ViewR: Architectural-Scale Multi-User Mixed Reality With Mobile Head-Mounted Displays

Florian Schier ⬤, Daniel Zeidler ⬤, Krishnan Chandran ⬤, Zhongyuan Yu ⬤, and Matthew McGinity ⬤

Abstract—The emergence of mobile head-mounted displays with robust "inside-out" markerless tracking and video-passthrough permits the creation of novel mixed reality (MR) experiences in which architectural spaces of arbitrary size can be transformed into immersive multi-user visualisation arenas. Here we outline ViewR, an open-source framework for rapidly constructing and deploying architectural-scale multi-user MR experiences. ViewR includes tools for rapid alignment of real and virtual worlds, tracking loss detection and recovery, user trajectory visualisation and world state synchronisation between users with persistence across sessions. ViewR also provides control over the blending of the real and the virtual, specification of site-specific blending zones, and video-passthrough avatars, allowing users to see and interact with one another directly. Using ViewR, we explore the transformation of large architectural structures into immersive arenas by creating a range of experiences in various locations, with a particular focus on architectural affordances such as mezzanines, stairs, gangways and elevators. Our tests reveal that ViewR allows for experiences that would not be possible with pure virtual reality, and indicate that, with certain strategies for recovering from tracking errors, it is possible to construct large scale multi-user MR experiences using contemporary consumer virtual reality head-mounted displays.

Index Terms—Visualization systems, mixed / augmented reality, virtual reality, collaborative systems, co-located systems.

I. INTRODUCTION

MIXED or augmented reality (MR, AR) can be defined as any interface that provides a seamless blend of virtual and real elements. This might range from the embedding of a single virtual object in the real world, to the blending of a real element in an otherwise virtual environment, or any point on a continuum between the two extremes [1]. A mixed reality system is said to be "multi-user" when two or more co-located users share a single coherent world. As opposed to multi-user virtual reality (MU-VR), in which users see only virtual representations of one another even if co-located in the same space, multi-user mixed reality (MU-MR) allows co-located users to see one another and their environs directly, thereby facilitating natural communication, interaction and movement.

We describe a mixed reality system as "architectural scale" or "building scale" when viewers are able to freely navigate around spaces much larger than a typical room, such as large halls or atria, long corridors or entire buildings. Mixed reality on such scales opens the door to a wide range of novel experiences, such as visualisation of large structures at a 1:1 scale. In particular, it allows for the affordances [2], [3] of architectural structures to be integrated into the experience, such as rooms and hallways, doors and windows, staircases and mezzanines, elevators and escalators, auditoria and viewing galleries. The combination of mixed reality and unbounded roaming allows for existing architectural spaces to be transformed into immersive multi-user visualisation arenas.

Until recently, development of such building-scale experiences has been hampered both by the need for elaborate tracking systems capable of accurately tracking all users over large areas, and the cost and limitations of mixed reality head-mounted displays available to date. In particular, the restricted field of view of the "optical see-through" waveguide displays in devices such as the *Microsoft Hololens*[1] or *Magic Leap*[2] inhibits the presentation of large virtual structures or immersive environments and can suffer from poor contrast and opacity [4], [5]. It is perhaps for these reasons that the potential of building-scale multi-user mixed reality systems has hitherto proven difficult to realise.

Three recent developments in consumer head-mounted displays (HMD), however, change the equation. The first is untethering, either through on-board computing or wireless streaming from a remote computer, and the addition of robust "inside-out" markerless tracking, which together remove limits on size and shape of the physical arena. The second is the introduction of video-passthrough on mobile head-mounted VR, allowing the user to see a customizable blend of the real and virtual world. Third, the advent of relatively low-cost headsets allows for multi-user experiences with large numbers of users without significant financial investment in devices that tend to be rendered obsolete within relatively short time spans.

Capitalising on these recent developments, we present ViewR, a platform for rapidly creating multi-user mixed reality experiences at architecture-sized scales in any given environment (see Fig. 1). The system is primarily intended as a test bed for

[1] Microsoft Hololens. https://www.microsoft.com/hololens, last retrieved 24.Jun.2023.

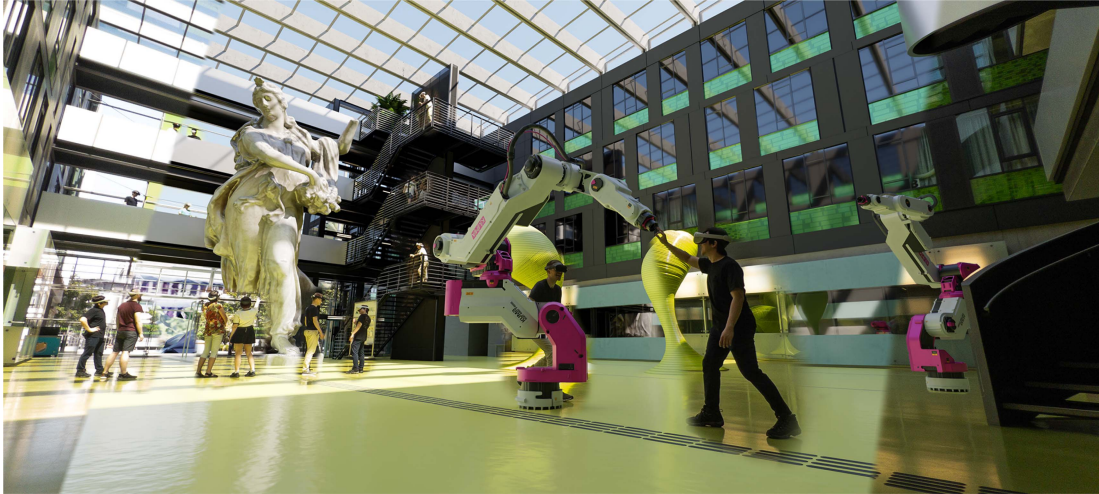[2] Magic Leap. https://www.magicleap.com/, last retrieved 24.Jun.2023.

Fig. 1. Illustration of ViewR, a framework for co-located multi-user mixed reality with mobile head-mounted displays, which allows architectural spaces of arbitrary size to be transformed into immersive multi-user visualisation arenas.

exploring the potential of such experiences. In developing the system, our primary goal is to answer the following questions: Can such experiences be constructed with contemporary mobile VR HMDs, and what are the biggest challenges? What kinds of experiences and applications do they enable? And how might the inclusion of real-world architectural affordances and structures enhance such applications?

The ViewR system provides a number of features of note. It uses highly affordable consumer head-mounted displays and does not require installation of large tracking infrastructure. It provides methods for aligning the real and virtual worlds with and without controllers, tracking-loss detection and recovery, world state synchronisation between users, tools for deployment and administration of multiple sites and sessions and tools for recording and visualising user motion and tracking performance. It allows for co-located users in a single site and the connection of users at remote sites. It supports manipulation of virtual objects both with hand tracking and controllers and persistence of world state across sessions, such that any modifications made to the virtual world in one session will remain in the future. ViewR also allows fine control over the blending of the real and the virtual, both through user control or the specification of blending zones. Most importantly, it provides "full body video-passthrough avatars" that permit users to see each other directly, while respecting occlusion of virtual objects.

What follows here is an outline of key elements of the ViewR system, with a focus on the real-to-virtual alignment strategies. An analysis of the tracking stability is provided and some observations on different strategies for blending the virtual and the real. To explore how ViewR can transform an existing architecture into an immersive arena, the kinds of experiences this might permit and how the architectural affordances of a site may enhance such experiences, we present a range of demonstration applications for a large 40 m × 20 m × 15 m foyer with three levels of mezzanines, stairs, elevators and gangways. We road-test the system by conducting two workshops during which participants were invited to develop a range of experiences for two large architectural sites. In addition, we publicly presented a mixed reality art installation, receiving 180 visitors over 5 days. We analyse the recalibration behaviour of 35 participants, collecting 21.8 km of user motion over 22 hours of use.

II. RELATED WORK

ViewR builds on previous work spanning the fields of virtual and augmented reality, tracking, multi-user systems, computer aided collaboration and interaction design. For the sake of brevity, we restrict this review to co-located multi-user immersive systems and applications, with a focus on HMD systems for more than two users.

With respect to architectural-scale head-mounted MU-MR, relatively few precedents can be found. We are unaware of any previous demonstrations of head-mounted MU-MR at these scales. Shifting our attention to room- or table- scale MU-MR experiences, however, reveals a rich history dating back three decades. The *Shared Space* project [6] and the *Studierstube* system [7] represent early research in this domain (1996, 1998, 2001), both systems capitalising on the *Virtual i-O i-glasses!* consumer optical see-through HMD (1995) that could be used in AR mode or covered with a plastic shield for VR [6]. This line of research continues today, using devices such as the *Microsoft Hololens*,[1] and a wealth of related work can be seen in a recent systematic review of collaborative MR [8], in which the authors identified 259 projects demonstrating MU-MR between 2013–2018. Among their results, they found 25% used head-mounted displays (such as *Microsoft Hololens*,[1] *Magic Leap*[2] or *HTC Vive*[3]), 40% used a co-located collaboration setup, and 10% used "a variable setup with both co-located and remote collaborations." Application areas were found to span architecture, engineering, construction, and operations (14%), education and training (29%), entertainment and gaming (28%), industrial (15%), medicine (8%), and tourism and heritage (6%).

