Generalitat de Catalunya and the European Social Fund (Juan Oro Grants to hire research staff in training FI 2022). This work has been also partially funded by the European Union's Horizon Europe research and innovation programme under the PRESENCE's project, grant agreement n° 101135025.

## REFERENCES

- [1] Montagud, M, et al. Towards socialVR: evaluating a novel technology for watching video together. *Virtual Reality* 26, 1593-1613, 2022
- [2] Fernández, S., Montagud, M., Cernigliaro, G., Rincón, D. Multiparty Holomeetings: Toward a New Era of Low-Cost Volumetric Holographic Meetings in Virtual Reality. *IEEE Access*, vol. 10, pp. 81856-81876, 2022
- [3] Fernández, S., Montagud, M., Cernigliaro, G., Rincón, D. Toward hyper-realistic and interactive social VR experiences in live TV scenarios. *IEEE Transactions on Broadcasting*, 68(1), 13-32. 2022
- [4] Montagud, M., et al. Social VR and multi-party holographic communications Opportunities, Challenges and Impact in the Education and Training Sectors. *arXiv:2210.00330*. 2022:
- [5] Jansen, J., et al. A pipeline for multiparty volumetric video conferencing: transmission of point clouds over low latency DASH. *ACM MMSys'20*. Istanbul (Turkey), 2020.

- [6] Debarba, H. G., Montagud, M., Chagué, S., Herrero, J. G. L., Lacosta, I., Langa, S. F., &Charbonnier, C. (2022). Content format and quality of experience in virtual reality. *Multimedia Tools and Applications*, 1-26.
- [7] J. van der Hooft, H. Amirpour, M. T. Vega, Y. Sanchez, R. Schatz, T. Schierl, and C. Timmerer, "A Tutorial on Immersive Video Delivery: From Omnidirectional Video to Holography," *IEEE Communications Surveys & Tutorials*, 2023
- [8] S. Fernández Langa, M. Montagud, D. Rincón Rivera, J. Moragues, and G. Cernigliaro, "Addressing Scalability for Real-time Multiuser Holoportation: Introducing and Assessing a Multipoint Control Unit (MCU) for Volumetric Video," in *ACM Multimedia (MM'23)*, 2023.
- [9] Y. Huang, H. Wang, X. Qiao, X. Su, Y. Li, S. Dustdar, and P. Zhang, "SCAXR: Empowering Scalable Multi-User Interaction for Heterogenous XR Devices," *IEEE Network*, 2023
- [10] Katona, Jozsef. "A review of human-computer interaction and virtual reality research fields incognitive InfoCommunications." *Applied Sciences* 11.6 (2021): 2646. Conference Name:ACM Woodstock conference
- [11] Kim, Y. M., Rhiu, I., & Yun, M. H. (2020). A systematic review of a virtual reality system from the perspective of user experience. *International Journal of Human-Computer Interaction*, 36(10), 893-910
- [12] Mazeas, A., Duclos, M., Pereira, B., & Chalabaev, A. (2022). Evaluating the effectiveness of gamification on physical activity: systematic review and meta-analysis of randomized controlled trials. *Journal of Medical Internet Research*, 24(1), e26779.