CoCreatAR: Enhancing Authoring of Outdoor Augmented Reality Experiences Through Asymmetric Collaboration

Nels Numan
Niantic Inc.
University College London
London, United Kingdom
nels.numan@ucl.ac.uk

Gabriel Brostow
Niantic Inc.
University College London
London, United Kingdom
g.brostow@cs.ucl.ac.uk

Suhyun Park
University College London
London, United Kingdom
pppshyun1226@gmail.com

Simon Julier
University College London
London, United Kingdom
s.julier@ucl.ac.uk

Anthony Steed
University College London
London, United Kingdom
a.steed@ucl.ac.uk

Jessica Van Brummelen
Niantic Inc.
London, United Kingdom
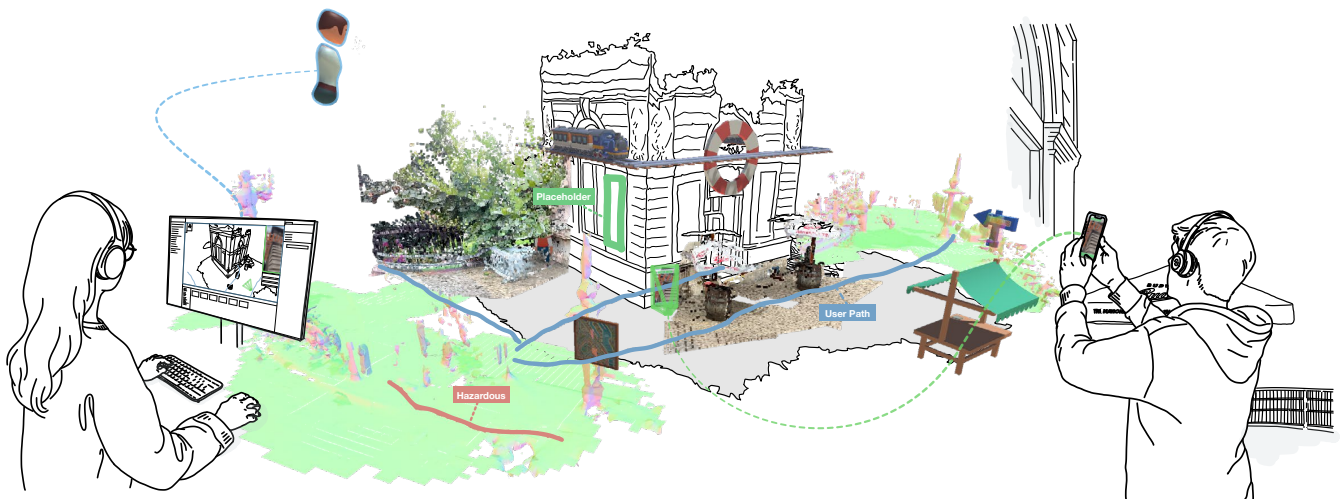jvanbrummelen@nianticlabs.com

Figure 1: CoCreatAR enables ex-situ developers (left) to collaborate with in-situ users (right) in the development of site-specific AR experiences. While iterating on the design and refining the user experience, the in-situ user can share critical real-world contextual information with their remote partner synchronously. This includes audio and video, meshes, RGB-D point clouds, annotations, as well as feedback on usability and performance. This additional anchored context enables the ex-situ developer to refine the prototype effectively using familiar tools, without requiring repeated visits to the target site.

Abstract

Authoring site-specific outdoor augmented reality (AR) experiences requires a nuanced understanding of real-world context to create immersive and relevant content. Existing ex-situ authoring tools typically rely on static 3D models to represent spatial information. However, in our formative study (*n*=25), we identified key limitations of this approach: models are often outdated, incomplete, or insufficient for capturing critical factors such as safety considerations, user flow, and dynamic environmental changes. These issues necessitate frequent on-site visits and additional iterations, making the authoring process more time-consuming and resource-intensive.

To mitigate these challenges, we introduce CoCreatAR, an asymmetric collaborative mixed reality authoring system that integrates the flexibility of ex-situ workflows with the immediate contextual awareness of in-situ authoring. We conducted an exploratory study (*n*=32) comparing CoCreatAR to an asynchronous workflow baseline, finding that it enhances engagement, creativity, and confidence in the authored output while also providing preliminary insights into its impact on task load. We conclude by discussing the implications of our findings for integrating real-world context into site-specific AR authoring systems.

CCS Concepts

• **Human-centered computing** → **Mixed / augmented reality**;
• **Computing methodologies** → **Mixed / augmented reality**; •
Software and its engineering → *Virtual worlds software*.

Keywords

collaborative mixed reality, augmented reality, co-creation, site-specific, authoring tools, reconstruction, context-aware systems

1 Introduction

Outdoor augmented reality (AR) experiences are often designed for specific locations, spanning a range of settings from controlled environments such as theme parks [45] to dynamic public urban spaces [8]. Developing these experiences is inherently complex because the user experience is closely intertwined with the physical environment. However, developers often work remotely, with only intermittent access to the target location.

Ex-situ (i.e., *off-site, remote*) development workflows typically rely on pre-captured 3D models of the environment obtained from prior scans or geospatial data to design and position AR content. Tools such as *Google Geospatial Creator* [21] and *Niantic Remote Content Authoring* [60] facilitate this process by providing site-specific environmental representations at a large scale. Additionally, they offer a visual positioning system (VPS) to precisely align virtual content with the real-world target environment. While these tools effectively support development, their reliance on static representations often leads to missing critical contextual information.

Context, broadly defined by Abowd et al. [1] as any information characterizing the situation of relevant entities, is particularly important when creating experiences for outdoor settings. For instance, dynamic elements such as lighting conditions, moving objects (e.g., vehicles, temporary structures), and pedestrian activity are typically absent from pre-captured 3D models. This absence can lead to inconsistencies between virtual elements and the real environment, ultimately degrading the user experience [54]. Consequently, developers frequently resort to repeated on-site visits, which can be costly, time-consuming, and logistically challenging.

In-situ (i.e., *on-site*) authoring tools, in contrast, enable AR content creation and testing directly within the target environment, providing immediate contextual awareness [2, 39, 86]. However, these tools often lack the flexibility and expressive power of remote development environments, particularly for large-scale outdoor experiences that require precise object placement, complex interactions, or considerations of user flow and safety. Moreover, mobile devices used for in-situ authoring often have limited computational resources and interaction capabilities, making it difficult to manage complex assets or scripting tasks [42, 49, 88].

To better understand the challenges of ex-situ authoring of site-specific outdoor AR experiences, we conducted a formative study comprising a survey (*n*=25) and follow-up interviews (*n*=5) with industry professionals. Our findings highlighted common issues, including missing or outdated environmental representations, time-consuming iterative testing cycles that necessitate site visits, and

challenges in capturing and communicating contextual information among team members.

Building on these insights, we explored collaborative approaches that integrate the advantages of both ex-situ and in-situ authoring. By enabling synchronous collaboration between ex-situ developers and in-situ collaborators, we sought to address a set of major challenges we identified. Inspired by the concept of pair programming [23], we structured collaborator roles so that the in-situ user captures and transmits real-time environmental context while the ex-situ developer remotely utilizes advanced authoring tools.

To support this workflow, we developed CoCreatAR, an asymmetric collaborative AR authoring system that facilitates real-time editing of site-specific AR content. CoCreatAR enables ex-situ developers to work synchronously with in-situ users by providing enhanced spatial information, anchored annotations, and communication tools to bridge the gap between remote development and the physical target environment.

We conducted a user study (*n*=32, in pairs) to compare the effect of synchronous (CoCreatAR) and asynchronous collaborative authoring approaches on task load, engagement, and confidence in the authored result. In the synchronous condition, ex-situ and in-situ collaborators used CoCreatAR to work together in real-time, whereas in the asynchronous condition, development proceeded sequentially without immediate interaction, reflecting current practices identified in our formative study. Our findings suggest that authoring with CoCreatAR improves the integration of real-world context, enhances developers' confidence in the accuracy and feasibility of their AR designs, and leads to greater engagement and creativity among team members. Overall, our contributions include:

- A formative study analyzing current developer workflows and the role of real-world context in site-specific outdoor AR experience development, based on a survey (*n*=25) and interviews (*n*=5) with industry professionals.
- The design and implementation of CoCreatAR, an asymmetric collaborative system that supports real-time, site-specific outdoor AR content authoring by integrating ex-situ and in-situ roles.
- An empirical evaluation comparing synchronous authoring with CoCreatAR and an asynchronous workflow baseline, demonstrating the benefits and trade-offs of each method through an exploratory user study (*n*=32).

2 Related Work

This section discusses related work on AR authoring tools, focusing on in-situ and ex-situ methods, collaborative workflows, and techniques for integrating dynamic environmental context.

2.1 Leveraging Real-World Context in AR Experience Authoring

AR authoring tools span a wide spectrum of fidelity levels, supporting different development stages and expertise levels. As categorized by Nebeling and Speicher [57], these range from low-fidelity prototyping tools that require little to no technical expertise [2, 56, 70] to high-fidelity authoring tools, such as Unity [85] and Unreal Engine [15], which are commonly used for developing production-ready applications.

Despite their utility, existing AR authoring tools have challenges that limit their effectiveness. Several surveys and studies have identified key issues, including difficulties in spatial navigation, adapting to changing environmental contexts, and integrating context-specific considerations into AR experiences [5, 36, 37, 54]. A central limitation, as noted by Nebeling [54], is the insufficient incorporation of real-world context during authoring, which can disrupt user experience and immersion. To incorporate this context, authoring tools generally adopt one of two approaches: *in-situ* authoring, which supports real-time creation in the target environment, or *ex-situ* authoring, which relies on proxy representations or recordings of the target environment.

2.1.1 In-situ authoring. Tools that incorporate in-situ interfaces, such as the mobile applications of Adobe Aero [2], Reality Composer [3], and Unity Mars [86] and other prior work [39, 98], build on the *WYXIWYG (What You Experience Is What You Get)* paradigm introduced by Lee et al. [43], which emphasizes the benefits of allowing users to experience and verify authored content in situ, such as the ability for immediate testing and adjustment.

In a user study, Lee and Kim [42] found that their in-situ system allowed for efficient and precise arrangement of virtual content in real-world settings. However, they noted a key limitation: abstract tasks, such as programming behaviors, were better supported by traditional ex-situ desktop environments. Langlotz et al. [39] developed a mobile AR authoring system with a focus on enabling spontaneous content creation in unprepared environments, including outdoor settings. While highlighting the low barriers to entry in-situ authoring provides, they also emphasized the importance of integrating it with ex-situ desktop-based interfaces for refining the content. Vargas González et al. [88] compared in-situ AR and ex-situ desktop authoring tools in scenario-based training contexts and similarly found that while both approaches offered comparable usability and task completion times, desktop tools were perceived as more efficient, particularly for tasks requiring global context.

2.1.2 Ex-situ authoring. Ex-situ authoring tools, such as *DART* [51], *ScalAR* [68], *DistanciAR* [93], and *Corsican Twin* [67], offer broader authoring capabilities than in-situ tools by providing 3D scene editors, scripting environments, flexible testing workflows, and asset integration. These tools represent environmental context using various forms of pre-captured data, including 3D models [10, 21, 60, 67, 68], video [47, 51], sensor data [51], or 360° footage [56]. For example, *DART* [51] facilitates recording and playback for authoring by synchronizing real-time video and sensor data capture, while *DistanciAR* [93] supports in-situ 3D reconstruction to capture environmental context and enables ex-situ authoring and testing through different model visualization modes. However, a key limitation of these approaches is their reliance on static environmental representations. These models capture only a fixed moment in time, falling short in accounting for dynamic changes. As a result, authoring errors and spatial misalignment may occur.

Recent ex-situ authoring tools for outdoor AR experiences, such as *Niantic Remote Content Authoring* [60] and *Google Geospatial Creator* [21], adopt similar approaches by providing 3D models of large outdoor sites. However, outdoor environments present even greater challenges due to a higher degree of dynamic change. Unlike indoor spaces, outdoor settings are continuously affected by shifting

lighting, weather, seasons, and moving elements like people and vehicles. These dynamic factors significantly impact the accuracy and relevance of pre-captured data, making static environmental representations particularly problematic for authoring outdoor AR content. While in-situ testing can help mitigate some challenges, it is often constrained by logistical factors such as time, cost, and accessibility. Moreover, in-situ authoring of outdoor AR experiences is complicated by the scale and complexity of outdoor spaces [9, 26, 62], making in-situ adjustments more difficult to execute.

Overall, our analysis indicates that current AR authoring tools remain limited in bridging the gap between ex-situ flexibility and in-situ contextual awareness. Ex-situ tools rely on static proxies that fail to capture the dynamic nature of outdoor spaces, while in-situ tools, despite offering real-time context, lack the expressiveness and flexibility required for comprehensive authoring, as we found in our formative study (Sec. 3). This disconnect forces authors to switch between ex-situ and in-situ development, leading to inefficiencies and challenges in maintaining contextual accuracy. To address this gap, we propose COCREATAR, a system that leverages the flexibility of ex-situ authoring while incorporating real-time environmental awareness from in-situ users. By enabling a synchronous collaborative workflow between ex-situ and in-situ roles, our approach supports contextually aware authoring without compromising the richness and advanced capabilities of ex-situ tools.

2.2 Collaborative Approaches to AR Experience Authoring

Given the complex and multidisciplinary nature of AR experience development, collaboration is integral to the authoring process, as noted by Krauß et al. [36]. Specifically, AR application development typically involves multiple roles, such as interaction designers, content creators, and developers, each contributing distinct expertise, as well as non-technical collaborators like clients or end-users, who can help shape the design and content [37]. Inspired by role-based collaboration, Nebeling et al. [55] introduced *XRDirector*, a system that enables users to manipulate virtual objects from different subjective viewpoints in AR and VR. Designed for indoor immersive storytelling, *XRDirector* enabled multiple contributors to author experiences from perspectives aligned with their respective roles.