---

[3]HTC Vive. https://www.vive.com/, last retrieved 24.Jun.2023.

Unfortunately, the review did not classify systems according to their scale (e.g. table, room), number of users, or where they landed on the reality-virtuality continuum [1].

The benefits of co-located mixed reality include a shared understanding of spatial relationships and permit natural, direct communication and interaction amongst users, visually, haptically and aurally. Radu et al. [9] propose a list of features important for successful collaboration in co-located multi-user HMD-based AR including a shared environment and shared manipulation of virtual objects, awareness and coordination of others' attention, intentions and activities and awareness of the past. A good example is PoseMMR [10], which demonstrates a MU-MR authoring tool for character animation using video-passthrough on HTC Vive. In the field of immersive analytics, the mechanics and benefits of co-located collaboration in MU-MR have been explored at table and room scales [11], [12], [13], [14]. Similar findings have been observed with co-located hand-held AR [15].

Broadening our review to include head-mounted co-located virtual reality, we find interesting demonstrations, such as CollaboVR [16] a system for collaborative sketching, but with only 4 users and tethered HMD. The most extensive examples of large scale MU-VR are found in commercial VR gaming centres, such as *Hologate*,[4] *The VOID*,[5] *Zero Latency*.[6] Such systems demonstrate very successful large scale MU-VR, but rely on installation of extensive tracking infrastructure and dedicated playing arenas.

A. Architectural-Scale Tracking

Systems for indoor tracking have been implemented with a wide variety of technologies, from Wi-Fi and Low-Energy Bluetooth (BLE) to RFID, ultrasound and various forms of cameras [17]. However, with the exception of visual tracking systems, most cannot yet provide the speed, accuracy and range required for mixed reality [18].

Visual tracking systems are often classified into one of two categories: "outside in," in which an array of stationary cameras look in on moving subjects, or "inside out," in which the camera is attached to the moving subject and views the a surrounding stationary environment. The number of cameras required for "outside in" configurations, such as those provided by *Vicon*, *Optitrak* or *ART*, rises with the size of the tracked arena and the potential for the line of sight between markers and camera to be occluded. A 30 m × 30 m arena might require upwards of 200 cameras mounted in a regular grid not far above the playing arena, with price in the region of $400 K–$500 K.[7] A recent "inside out" alternative is the *Antilatency* optic+IMU system, which potentially reduces cost ten-fold[8] but still requires the installation of LED markers across the arena floor or ceiling. It

can be seen that both classes of tracking systems involve considerable investment and long-term installation of infrastructure that may significantly impact the appearance of a site.

Simultaneous Localization and Mapping (SLAM), in which the 6-DOF pose of a sensor can be determined by incrementally constructing a map of its environment, has been the subject of a great deal of research since its formulation in the 1980s [19]. SLAM methods rely on natural features in the environment for tracking, relieving the need for the placement of explicit markers. Levels of accuracy, robustness and speed necessary for VR and AR arrived with the emergence of visual-inertial SLAM techniques, in which the high frequency low-latency measurements of an IMU are tightly coupled with visual odemetry. See Servières et al. [19] and Barros et al. [20] for comprehensive histories and state-of-the-art of SLAM tracking and Jinyu et al. [21] for a survey and evaluation of visual-inertial SLAM specifically for augmented reality. As a form of markerless inside-out tracking, these systems have arrived at a level of robustness and accuracy suitable for MR and VR [22], freeing us from the finite confines of tracking arenas.

However, when used for MU-MR, contemporary implementations of markerless inside-out tracking present two shortcomings that must be addressed. First, they do not typically offer an absolute measure of position or orientation relative to a known reference frame. As such, when using inside-out tracking for mixed reality, methods for aligning the real and virtual worlds are required. Second, such tracking methods are susceptible to drift or occasional "tracking losses" when used over large areas [21], and so solutions for rapidly re-aligning the virtual and real reference frames may be needed.

When discussing real-to-virtual alignment, we differentiate between *User-to-User* and *User-to-World* alignment methods. Reimer et al. [23] evaluate *User-to-User* alignment using hand-tracking, where all users observe the same pair of hands, reporting errors of the order of ~1 cm for position, but make no mention of rotational accuracy.

For our intended use case, however, *User-to-World* alignment is necessary. This is sometimes achieved using physical fiducial markers, such as ARTag [24] or ArUco-maker [25], as in [23]. Alternatively, markerless approaches to absolute positioning, in which pre-defined spatial patterns are identified in the environment, are possible. Examples of these are *Vuforia* Area targets[9] or *Niantic Lightship's* Visual Positioning System.[10]

At the time of writing however, our target devices support neither marker-based nor markerless absolute positioning, and so an alternative solution must be found.

Existing toolkits for accelerating the development multi-user VR & MR experiences include UBIQ [26], and ARENA [27]. UBIQ is primarily a networking library for the Unity[11] game engine, designed for MU-VR applications and thus neither requires nor considers alignment tools. ARENA, on the other hand, leverages WebXR to support for wide range of devices, with a focus on experiences that unite both co-located and remote users.

---

[4]Hologate. https://www.hologate.com/, last retrieved 24.Jun.2023.

[5]The VOID. https://www.thevoid.com/ last retrieved 24.Jun.2023.

[6]Zero Latency. https://zerolatencyvr.com/, last retrieved 24.Jun.2023.

[7]Price estimate from optitrack.com for 30 m × 30 m which recommends 192 PrimeX-13w or SlimX-13 cameras, retrieved 04.Jun.2022

[8]$50,000 for 10 users in 30 m × 30 m arena, estimated on antilatency.com, retrieved 04.Jun.2022

[9]Vuforia. https://www.vuforia.com/, last retrieved 29.Jun.2023.

[10]Lightship Visual Positioning System. https://lightship.dev/products/vps, last retrieved 29.Jun.2023.

[11]Unity. https://unity.com/, last retrieved 06.Aug.2022.

ARENA provides real-to-virtual alignment through the use of optical fiducial markers, but does not currently offer this feature for video-passthrough HMDs.

III. Design

Upon reviewing previous work and given prior experience in the creation of multi-user VR communication for education (TeachInVR [28]), the following design criteria were devised for the development of the ViewR system:

- **D1.** Multi-user co-located mixed reality. Multiple users perceive and interact with virtual entities and one another in a unified, coherent mixed reality environment.
- **D2.** Indoor locations of arbitrary size and structure. To begin, we restrict our attention to indoor locations as they tend to provide WiFi access, controlled lighting and mostly regular, static geometries.
- **D3.** No installation of infrastructure necessary. Specifically, we seek a system that does not require installation of large-scale tracking systems or any significant modification of the site itself.
- **D4.** Augmented and virtual reality. The system should provide flexible control over the blending of virtual and real elements, allowing a smooth transition from pure virtual to partially virtual views. In particular, it should allow real-time manual or programmable modification of the appearance and real-virtual blending of users, objects and the environment.
- **D5.** Easy session, site and scene management.
- **D6.** Persistence. We are particularly intrigued by the possibilities arising from mixed reality spaces that endure over multiple sessions. That is, the state of the shared virtual world should persist between sessions.
- **D7.** Fast and versatile. The system should allow rapid creation of MU-MR experiences and applications in a wide range of domains.
- **D8.** Affordable and scalable. By affordable, we mean cost-per-user to be no more than the cost of a contemporary consumer-grade VR device. By scalable, we aim for a system that can support an arbitrary number of users, and scales linearly in cost with number of users.
- **D9.** Reliable. Ultimately, the system must be sufficiently robust to support public exhibition.

IV. Implementation

A. Platforms and Toolkits

ViewR was developed initially to support Meta Quest head-mounted VR displays. These are low-cost consumer devices with widespread availability, satisfying *affordability and scalability* (D8). Furthermore, they support video-passthrough allowing users to perceive their surroundings, themselves, and one another, effectively transforming the HMDs into a mixed reality device (D4). Most importantly, these devices are equipped with a robust markerless inside-out tracking system, removing the need for laborious, site-specific, and expensive infrastructure
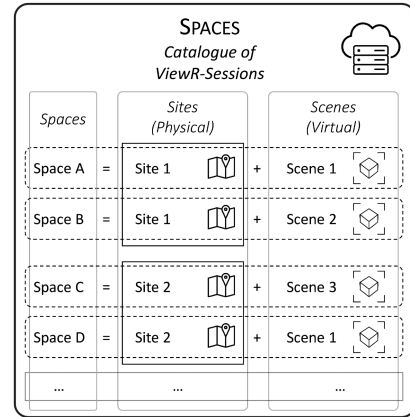

Fig. 2. Concept of *spaces* in ViewR. A *space* consists of a specific physical *site*, in which a virtual *scene* is anchored.

setups (D3, D9, D5). Initial tests suggested they could support tracking over large distances (D2). Lastly, the camera-based hand tracking removes the need for controllers, freeing the hands for real-world interactions and manual communication amongst users.

