2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)

2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct) | 978-1-6654-1298-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISMAR-Adjunct54149.2021.00103

Hybrid Conference Experiences in the ARENA

Nuno Pereira* Anthony Rowe* Michael W Farb* Ivan Liang* Edward Lu* Eric Riebling*

Carnegie Mellon University

ABSTRACT

We propose supporting hybrid conference experiences using the Augmented Reality Edge Network Architecture (ARENA). ARENA is a platform based on web technologies that simplifies the creation of collaborative mixed reality for standard Web Browsers (Chrome, Firefox) in VR, Headset AR/VR Browsers (Magic Leap, Hololens, Oculus Quest 2), and mobile AR (WebXR Viewer for iOS, Chrome with experimental flags for Android, and our own custom WebXR fork for iOS). We use a 3D scan of the conference venue as the backdrop environment for remote users and a model to stage various AR interactions for in-person users. Remote participants can use VR in a browser or a VR headset to navigate the scene. In-person participants can use AR headsets or mobile AR through WebXR browsers to see and hear remote users. ARENA can scale up to hundreds of users in the same scene and provides audio and video with spatial sound that can more closely capture real-world interactions.

Index Terms: Computing methodologies—Computer graphics—Graphics systems and interfaces Mixed / augmented reality Information Interfaces and Presentation [Miscellaneous]

1 INTRODUCTION

Current video telepresence systems either flatten user attention equally across all participants with a grid of videos or focus attention on a single active speaker. This falls short of real-world conversations that benefit from gaze queues and parallel conversations. Notably, it induces a Lombard effect where speakers compensate for poor acoustics and rapidly leads to fatigue.

We present an alternative hybrid conference experience where remote attendees can interact (in VR) with attendees wearing AR headsets and tablets. By using a 3D model of the venue (obtained using an off-the-shelf laser scanner) and registering digital objects with the real world, remote attendees have a similar sense of the space as those physically present. As virtual attendees explore the venue, spatial audio (distance and direction-aware) more closely captures natural conversations with concurrency among different groups of people. The video camera feed of attendees is also shown as a video cube that rotates to match their view of the environment. Remote attendees can see other remote attendees and also physically located ones that are using mobile AR devices. By default, all attendees can see the same content, such as other attendees, signage, presentation material, videos, screen-shares, or 3D models. Some content, such as the scanned venue model, is marked as not visible in AR.

To illustrate the conference experience proposed, in Section 3, we provide some insight into the setup of a 3D environment for the venue and describe a few common conference scenarios. We built our hybrid conference eaxperience using the Augmented Reality Edge Networking Architecture (ARENA), and we discuss a few notable platform features in the next section.

*e-mail:{npereira, agr, mwfarb, hiliang, elu2, er1k}@andrew.cmu.edu.

2 THE ARENA PLATFORM

ARENA simplifies building multi-user cross-platform mixed reality applications with built-in support for geographic content lookup, accurate relocalization, access control, the ability to host "hot-loadable" programs and manage user audio and video feeds at scale. We provide an overview of some of the main ARENA features, but a more in-depth technical description can be found in [1].

AR/VR Support. ARENA supports a seamless mixture of AR and VR interactions. Users can view a scene in AR or VR using a headset or within a 2D projection of a 3D environment in a desktop browser window. Figures 1-4 present several examples of the same 3D environment displayed in both AR and VR. Users in the physical space can see the same 3D content anchored to the physical world in AR, and the properties of all objects (and state of users) are networked to provide a consistent real-time view from any device.

3D Exploration. Users can move through the 3D environment with the mouse, keyboard arrow, and WASD keys or with touch-screen swipes, long press, and accelerometer rotations for mobile devices and VR headsets. In addition, in VR, users can effectively 'fly' (unlock their movement height and travel high above or below the ground plane) and 'teleport' (jump to near a user or location). AR users physically move through the environment tracked by a headset relocalization system or with referenced optical markers on mobile phone and tablet devices.

User Presence. By default, users appear represented in the 3D environment as a neutral robot head, which provides an indication of the pose of their camera to other users. Users can enable or disable their microphone, video, or facial landmarks used to rig an avatar. Sound projecting from each user is spatial, so nearby users can be heard louder than users far away (there are controls for the sound drop-off characteristics). When turned on, a video-based avatar of the user is created by texture-mapping the live video onto the surfaces of a 3D cube, with the front side of the box highlighted, providing an important visual indication to other users about the direction the user is facing. On the user side, a video preview box shows the user what camera view will be transmitted.

3D Content and Programs. ARENA includes capability to display arbitrary 3D models, including scanned models of real-world spaces. Large 3D models, scanned real-world spaces, and panoramic 360-degree photography are all supported depending on the desired degree of augmented reality or simulated virtual reality. Users can also share their screen to present slides or other material. The shared screen can be mapped onto one or multiple 3D surfaces. It is also possible to create programs to manipulate the 3D environment and create highly customized interactive environments.

3 THE HYBRID CONFERENCE

This section presents details of our setup process and describes three conference interactions.