Beyond role-based collaboration, prior work has also examined different modes of collaboration when authoring AR experiences. Guo et al. [22] explored synchronous and asynchronous collaboration in AR authoring through their system, *Blocks*. Their findings indicated that synchronous collaboration fostered higher engagement, particularly when users collaboratively created shared AR structures in real time. Conversely, asynchronous collaboration, while offering greater flexibility, introduced challenges related to maintaining awareness of others' contributions.

While prior systems demonstrate different approaches to collaborative authoring, they do not adequately address the challenges of site-specific AR development, particularly in outdoor environments where ex-situ authors must consider in-situ perspectives. In designing COCREATAR, we sought to support this workflow by facilitating collaboration between ex-situ developers and in-situ collaborators without requiring specialized technical skills beyond familiarity

with handheld AR. Our system leverages subjective views, similar to *XRDirector*, but applies them to gathering in-situ context and feedback in outdoor, site-specific authoring. Additionally, we compare synchronous and asynchronous workflows, as examined in *Blocks*, including specific hypotheses on engagement and task load (Sec. 5).

2.3 Environmental Context Awareness in Collaborative Mixed Reality

Effective integration of real-world context is crucial for collaborative AR authoring, particularly when bridging ex-situ and in-situ roles. While previous work has explored a wide range of methods to represent remote environments, most focused on indoor settings or static outdoor spaces, leaving the dynamic nature of outdoor environments largely unaddressed.

Early systems such as *MARS* [25] and *Tinmith-Metro* [66] enabled interaction with virtual models of physical spaces in both indoor and outdoor contexts. However, they lacked visual fidelity, real-time updates, and dynamic environmental data, which limited their effectiveness for site-specific AR authoring.

To enhance environmental awareness, more recent approaches utilize higher-fidelity representations such as 360°-video for immersive panoramic views [31, 76, 81]. For example, *360Anywhere* [76] supports multi-user collaboration with gaze awareness and annotation tools. While offering comprehensive visual coverage, these methods often lack depth information and do not fully capture the spatial complexity of outdoor environments.

Volumetric representations, including static photogrammetry-based 3D reconstructions and fused RGB-D streams, offer richer spatial representations. Photogrammetry-based methods generate detailed 3D models from images [20, 63, 84], though their creation is time-consuming and not suitable for capturing dynamic changes. Systems employing fused RGB-D data, such as *Remixed Reality* [48] and *TransceiVR* [83], integrate real-time color and depth information for interactive and dynamic 3D scene representations. However, these approaches commonly rely on specific sensors, system configurations, and controlled settings, which can limit their application in outdoor environments.

Combining multiple spatial representation techniques has proven effective in enhancing environmental awareness in collaborative settings [50, 71, 73, 81, 82, 84, 96]. For example, Thoravi Kumaravel et al. [82] introduced *Loki*, integrating live video feeds, 3D models, and spatial annotations for remote instruction in indoor controlled environments. Focused on representing outdoor environments, Du et al. [14] projected *Google Street View* imagery onto building geometries derived from *OpenStreetMap* to place and view geotagged information in *Geollery*. While effective within their respective contexts, these systems generally do not account for the challenges posed by highly dynamic, large-scale outdoor environments.

Building on prior work, we aim to address these challenges by enhancing environmental awareness in outdoor environments through CoCreatAR. We integrate multiple forms of spatial capture methods to supplement pre-captured 3D meshes of outdoor locations used for authoring site-specific AR content. Specifically, CoCreatAR enables in-situ users to capture single-frame RGB-D data for detailed short-range capture and coarse 3D meshes for broader geometric context. This approach provides ex-situ authors with up-to-date, targeted, real-world spatial context to incorporate into their authoring process.

While several individual components of our proposed system—such as RGB-D capture, mesh reconstruction, and annotation tools—have been explored in prior research, their combined application in highly dynamic, large, and diverse outdoor environments within an AR authoring context represents a novel exploration. By integrating these spatial capture methods into a collaborative workflow, CoCreatAR aims to bridge the gap between ex-situ and in-situ roles, enhancing environmental awareness and supporting effective real-time collaboration in developing site-specific AR experiences.

3 Formative Study

To inform the design of CoCreatAR, we conducted formative surveys and interviews with experienced site-specific AR developers, artists, designers, and testers. Inspired by other investigations into AR development workflows [5, 36, 77], we sought to determine current workflows and key challenges of outdoor site-specific AR experience development.

3.1 Method & Participants

We recruited 25 survey respondents through internal mailing lists, AR development social media channels, and targeted emails to professional outdoor AR agencies. The survey included both quantitative and qualitative questions about the projects respondents had worked on, the tools they used, and their typical workflow. One survey question prompted respondents to list issues encountered during in-situ testing. To provide an initial set of response options, we defined a set of issue types through a brainstorming session and included an open-ended question for respondents to report additional issues they had come across.

The resulting issue types were as follows: (A) Physical constraints (e.g., blocked paths and restricted areas), (B) User safety (e.g., hazardous environments), (C) Misaligned AR elements (e.g., misplaced anchors, occlusion, or perspective issues), (D) User context (e.g., environmental noise or lighting conditions), (E) Hardware performance (e.g., device processing and rendering power limitations), (F) Registration issues (e.g., localization and tracking errors), (G) User flow (e.g., user position or perspective affecting AR interaction), (H) Socio-cultural appropriateness of the area or experience, and (I) Semantic interaction (e.g., whether the system correctly recognizes objects in the world). We refer to Fig. 4 for visual illustrations of these issues.

Five of the survey respondents agreed to complete an interview, in which we asked semi-structured follow-up questions to obtain further details. Two authors of this paper conducted a reflexive thematic analysis to analyze the qualitative results. After developing initial codes, the authors iteratively refined the codebook while reviewing transcripts, following the methods described by Braun et al. [7]. Each interview participant received a gift card valued at £25, and survey respondents had a 1-in-10 chance of winning a gift card valued at £30. The survey and interview materials are available in the Supplementary Materials.

3.2 Findings

We structure our findings by first outlining typical workflows, followed by a review of major hurdles, and concluding with a discussion of future tools that participants considered useful. These findings inform our understanding of how users typically develop site-specific AR, support us to establish requirements for our proposed system (Sec. 4) and a baseline for our study (Sec. 5).

3.2.1 Workflow walkthrough and tools. One of the first steps participants took was to **acquire context about the target location** for their AR experience. This typically involved downloading publicly available pre-captured meshes or visiting the site to scan the location themselves. Participants primarily used *Niantic Geospatial Browser* [59] (64.0%) or *Google Geospatial Creator* [21] (28.0%) to obtain pre-captured meshes, which provided spatial information about the location to support the design and development process. We refer to these pre-captured meshes as *location meshes* throughout the rest of this paper. These representations helped them place and carefully align virtual content with real-world structures and surfaces. Later, when at the site, end-users could use a VPS to localize and interact with the experience.

Following this, participants typically **worked ex situ to design and build the experience** based on the *location mesh*. For development, they most often used *Unity* (96%) and *Blender* (76%). Additionally, they employed AR frameworks to support features such as plane detection, occlusion meshes, and semantic segmentation. The most commonly used framework was *Niantic SDK for Unity* (80.0%), followed by *ARFoundation* (76.0%). A summary of the tools participants used is available in the Supplementary Materials.

Since ex-situ tools could not provide full contextual awareness of the location (e.g. dynamic objects, crowding, and lighting conditions), **participants or their team members often had to travel back and forth** between their ex-situ development location and the target environment. When in situ, participants logged information about issues with their AR experience to address them later or relay them to their team. On average, they reported spending 36% (SD=24.7) of their time testing the experience in situ and had between 0 and 12 team members that worked with them (M=4.16; SD=2.97). Notably, one interview participant (P4) mentioned considering bringing their "computer with [their] phone" to the site to address alignment issues in situ but noted that they "do not think [they] should" for practical reasons, such as environmental conditions like weather and crowds, which were also cited by other participants (P1, P2, P5).

In terms of the issues described in Sec. 3.1, *Physical Constraints*, *User Safety*, and *Misaligned AR Elements* were the most commonly reported[1] (see Fig. 3). To record these issues in situ, participants typically took screenshots, screen recordings, or typed notes (see Fig. 2). Some also received recordings or notes from team members who visited the site. After initial context gathering, participants or their team members continued iterating on their AR experiences, traveling between locations as needed until completion.

3.2.2 Challenges with context capture. In the interviews, all five participants mentioned **difficulties with pre-captured meshes**,

---

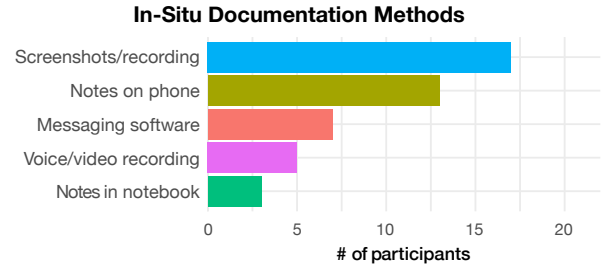[1]Only one participant described an additional issue type in the survey, "Game Flow," which we categorize under "User Flow."

In-Situ Documentation Methods

Figure 2: The number of participants who reported using various methods to document issues in their AR experiences when in situ.
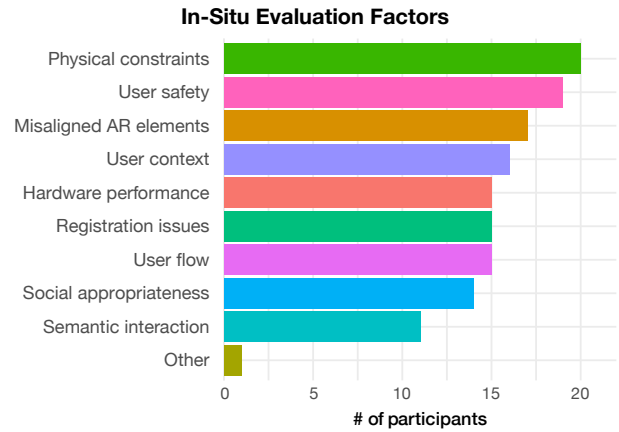

In-Situ Evaluation Factors

Figure 3: The number of participants who reported looking for various issues in their AR experiences when in situ. See Sec. 3.1 for descriptions of the issues.

whether obtained from geospatial browsers or personally scanned. For instance, P2 described pre-captured *location meshes* as "old" and not up to date with the current state of the real world. P4 noted that *location meshes* were "always a challenge because there are mesh holes everywhere [...] so the geometry falls through." Similarly, P5 stated that "the maps are very incomplete," but the alternative—scanning the location manually—was "quite labor intensive."

Participants also mentioned **difficulties with scanning**. P1 described how "often the test scans do not come out great" and how they sometimes needed to select alternative target locations due to poor scan quality. P5 noted that scan processing times could be lengthy, which reduced iteration efficiency. Two participants also commented on the difficulty of the scanning process due to social awkwardness (P3) or poor weather conditions (P5).

Poor-quality meshes contributed to **alignment issues** between virtual objects and the real world, both because degraded meshes reduced localization accuracy and because poor mesh geometry hindered precise placement when working ex situ. For example, P5 described this limitation: "It's like you try to figure out where [the virtual object] has to be [ex-situ], and then every time you go [in-situ], your viewpoint is a bit different because you are looking at it from street level or it's slightly tilted [...] we always ended up

Figure 4: The five main site-specific AR issue types we aimed to address with CoCreatAR. We refer to Sec. 3 for further details on the full range of issues identified in our formative study.

having to reposition the thing compared to the initial [placement]." All participants mentioned alignment issues as a recurring problem.

3.2.3 Missing contextual information. All five participants also mentioned **challenges in acquiring contextual information missing from meshes**. For example, P3 and P4 described how locations found online might not actually be accessible, or how previously accessible locations might become restricted over time. Three participants mentioned how some locations were hard to conceptualize ex situ without visiting in person—either because they were too complex or large (P1) or too cramped (P2, P3).

P1, P2, and P4 mentioned challenges related to **lighting conditions**, particularly at different times of the day. This concern arose both from an end-user experience perspective (e.g. P1: "You do not want someone staring into the sun") and in terms of VPS localization accuracy (e.g. P4: "we generally test throughout the day to try to make sure that it is going to work under any lighting"). P1 also highlighted "audio contamination" as a factor that could affect the experience, but which could not be discovered without being at the target location. Additionally, P1, P2, and P5 mentioned adverse weather conditions affecting in-situ testing.

One crucial piece of contextual information that all five participants mentioned gathering in situ was **the area's safety**. This included assessing foot and vehicle traffic, as well as the potential for theft. For example, P3 described how even though an area might be plenty of areas where things are walkable [... this still] means very heavy foot traffic, [so] you still have to be careful." P1 pointed out that if users are "getting jostled, or [using] a very expensive phone, then someone could just snatch it."

All participants also mentioned the need to assess **the level of crowding at a location**, as high foot traffic could impede both the user experience and VPS localization. P2 even described having to change the target site due to crowding. **Moving objects** also posed a challenge for both the experience and localization. As objects at the location changed—such as cars, signs, or even "posters" (P1)— VPSs struggled to match the environment with the *location meshes*.

3.2.4 User experience challenges. All five participants reported **difficulties identifying end-user experience challenges**, particularly when developing ex situ. P1 and P2, for example, emphasized the importance of understanding the "human scale" to ensure objects were appropriately reachable and scaled, which was difficult to

achieve without being in situ. Participants also noted visibility challenges. P1, P2, and P3 described how "tight FOVs," long distances, or sun direction could affect visibility. Gameplay considerations also required in-situ testing; P2 stated that users should not have to run during the experience, while P1 mentioned ensuring that gameplay events were not "too close" or "too far away."

3.2.5 Workflow challenges. Due to these difficulties, four of the five participants described **traveling to target locations as a significant challenge**. They characterized these trips as "inefficient" (P1), "a lot of effort" (P3), and "costly" (P5). These trips could also be extensive, with participants traveling to different cities (P2, P3) and even different countries (P5) for testing.

Other workflow challenges included collecting and sharing feedback from in-situ testing. P1 described difficulties in physically taking notes while testing the experience, explaining issues over audio, and annotating video recordings. P2 and P5 described communication challenges when working with team members from different backgrounds. P2 also mentioned "hassles" related to uploading, downloading, and sharing information while in situ.