Unity was selected as software development environment, primarily for its richness of support for VR and, more specifically, the Meta Quest devices. The tools utilized to implement ViewR encompass Normcore[12] for networking and the Oculus Integration SDK[13] for interfacing with the HMD.

B. Session, Site and Scene Management

Using ViewR, physical *sites* can be augmented with virtual content defined as *scenes*, which contain the virtual content the user may interact with. Together, a specific *site* combined with a specific *scene* define a *space* (Fig. 2). Based on this concept, ViewR can be used to manage multiple simultaneous sites, scenes and spaces, even allowing the coexistence of different scenes at the same site.

Scenes within a *site* can be changed during runtime, allowing users to visit various experiences within a single session. Without leaving MR/VR, users can activate scenes, either for themselves or for everyone at the site. We implemented an (optionally) synchronised component, allowing developers to quickly manipulate scene behaviours, scene representations and vary additive loaded Unity-scenes. The latter allowing for easy exchange of high fidelity baked lighting, matching each scenes content. Additionally, these *spaces* may be persistent, allowing users to revisit various spaces, or enabling developers to create truly parallel worlds (D4, D5, D6).

To ease remote control and administration of large numbers of headsets, tools for batch installation, launching and stopping

[12]Normcore. https://normcore.io/, last retrieved 12.Oct.2022.
[13]Oculus Integration SDK. https://developer.oculus.com/downloads/package/unity-integration, last retrieved 12.Oct.2022.

ViewR across multiple headsets were developed using the Android Debug Bridge.[14] Similarly, automated scripts for remotely activating wireless screen-mirroring and recording of the view of any particular headset prove very useful. To make these tools accessible to the community, these have been open-sourced in their respective repository.[15]

C. Virtual Digital Twin

ViewR uses a 3D "digital twin" of the real environment to provide occlusion, shadows, reflections, collisions and other interactions between the real and virtual entities (D4). The required amount of detail and accuracy of this polygonal model depends greatly on both the structure of the physical site and the type of experience that is to be created. For example, a largely empty room may require only a rudimentary box-like approximation of the floors and walls, or for some applications, no model at all (D7). However, when the physical environment offers many opportunities for occlusion of virtual objects or interactions between virtual and physical entities, the geometric accuracy of the digital twin becomes increasingly important. If users are to navigate a complex environment in mostly "virtual reality" mode, then the geometric accuracy and alignment of the digital twin becomes paramount. Failing to provide an accurate virtual model can lead to visual incoherence or, in the worst case, physical harm. To mitigate these risks, ViewR offers a range of safety features (see Section IV-J).

In addition, the digital twin of the real world is used for creating accurate lighting and shadowing of virtual objects (baked or realtime shadows, light probes and reflection maps), so matching the lighting and colour is an effective method for achieving high realism at low performance costs.

D. Real-to-Virtual Alignment

Contemporary implementations of inside-out tracking present two challenges when attempting MU-MR: alignment of the real and virtual, and occasional tracking loss. Here we describe strategies for dealing with both.

Provided with a sufficiently accurate, site- and application-specific digital twin of the real environment, we need to establish sound alignment between this virtual model and the real surroundings. We refer to this as *real-to-virtual alignment*. A consistently accurate alignment method is vital for single-user MR applications, and becomes crucial for coherent MU-MR use cases. Only by achieving results consistently can we ensure all users see the same virtual objects in the same real-world positions. For large arenas specifically, consistently accurate rotational alignment quickly outweighs accuracy in positional alignment.

Importantly, when applying such alignments, ViewR always re-positions the virtual user to have the same relationship with the virtual world as they have with the real world. Thereby, we

maintain a constant virtual-world coordinate system across all users, subsequently simplifying the process of synchronisation.

To align the virtual with the real, we must determine the position and orientation of the user relative to some known frame of reference. This is achieved through the measurement of known "calibration points". However, as the Meta Quest (and other similar HMD) can effectively only "see" controllers and hands, we must use either the hands or controllers as measurement devices. We describe here methods for both, with a focus on speed, ease-of-use and accuracy.

The general alignment strategy is to measure the position of two or more calibration points in the real world and then align these with corresponding points in the virtual world. As the HMD has an accurate internal measure of "up", only the azimuth and position are needed, which can be obtained from measuring the position of just two points grouped into a "2-point calibration station".

With both hands and controllers, there are three sources of error that can lead to misalignment.

1) Placement error E_P: a discrepancy between the position of calibration points in the real and virtual worlds, arising from inaccurate virtual models of the environment (e.g. parallel walls are not truly parallel in the real world), or difficultly in accurately placing calibration points in the real world, which might arise from lack of environmental affordances for accurately placing calibration points. This error can be mitigated by the use of laser scanners during the creation of the digital twin, and laser guides and measures when placing calibration points.
2) Measurement error E_M: error in the measurement of the calibration points using hands or controllers.
3) Tracking drift: accumulative error in the tracking system, in which the user's virtual trajectory increasingly deviates from their real world path. In a small 5 m × 5 m, static well-controlled arena, drift of the order of 0.4 cm and 0.1° is observed [29]. However, over large areas, drift can be much more pronounced.

Controller-Based Calibration: A controller-based calibration station is a physical template onto which two controllers can be accurately placed. Before a session, one or more stations are placed in the physical environment in easily reached locations, and their virtual counterparts placed in identical positions in the virtual world. To calibrate, the user simply places their controllers in the circular slots of the calibration station and holds two buttons for 2 seconds. To further increase convenience and accuracy, we provide a 3D-printable template for calibration stations specifically for the Quest controllers (see Fig. 3). The system can distinguish various stations based on both controller's orientation. Varying the orientation in which each controller fits the 3D-printed station results in a multitude of unique combinations.

Carnevale et al. assess the accuracy of the Quest controllers at different distances from the HMD [30], finding absolute position errors of the order of 1–2 mm at distances of 20–40 cm from the HMD. However, as the controllers are tracked by cameras mounted within the HMD, we anticipate head motion

---

[14]Android SDK Platform Tools. https://developer.android.com/studio/releases/platform-tools, last retrieved 04.Aug.2022

[15]Repository: https://github.com/ixlabTUD/MultiQuest-Management

Fig. 3. 3D printed calibration station for controllers. Modular design allows printing left *(here: purple)* and right *(here: green)* elements with different socket orientations, thereby creating various configurations that facilitate distinctive identification features for each individual station.

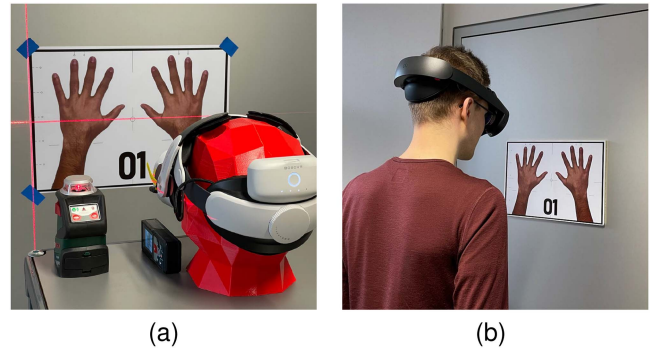(a)   (b)

Fig. 4. Hand-tracking based calibration stations. (a) Experimental setup to identify the most consistent points in our hand based calibration system. (b) User at an alignment station.
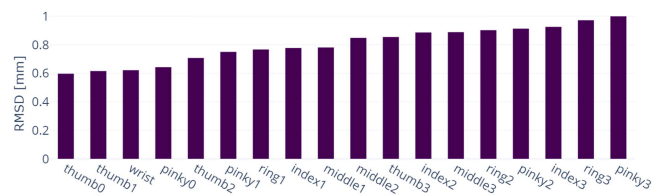

Fig. 5. RMSD for the recorded hand joint positions ($n = 237,637$ per joint). The wrist joints as well as the metacarpophalangeal and interphalangeal joints of the thumb (thumb0 and thumb1) had the least jitter. Refer to the Oculus SDK documentation[13] for further information on the joint naming convention.

to have some influence on measurements. To assess this, we recorded static controller positions with a number of different head motions, ranging from completely stationary to rotating and translating. For stationary test, the headset was mounted on a dummy head. The root mean square deviation (RMSD) of 3D positions was used a measure of precision.

Jitter was minimal when the headset was completely stationary, with RMSD of 1.9 mm. We note a degradation in the precision when the headset was subject to swift rotations in the yaw axis, increasing RMSD to 5 mm. For cases with natural but small head motions however, we observe an average precision of 2 mm.

When deciding on the distance between the two calibration points, we must balance two factors. On one hand, the greater the separation, the less the rotational misalignment arising from measurement error. On the other hand, positional error of the controllers is known to increase with distance from the HMD, so as we increase the separation distance, we also increase measurement error. Based on the data in [30], we estimate the highest rotational accuracy to occur with a separation of 30 cm between controllers, where the z-position error is 1.85 mm.