3.1 Capturing a 3D Model of the Venue

Creating a 3D model of the venue is an important step that allows remote participants to have a better sense of presence. The latest (2020) iPad and iPhone Pro models have laser scanners convenient for quick scans. For higher fidelity models, we use the Leica BLK360, a terrestrial laser scanner (TLS) with registered 360 color images. Using the BLK360, along with 3D reconstruction software (e.g. Leica's

978-1-6654-1298-8/21/$31.00 ©2021 IEEE
DOI 10.1109/ISMAR-Adjunct54149.2021.00103

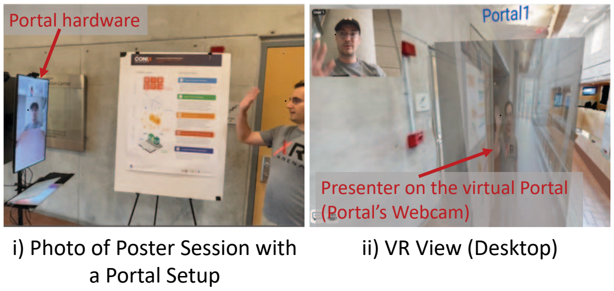

Figure 1: Poster Session with a *Portal*

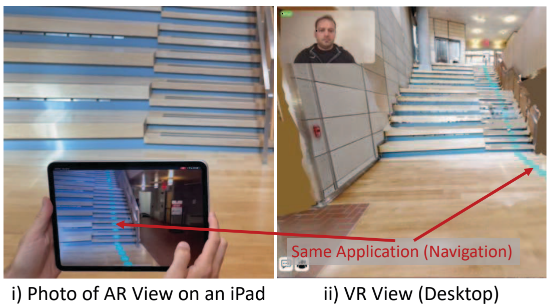

Figure 2: Way-finding application viewed from (i) AR on an iPad and (ii) desktop VR.

Cyclone FIELD 360, or Matterport), we create 3D models of physical spaces that can easily be imported into ARENA. We place the laser scanner at different spots around the venue that are merged to create the final model. Each scan can take 30 seconds, or up to 2 minutes, depending on the resolution of the scan (we often use the fastest setting). The scan density might vary significantly from space to space, where large open spaces can be captured with few scans, and more complex areas require more scans due to occlusion. As an example, a model[1] with over 400 square meters took less than 30 minutes to capture. Before loading the model into ARENA, manual adjustments to simplify the model can be made using, e.g., Blender.

3.2 Registering with the Real-World

ARENA provides several mechanisms to help streamline the management and sharing of anchor data and simplify the process of combining multiple tracking technologies into a uniform coordinate system. One simple, infrastructure-free, way of registering 3D content is using AR markers (such as AprilTags [2]) that can be placed in the venue and set as static or dynamic to determine if clients should use them for relocalization or to provide location information for the tag. For example, a 3D environment might contain several AR markers that have GPS coordinates as well as local coordinates referenced from the origin of the 3D virtual environment. ARENA's current client can decode AprilTags in browsers that allow camera access (e.g. Mozilla WebXR Viewer, and soon Chrome). If the client decodes a static marker, it uses the location data to compute the pose of the device's camera.

3.3 Example Experiences

Figures 1-4 present some example scenarios for a hybrid conference experience as described below.

[1] Shown here: https://arenaxr.org/agr/cic-lobby

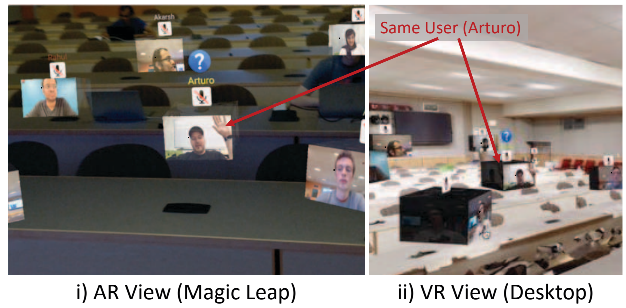

Figure 3: Talk attendees viewed from (i) the speaker point of view, with a Magic Leap headset and (ii) desktop VR.

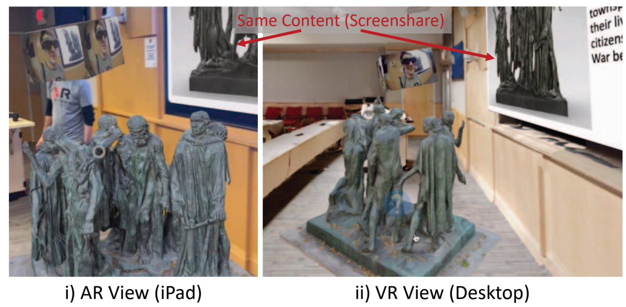

Figure 4: Content (model of statue and screenshare) viewed from (i) AR on an iPad and (ii) desktop VR.

Event Navigation: Figure 2 shows a way-finding application that helps attendees navigate to a conference session in VR and AR. The AR component provides an incentive to connect people that might be interacting in the VR space.

Attending a Talk: In Figure 4, we see a speaker presenting slides with screenshare and addresses a 3D interactive model of a statue. The speaker can look out and see both AR and VR users in the audience (shown in Figure 3).

Poster Session: We find that having a double-sided *AR/VR portal* as shown in Figure 1 is a natural way to interact across the virtual/physical boundary. Remote participants can explore the poster session by seeing digital versions of the posters overlaid in the 3D model of the venue. They can peer through portals that allow participants to interact with speakers and see a glimpse into what is going on at the physical location. Since the portal is registered in the 3D scan, the virtual and real-world geometry are aligned.

4 Conclusion

We present a hybrid conference experience for remote telepresence through an AR/VR environment that provides all users (remote and physical) with a shared sense of the conference venue. In the future, we will focus on portal hardware to include high resolution 360 video and more immersive and practical displays.

References

[1] N. Pereira, A. Rowe, M. Farb, I. Liang, E. Lu, and E. Riebling. ARENA - The Augmented Reality Edge Networking Architecture. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Oct 2021.

[2] J. Wang and E. Olson. Apriltag 2: Efficient and robust fiducial detection. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4193–4198. IEEE, 2016.