3.2.6 Desired features for future tools. As participants described these challenges, they also suggested potential solutions. To address poor context capture, P2 and P4 proposed **methods to tag areas with semantic or other labels**. Others suggested **collaborative approaches to context gathering**. For instance, P1, P4, and P5 suggested streamlined methods for sharing and viewing *location meshes* with team members. P4 also proposed a way to combine and edit multiple team members' meshes to enhance their accuracy. Additionally, P2 and P4 expressed interest in a system that allowed team members to collaborate in situ and ex situ via live streaming of location data and voice communication.

Participants also expressed interest in tools that improve the authoring process. Four of the five participants expressed interest in **editing objects in situ**. P2 and P4 suggested **procedural placement of objects based on semantic data**. P3, P4, and P5 wanted better methods for simulating the site ex situ, with P5 stating, "The more you can do without actually being on the spot, the better."

4 CoCreatAR System

Based on our formative study, we developed CoCreatAR, a collaborative authoring system for outdoor AR experiences, designed to facilitate real-time interaction between ex-situ developers and

Figure 5: In-situ user interface of CoCreatAR. (A) Users can scan the environment to obtain a *Coarse Mesh* of the surroundings; (B) Users can tap Capture to take a *3D Snapshot*; (C) Users can tap Capture or Annotate to access feature sub-menus from the main menu and tap and hold on the screen to spawn a *3D Cursor*; (D) Users can create drawings projected onto surfaces with *Surface Draw*; (E) Users can create a trajectory or a 3D drawing with *Air Draw* by moving their smartphone.

in-situ collaborators. The system enables ex-situ creators, who typically design and develop the experience within Unity [85] asynchronously, to synchronously collaborate with in-situ users who experience the AR content directly in the field. By integrating real-time communication, contextual reference tools, and spatial data capture, CoCreatAR aims to reduce the need for repeated on-site visits during the iterative design of site-specific AR experiences. While CoCreatAR supports multiple ex-situ and in-situ users, in this section and our user study, we focus on a two-user scenario in which a single ex-situ user collaborates with one in-situ partner.

CoCreatAR primarily utilizes two frameworks: the *Niantic SDK for Unity* [60], identified in the formative study as the most widely used outdoor AR development kit, and *Ubiq* [17], an open-source library for collaborative mixed reality. The *Niantic SDK for Unity* provides VPS localization, depth estimation, meshing, and occlusion handling, whereas *Ubiq* facilitates real-time networking, avatar representation, and peer-to-peer communication, including WebRTC-based audio and video streaming.

4.1 System Requirements

Based on challenges reported by participants in our formative study (Sec. 3) and previous work, we identified five requirements for the development of CoCreatAR.

R1: Reduce the development and testing iteration loop caused by frequent on-site visits. Our formative study revealed that participants required frequent on-site visits to gather real-world context and feedback. CoCreatAR aims to reduce the travel burden by allowing ex-situ users to collaborate with in-situ partners

synchronously, enabling adjustments to be made remotely based on immediate feedback and live validation.

R2: Support incorporation of additional or updated real-world spatial information. Interviewees highlighted the limitations of static meshes, which are often outdated or incomplete, leading to alignment issues. CoCreatAR aims to address this by enabling in-situ users to capture and relay real-time updates about the site, such as changes in the environment.

R3: Facilitate the capture of real-world context beyond spatial data. Participants emphasized the importance of capturing additional real-world context, including lighting conditions, pedestrian flow, and safety concerns. CoCreatAR incorporates tools that enable in-situ users to capture this contextual information through annotations, real-time audio, snapshots, and a live video feed.

R4: Enable live persistent scene adjustments and additions during testing and prototyping. Participants described having to return to their workplace to make adjustments after in-situ testing as challenging. To mitigate this, CoCreatAR supports live modifications during testing sessions, allowing developers to make immediate adjustments based on in-situ feedback.

R5: Integrate captured data into existing development workflows. Participants reported difficulties in incorporating in-situ feedback into the virtual scene representations within their development tools, particularly in Unity. CoCreatAR addresses this by integrating captured spatial data and annotations directly into Unity and anchored to the *location mesh*, enabling developers to incorporate feedback without manual import procedures.

Figure 6: Ex-situ user interface of CoCreatAR. (A) All objects under `NetworkedScene` are automatically synchronized between ex-situ and in-situ users; (B) *3D Snapshots* captured by the in-situ user; (C) The *location mesh* of *Location* A; (D) *Coarse Mesh* captured by the in-situ user; (E) Live feed of the in-situ user's screen, including AR content; (F) The *3D Cursor* of the ex-situ user, projected into world space; (G) Close-up of the *3D Cursor* of the ex-situ user as seen in the scene view; (H) List of annotations and spatial captures, persistently saved in the scene for later review; (I) Sample assets that can be added to the scene at runtime.

4.2 Scene View and Localization

CoCreatAR enables both users to view the same AR content from their respective familiar perspectives. In-situ users can interact with the AR experience and a CoCreatAR UI layer within their real-world environment via a smartphone (Fig. 5), whereas ex-situ users observe the scene remotely through Unity's scene editor (Fig. 6).

At launch, in-situ users complete a step-by-step VPS-based localization process to align the AR experience with their surroundings. The system continuously localizes, prompting the in-situ user to confirm alignment by overlaying a semi-transparent pre-captured *location mesh* onto the real world. The in-situ user can adjust its transparency to verify overlap before finalizing localization. Once confirmed, the full scene, including the in-situ user's avatar, loads into the ex-situ user's scene view. The ex-situ user can similarly adjust the *location mesh* transparency to better view AR elements, spatial captures, and annotations throughout the collaboration process (Fig. 6J). If needed, localization can be restarted via the settings panel to correct for drift that may occur during collaboration.

4.3 Real-Time In-Situ View Streaming and Audio Communication

CoCreatAR supports real-time peer-to-peer video and audio communication to facilitate collaboration between users. A peer-to-peer audio channel enables verbal communication. Additionally, a live

feed of the in-situ user's screen is streamed to the ex-situ user, who can view this feed in a floating panel within the Unity scene view (Fig. 6E). This feed presents both the real-world camera view and the rendered AR elements, enabling the ex-situ user to understand the interaction between AR content and the physical environment.

Ex-situ users can also capture screenshots of the in-situ user's screen, which are saved and displayed in the Unity scene and anchored to the *location mesh* (Fig. 6K). These captures align with the in-situ user's perspective and can mark specific moments, errors, or areas of interest for later review.

4.4 Spatial References with 3D Cursors

To support spatial referencing during the collaborative process, CoCreatAR incorporates networked *3D Cursors* that each user can control. These cursors are projected into shared world space based on screen inputs. In-situ users can tap and hold anywhere on their screen to spawn and control a cursor in the 3D environment, while ex-situ users can hover over the live in-situ view panel within Unity to achieve the same. Cursors appear as semi-transparent spheres with axis indicators, aiding spatial referencing. Green represents in-situ cursors, while blue represents ex-situ cursors (see Fig. 6F).

Both users can create persistent cursor markers: ex-situ users by clicking within the Unity editor and in-situ users by double-tapping on their screen. This functionality supports alignment tasks and spatial referencing during collaboration.

4.5 Auto-Networked Objects and Scene Updates

A core feature of CoCreatAR is the automatic synchronization of all scene changes (Fig. 6A). Any modifications made by the ex-situ user in Unity—such as adjusting the position, rotation, or properties of AR objects—are immediately reflected in the in-situ user's view. This functionality extends Ubiq's networking capabilities to transmit network messages whenever changes in object transforms, materials, or user-controllable serializable script parameters are detected. CoCreatAR persistently saves all modifications within the Unity scene file, enabling continued iteration and version control after the collaborative session concludes.

4.6 Spatial Context Capture with Coarse Meshes and 3D Snapshots

In-situ users have two tools for capturing additional spatial context to supplement the ex-situ user's representation of the real world: *3D Snapshots*, which are point clouds based on world-projected RGB-D frames, and *Coarse Meshes*, which are untextured 3D meshes.

The *3D Snapshot* feature utilizes the depth submodule of *Niantic SDK* to capture a depth image alongside the current camera frame from a LiDAR-enabled iPhone model. However, we note that *Niantic SDK* also supports 3D data capture on other devices and can extend beyond the LiDAR sensor's range. Following this, the obtained depth data is combined with RGB information to generate a point cloud, which is transmitted to the ex-situ user's Unity scene. There, it is processed and visualized as a colored world-space point cloud anchored to the *location mesh*. For in-situ users, the point cloud appears in white and is semi-transparent to provide feedback without distracting from the overall experience. This feature is particularly useful for capturing smaller objects or areas requiring finer detail, such as user interfaces or situated virtual objects that demand precise positioning.

For larger-scale spatial data capture, CoCreatAR provides a coarse 3D mesh capturing feature, extending the *Niantic SDK's* meshing submodule. Unlike *3D Snapshots*, the *Coarse Mesh* feature captures the environment's geometry without color information. This method is well-suited for representing larger structures, such as buildings, or open areas, like streets, where surface texture may be less relevant. Mesh updates are streamed to the ex-situ user's Unity editor in real time, with network message coalescing applied to reduce network load. The meshing process automatically terminates after 15 seconds to prevent excessive network usage but can be restarted at any time. The colored mesh material represents surface normals to enhance shape visualization.

Both *3D Snapshots* and *Coarse Meshes* are persistently saved within the Unity scene for future reference, organized into separate timestamped objects for each collaborative session (Fig. 6H).

4.7 Annotation of the Real World with Surface Draw and Air Draw

CoCreatAR also enables in-situ users to provide contextual feedback through annotations. The system supports two types of annotations: *Surface Draw* and *Air Draw*. *Surface Draw* projects the in-situ user's sketches onto real-world surfaces using depth estimates, while *Air Draw* allows users to draw freely in 3D space by moving their smartphone through the environment.

Annotations are color-coded and can be tagged with predefined labels, such as *hazard* or *user flow*, or with custom labels which are automatically assigned distinct colors for easy identification. All annotations are streamed to the ex-situ user's Unity editor in real time, where they appear as world-space objects and are saved for future iterations. Like spatial captures, annotations are persistently stored in the Unity scene, allowing contextual information to be revisited and informing subsequent development steps (Fig. 6H).

5 Exploratory User Study

In this study, we aimed to explore how paired users co-author site-specific AR experiences using CoCreatAR compared to a typical authoring workflow. In addition to assessing effectiveness and user experience, we sought to identify key directions for future research.

We employed a within-subjects counterbalanced design, comparing *two collaboration modes*: Async (baseline) and Sync (CoCreatAR, ours). For each session, we assigned one participant to the ex-situ role and the other to the in-situ role, based on their background experience. Participants were compensated with gift cards valued at £40 for their study participation, which lasted around two hours. The study was approved by the UCL Computer Science Research Ethics Committee (Study ID UCL/CSREC/R/17).

To control for order and environmental effects, we counterbalanced both *collaboration mode* (Sync or Async) and *location* (*Location* A or *Location* B). Each pair began by completing the task using either the Sync or Async *collaboration mode*, with the in-situ user physically at either *Location* A or *Location* B, while the ex-situ user participated remotely. They then completed the task for a second time using the alternate *collaboration mode*, with the in-situ user at the alternate *location*.

We focused on evaluating the below hypotheses, which emerged based on the formative study and prior work [22, 36, 58, 90]:

HYPOTHESIS 1. *Ex-situ users will report a **lower** level of task load for Sync compared to Async.*

HYPOTHESIS 2. *In-situ users will report a **higher** level of task load for Sync compared to Async.*

HYPOTHESIS 3. *Ex-situ and in-situ users will report a **higher** level of engagement for Sync compared to Async.*

HYPOTHESIS 4. *Ex-situ users will report a **higher** level of confidence in authored experiences for Sync compared to Async.*

5.1 Baseline

Our baseline condition, Async, was designed to reflect the typical workflow identified in the formative study (Sec. 3). In this workflow, developers commonly receive or create a set of notes, recordings, and screenshots with insights on improving the AR experience. To align with this, we included these elements in the baseline condition. In our study, the in-situ participant gathered this feedback using a series of custom iPhone *shortcuts*[4], activated via the iOS *AssistiveTouch* button. These shortcuts allowed for the creation of voice recordings, annotated screenshots, and screen recordings of the AR application, each automatically saved to the iOS *Notes* app. The user interface of this baseline system is shown in Fig. 7.