Hand-Based Calibration: ViewR also provides a controller-free calibration method that exploits the in-built visual hand tracking system by using images of hands as calibration markers. To calibrate, the user need only move their own hands out of sight of the HMD (behind the head or back) and focus on a calibration marker until the process is complete. The user can easily calibrate their device without the need for complicated instructions or additional equipment.

In order the reduce the calibration error, we run a series of tests to determine the areas of the hands with the highest tracking consistency in the Meta Quest 2 hand tracking system.

Fig. 4(a) shows how we arranged the calibration marker in the visual field of the HMD to determine the joints with the highest positional consistency. We performed the data recording at a sampling rate of 90 fps.

Based on our data, we observed that the wrist joint has least amount of jitter alongside the metacarpophalangeal and interphalangeal joints of the thumb (see Fig. 5), considerably less than the fingertips. We decided to use the wrist joints for our calibration marker design.

Similar to controllers, hand tracking error is known to increase with distance from the HMD. Abdlkarim et al. report a difference in positional error between a fingertip in the centre of view

($M = 0.75$ cm, $SD = 0.1$ cm) and a fingertip 60 cm to the left ($M = 1.3$ cm, $SD = 0.1$ cm) [31]. Again, for convenience, we choose a hand separation of 24 cm, allowing two life-size hands to fit on a single A3 sheet. Occasionally, we observe that the tracking system will mistakenly confuse the left and right hands. Tests found that rotating the hands inwards by 15° effectively eliminated such errors.

Calibration is triggered automatically when certain conditions are met for a sufficient time period. Both hands must be visible and have high tracking confidence, stationary with palms facing away from the user and lie within a volume 50 cm × 70 cm × 50 cm in front of the head. When the hands satisfy these constraints for 0.5 seconds, the smoothed positions of the hands are used to perform alignment.

2-Point Calibration Consistency: We examine consistency of calibration results between repeated detection and loss of the hands. Four different conditions were tested: A) baseline - repeated uninterrupted calibrations with a completely static HMD, B) hand tracking deactivated and reactivated with static HMD, C) natural head motion, with repeated head yaw rotations away and back to the hand markers, and D) random walk away and back to the hand markers. In each case, we measure the average shift in position and rotation after subsequent calibrations.

Analysis reveals no significant differences between A, B and C (1-way ANOVA $F(2, 74) = 0.301, p = 0.741$), indicating that the hand-based calibration system tends to converge on the same result after losing and regaining sight of the hands (see Fig. 6). The positional precision is 0.044 cm ($SD = 0.029$ cm)
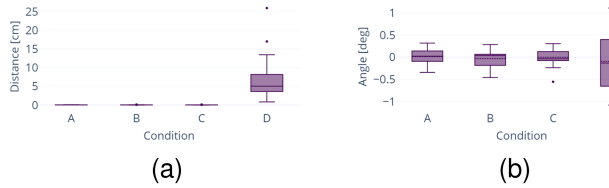
Fig. 6. Consistency of the 2-point calibration system between successive calibration attempts. Condition A: Baseline ($n = 27$), Condition B: Hand tracking reactivated between calibrations ($n = 27$), Condition C: Natural head motion ($n = 23$), Condition D: Random user movement between calibrations ($n = 21$). (a) Posititional correction. (b) Rotational correction. Conditions A, B, and C tend to converge on the same result and condition D indicates the effect of a tracking drift.

Fig. 8. Errors of 2-point, 4-point and conditional algorithm (each $n = 8$). (a) Positional error. (b) Rotational error. Using the conditional 4-point approach resulted in lower positional and rotational errors.
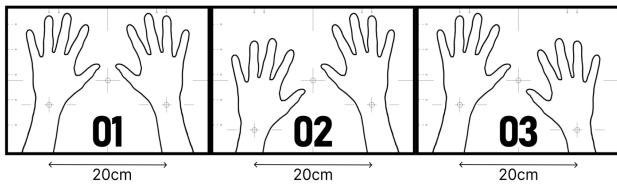

Fig. 7. Calibration stations for hand-based alignment. A set of marker variants that can be distinguished by varying the vertical positions of the depicted hands. Additionally, more stations can be created by varying the horizontal distance between the hands.

and rotational precision is $0.14°$ (SD $= 0.11°$). In condition D, however, we clearly see the effect of tracking drift.

Rotational Accuracy: In summary, our two 2-point calibration methods provide very simple, easy-to-use alignment strategies. The positional accuracy of both methods is sufficient for many use cases, with 2 mm accuracy for controllers and 6 mm for hands. Based on the absolute position errors for controllers and hands given in [30] and [31] and our own noise measurements, we calculate the rotational accuracy for controller-based calibration to be of the order of $0.5° - 0.8°$ degrees, and ~$1.7°$ for hands. The importance of this rotational misalignment grows with the size of the arena. At 5 m from a calibration point, an error of $1°$ will result in an 8.7 cm misalignment, but at 50 m this grows to 87 cm which for most use cases is likely unacceptable.

Multiple Calibration Stations: One strategy for reducing rotational misalignment is to use multiple calibration stations. However, this presents two challenges: uniquely identifying calibration stations and dealing with potential tracking drift as the user moves between stations.

To identify different calibration stations, we vary the horizontal and vertical offset between hands, as shown in Fig. 7. Given the accuracy of the hand tracking, we recommend offset intervals of 2 cm to minimise the risk of mis-identification. Controllers, on the other hand, are distinguished by their orientation (see Fig. 3).

While rotational error can be reduced by increasing distances between calibration points and increasing numbers of points, this is offset by the increased risk of tracking drift. To accommodate this risk, we adopt the following 3-step algorithm:

1) Calculate target position and rotation using 2 points from the current calibration station only.
2) Calculate target rotation using 4 points from current and previous stations, using the Kabsch algorithm [32].
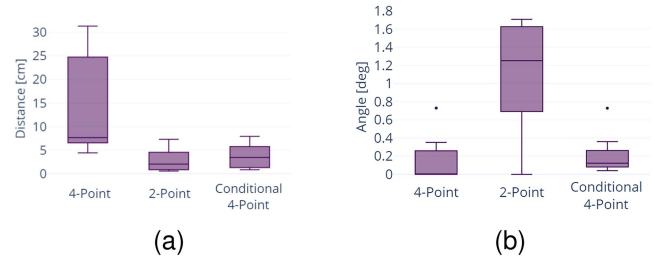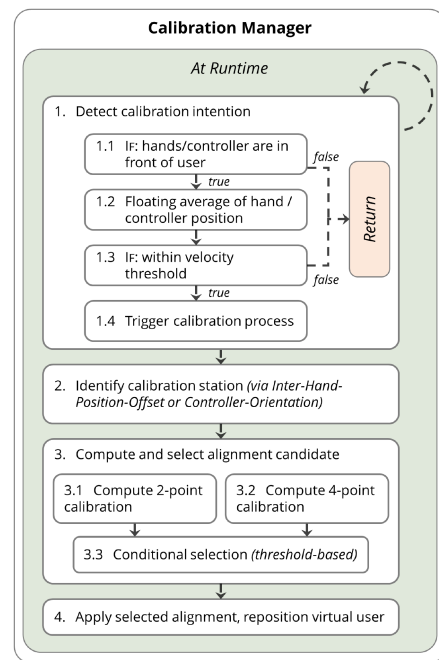

Fig. 9. Implementation of *User-to-World* calibration procedure. Hands must be visible, stationary, have high tracking confidence, lie within a volume of 50 cm × 70 cm × 50 cm in front of the head with palms facing away from the user for a duration of 0.5 seconds. The specific calibration station is identified according to the horizontal and vertical offset between hands. Current and previous station is used for 4-point calibration only if it returns results within the error bounds of the single-station calibration.

3) If results from step 2 are within the bounds of error of the rotation achieved in step 1 ($4°$), then use it; otherwise assume drift has occurred and use the result from step 1.

When no drift occurs, this approach results in considerably higher rotational accuracy, as shown in Fig. 8, while maintaining the same high level of positional accuracy as the 2-point system. Fig. 9 summarizes the resulting calibration system.

E. Tracking Loss Detection and Recovery

When used over large areas, the Quest HMD can suffer from occasional "tracking loss" - sudden large positional jumps that can severely affect the user experience and compromise user safety. Such tracking loss is often preceded by a "tracking freeze", where over a certain number of frames the head position
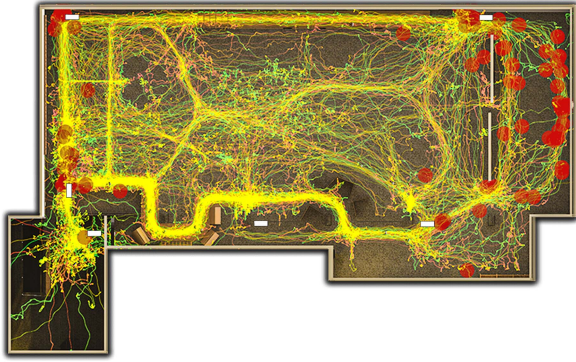
Fig. 10. Screenshot of user trajectory visualisation in ViewR, shown here with the paths of 35 users in a 30 m × 20 m venue. Red circles indicate tracking loss events. White rectangles indicate calibration stations.