Once the in-situ participant finished gathering information about the experience, the notes were sent to the ex-situ participant. To
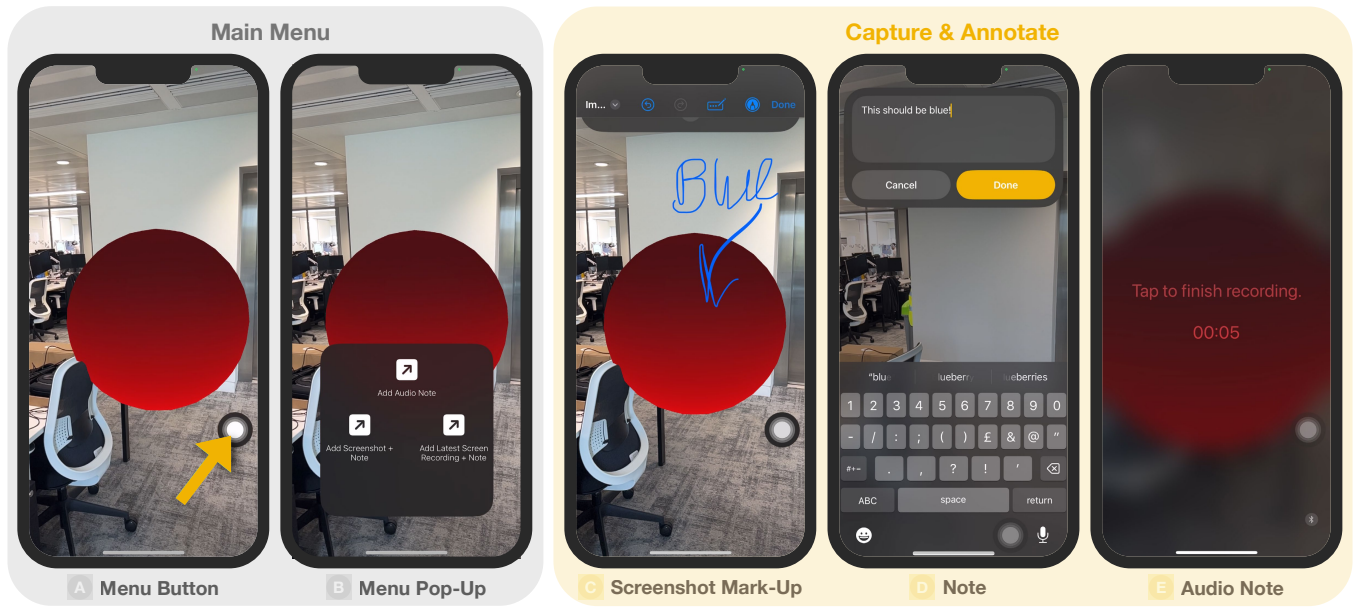
Figure 7: *In-situ* user interface of the baseline system. (A) The menu button (highlighted by a yellow arrow), which is the iOS *AssistiveTouch* button; (B) The menu where users can easily access screenshot, audio note, and screen recording features; (C) Interface for annotating screenshots; (D) Interface for attaching typed notes to screenshots or recordings; (E) Interface for recording audio notes. *Note: the interface for screen recordings is not shown, as the UI is simply a red dot in the upper right corner.*

recreate a typical developer environment, the baseline system for the ex-situ user was based on Unity [85] and *Niantic SDK for Unity* [60]. The Unity editor contained the *location mesh*, an initial prototype of a site-specific AR experience, and a set of sample assets, including objects and materials from the public domain[2]. More information on the study procedure is available in Sec. 5.4.

5.2 Participants

We recruited 32 participants (16 pairs) through internal mailing lists and social media platforms. Participants self-reported their gender, with 19 identifying as men and 13 as women. The mean age of participants was 30.16 years (SD = 9.78). Participants were primarily students and researchers from fields related to computer science, virtual and augmented reality, and interactive media, with a few professionals from sectors such as government, consulting, and game development. For in-situ participants, no prior experience was necessary. For ex-situ participants, basic Unity experience was required. To increase our participant pool and obtain a broad range of experience in Unity—similar to the demographics identified in our formative study—we enabled ex-situ participants to participate remotely from anywhere in the world, using a high-speed internet connection. All participants collaborated in English.

Given the exploratory nature of our study, we selected a sample of 32 participants, consistent with related research on collaborative and authoring systems [55, 82, 83]. Moreover, the requirement for participants with Unity experience imposed recruitment constraints, limiting the feasibility of a larger sample [38]. Given the study's scale and exploratory nature, the statistical results should

be interpreted with due consideration of potential biases and limitations, as outlined in Sec. 8.

5.3 Task

The task involved a two-part collaborative process aimed at improving and expanding predefined site-specific outdoor AR prototype experiences, which are described in Sec. 5.3.1.

In *Phase 1: Feedback Collection & Refinement*, participants focused on identifying and addressing issues within the prototype. The in-situ participant primarily gathered feedback and contextual insights through real-world interaction with the prototype, while the ex-situ participant refined the prototype by implementing necessary corrections based on the feedback provided. In the Async condition, the feedback gathered by the in-situ participant was available to the ex-situ participant *after* the feedback collection phase. To accommodate this order of actions in the Async condition, the ex-situ participant had a short break prior to starting this phase while the in-situ participant gathered feedback. In the Sync condition, this process occurred synchronously.

In *Phase 2: Ideation & Prototyping*, which lasted six minutes, participants transitioned to brainstorming and implementing extensions or enhancements to the prototype. During the Async condition, the in-situ participant ideated independently, noting ideas and gathering relevant contextual artifacts, which were then shared with the ex-situ participant for prototyping. In contrast, in the Sync condition, ideation and prototyping occurred synchronously.

5.3.1 Prototype AR experiences. The prototype experiences used in the study were designed based on the environmental characteristics
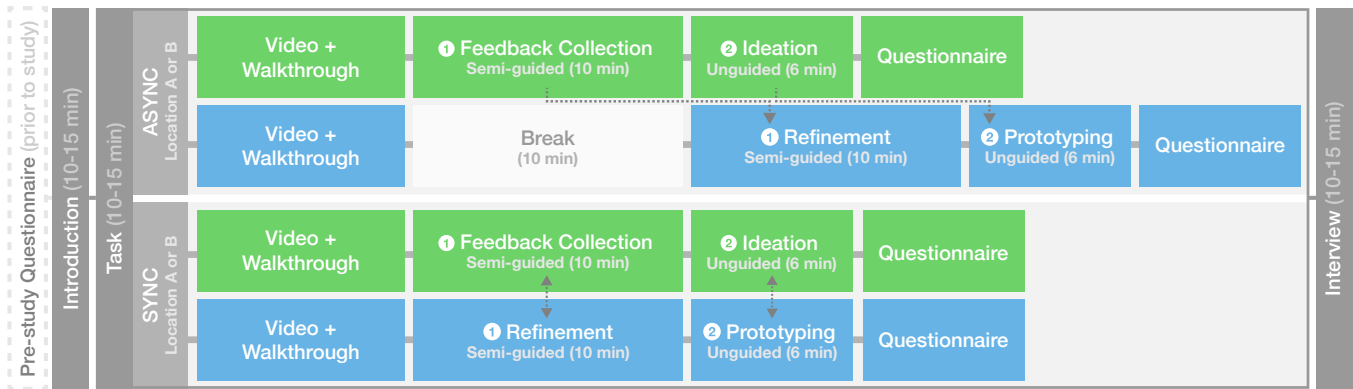
---

[2]https://kenney.nl/assets/

Figure 8: Overview of the study procedure. The process begins with a general introduction (left), followed by the task phase, performed under two conditions (SYNC (i.e., COCREATAR) and ASYNC), and concludes with a post-study interview (right). Green represents the procedure components of the in-situ participants, blue represents the procedure components of the ex-situ participants, and grey blocks represent components completed by both. ❶ represents Phase 1 (Feedback Collection & Refinement) and ❷ represents Phase 2 (Ideation & Prototyping) of the task. The arrows represent the exchange of task-related information across pairs. Note: both *collaboration modes* (SYNC and ASYNC) and *locations* were counterbalanced across pairs.

of the two selected outdoor locations (*Location* A and *Location* B), informed by both the *location mesh* and supplementary information from *Google Maps* and *Google Street View*. To maintain consistency across locations and ensure generalizability to other site-specific AR applications, the prototypes adhered to a set of predefined design dimensions. These dimensions, grounded in prior work and insights from our formative study, were applied to create similar AR prototypes for each location, ensuring participants could address comparable issues across both locations.

(1) **Design Patterns** *used as design guidelines for the representation of each AR element, based on [41]:* Glyphs (navigation elements, indicative of actions with spatial alignment to referents), Decals (information related to referent context), Trajectories (e.g., paths, outlines, arrows), Labels (contextual annotations), Ghosts (overlays linked to referents), and Audio (non-visual cues providing additional context).
(2) **Referent Types** *used as design guidelines for the physical referent of each AR element:* Small objects (e.g., mobile elements such as signs or markers), Large objects (e.g., statues, lampposts, or trees), Large planes (e.g., open spaces or flat surfaces), and Building structures (static physical structures with specific entry points or facades).
(3) **Alignment Types** *used as design guidelines for the placement of each AR element:* Overlap (directly aligned with referents), Proximity (placed relative to nearby referents), and Surface (mapped directly onto physical surfaces).
(4) **Issue Types** *used as a design guideline to incorporate common design issues, based on the five most prominent issues identified in Sec. 3:* Physical constraints, User safety, Misaligned AR elements, User context, and User flow, as visualized in Fig. 4.

Both selected locations were within walking distance of University College London (*Location* A and *Location* B). *Location* A, located near a major transit station, featured high foot traffic, a nearby road, and elevated levels of environmental noise. The theme of the AR

experience at *Location* A was "*Welcoming Tourists to the City*," reflecting its proximity to a transit hub. *Location* B was situated along a narrow street with a local cafe, characterized by significant foot and bicycle traffic and lower environmental noise. The theme for this location was "*Promoting and Celebrating the Upcoming Cookie Party*," aligning with the atmosphere of nearby restaurants and cafes. For each location, seven issues covering the range of issue types, as previously defined, were identified and verified through on-site testing prior to the user study. An overview of the prototypes is shown in Fig. 9, with detailed descriptions available in the Supplementary Materials.

To offer both participants hints regarding issues in the prototype, both in-situ and ex-situ participants were provided with a list of seven *task clues* in Phase 1. This list contained fictional observations made by "early testers" of the prototype experience, offering participants a semi-guided path through the experience while still requiring further investigation to ensure consistency across pairs. Both task clue lists are available in the Supplementary Materials. Participants were allowed to start at any point on the list and could skip or revisit clues as they desired.

5.4 Procedure

An overview of the study procedure is shown in Fig. 8. Prior to the study session, participants completed a questionnaire covering demographics and their background in smartphone-based AR, 3D game engines, and 3D editors.

For each session, the in-situ participant took part in person using an iPhone 13 Pro, while the ex-situ participant participated remotely by connecting to a PC with all necessary tools pre-installed via the low-latency screen-sharing software *Parsec*[3].

Upon arrival, the in-situ participant was welcomed by an experimenter, while another experimenter connected with the ex-situ participant via video call. The in-situ participant was then brought

---

[3]https://parsec.app
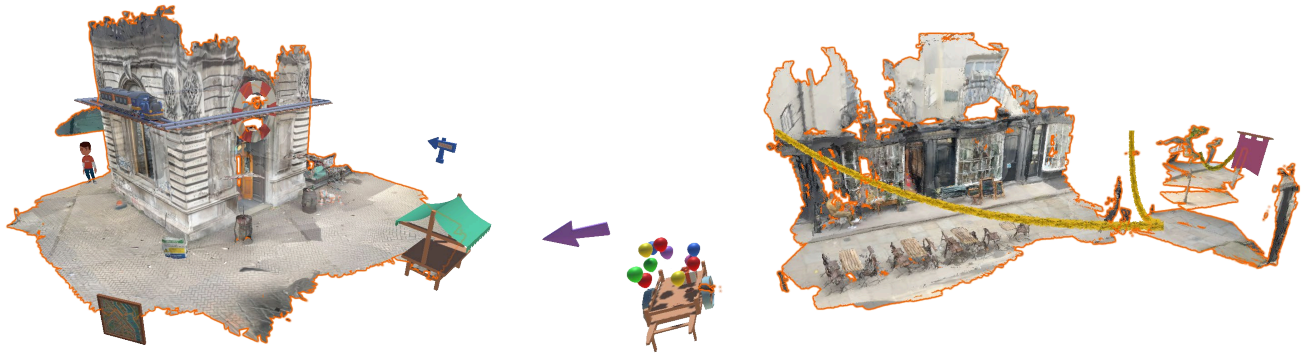
Location A

Location B

Figure 9: Screenshots of the prototypes designed for *Location* A and *Location* B. The *location mesh* is outlined in orange, to which the visible virtual low-poly AR elements are anchored.

to a room where they joined the same video call. The experimenter provided a general introduction to the study context and procedure. Before the first condition, the ex-situ participant was guided through a setup document containing log-in information and instructions for forwarding their microphone to the PC. Once the ex-situ participant connected, latency statistics were recorded, ranging between ~35–120ms (including network, decoding, and encoding latency) across participants.

Each pair completed the task twice: once under the Sync and once under the Async *collaboration mode* conditions, with one condition occurring in *Location* A and the other in *Location* B. At the start of each condition, participants watched a tutorial video tailored to their role and condition. The in-situ participant then proceeded to either *Location* A or *Location* B, assisted by an experimenter, where they received a brief walkthrough of the Sync or Async system. Meanwhile, the ex-situ participant received a similar walkthrough remotely, along with an introduction to the prototype experience, providing background information on its components and initial design decisions.

In the Async condition, during Phase 1 (Feedback Collection & Refinement), the in-situ participant had ten minutes to gather feedback on the prototype experience and any relevant real-world context necessary to address the issues listed on the task clue list. The in-situ participant then proceeded to Phase 2 (Ideation & Prototyping), where they collected artifacts related to new ideas for improving or extending the prototype. Concurrently, the ex-situ participant began Phase 1 by refining and correcting the prototype based on the other participant's generated feedback artifacts. After six minutes, the in-situ participant's Phase 2 concluded, and their artifacts were made available to the ex-situ participant for Phase 2, which commenced after their refinement phase (Phase 1) was completed. The in-situ participant then returned indoors, guided by the experimenter, and both participants completed post-task questionnaires separately.

The same process was followed for the Sync condition, except that the in-situ and ex-situ participants progressed through both

phases simultaneously. Upon completing both conditions, participants were separately interviewed to discuss their experiences across the conditions.

5.5 Measures and Analysis

Each collaborative session was recorded from the perspective of the ex-situ participant. In the Sync condition, this recording included the streamed view of the in-situ participant along with the audio captured from both participants. The recordings were coded to track each participant's feature usage throughout the session. An overview of feature usage is provided in the Supplementary Materials and is discussed in Sec. 7.1.

The post-task questionnaire included the NASA Task Load Index (TLX) [24] to assess workload across conditions using a 7-point Likert scale. All TLX items were included for in-situ participants, whereas the item measuring physical demand ("How physically demanding was the task?") was excluded for ex-situ participants due to their seated, computer-based role in both conditions. An overall task load score was calculated by averaging the responses to the included items.

To measure engagement, we employed the short-form version of the User Engagement Scale (UES-SF) [64], focusing on the subscales of *focused attention*, *perceived usability*, and *reward*. The *aesthetic appeal* subscale was omitted as it was deemed irrelevant. An overall engagement score was obtained by averaging the 5-point Likert scale responses across subscales, adjusting for reverse-coded items.

Additionally, we included several custom items in the post-task questionnaire, which are provided in the Supplementary Materials. For ex-situ participants, custom items assessed their *overall confidence* in the final AR experience as end-users would encounter it in situ, confidence in the successful execution of task components that influence end-user experience, and specific feedback on the goals of the different task phases. Both participant groups also rated perceived overall task performance using a custom item.

Before concluding the study, both participants were interviewed separately. The post-study interviews focused on evaluating their experiences and preferences regarding the systems and workflows in the Sync and Async conditions. Key topics included overall

preference, confidence, communication quality, advantages and limitations of each method, and suggestions for improving system features. The full interview scripts are available in the Supplementary Materials. Two authors of this paper used an affinity diagramming approach to analyze and synthesize themes from the interview transcripts.

6 Exploratory User Study Results

Building on the formative study and resulting requirements, we present the results of our evaluation of CoCreatAR following the user study design detailed in Sec. 5. We first report the self-reported questionnaire results, followed by four themes derived from a thematic analysis of interviews with ex-situ and in-situ participants. Plots that show an overview of the usage of specific CoCreatAR features are provided in the Supplementary Materials.

6.1 Questionnaire Results

We conducted statistical analyses using linear mixed-effects models, accounting for participant variability as a random effect and controlling for confounding factors such as task location and prior experience. All models converged, and model assumptions were verified (linearity, normality, and homoscedasticity). Where applicable, we report the estimated coefficients (β), standard errors (SE), p-values (p), and 95% confidence intervals (CI). Unless stated otherwise, each analysis includes 32 observations (n=32). In this section, we provide full details for (marginally) statistically significant results, whereas an overview of all statistical results is available in the Supplementary Materials.

Cronbach's alpha values indicated acceptable internal consistency for both engagement and task load measures. Engagement items yielded alpha values of 0.740 for in-situ users and 0.764 for ex-situ users. Task load items had alpha values of 0.812 and 0.772 for in-situ and ex-situ users, respectively.

6.1.1 Engagement. The analysis revealed a significant effect of *collaboration mode* on self-reported engagement for both ex-situ and in-situ participants. Ex-situ participants reported significantly higher engagement in the SYNC condition ($\beta = 0.500$, SE = 0.140, $p < 0.001$, 95% CI [0.226, 0.774]). A similar effect was observed for in-situ participants, who reported significantly increased engagement in the SYNC condition ($\beta = 0.340$, SE = 0.130, $p = 0.009$, 95% CI [0.085, 0.595]). A plot of engagement scores is shown in Fig. 10A.

While no significant effect of *location* was found for ex-situ participants ($p = 0.196$), in-situ participants reported significantly higher engagement at *Location* B ($\beta = 0.299$, SE = 0.130, $p = 0.022$, 95% CI [0.044, 0.554]). *Prior experience* with AR or 3D editing tools did not significantly influence self-reported engagement for either group (ex-situ: $p = 0.093$; in-situ: $p = 0.231$).

6.1.2 Task load. Self-reported task load showed no significant difference between SYNC and ASYNC *collaboration modes* for either ex-situ ($\beta = 0.313$, SE = 0.309, $p = 0.312$) or in-situ participants ($\beta = 0.213$, SE = 0.345, $p = 0.538$). A plot of task load scores is shown in Fig. 10B.

Location did not significantly impact self-reported task load for ex-situ participants ($p = 0.419$). However, in-situ participants reported a marginally significant decrease in task load at *Location*
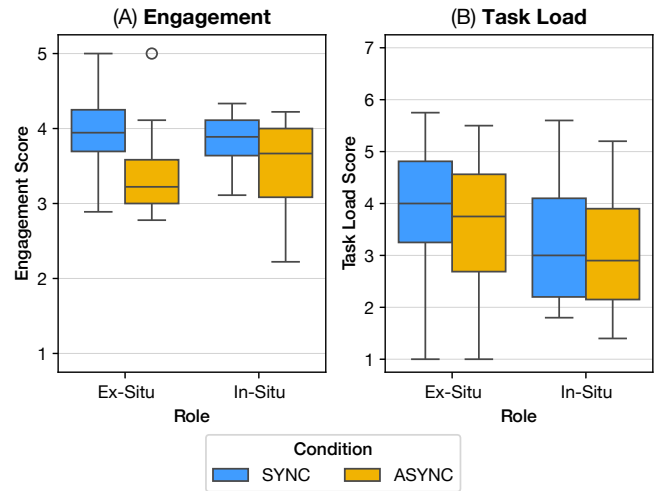

Figure 10: (A) Engagement scores for each role per condition, with significantly higher engagement in the SYNC condition for both in-situ and ex-situ participants. (B) Task load scores for each role (ex-situ and in-situ) per condition. No significant differences were found between conditions.
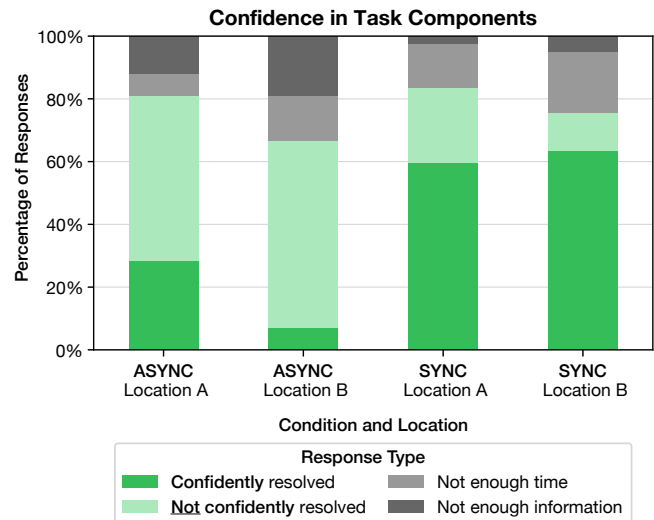

Figure 11: Stacked bar plot illustrating response proportions where participants indicated perceived performance and confidence for each task component across *collaboration mode* and *location* combinations.

B compared to *Location* A ($\beta = -0.637$, SE = 0.345, $p = 0.065$, 95% CI [−1.314, 0.039]). *Prior experience* did not significantly affect self-reported task load for ex-situ participants ($p = 0.162$). However, in-situ participants showed a marginally significant association between lower task load and more AR experience ($\beta = -0.320$, SE = 0.176, $p = 0.070$, 95% CI [−0.666, 0.026]).

6.1.3 Confidence in authored results. Confidence in the number of issues fixed and overall self-reported confidence were both significantly influenced by *collaboration mode.* Participants in the Sync condition reported fixing more issues with confidence than those in the Async condition ($\beta = 3.000$, SE = 0.373, $p < 0.001$, 95% CI [2.270, 3.730], $n = 24$). Additionally, overall confidence scores were significantly higher in the Sync condition ($\beta = 1.125$, SE = 0.387, $p = 0.004$, 95% CI [0.367, 1.883]). An overview of the proportion of participant responses regarding perceived performance and confidence for each task component is shown in Fig. 11.

A marginal effect of *location* was observed for ex-situ participants, with fewer issues confidently fixed at *Location* B ($\beta = -0.667$, SE = 0.373, $p = 0.074$, 95% CI [−1.397, 0.064], $n = 24$). However, *location* did not significantly affect overall confidence scores ($p = 0.747$). *Prior experience* did not significantly impact the number of confidently fixed issues ($p = 0.096$) or overall confidence scores ($p = 0.466$).

6.2 Interview Results

At the start of the interview, each participant was asked which system they would prefer to use if they were to perform a similar task again. Among in-situ participants, the majority (12 out of 16) expressed a preference for the Sync condition, citing advantages such as real-time feedback, improved communication, and enhanced engagement. Two participants indicated that their preference would depend on the scenario, while two favored the Async condition, highlighting the ability to focus better without the distractions of real-time interaction.

Similarly, ex-situ participants also largely favored the Sync condition (12 out of 16), with many emphasizing the value of real-time collaboration and immediate feedback. Three participants indicated that their preference depended on the context, and one favored the Async condition due to the slower pace, which allowed more time to refine their contributions.

Our thematic analysis of interview transcripts resulted in four main themes comparing the synchronous CoCreatAR and asynchronous baseline experiences. We denote in-situ participant quotes with an 'I' (e.g., I2) and ex-situ participants with an 'E' (e.g., E8).

6.2.1 Collaboration with in-situ users facilitated the integration of real-world context into ex-situ users' design process.
In the Async condition, all ex-situ participants struggled to integrate real-world context into their design decisions due to missing information. Ex-situ users were uncertain about object placement and its interaction with the environment, as they could not verify details without live input. One participant noted, "I'm not 100% sure about the real way, if it's low enough or is this blocking other ways because I need to see it from the real end" (E7). Lacking live visual and spatial information, ex-situ participants relied on limited artifacts provided by in-situ participants, restricting their ability to account for dynamic environmental elements. One participant explained, "I tried to put the map onto that light post but I'm not sure whether that's on the post or not" (E7).

In contrast, the Sync condition enabled participants to gain a clearer understanding of real-world context, leading to more informed decision-making in the design process (E12, E10, E9, I5, E17). One ex-situ participant emphasized the importance of real-time

observation, particularly for user flow, explaining that although an object appeared well-positioned, real-world feedback revealed that it interfered with pedestrian flow: "We had a situation where people were walking through one of the objects. It looked like an ideal location for the object on the map [*location mesh*], but in real time and in real life, it wasn't going to work because it was in a walkway" (E10).

The integration of real-world context through synchronous communication also extended to addressing safety and navigation concerns. One ex-situ participant described uncertainty about whether a virtual cart placed in the scene would interfere with foot and cycling traffic, as there were two real obstacles on either side of it. They noted that this uncertainty was mitigated by being able to check the scene in real time with the in-situ user (E12). Another participant highlighted the value of hearing environmental noise in real time, which influenced decisions about setting audio levels in the AR experience: "I need to know the population density there, like what's the environment like, should I put the speaker very loud or not at all" (E9).

Participants also reported that changes in the environment were easier to account for in the Sync condition. One ex-situ user noted that objects had shifted between the time the location mesh was captured and their session, something they would not have noticed without real-time feedback. As they explained, "The barrel, [...] had actually moved... I wouldn't have been able to tell that from the picture or the scan [location mesh]" (E17). Following this realization, they asked the in-situ participant to place cursor markers indicating the correct position of the objects on the barrel.

Ex-situ participants also emphasized that synchronous collaboration allowed them to indirectly experience contextual elements of the real-world environment. One participant underscored the importance of experiencing real-world audio and visual elements: "You can actually hear what they're hearing as well, which is quite important" (E7). This sense of presence contributed to an increased understanding of spatial relationships within the environment. One participant mentioned, "I got a much better idea of the space when the other person was there," as the in-situ user provided real-time feedback and visual information that could not have been fully captured asynchronously (E16). The combination of real-time video, real-time 3D captures of the environment, and verbal feedback gave ex-situ users a more comprehensive view of the environment, enhancing their understanding of how the design fit within real-world context (E7, E9, E10, E12, E16, E17).

6.2.2 Immediate feedback and changes supported confidence in design decisions, mutual understanding, and perceived accuracy of the outcome.
All ex-situ participants noted that the Sync condition facilitated more informative feedback and iterative adjustments, with many additionally highlighting that this strengthened their confidence in the authored outcome (I2, E3, E5, E7, I10, E12-17). The ability to communicate in real time facilitated quicker, informed decisions, with several ex-situ participants noting that the immediate feedback loop significantly reduced the guesswork involved in refining the AR prototype. As E16 stated, "I think that, very strongly, I'd be more confident that the positions that we put those objects, they actually line up with the real world a lot better." This confidence was echoed by other ex-situ participants

who mentioned that making the changes in real-time felt "much, much quicker" (E3), "was a lot more dynamic" (E12), and that "it felt like, once we were done, it was very complete" (E10).

The ability to communicate synchronously and see changes immediately not only built confidence but also improved mutual understanding among pairs. Participants emphasized that the visual nature of the real-time interaction removed ambiguity, leading to more accurate placement of virtual objects. Participant I14 noted that discrepancies were minimized: "I don't think we had a lot of discrepancies in what we discussed about [in the Sync condition], which means probably we had a good sync about the scene."

Beyond the real-time feed, CoCreatAR's additional features — such as the *3D Cursor*, annotation tools, and the ability to capture *Coarse Meshes* and *3D Snapshots* — enhanced mutual understanding between ex-situ and in-situ users. Figure 12 presents examples of feature usage throughout the user study. Participant I16 noted that using the annotation tool to mark exact locations allowed them to "draw exactly where" changes were needed, acting as "the bridge between being there in reality and the scene that [ex-situ user] is seeing." Similarly, I5 emphasized that "drawing in real time is much easier than just explaining it," particularly when identifying hazardous areas. The 3D cursor helped reduce ambiguity by allowing participants to visually reference key elements and create placeholders for alignment (I2, I5, I6, I8–10, I12, I14, I16–17, E6, E8, E10–12, E14, E16–17). Additionally, E3 highlighted how the ability to capture *Coarse Meshes* and *3D Snapshots* enabled the in-situ user "to add more detail to the scene," helping ex-situ users make more informed design decisions.

In contrast, the Async condition often resulted in lower confidence due to the lack of immediate feedback and reliance on static information (E9, E12). As E12 described, "With the asynchronous one, I don't have any information in terms of how much should I move... there's no feedback whatsoever." The absence of real-time interaction forced participants to work with limited information, leading to hesitant decisions. E5 emphasized this challenge: "There was some ineffective information for me in the notes," indicating that static instructions without live verification did not provide the necessary context for accurate design decisions. This lack of real-time verification made it difficult to gauge whether changes were correctly applied. E4 highlighted this issue, stating, "You can't see the change you made, like, in the real world."

In-situ participants also expressed uncertainty about whether they had captured the right content or communicated their observations clearly (I6, I12, I14, I17). One in-situ participant stated, "I wasn't sure it was clear enough" (I14), while another noted that the lack of immediate dialogue required them to provide excessive detail, which still might not ensure accurate interpretation (I17).