Fig. 11. Schematic illustration of selected features associated with the networked avatar. Requires initial information about user's on-site or remote participation. The avatars appearance can be synchronised; for visualization examples see Fig. 12. Control about this variable depends on the desired implementation. Additionally, features such as voice transmission and remote visualization of the input devices are controlled automatically by the system, governed by certain events.

values are constant (at the position before the tracking loss occurred) and the head rotation values are consistently zero. We implement a filter to detect such tracking events, checking for abnormally high instantaneous velocities. When a tracking loss event is detected, full passthrough mode is automatically activated and the user is requested to perform a recalibration. In addition, we have also implemented an experimental automatic recovery system that, after tracking loss is detected, simply relocates the user to the last known "good" pose. Note, however, that this can only be done when no significant tracking freeze has occurred, as the user will have moved during the freeze. This feature is subject to further testing.

In the event of tracking errors or tracking drift, the user is able to recalibrate at any time without leaving the experience through either of methods described above. For example, if a user detects a certain misalignment between the virtual and real, they simply walk to the closest calibration station and perform a re-alignment (D9). ViewR also provides, via a menu, the option to "grab" the virtual world with hands or controllers and move it manually into alignment with the real world.

To help identify areas of the environment with poor tracking performance, ViewR provides tools for recording and visualising trajectories of users and occurrence of calibration and tracking loss events (see Fig. 10).

F. Networking, Synchronisation and Persistence

ViewR employs a distributed networking model to ensure every user is presented with the same virtual world. For networking and world-state synchronisation, ViewR utilizes *Normcore*[13]. *Normcore* employs a central server to maintain a world model, allowing users to join a session at any time whilst ensuring non-diverging spaces. *Normcore* also provides world-state persistence, meaning any changes made by users will endure between sessions. This powerful feature allows the creation of experiences that persist in both space and time, and facilitating synchronous and asynchronous collaboration, similar to the approach of Guo et al. [15].

ViewR does not automatically distribute the state of every element in the virtual scene, rather, when creating experiences
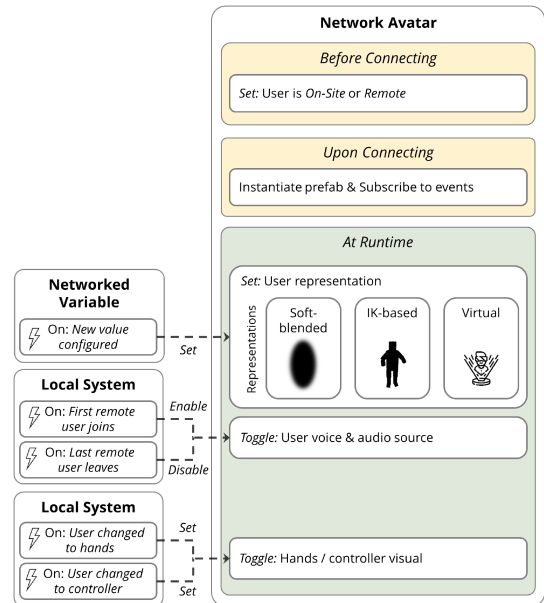
with ViewR, developers must decide which virtual entities are to have their state synchronised or not. As we only broadcast state changes, transmission data-rates are dependent on the number of state changes per frame.

Voice: ViewR also allows voice transmission. While this functionality remains deactivated for sessions comprised solely of on-site users, the system will enable it automatically should any user connect remotely (see Fig. 11). This will activate the microphones on all users, remote and on-site, facilitating a channel communication between them. To prevent potential sound artefacts, however, on-site devices will only process and output audio from remote users, as they can naturally hear other on-site users. In order to maintain transparency, the user is notified, should the system enable or disable their microphone input.

G. Interaction and Control

ViewR supports both hand and controller based input, and provides simple object manipulation such as direct and remote grabbing, releasing, rotation and translation and (two-handed) scaling of objects. Such manipulations as well as the input modality are synchronised with other users (see Fig. 11).

At any time, the user can summon a control menu with either a press of a button or a left-hand pinch gesture. In accordance with common VR best practices, motion such as the movements user interfaces are smoothly animated, avoiding incomprehensible or sudden changes in the users field of view. Additionally, the user has control over many of such scene behaviours and can configure them to their preferences.
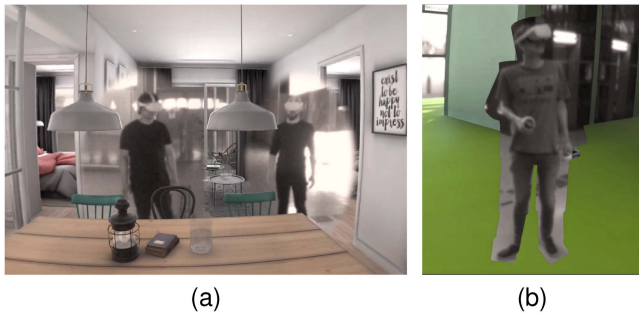
Fig. 12. Visual appearance of ViewR's passthrough avatars. (a) Soft-blended avatar. (b) IK based avatar.

H. User Representation

The user representation defines how the user is visually displayed to other users in the same space. In many VR applications, this usually implies the representation using a virtual avatar.

With ViewR, we use video-passthrough to present other users directly. Users are able to select their preferred user representation of other users, as illustrated in Fig. 11. At the time of writing, available options are

- a fully virtual avatar comprising a simple head and hands,
- a video-passthrough avatar, represented by an edge-blended video-cut-out where other users are (see Fig. 12(a)), or
- an IK video-passthrough avatar, in which an articulated full-body silhouette of the user is estimated using inverse kinematics[16] (see Fig. 12(b)).

Both passthrough options allow direct view of co-located users (D4), while for remote users only the VR avatar is available. Most importantly, both *video* and *virtual* avatars are 3D objects, allowing users to occlude virtual objects, and vice versa. In addition, when hand-tracking is active, for all 3 avatar types, we enable video-pass through on a 3D proxy of the users hands, allowing for interesting manual interactions between users. Being a sandbox environment, ViewR allows the developers to easily exchange and experiment with the various user representations (D7).

I. Remote and Co-Located Use

Following the core idea of mixed reality, an experience in ViewR that is shared with other co-located users allows the user to see others directly (see Fig. 12). However, ViewR also allows users to collaborate remotely. While the focus of most features lies on co-located use (D1), it can also facilitate entirely remote shared experiences.

J. Blending Modes and Blending Zones

Similar to Gruenefeld et al. [33], ViewR allows gradual transition from AR to VR.

Blending Modes: To allow users' interactive control over the degree of virtual or real elements visible in the scene,

---

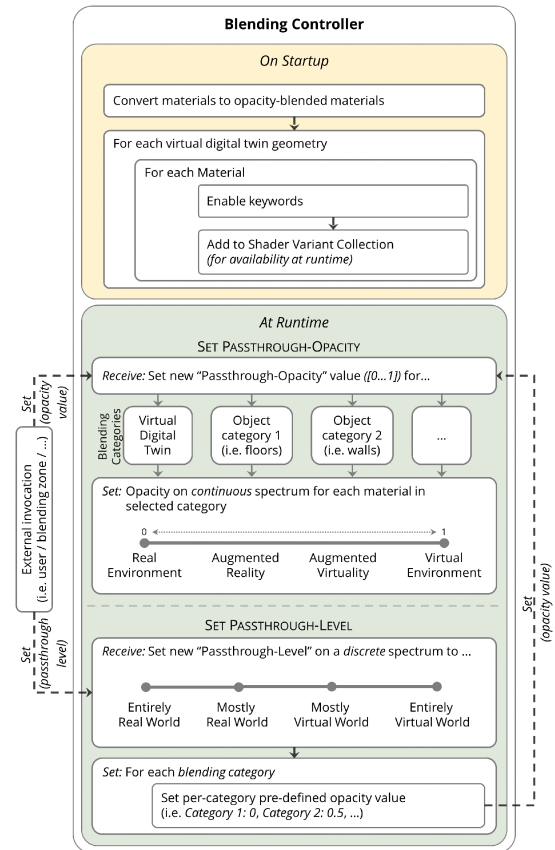[16]Rootmotion *FinalIK*. http://root-motion.com/, last retrieved 04.Jun.2022.

Fig. 13. Illustration of the passthrough-blending. Initializes by converting relevant materials such that blending is possible. Depending on the implementation, developers can choose to design different procedures within their application to control the Passthrough-Opacity and the Passthrough-Level. The first provides a *continuous* control along the reality-virtuality continuum [1], selectively applicable to any chosen, configurable category. The latter allows for the configuration of Passthorugh-Levels across the entire space, setting each category of objects to its pre-defined opacity value for the respective level.

we provide a simple system for controlling the visibility and appearance of virtual and real elements, as illustrated in Fig. 13. In Unity, virtual entities are grouped into categories whose visibility and appearance can then be controlled interactively during the experience. For example, entities might be placed in categories "Walls", "Floor", "Sculptures" or any such collection. Each category is then given a "visibility opacity". During the experience, all users are able to set a visibility level, similar to the concept of the AR-VR continuum, controlling which categories of objects are visible.