6.2.3 *Synchronous authoring increased engagement and encouraged creative exploration through collaborative interaction.* In the Sync condition, participants frequently reported higher levels of engagement, often attributing this to the sense of real-time collaboration and mutual decision-making (I3, I8, I12, I15, E3, E5, E12, E14, E17). For instance, I10 described the synchronous condition as "a lot more fun and engaging" due to the opportunity to work together with another person. Similarly, I14 noted that the synchronous session was "more enjoyable" and led to "higher

engagement" because of the collaborative nature of the task. This sentiment was echoed by E17, highlighting the satisfaction of "working together" and described the experience as akin to "live game testing," suggesting that seeing immediate reactions and feedback from the in-situ user created a more interactive and stimulating process. The ability to see their ideas come to life in real time enriched the creative aspect of the experience, as noted by multiple participants (E9, I17).

Synchronous collaboration also drove a sense of teamwork and shared ownership of the final result, contributing to a more engaging authoring process (E7, I10, I12, E16). One participant explained that "it was fun to chat to another person while I was doing it," which made the task feel more like a collaborative endeavor rather than an individual effort (I6). This increased engagement was reflected in the enthusiasm with which participants approached the synchronous conditions, with one participant stating that they felt "buzzing" with excitement after completing the task together with their collaborator (I12).

In contrast, the Async condition was perceived as less engaging by the majority of participants (I3, I4, E4, I6, E7, I8–I10, I12, E12, I14–I17, E14, E16). Participant I12 compared the two conditions, explaining that the asynchronous condition felt "more like a task as opposed to fun," a sentiment echoed by other participants who described the asynchronous process as more isolated and procedural (I4, I7, E14, I17). E12, for example, mentioned that the asynchronous condition felt "quicker but less satisfying," as there was no immediate feedback or dynamic interaction.

In-situ participants frequently expressed dissatisfaction with working toward a goal without seeing the final result (I3, I4, I5, I6, I7, I12, I16). For example, E16 stated, "It was a weird feeling leaving and then not seeing it change." The lack of real-time collaboration in the Async condition also constrained opportunities for creative exploration. One participant explained that without the ability to brainstorm with another person, the task became "more about checking a box" rather than experimenting with new ideas (I10).

Collaborative brainstorming emerged as a key driver of engagement in the synchronous sessions. Participants reported that having someone to exchange ideas with in real time not only enhanced the creative process but also led to more diverse and spontaneous solutions (I9, I10, I17, E14). Participant I9 emphasized the ease of "drafting concepts" together in the synchronous condition, stating that "we could brainstorm easily" and quickly iterate on suggestions. These interactions not only facilitated creative exploration but also enhanced the sense of shared authorship, which several participants valued (I17, E10).

6.2.4 *Multitasking overwhelmed some participants in synchronous collaboration, while asynchronous workflows were seen as more suitable for certain scenarios.* Although the synchronous condition generally resulted in increased engagement and collaboration, some participants reported that the increased complexity of multitasking within the synchronous workflow was overwhelming (I8, E16, I17). Some participants also noted that the Sync condition increased pressure, as they felt their collaborators had to wait for them to complete tasks (I4, E12, E14).

Participant E16 pointed out that managing multiple tasks simultaneously — such as communicating with the in-situ user, observing
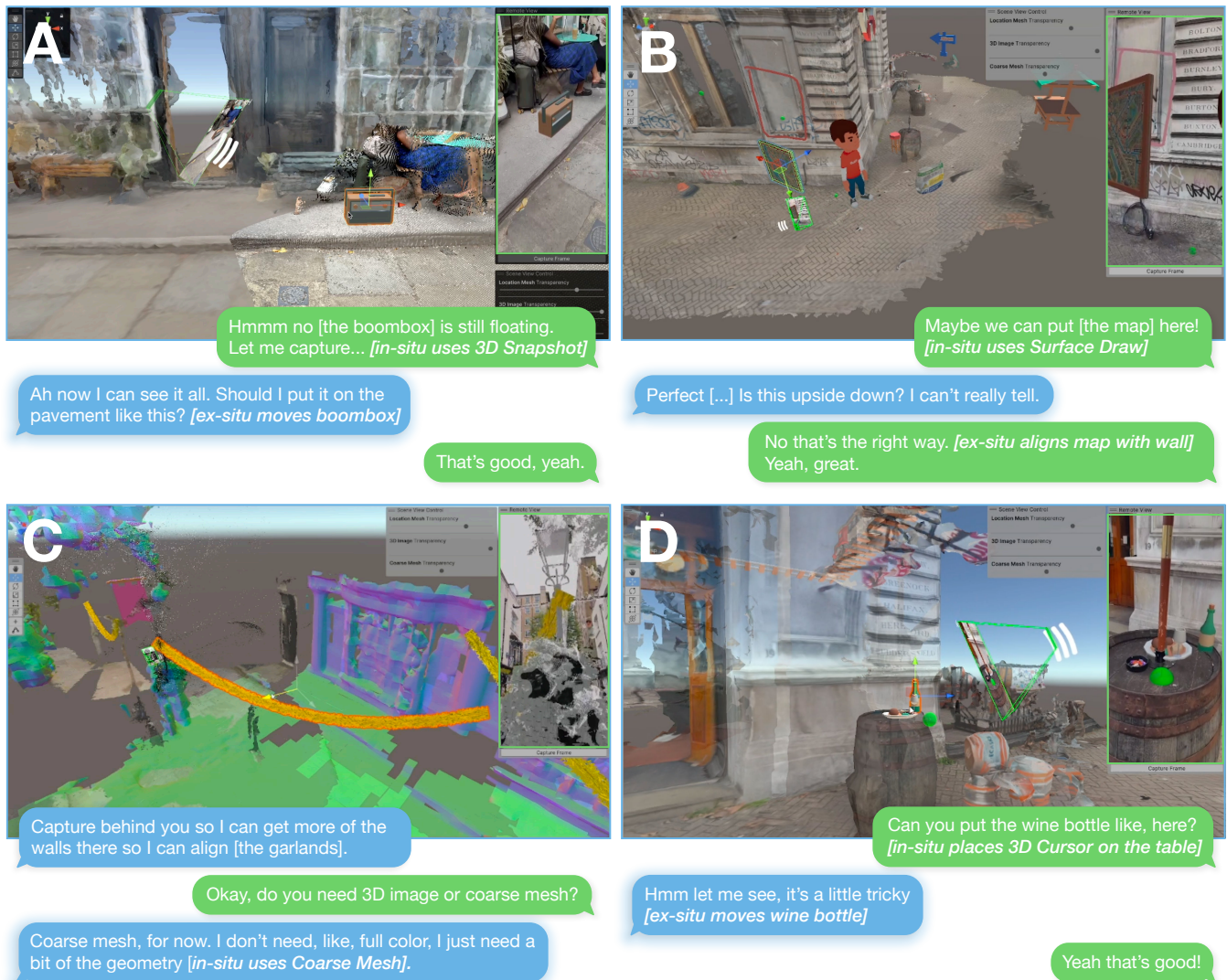
Figure 12: Overview of CoCreatAR feature usage during Phase 1, shown as *ex-situ* perspective screenshots. Participant conversations are shown in color-coded speech bubbles: in-situ (green) and ex-situ (blue). Speech bubbles with a *glow* indicate utterances made at the moment of the screenshot. (A) Alignment of a boombox based on spatial context captured using the *3D Snapshot* feature; (B) The *ex-situ* participant moving the map to a position on the wall as specified by the *in-situ* participant through *Surface Drawing*; (C) Alignment of a garland to a previously unmapped region of the street using the *Coarse 3D Mesh* feature; (D) Alignment of misplaced food items based on *in-situ* input using the *3D Cursor*.

the scene, and making design adjustments — led to cognitive overload: "I felt like I was doing too many things at once. Talking to [the in-situ user] and trying to watch the environment was sometimes just too much" (E16). Similarly, I10 expressed that for users unfamiliar with the system, synchronous interactions might feel overwhelming, particularly due to the need to navigate while processing real-time feedback: "For someone who's not experienced, it was just a lot. You're trying to follow directions, but there's so much going on" (I10). This aligns with the feedback of several other participants less familiar with AR, who wished they had more time to practice ahead of the task (I3, I6, I10, I13, E10, E11, E16).

Several participants indicated that the synchronous condition was ideal for scenarios where immediate feedback was critical, but the asynchronous condition was more suitable when working in overwhelming environments. Participant I10 described how synchronous interaction could become overwhelming in crowded environments with high noise levels, explaining that "if you're dealing with a busy place and someone is talking in your ear, it gets really overwhelming" (I10).

Participants also recognized that both synchronous and asynchronous workflows had their strengths depending on the scenario. While some participants felt overwhelmed by multitasking in the

synchronous condition, they still acknowledged its value for tasks requiring rapid decision-making or creative brainstorming. E12 noted, "It's definitely harder when you have to do everything at once, but I still think the synchronous one is better when you're trying to bounce ideas off someone" (E12). Participant I16 described how they viewed the SYNC condition as most applicable for complex experiences, whereas the ASYNC condition might be more suitable "if it was, like, a smaller experience with one object." E9 also noted scenarios where asynchronous workflows could be beneficial, such as when in-situ participants already have "enough information to already produce what I want" (E9).

Moreover, participants reflected on how asynchronous and synchronous workflows or features could be complementary and applied at different stages of the development process (I5, E15, E16). For example, E15 highlighted a hybrid approach: "I kind of see it as you start with asynchronous, then you do synchronous to refine it, to collect feedback on your experience." Other participants saw potential in a hybrid system that flexibly integrates asynchronous and synchronous workflows while incorporating all of CoCre-atAR's capturing and annotation features (I5, E15) and enabling lightweight in-situ editing as in addition (E16).

7 Discussion

In this section, we reflect on the results of the user study, discuss our findings in the context of our formative study and related work, and offer recommendations for future authoring systems of site-specific outdoor AR experiences.

7.1 Enhancing Environmental Understanding Through Capture and Communication Tools

Our findings indicate that synchronous collaboration using CoCre-atAR significantly enhanced the ability of ex-situ participants to integrate real-world context into their design process. In particular, ex-situ participants in the synchronous condition demonstrated a more comprehensive understanding of environmental dynamics, such as user flow and spatial relationships. In this subsection, we contextualize how each feature group contributed to this outcome, drawing on insights from post-study interviews and observations made when coding feature usage counts (see Supplementary Materials for plots).

Participant feedback emphasized that real-time video and audio communication was the most critical feature of CoCreatAR. It provided ex-situ participants with immediate visual context from the in-situ user's perspective and supported stable communication. This was especially valuable when in-situ participants felt overwhelmed, enabling them to carry out the task nonetheless. Notably, background noise captured by the in-situ user's microphone proved useful for decisions regarding audio components (Sec. 6.2.1) — an issue that, according to P1 of our formative study, could only be resolved when in situ (Sec. 3.2.3). Additionally, some ex-situ participants mentioned that background audio increased their sense of presence in the in-situ environment.

Surface Draw was commonly used due to its task relevance, as objects were typically tied to physical referents (e.g., the map on the wall in Fig. 12B). *Air Draw* was primarily used to indicate user

paths and, occasionally, for the creation of 3D wireframe place-holder objects during brainstorming. Labeled annotations were used less often than color annotations, likely because ex-situ participants were not expected to revisit annotations asynchronously at a later stage during the study, unlike in real-world scenarios where annotations might be revisited over time.

Coarse Mesh was often used to capture large areas (e.g., buildings in Fig. 12C), while the *3D Snapshot* feature was favored for specific points of interest (e.g., a curb and table set in Fig. 12A). Although the number of *3D Snapshot* captures was higher overall, the longer capture time required for *Coarse Meshes* balanced the frequency of their use. The lack of color in *Coarse Meshes* was not described as a major issue, though one participant creatively captured over ten *3D Snapshots* to obtain a large colored point cloud. *3D Cursors* were mainly used by in-situ participants for object alignment (Fig. 12D) and for supporting deictic references, while ex-situ participants used *3D Cursors* to guide in-situ users' movements or capture actions.

Recommendations for future work. Enhancing spatial representations and integrating richer metadata layers could expand CoCreatAR's potential for complex real-world locations. Future work could leverage 3D Gaussian splatting techniques [32] and their temporal extensions [53, 94] to support more intricate design tasks and precise alignment. Automatically capturing metadata such as pedestrian flow [75] or hazards [80] during on-site visits could streamline context handling for both synchronous and asynchronous ex-situ authoring.

Future work could also leverage insights from Fussell et al. [18] to enhance collaboration in CoCreatAR. For instance, visualizing collaborators' actions, such as object selection or transformation, could improve task awareness. Alternative in-situ interfaces, including head-mounted displays with larger fields of view, might offer a more comprehensive perspective and facilitate natural interactions. Furthermore, hand tracking could enable intuitive gestures, as observed in our study when participants occasionally employed deictic gestures in front of the device camera [19, 34].

7.2 Task Load and Multitasking Challenges

While our results regarding task load are inconclusive, they offer early insights into the impact of task context, multitasking demands, and individual experiences. A slight trend visible in Fig. 10B suggests a higher task load in the SYNC condition for both roles, although no statistically significant differences were observed across conditions. Notably, qualitative feedback revealed that a subset of participants, particularly those with less prior experience, expressed feeling overwhelmed during the SYNC condition. This feedback provides preliminary insights that challenge **Hypothesis 1**, suggesting that ex-situ participants may experience higher, rather than lower, task load in the SYNC condition. At the same time, qualitative observations align with **Hypothesis 2**, as in-situ participants reported multitasking challenges indicative of higher task load in the SYNC condition. Based on the interviews, we conclude that an overall increase in task load during the SYNC condition could have arisen from the substantial volume of information flow combined with the demands of concurrent coordination and communication. For in-situ users, this challenge appeared compounded by the need to

safely navigate the real world—an activity previously recognized as inherently demanding [52].

Some participants reported that limited familiarity with the system contributed to their sense of being overwhelmed. We hypothesize that these experiences may be influenced by participants' background knowledge and individual personality traits [29]. As we detail in Sec. 8, follow-up research could explore these hypotheses further in larger-scale studies.

Participants' qualitative feedback also highlighted potential *process losses*, defined as productivity reductions that occur when individuals collaborate synchronously rather than working independently [78]. Specifically, some participants reported that the multitasking demands of the Sync condition, such as communicating with their collaborator while performing tasks, led to perceived inefficiencies. Related challenges included waiting for collaborators to complete tasks (e.g., annotation by an in-situ user or scene editing by an ex-situ user) and feeling pressured to avoid blocking their partner's progress (Sec. 6.2.4). Guo et al. [22] similarly observed potential process losses in synchronous collaboration during an authoring task, specifically noting that skill mismatches within pairs could limit creativity. While we did not observe direct evidence of this particular form of process loss, we theorize that the clearly defined roles and responsibilities assigned to participants, along with the semi-guided nature of the task, may have mitigated these effects.

Recommendations for future work. Future work could benefit from more extensive onboarding processes to familiarize users with system features and clarify collaborators' roles. Additionally, future iterations of CoCreatAR could address challenges related to task load by streamlining workflows. For instance, simplified scene editing for ex-situ users could be achieved through modular templates [30] or AI-driven authoring tools [16, 68, 74]. In parallel, in-situ users could benefit from AI-assisted annotation methods, such as transcribed audio notes [35, 40].

To improve orientation and reduce cognitive load for in-situ users in complex environments, future systems might integrate visual aids such as mini-maps [79] or directional cues [44], in addition to enhanced visual indicators for collaborator actions as noted in Sec. 7.1. Lastly, for ex-situ users, features such as user-perspective scene rendering [6] and selective visualization of information layers [33, 89] could help reduce visual overload.

7.3 Engagement and Creative Exploration

Both ex-situ and in-situ participants reported higher engagement in the Sync condition, supporting **Hypothesis 3**. Real-time interaction supported effective teamwork and creative exploration, which the vast majority of participants described as more enjoyable than the Async workflow. This finding aligns with prior research on *social flow*, which found that interdependent team settings amplify engagement compared to solitary workflows [90].

Prior work on in-situ AR game level editing [58] found that users were highly engaged in both creating AR content and observing others interact with it. Similarly, in our study, in-situ users were engaged by experiencing and interacting with the AR content as it materialized in their environment, while ex-situ users likened the

process to live game testing, emphasizing how immediate in-situ feedback served as a source of inspiration (Sec. 6.2.3).

The advantages of synchronous collaboration observed in our study are further supported by findings from Guo et al. [22], who identified synchronous authoring as a key driver of engagement and creativity compared to asynchronous authoring. Notably, their work additionally revealed a user preference for *co-located* over *remote* synchronous collaborative authoring, which relied on standard video conferencing. This preference underscores the importance of designing remote collaboration tools that replicate the dynamics of co-located experiences more closely. For instance, I3 expressed a desire to see their collaborator, particularly when working together for the first time.

Recommendations for future work. While CoCreatAR currently employs avatars to simulate the embodied perspective of co-located collaboration, future research could investigate enhancing interpersonal connection through spatial video-based avatars, as demonstrated in recent work [69, 87]. Another avenue for future research is to explore how creativity can be supported in larger groups, which offer potential for richer creative outcomes [65] and greater productivity. Although larger groups may face greater coordination overhead, strategies such as role specialization, as employed by XRDirector [55], could streamline the workflows of larger teams.

7.4 Confidence in End-User Experience Through In-Situ Feedback

Our findings demonstrate that participants in the Sync condition reported significantly higher confidence in the authored AR experiences than those in the Async condition, supporting **Hypothesis 4**. Participant feedback indicated that this confidence stemmed from the ability to iteratively verify the designed experience in real time, ensuring alignment with the spatial and contextual dynamics of the real-world environment. In contrast, participants in the Async condition struggled with uncertainty due to reliance on static data and delayed or incomplete feedback, which often led to hesitation in their design decisions.

This positive influence of representative in-situ feedback aligns with prior research in ubiquitous computing that emphasized the critical role of situated evaluation in achieving a reliable understanding of how applications behave in real-world contexts [12, 72], an approach that our work both embraces and extends.

Notably, much of this research has focused on asynchronous techniques, such as event logging and recordings, to collect in-situ feedback [61, 72]. Our findings contribute a new perspective by highlighting the opportunity and value of evaluating and adapting an application simultaneously, a method that our user study revealed to be particularly effective for the iterative refinement of site-specific outdoor AR experiences.

Recommendations for future work. Future work could explore how author confidence and the perceived quality of AR end-user experiences evolve over multiple revisits of a particular site, especially under dynamic environmental conditions. Additionally, research could focus on developing quantitative metrics that represent the quality of AR experiences, as explored by ARCHIE [46], to

provide authors with additional objective insights. While quantitative metrics typically require controlled environments or ground truth sources, metrics derived from user behaviors (e.g., frequency of manual relocalization triggers) could be a practical alternative. Finally, author confidence could be supported by incorporating constraints for contextual adaptation, such as semantic referents and rules [68, 86], which have been shown to improve confidence when used alongside virtual simulation environments for indoor AR tutorial authoring [68].

7.5 Enabling Hybrid Approaches to AR Experience Authoring

Despite the advantages of synchronous collaboration, participants identified some scenarios where asynchronous workflows were desired. In contexts with overwhelming environmental conditions or when tasks did not require immediate feedback, asynchronous methods offered more flexibility and reduced pressure. This insight underscores the importance of providing flexible authoring workflows that can adapt to the diverse needs of authoring teams [36], including collaboration across *place* and *time* [28].

Recommendations for future work. Studying the impact of switching between synchronous and asynchronous modes on collaborative processes and authoring outcomes could reveal opportunities for optimization. Furthermore, to support flexible workflows, future work could explore designing hybrid systems that integrate session management and version control for iterative collaborative authoring [95, 97]. Additionally, techniques to visualize synchronous collaboration sessions as artifacts for asynchronous follow-up work could increase and prolong the value of collaborative sessions [11, 27, 91, 92].

8 Limitations

While our study provides valuable early insights into synchronous and asynchronous collaborative authoring for site-specific outdoor AR experiences, several limitations should be noted.

First, there was variability between the two locations used in the study, which differed in environmental characteristics such as foot traffic, noise levels, and spatial layout. Although we counterbalanced the assignment of locations to conditions, maintained a consistent task across locations, and did not identify a major impact of location in our analyses, these differences may still have influenced participants' experiences and performance.

Second, the sample size of our user study was small, with 32 participants forming 16 pairs. This limited the power of our statistical analysis to detect certain effects. For instance, while qualitative data indicated that some novice participants reported increased task load, this observation may not have been fully reflected in the quantitative data due to the small sample and variability in participants' experience levels. These findings highlight the need for future research with larger sample sizes and specific target groups to examine the role of prior experience and explore strategies for lowering barriers to AR authoring, in line with the recommendations of Ashtari et al. [5].

Third, the novelty of the system compared to current methods may have led to participant response bias [13], potentially influencing measures such as engagement. Longitudinal studies with extended and repeated exposure could help isolate these effects.

Finally, our study was limited to a specific type of site-specific AR experience and a particular set of tasks. Consequently, the findings may not generalize to other types and scales of AR application development, such as those involving larger teams or multiple co-located users collaborating on extensive projects. Future research could examine a broader range of AR authoring scenarios that vary in task type and group size, including those with multiple in-situ and ex-situ users engaged in synchronous distributed authoring (e.g., guided tours over larger areas) and environmental exploration.

9 Conclusion

This paper examined the challenges of authoring site-specific outdoor AR experiences, which are often constrained by incomplete and outdated world representations and limited access to evolving real-world conditions. Our formative study revealed that developers and designers frequently encounter these limitations, necessitating costly and time-consuming on-site visits to capture environmental details, assess user flow, and ensure contextual relevance. Based on these insights, we identified key requirements for integrating real-world context into remote authoring workflows, leading to the development of CoCreatAR, an asymmetric collaborative authoring system that facilitates synchronous collaboration between ex-situ (i.e., *off-site*, *remote*) developers and in-situ (i.e., *on-site*) collaborators.

Our exploratory user study demonstrated that this approach mitigates key challenges by enhancing confidence in authored results, stimulating engagement and creativity, and enabling direct iterative refinements informed by up-to-date environmental data. At the same time, our findings highlight multitasking demands as a challenge in synchronous collaboration and emphasize the need for a balanced integration of synchronous and asynchronous workflows. Situating these findings within the broader landscape of AR authoring and remote collaboration, we provided recommendations for future work. We hope this research motivates further exploration of methods for building, testing, and evaluating site-specific AR experiences, contributing to the future of immersive and contextually grounded interactive applications.

Acknowledgments

References

[1] Gregory D. Abowd, Anind K. Dey, Peter J. Brown, Nigel Davies, Mark Smith, and Pete Steggles. 1999. Towards a Better Understanding of Context and Context-Awareness. In *Handheld and Ubiquitous Computing*, Hans-W. Gellersen (Ed.).

Springer, Berlin, Heidelberg, 304–307. https://doi.org/10.1007/3-540-48157-5_29

[2] Adobe. 2024. Adobe Aero: Augmented Reality. https://www.adobe.com/uk/products/aero.html

[3] Apple. 2024. Reality Composer: Build Augmented Reality. https://apps.apple.com/us/app/reality-composer/id1462358802

[4] Apple. 2024. Shortcuts User Guide. https://support.apple.com/en-gb/guide/shortcuts/ios

[5] Narges Ashtari, Andrea Bunt, Joanna McGrenere, Michael Nebeling, and Parmit K. Chilana. 2020. Creating Augmented and Virtual Reality Applications: Current Practices, Challenges, and Opportunities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–13. https://doi.org/10.1145/3313831.3376722

[6] Domagoj Baričević, Cha Lee, Matthew Turk, Tobias Höllerer, and Doug A. Bowman. 2012. A hand-held AR magic lens with user-perspective rendering. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 197–206. https://doi.org/10.1109/ISMAR.2012.6402557

[7] Virginia Braun, Victoria Clarke, Nikki Hayfield, and Gareth Terry. 2019. Thematic Analysis. In *Handbook of Research Methods in Health Social Sciences*, Pranee Liamputtong (Ed.). Springer, Singapore, 843–860. https://doi.org/10.1007/978-981-10-5251-4_103

[8] BroadcastTech. 2023. Disney & Snapchat create Kingdom Of The Planet Of The Apes AR lenses. https://www.broadcastnow.co.uk/tech/disney-and-snapchat-create-kingdom-of-the-planet-of-the-apes-ar-lenses/5193320.article

[9] Jacky Cao, Kit-Yung Lam, Lik-Hang Lee, Xiaoli Liu, Pan Hui, and Xiang Su. 2023. Mobile Augmented Reality: User Interfaces, Frameworks, and Intelligence. *Comput. Surveys* 55, 9 (Sept. 2023), 1–36. https://doi.org/10.1145/3557999

[10] Marco Cavallo and Angus G. Forbes. 2019. CAVE-AR: A VR Authoring System to Interactively Design, Simulate, and Debug Multi-user AR Experiences. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Osaka, Japan, 872–873. https://doi.org/10.1109/VR.2019.8798148

[11] Hyunsung Cho, Matthew L. Komar, and David Lindlbauer. 2023. RealityReplay: Detecting and Replaying Temporal Changes In Situ Using Mixed Reality. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 3 (Sept. 2023), 1–25. https://doi.org/10.1145/3610888

[12] Andy Crabtree, Alan Chamberlain, Rebecca E. Grinter, Matt Jones, Tom Rodden, and Yvonne Rogers. 2013. Introduction to the Special Issue of "The Turn to The Wild". *ACM Trans. Comput.-Hum. Interact.* 20, 3 (July 2013), 13:1–13:4. https://doi.org/10.1145/2491500.2491501

[13] Nicola Dell, Vidya Vaidyanathan, Indrani Medhi, Edward Cutrell, and William Thies. 2012. "Yours is better!": participant response bias in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1321–1330. https://doi.org/10.1145/2207676.2208589

[14] Ruofei Du, David Li, and Amitabh Varshney. 2019. Geollery: A Mixed Reality Social Media Platform. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–13. https://doi.org/10.1145/3290605.3300915

[15] Unreal Engine. 2024. Unreal Engine: We make the engine. You make it unreal. https://www.unrealengine.com/

[16] João Marcelo Evangelista Belo, Mathias N. Lystbæk, Anna Maria Feit, Ken Pfeuffer, Peter Kán, Antti Oulasvirta, and Kaj Grønbæk. 2022. AUIT – the Adaptive User Interfaces Toolkit for Designing XR Applications. In *The 35th Annual ACM Symposium on User Interface Software and Technology*. ACM, Bend OR USA, 1–16. https://doi.org/10.1145/3526113.3545651

[17] Sebastian J Friston, Ben J Congdon, David Swapp, Lisa Izzouzi, Klara Brandstätter, Daniel Archer, Otto Olkkonen, Felix Johannes Thiel, and Anthony Steed. 2021. Ubiq: A System to Build Flexible Social Virtual Reality Experiences. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology (VRST '21)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3489849.3489871

[18] Susan R. Fussell, Robert E. Kraut, and Jane Siegel. 2000. Coordination of Communication: Effects of Shared Visual Context on Collaborative Work. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work (CSCW '00)*. Association for Computing Machinery, New York, NY, USA, 21–30. https://doi.org/10.1145/358916.358947

[19] Susan R. Fussell, Leslie D. Setlock, Jie Yang, Jiazhi Ou, Elizabeth Mauer, and Adam D. I. Kramer. 2004. Gestures Over Video Streams to Support Remote Collaboration on Physical Tasks. *Human–Computer Interaction* 19, 3 (Sept. 2004), 273–309. https://doi.org/10.1207/s15327051hci1903_3

[20] Lei Gao, Huidong Bai, Weiping He, Mark Billinghurst, and Robert W. Lindeman. 2018. Real-time visual representations for mobile mixed reality remote collaboration. In *SIGGRAPH Asia 2018 Virtual & Augmented Reality (SA '18)*. Association for Computing Machinery, New York, NY, USA, 1–2. https://doi.org/10.1145/3275495.3275515

[21] Google. 2024. Geospatial Creator: Make the world your canvas. https://developers.google.com/ar/geospatialcreator

[22] Anhong Guo, Ilter Canberk, Hannah Murphy, Andrés Monroy-Hernández, and Rajan Vaish. 2019. Blocks: Collaborative and Persistent Augmented Reality Experiences. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (Sept. 2019), 83:1–83:24. https://doi.org/10.1145/3351241

[23] Jo E. Hannay, Tore Dybå, Erik Arisholm, and Dag I. K. Sjøberg. 2009. The effectiveness of pair programming: A meta-analysis. *Information and Software Technology* 51, 7 (July 2009), 1110–1122. https://doi.org/10.1016/j.infsof.2009.02.001

[24] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*, Peter A. Hancock and Najmedin Meshkati (Eds.). Human Mental Workload, Vol. 52. North-Holland, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[25] Tobias Höllerer, Steven Feiner, Tachio Terauchi, Gus Rashid, and Drexel Hallaway. 1999. Exploring MARS: developing indoor and outdoor user interfaces to a mobile augmented reality system. *Computers & Graphics* 23, 6 (1999), 779–785. https://doi.org/10.1016/S0097-8493(99)00103-X Publisher: Elsevier.