This can be controlled individually or overridden centrally programmatically if a homogeneous experience is desired. By default, all settings available to the user are synchronised across clients. However, users may choose to decouple and use local settings instead, and to return to the synced settings at a later point.

The mix of real and virtual can be controlled in two ways. "Passthrough-Levels" is a simple method for showing or hiding different categories of objects. "Passthrough-Opacity", on the other hand, allows fine continuous control over the blending of virtual and real layers (i.e. percentage of opacity of video on the floor). For example, setting the background environment to
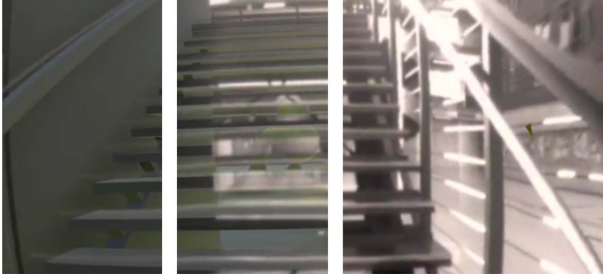
Fig. 14. Passthrough blending zone automatically actives full passthrough as the user enters the staircase. Left to right: Virtual, blended, and pure passthrough state on the relevant geometry.
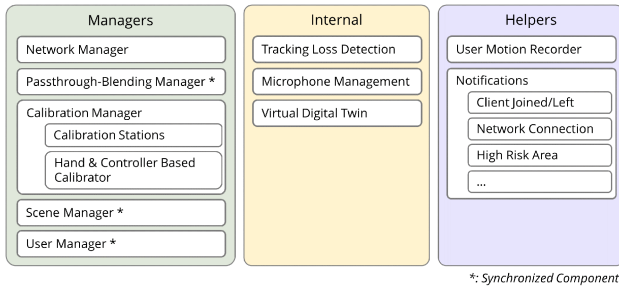

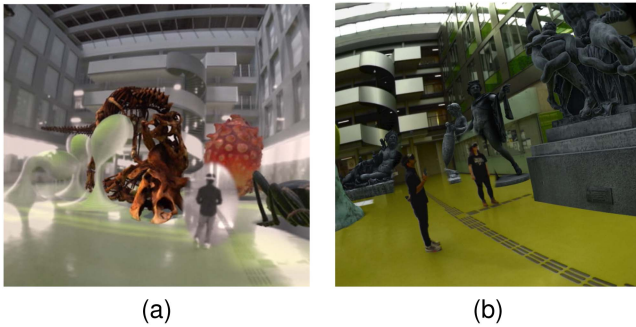Fig. 15. High-level overview of selected ViewR features.

Fig. 16. Exhibition curated with models for a museum. Visitors can be given various permissions and may be guided through the space of the museum by a guide. Selective blending allows real-world reflections, shadows and lighting to enhance the virtual worlds fidelity. (a) Natural History Museum *(Meta Quest 2, 50% passthrough on digital twin, 100% opacity on scene content)*. (b) Sculpture Gallery *(Meta Quest Pro, 100% passthrough on digital twin, 100% opacity on scene content)*.

50% passthrough, but all the foreground entities to 100% opaque proves to be a very effective configuration (see Fig. 16(a)).

Blending Zones: ViewR also allows the specification of blending zones, fixed volumes or surfaces in the world that define particular passthrough configurations. For example, a "passthrough zone" can be defined around staircases, automatically transitioning to full passthrough as the user enters a staircase (see Fig. 14).

Passthrough Surfaces: ViewR also allows parts of the underlying Digital Twin to be used as passthrough surfaces. These will display the passthrough, regardless of the current configuration on the AR-VR spectrum. This has proven helpful in areas of increased traffic and uncertainty, such as elevators.

General Areas of Increased Risk: Naturally, the real world provides certain components that are prone to increased risk. Two such examples are staircases, which require a guaranteed alignment precision, and moving objects such as swinging doors. Both present a risk of harm. As described, ViewR allows the definition of areas and zones in which the system automatically transitions to passthrough-mode (see Fig. 14). Additionally, ViewR provides a notification system, which allows developers to send messages and warnings.

K. Summary of Main Modules

A selection of features is summarized in Fig. 15. By design, these modules are largely isolated from one another.

V. APPLICATION AND EVALUATION

In order to evaluate the system and explore the potential of architectural-scale MU-MR, a range of test scenarios were developed for three different sites. The sites were chosen for their size and architectural affordances: Site A an atrium measuring $40\,\mathrm{m} \times 20\,\mathrm{m}$ and $15\,\mathrm{m}$ high, with three levels of balconies, stairways, elevators and gangways; Site B a foyer of a concert hall, measuring $76\,\mathrm{m} \times 20.5\,\mathrm{m} \times 6\,\mathrm{m}$, with stairs and mezzanines, and Site C an exhibition hall, measuring $30\,\mathrm{m} \times 20\,\mathrm{m} \times 5\,\mathrm{m}$ (see Fig. 19). Digital twins for the locations were created by scanning the locations with LIDAR scanner (Leica BLK360), and then, using the resulting point cloud scans as a guide, low-polygon models of the sites were modelled by hand. Note that the digital twin of the site need only be rudimentary, but must be geometrically accurate.

A. Application Prototypes

Virtual Museums: Two virtual museum scenarios were developed (see Fig. 16). The first depicts a collection of sculptures, in which life-size or larger-than-life-size 3D scans of sculptural forms are presented. The second demonstrates a natural history museum, in which 3D scans of insects, animals and fossils are presented. We demonstrate the usability of the scenario by performing "guided tours" and "immersive lectures", in which a guide or teacher leads a group of students on an educational tour of the exhibit. In both scenarios, images and text presented on interactable "AR slides" provide contextual information, in a form that can be freely moved and scaled, allowing the teacher or guide to easily perform an immersive slide-show. Additionally, as the state of the virtual world in ViewR persists between sessions, the teacher can enter the virtual space beforehand to prepare the content.

Architectural Visualisation: An obvious benefit of architectural-scale MU-MR is the ability to present entire buildings at 1:1 scales. We demonstrate this by simulating an architectural walk-through of a full-scale apartment (see Fig. 12(a)).

Factory Digital Twin and Human-in-the-loop Robotics: To demonstrate the potential of MU-MR in a dynamic scene, we simulate a robotic production line, complete with interactive industrial robots (see Fig. 17). Users can manipulate a robotic
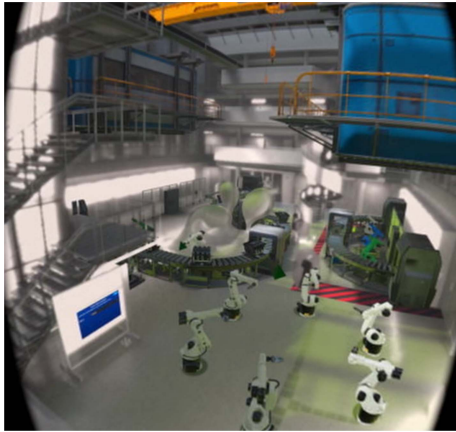
Fig. 17. Dynamic full-scale simulation of a robotic factory floor. This demonstrates that ViewR can be used to prototype human–robot interaction simulations. Here we see the viewpoint of a user on an upper gangway.

Fig. 18. User stepping on a physical wooden plank, emphasizing the perceived height. Left: User 1. Right: User 2. Center: Real-world view.

arm or interact with mobile robots as they navigate the space. Here, our goal is to demonstrate how autonomous machines designed for operation near and with humans might be tested safely in a mixed reality setting. A second use case is the training of employees to manage and control such machines.

Large Vehicle Training Simulation: In this scenario, we demonstrate how the system can emulate the real-time operation of a large crane at 1:1 scale. While a trainee remote controls a real-time simulation of crane, other immersed users guide them in controlling the vehicle.

Hybrid MU-MR Lessons: Since the start of the COVID-19 pandemic, lectures were moved online in many countries. With ViewR, we propose a tool for exploratory creation and evaluation of new teaching concepts, and support on-site, hybrid and remote sessions. With an early study prototype we currently explore the potential benefits of spatial association of information during lectures, which is inspired by earlier work assessing the Method of Loci in VR by Krokos et al. [34].

Multi-User Therapy: In the context of another project, we begin using ViewR to develop MU-MR anxiety therapies, in which patient and therapist join together in the virtual world. The benefits of MU-MR here allow therapy to remain a communicative relationship between patient and therapist. Fig. 18 shows users sharing a virtual heights scenario.
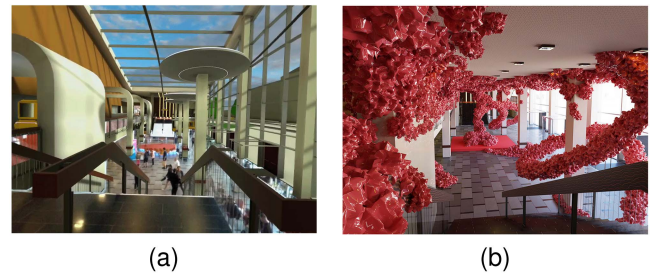

Fig. 19. Prototypes from the student workshops. (a) Transformation of a concert hall foyer into a cable-car station for autonomous pod vehicles. (b) MR experience for the same foyer with transformation of music into 3D structures.

B. Developer Workshops

To evaluate the design of the ViewR as a developer tool, we conducted three student workshops in which altogether 30 students used the system to develop MU-MR experiences. These workshops demanded a great deal from the ViewR platform, in terms of both ease of development and run-time performance, and proved invaluable in the development and design of ViewR.

The first workshop, run intensively over 2 weeks, had computer science students develop a mixed-reality experience for the foyer of a concert hall. Here, the focus was the real-time transformation of sound and music into 3D structures.

The second workshop, run over 5 months, invited computer science and design students to use the same concert hall foyer to communicate concepts on future mobility. The resulting prototype transformed the foyer into a cable-car station for autonomous pod vehicles, see Fig. 19(a).

Workshop participants found ViewR useful for multi-user MR experiences. In a follow-up survey ($n = 6$), participants were asked to reflect on their experience. The majority of the participants had limited experience with Unity before attending the workshop. Despite this, they did not face severe challenges directly related to ViewR while developing. All reported that the amount of time required to develop their experience from scratch would have been considerably higher without ViewR, with answers ranging from "at least 200 h" to "around 2 months" of additional time required. Participants also suggested the need for a local network solutions due to site specific connectivity issues, "additional [beginner-friendly] documentation", and a "[more] bare-bones template". One participant of the first workshop also asked for system notifications "when the user has to recalibrate", which, by now, has partly been addressed by the tracking loss recovery.

The third workshop, run over 2 months, united computer science with architecture students to create hybrid physical/virtual sculptures for a 30 m × 20 m exhibition hall.

The results of this workshop were publicly exhibited as *Hybrid Spaces*, receiving over 180 visitors over 5 days. Architectural installations including wooden walkways and large modular inflatable structures provided passive haptics, which were used as the basis for six virtual worlds, each providing a different visual interpretation and augmentation of the physical

TABLE I
SUMMARY STATISTICS OF USER MOVEMENT DATA DURING THE EXHITION
(N = 35)

	min	*max*	*median*	*mean*
Tracking Losses (count)	0	8	2	2.2
Recalibrations (count)	2	34	15	16.08
Time Spent (mins)	18.63	77.22	35.44	38.38
Distance Travelled (m)	252.39	1561.18	611.11	623.76

structures. A virtual elevator transported visitors between these different worlds.

On the final day of the exhibition, we recorded the movement data of 35 users to gain insight into calibration behavior and tracking loss. User trajectory data was recorded for a combined total of 21,832 meters over 1,343 minutes and is depicted in the Fig. 10. The figure also shows location of tracking loss events, revealing problematic areas in the site. Table I provides a summary of the statistics collected during the exhibition, including the number of visitors, distance traveled, and the time spent in the multi-user MR experience.

We also collected feedback from the visitors, in the form of a questionnaire ($n = 35$). Overall, the feedback was positive and highlighted that they enjoyed the experience. 66% of the users reported that the calibration system was easy to use. Some users (45%) noted that they were occasionally worried about colliding with physical obstacles, as most experiences were primarily virtual, occluding much of the real world. However, more than 60% of users reported that they became increasingly confident with moving through the experience over time.

VI. DISCUSSION

A. Mixed Reality

Development of the ViewR system began with the arrival of the first generation of mobile HMD with video-passthrough capabilities. Early experiments found the addition of video-passthrough, even if monochrome and low resolution, to be transformative, allowing for experiences not possible in pure VR. Being able to see their physical surroundings, we observe in users a far greater tendency for confident, natural exploration of large spaces without fear of colliding with the real world. Further, being able to see fellow users allowed natural social interactions to emerge among users. Users tended to explore spaces together, with a heightened awareness of being in a shared virtual world.

In designing and testing the different scenarios described above, we observe one universal rule concerning the blending of the real and virtual. For anxiety free motion, the user must always be able to see the real floor. In many cases, we find showing the real world ground in an otherwise entirely virtual world is sufficient for anxiety free navigation.

The gradual blending modes on selective scene geometry described in Section IV-J have proven to be a sound method to extend the colourless passthrough with colours. An unexpected discovery was an increase in perceived realism gained by blending the video-passthrough with the simple colour virtual model of the environment. The monochrome passthrough provided real-world reflections, shadows and lighting while the virtual model provided colour and sharpness (see Fig. 16(a)).

B. Architectural Affordances

All the prototypes and experiments described above demonstrate the potential of transforming architectural sites into visualisation arenas and leveraging the architectural affordances provided by the site. With building-scale MR, the integration of balconies and mezzanines, walkways and passages, staircases and even elevators provide new ways of navigating virtual content. The success of the elevators, in which tracking was maintained due to a large glass window, was a particularly surprising outcome.

C. Limitations

To date, ViewR has only been used with Meta Quest devices, which are only able to track the pose of users' heads, hands and controllers. As such, it is blind to the physical motion of objects in the real world.

The inside-out tracking proves surprisingly robust even when used over much larger scales than recommended, but is nonetheless still susceptible to drift and occasional tracking loss. As such, methods for rapid re-alignment remain necessary.

D. Future Work

We are currently evaluating the use of the recently released "shared spatial anchors" in the Oculus SDK. If sufficiently precise and robust, spatial anchors may remove the need for physical calibration markers. If spatial anchors could be programmatically distributed across a large site at locations likely to provide uniquely identifiable optical features (using the digital twin as a guide), they may be beneficial for both tracking loss recovery and real-world alignment.

We are also currently evaluating different strategies for interactively or automatically altering the blending between the real and virtual in response to the situation or user's desires.

VII. CONCLUSION

We present here an open-source framework for rapidly constructing and deploying multi-user MR experiences. ViewR includes tools for controlling blending of virtual and real elements and video-passthrough avatars, rapid alignment of real and virtual worlds, tracking loss detection and recovery, user trajectory visualisation and world state synchronisation between users with session persistence. Our tests reveal that ViewR allows for experiences that would not be possible with pure virtual reality, and indicate that it is, under certain conditions, possible to construct architectural-scale multi-user mixed reality experiences using contemporary consumer virtual reality head-mounted displays. However, tracking loss and drift must be expected, and so methods for easy and rapid re-alignment remain necessary. ViewR is currently being used to develop applications in scientific and architectural visualisation, education, artistic and psychotherapy domains.

REFERENCES

[1] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino, "Augmented reality: A class of displays on the reality-virtuality continuum," in *Telemanipulator and Telepresence Technologies*, vol. 2351. Bellingham, WA, USA: SPIE, 1995, pp. 282–292. [Online]. Available: https://doi.org/10.1117/12.197321

[2] J. J. Gibson, *The Ecological Approach to Visual Perception*. Boston, MA, USA: Houghton Mifflin, 1979.

[3] J. R. A. Maier, G. M. Fadel, and D. G. Battisto, "An affordance-based approach to architectural theory, design, and practice," *Des. Stud.*, vol. 30, no. 4, pp. 393–414, 2009. [Online]. Available: https://doi.org/10.1016/j.destud.2009.01.002

[4] D. C. Rompapas, "Designing for large scale high fidelity collaborative augmented reality experiences," Ph.D. dissertation, 2019.

[5] J. Miller, M. Hoover, and E. Winer, "Mitigation of the Microsoft Hololens' hardware limitations for a controlled product assembly process," *Int. J. Adv. Manuf. Technol.*, vol. 109, no. 5, pp. 1741–1754, 2020. [Online]. Available: https://doi.org/10.1007/s00170-020-05768-y

[6] M. Billinghurst, S. J. Weghorst, and T. A. Furness, "Shared space: An augmented reality approach for computer supported collaborative work," *Virtual Reality*, vol. 3, pp. 25–36, 1998. [Online]. Available: https://doi.org/10.1007/BF01409795

[7] Z. Szalavári, D. Schmalstieg, A. Fuhrmann, and M. Gervautz, ""Studierstube": An environment for collaboration in augmented reality," *Virtual Reality*, vol. 3, no. 1, pp. 37–48, 1998. [Online]. Available: https://doi.org/10.1007/BF01409796

[8] R. A. J. de Belen, H. Nguyen, D. Filonik, D. D. Favero, and T. Bednarz, "A systematic review of the current state of collaborative mixed reality technologies: 2013–2018," *AIMS Electron. Elect. Eng.*, vol. 3, no. 2, pp. 181–223, 2019. [Online]. Available: https://doi.org/10.3934/ElectrEng.2019.2.181

[9] I. Radu, T. Joy, Y. Bowman, I. Bott, and B. Schneider, "A survey of needs and features for augmented reality collaborations in collocated spaces," in *Proc. ACM Hum.-Comput. Interact.*, vol. 5, no. CSCW1, pp. 1–21, Apr. 2021. [Online]. Available: https://doi.org/10.1145/3449243

[10] Y. Pan and K. Mitchell, "PoseMMR: A collaborative mixed reality authoring tool for character animation," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces Abstr. Workshops*, 2020, pp. 758–759. [Online]. Available: https://doi.org/10.1109/VRW50115.2020.00230

[11] M. Cavallo, M. Dholakia, M. Havlena, K. Ocheltree, and M. Podlaseck, "Dataspace: A reconfigurable hybrid reality environment for collaborative information analysis," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces*, 2019, pp. 145–153. [Online]. Available: https://doi.org/10.1109/VR.2019.8797733

[12] T. Mahmood, W. Fulmer, N. Mungoli, J. Huang, and A. Lu, "Improving information sharing and collaborative analysis for remote GeoSpatial visualization using mixed reality," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2019, pp. 236–247. [Online]. Available: https://doi.org/10.1109/ISMAR.2019.00021

[13] B. Lee, X. Hu, M. Cordeil, A. Prouzeau, B. Jenny, and T. Dwyer, "Shared surfaces and spaces: Collaborative data visualisation in a co-located immersive environment," *IEEE Trans. Vis. Comput. Graph.*, vol. 27, no. 2, pp. 1171–1181, Feb. 2021. [Online]. Available: https://doi.org/10.1109/TVCG.2020.3030450

[14] W. Luo, A. Lehmann, H. Widengren, and R. Dachselt, "Where should we put it? Layout and placement strategies of documents in augmented reality for collaborative sensemaking," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2022, pp. 1–16. [Online]. Available: https://doi.org/10.1145/3491102.3501946

[15] A. Guo, I. Canberk, H. Murphy, A. Monroy-Hernández, and R. Vaish, "Blocks: Collaborative and persistent augmented reality experiences," in *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 3, no. 3, pp. 1–24, Sep. 2019. [Online]. Available: https://doi.org/10.1145/3351241

[16] Z. He, R. Du, and K. Perlin, "CollaboVR: A reconfigurable framework for creative collaboration in virtual reality," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2020, pp. 542–554. [Online]. Available: https://doi.org/10.1109/ISMAR50242.2020.00082

[17] W. He, M. Xi, H. Gardner, B. Swift, and M. Adcock, "Spatial anchor based indoor asset tracking," in *Proc. IEEE Virtual Reality 3D User Interfaces*, 2021, pp. 255–259. [Online]. Available: https://doi.org/10.1109/VR50410.2021.00047

[18] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Commun. Surveys Tut.*, vol. 21, no. 3, pp. 2568–2599, Third Quarter 2019. [Online]. Available: https://doi.org/10.1109/COMST.2019.2911558

[19] M. Servières, V. Renaudin, A. Dupuis, and N. Antigny, "Visual and visual-inertial SLAM: State of the art, classification, and experimental benchmarking," *J. Sensors*, vol. 2021, pp. 1–26, 2021. [Online]. Available: https://doi.org/10.1155/2021/2054828

[20] A. M. Barros, M. Michel, Y. Moline, G. Corre, and F. Carrel, "A comprehensive survey of visual SLAM algorithms," *Robot.*, vol. 11, no. 1, 2022, Art. no. 24. [Online]. Available: https://doi.org/10.3390/robotics11010024

[21] L. Jinyu, Y. Bangbang, C. Danpeng, W. Nan, Z. Guofeng, and B. Hujun, "Survey and evaluation of monocular visual-inertial slam algorithms for augmented reality," *Virtual Reality Intell. Hardware*, vol. 1, no. 4, pp. 386–410, 2019. [Online]. Available: https://doi.org/10.1016/j.vrih.2019.07.002

[22] V. Holzwarth, J. Gisler, C. Hirt, and A. Kunz, "Comparing the accuracy and precision of steamVR tracking 2.0 and Oculus quest 2 in a room scale setup," in *Proc. ACM 5th Int. Conf. Virtual Augmented Reality Simul.*, ser. ICVARS 2021. New York, NY, USA, 2021, pp. 42–46. [Online]. Available: https://doi.org/10.1145/3463914.3463921

[23] D. Reimer, I. Podkosova, D. Scherzer, and H. Kaufmann, "Colocation for SLAM-tracked VR headsets with hand tracking," *Computers*, vol. 10, no. 5, 2021, Art. no. 58. [Online]. Available: https://doi.org/10.3390/computers10050058

[24] M. Fiala, "Designing highly reliable fiducial markers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 7, pp. 1317–1324, Jul. 2010. [Online]. Available: https://doi.org/10.1109/TPAMI.2009.146

[25] S. Garrido-Jurado, R. Muñoz-Salinas, F. Madrid-Cuevas, and M. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognit.*, vol. 47, no. 6, pp. 2280–2292, 2014. [Online]. Available: https://doi.org/10.1016/j.patcog.2014.01.005

[26] S. J. Friston et al., "Ubiq: A system to build flexible social virtual reality experiences," in *Proc. 27th ACM Symp. Virtual Reality Softw. Technol.*, 2021, pp. 1–11. [Online]. Available: https://doi.org/10.1145/3489849.3489871

[27] N. Pereira, A. Rowe, M. W. Farb, I. Liang, E. Lu, and E. Riebling, "ARENA: The augmented reality edge networking architecture," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2021, pp. 479–488. [Online]. Available: https://doi.org/10.1109/ISMAR52148.2021.00065

[28] F. Schier, K. Chandran, and M. McGinity, "TeachInVR: A virtual reality classroom for remote education," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces Abstr. Workshops*, 2022, pp. 283–286. [Online]. Available: https://doi.org/10.1109/VRW55335.2022.00064

[29] G. Mulvany, S. Taylor, and S. Greuter, "Target: Zero drift: Minimizing spatial drift in VR optical tracking systems such as Oculus insight," in *Proc. ACM Annu. Symp. Comput.-Hum. Interaction Play*, ser. CHI PLAY '20. New York, NY, USA, 2020, pp. 321–325. [Online]. Available: https://doi.org/10.1145/3383668.3419875

[30] A. Carnevale et al., "Virtual reality for shoulder rehabilitation: Accuracy evaluation of Oculus quest 2," *Sensors*, vol. 22, no. 15, 2022, Art. no. 5511. [Online]. Available: https://doi.org/10.3390/s22155511

[31] D. Abdlkarim et al., "A methodological framework to assess the accuracy of virtual reality hand-tracking systems: A case study with the Meta Quest 2," *Behav. Res. Methods*, pp. 1–12, 2023. [Online]. Available: https://doi.org/10.3758/s13428-022-02051-8

[32] W. Kabsch, "A solution for the best rotation to relate two sets of vectors," *Acta Crystallographica Sect. A, Cryst. Phys., Diffraction, Theor. Gen. Crystallogr.*, vol. 32, no. 5, pp. 922–923, 1976. [Online]. Available: https://doi.org/10.1107/S0567739476001873

[33] U. Gruenefeld et al., "VRception: Rapid prototyping of cross-reality systems in virtual reality," in *Proc. ACM CHI Conf. Hum. Factors Comput. Syst.*, 2022, pp. 1–15. [Online]. Available: https://doi.org/10.1145/3491102.3501821

[34] E. Krokos, C. Plaisant, and A. Varshney, "Virtual memory palaces: Immersion aids recall," *Virtual Reality*, vol. 23, no. 1, pp. 1–15, 2018. [Online]. Available: https://doi.org/10.1007/s10055-018-0346-3

[17]Sketchfab. https://sketchfab.com/, last retrieved 12.Oct.2022.

Florian Schier received the master's degree in computational modeling and simulation from the Technische Universität Dresden, Germany. He is a research associate with Technische Universität Dresden, Germany, where he is currently working toward the PhD degree with the Immersive Experience Lab. He is currently exploring co-located collaborative multi-user MR interfaces, interactions, and various potential applications thereof.

Zhongyuan Yu is currently working toward the PhD degree with Immersive Experience Lab, School of Computer Science, Technische Universität Dresden, Germany. His research interests include visualization techniques and advanced user interfaces in virtual/augmented reality.

Daniel Zeidler is currently working as a research assistant with the Immersive Experience Lab, School of Computer Science, Technische Universität Dresden, Germany. His research interests include full-body avatars and multi user experiences in virtual/augmented reality.

Matthew McGinity received the PhD degree from the University of New South Wales in 2014. He is junior professor for Immersive Media Design with TU Dresden, where he heads the Immersive Experience Lab. Prior to joining TU Dresden, he developed immersive media systems in artistic, industrial and academic domains, ranging from immersive artworks to astronaut training simulations at the European Space Agency.

Krishnan Chandran received the master's degree in intelligent adaptive systems, from the University of Hamburg, Germany in 2019. He is a research associate with the chair of Immersive Media, the Technische Universität Dresden, Germany. Previously, he was a student researcher with INRIA, Paris and German Aerospace Center, Braunschweig. He is currently exploring the pedagogical implications of Multi-User Computer Mediated Realities.