[26] Hyekyung Imottesjo, Liane Thuvander, Monica Billger, Peter Wallberg, Gustav Bodell, Jaan-Henrik Kain, and Stig Anton Nielsen. 2020. Iterative Prototyping of Urban CoBuilder: Tracking Methods and User Interface of an Outdoor Mobile Augmented Reality Tool for Co-Designing. *Multimodal Technologies and Interaction* 4, 2 (June 2020), 26. https://doi.org/10.3390/mti4020026 Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

[27] Andrew Irlitti, Ross T. Smith, Stewart Von Itzstein, Mark Billinghurst, and Bruce H. Thomas. 2016. Challenges for Asynchronous Collaboration in Augmented Reality. In *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*. 31–35. https://doi.org/10.1109/ggfv62

[28] Robert Johansen. 1988. *GroupWare: Computer Support for Business Teams*. The Free Press, USA.

[29] Oliver P. John and Sanjay Srivastava. 1999. The Big Five Trait taxonomy: History, measurement, and theoretical perspectives. In *Handbook of personality: Theory and research, 2nd ed.* Guilford Press, New York, NY, US, 102–138.

[30] Audrius Jurgelionis, Leif Oppermann, Lisa Blum, and Richard Wetzel. 2012. Shapes, marbles and pebbles: Template-based content creation for location-based games. (July 2012). https://repository.lincoln.ac.uk/articles/conference_contribution/Shapes_marbles_and_pebbles_template-based_content_creation_for_location-based_games/25169999/2 Publisher: University of Lincoln.

[31] Shunichi Kasahara and Jun Rekimoto. 2015. JackIn head: immersive visual telepresence system with omnidirectional wearable camera for remote collaboration. In *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology*. ACM, Beijing China, 217–225. https://doi.org/10.1145/2821592.2821608

[32] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkuehler, and George Drettakis. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Trans. Graph.* 42, 4 (July 2023), 139:1–139:14. https://doi.org/10.1145/3592433

[33] Mik Kersten and Gail C. Murphy. 2006. Using task context to improve programmer productivity. In *Proceedings of the 14th ACM SIGSOFT international symposium on Foundations of software engineering (SIGSOFT '06/FSE-14)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/1181775.1181777

[34] Daehwa Kim, Vimal Mollyn, and Chris Harrison. 2023. WorldPoint: Finger Pointing as a Rapid and Natural Trigger for In-the-Wild Mobile Interactions. *Proceedings of the ACM on Human-Computer Interaction* 7, ISS (Oct. 2023), 357–375. https://doi.org/10.1145/3626478

[35] Tae Soo Kim, Seungsu Kim, Yoonseo Choi, and Juho Kim. 2021. Winder: Linking Speech and Visual Objects to Support Communication in Asynchronous Collaboration. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–17. https://doi.org/10.1145/3411764.3445686

[36] Veronika Krauß, Alexander Boden, Leif Oppermann, and René Reiners. 2021. Current Practices, Challenges, and Design Implications for Collaborative AR/VR Application Development. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3411764.3445335

[37] Veronika Krauß, Michael Nebeling, Florian Jasche, and Alexander Boden. 2022. Elements of XR Prototyping: Characterizing the Role and Use of Prototypes in Augmented and Virtual Reality Design. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–18. https://doi.org/10.1145/3491102.3517714

[38] Daniël Lakens. 2022. Sample Size Justification. *Collabra: Psychology* 8, 1 (March 2022), 33267. https://doi.org/10.1525/collabra.33267

[39] Tobias Langlotz, Stefan Mooslechner, Stefanie Zollmann, Claus Degendorfer, Gerhard Reitmayr, and Dieter Schmalstieg. 2012. Sketching up the world: in situ authoring for mobile Augmented Reality. *Personal and Ubiquitous Computing* 16, 6 (Aug. 2012), 623–630. https://doi.org/10.1007/s00779-011-0430-0

[40] Tobias Langlotz, Holger Regenbrecht, Stefanie Zollmann, and Dieter Schmalstieg. 2013. Audio stickies: visually-guided spatial audio annotations on a mobile augmented reality platform. In *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration (OzCHI '13)*. Association for Computing Machinery, New York, NY, USA, 545–554. https://doi.org/10.1145/2541016.2541022

[41] Benjamin Lee, Michael Sedlmair, and Dieter Schmalstieg. 2023. Design Patterns for Situated Visualization in Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics* (2023), 1–12. https://doi.org/10.1109/TVCG.2023.3327398

[42] Gun A. Lee and Gerard J. Kim. 2009. Immersive authoring of Tangible Augmented Reality content: A user study. *Journal of Visual Languages & Computing* 20, 2 (2009), 61–79. https://doi.org/10.1016/j.jvlc.2008.07.001

[43] Gun A. Lee, Gerard J. Kim, and Mark Billinghurst. 2005. Immersive authoring: What You eXperience Is What You Get (WYXIWYG). *Commun. ACM* 48, 7 (July 2005), 76–81. https://doi.org/10.1145/1070838.1070840

[44] Jaewook Lee, Fanjie Jin, Younsoo Kim, and David Lindlbauer. 2022. User Preference for Navigation Instructions in Mixed Reality. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 802–811. https://doi.org/10.1109/VR51125.2022.00102 ISSN: 2642-5254.

[45] Legoland. 2023. Augmented reality experience. https://www.legoland.co.uk/explore/theme-park/whats-new/augmented-reality-experience/

[46] Sarah M. Lehman, Haibin Ling, and Chiu C. Tan. 2020. ARCHIE: A User-Focused Framework for Testing Augmented Reality Applications in the Wild. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Atlanta, GA, USA, 903–912. https://doi.org/10.1109/VR46266.2020.00013

[47] Germán Leiva, Jens Emil Grønbæk, Clemens Nylandsted Klokmose, Cuong Nguyen, Rubaiat Habib Kazi, and Paul Asente. 2021. Rapido: Prototyping Interactive AR Experiences through Programming by Demonstration. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. ACM, Virtual Event USA, 626–637. https://doi.org/10.1145/3472749.3474774

[48] David Lindlbauer and Andy D. Wilson. 2018. Remixed Reality: Manipulating Space and Time in Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–13. https://doi.org/10.1145/3173574.3173703

[49] Wei Hong Lo, Stefanie Zollmann, and Holger Regenbrecht. 2021. From off-site to on-site: A Flexible Framework for XR Prototyping in Sports Spectating. In *2021 36th International Conference on Image and Vision Computing New Zealand (IVCNZ)*. IEEE, Tauranga, New Zealand, 1–6. https://doi.org/10.1109/IVCNZ54163.2021.9653277

[50] Ziwen Lu, Jingyi Zhang, Kalila Shapiro, Nels Numan, Simon Julier, and Anthony Steed. 2023. Reviving the Euston Arch: A Mixed Reality Approach to Cultural Heritage Tours. IEEE Computer Society, 821–826. https://doi.org/10.1109/ISMAR-Adjunct60411.2023.00181

[51] Blair MacIntyre, Maribeth Gandy, Steven Dow, and Jay David Bolter. 2004. DART: a toolkit for rapid design exploration of augmented reality experiences. In *Proceedings of the 17th annual ACM symposium on User interface software and technology (UIST '04)*. Association for Computing Machinery, New York, NY, USA, 197–206. https://doi.org/10.1145/1029632.1029669

[52] Munir Makhmutov, Timur Asapov, and Joseph Alexander Brown. 2021. Safety Risks in Location-Based Augmented Reality Games. In *Entertainment Computing – ICEC 2021*, Jannicke Baalsrud Hauge, Jorge C. S. Cardoso, Licínio Roque, and Pedro A. Gonzalez-Calero (Eds.). Springer International Publishing, Cham, 457–464. https://doi.org/10.1007/978-3-030-89394-1_39

[53] Marko Mihajlovic, Sergey Prokudin, Siyu Tang, Robert Maier, Federica Bogo, Tony Tung, and Edmond Boyer. 2024. SplatFields: Neural Gaussian Splats for Sparse 3D and 4D Reconstruction. In *Computer Vision – ECCV 2024: 18th European Conference, Milan, Italy, September 29–October 4, 2024, Proceedings, Part II*. Springer-Verlag, Berlin, Heidelberg, 313–332. https://doi.org/10.1007/978-3-031-72627-9_18

[54] Michael Nebeling. 2022. XR tools and where they are taking us: characterizing the evolving research on augmented, virtual, and mixed reality prototyping and development tools. *XRDS: Crossroads, The ACM Magazine for Students* 29, 1 (Oct. 2022), 32–38. https://doi.org/10.1145/3558192

[55] Michael Nebeling, Katy Lewis, Yu-Cheng Chang, Lihan Zhu, Michelle Chung, Piaoyang Wang, and Janet Nebeling. 2020. XRDirector: A Role-Based Collaborative Immersive Authoring System. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–12. https://doi.org/10.1145/3313831.3376637

[56] Michael Nebeling, Janet Nebeling, Ao Yu, and Rob Rumble. 2018. ProtoAR: Rapid Physical-Digital Prototyping of Mobile Augmented Reality Applications. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–12. https://doi.org/10.1145/3173574.3173927

[57] Michael Nebeling and Maximilian Speicher. 2018. The Trouble with Augmented Reality/Virtual Reality Authoring Tools. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 333–337. https://doi.org/10.1109/ISMAR-Adjunct.2018.00098

[58] Gary Ng, Joon Gi Shin, Alexander Plopski, Christian Sandor, and Daniel Saakes. 2018. Situated Game Level Editing in Augmented Reality. In *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction*. ACM, Stockholm Sweden, 409–418. https://doi.org/10.1145/3173225.3173230

[59] Niantic. 2024. How to Use the Geospatial Browser. https://lightship.dev/docs/ardk/how-to/vps/tooling/use_geospatial_browser/

[60] Niantic. 2024. Niantic SDK for Unity. https://www.nianticspatial.com/augment/sdk-for-unity

[61] Júlia Carneiro de Almeida Nogueira, Alex Sandro Gomes, Amadeu Sá de Campos Filho, and Fernando Moreira. 2023. Effectiveness of embodied evaluation of mobile applications: A qualitative study. *Heliyon* 9, 6 (June 2023), e17043. https://doi.org/10.1016/j.heliyon.2023.e17043

[62] Nels Numan, Ziwen Lu, Benjamin Congdon, Daniele Giunchi, Alexandros Rotsidis, Andreas Lernis, Kyriakos Larmos, Tereza Kourra, Panayiotis Charalambous, Yiorgos Chrysanthou, Simon Julier, and Anthony Steed. 2023. Towards Outdoor Collaborative Mixed Reality: Lessons Learnt from a Prototype System. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. Shanghai, China, 113–118. https://doi.org/10.1109/VRW58643.2023.00029

[63] Nels Numan and Anthony Steed. 2022. Exploring User Behaviour in Asymmetric Collaborative Mixed Reality. In *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology*. ACM, Tsukuba, Japan, 11. https://doi.org/10.1145/3562939.3565630

[64] Heather L. O'Brien, Paul Cairns, and Mark Hall. 2018. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies* 112 (April 2018), 28–39. https://doi.org/10.1016/j.ijhcs.2018.01.004

[65] Paul B. Paulus and Bernard A. Nijstad (Eds.). 2003. *Group creativity: Innovation through collaboration*. Oxford University Press, New York, NY, US. https://doi.org/10.1093/acprof:oso/9780195147308.001.0001 Pages: xiii, 346.

[66] W. Piekarski and B.H. Thomas. 2001. Tinmith-Metro: new outdoor techniques for creating city models with an augmented reality wearable computer. In *Proceedings Fifth International Symposium on Wearable Computers*. 31–38. https://doi.org/10/ctz243 ISSN: 1530-0811.

[67] Arnaud Prouzeau, Yuchen Wang, Barrett Ens, Wesley Willett, and Tim Dwyer. 2020. Corsican Twin: Authoring In Situ Augmented Reality Visualisations in Virtual Reality. In *Proceedings of the International Conference on Advanced Visual Interfaces (AVI '20)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3399715.3399743

[68] Xun Qian, Fengming He, Xiyun Hu, Tianyi Wang, Ananya Ipsita, and Karthik Ramani. 2022. ScalAR: Authoring Semantically Adaptive Augmented Reality Experiences in Virtual Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3491102.3517665

[69] Xun Qian, Feitong Tan, Yinda Zhang, Brian Moreno Collins, David Kim, Alex Olwal, Karthik Ramani, and Ruofei Du. 2024. ChatDirector: Enhancing Video Conferencing with Space-Aware Scene Rendering and Speech-Driven Layout Transition. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3613904.3642110

[70] Shwetha Rajaram and Michael Nebeling. 2022. Paper Trail: An Immersive Authoring System for Augmented Reality Instructional Experiences. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3491102.3517486

[71] Benjamin Reynolds, James Hobin, Yuanzhi Cao, Martin Klaudiny, Daniel A Dangond, Kaize Xie, Keranie Theodosiou, Rik De Leeuw, Shiva Kashalkar, and Valentin Heun. 2024. The Pop-Up Metaverse: A Multi-User, Multi-Tasking Spatial Computing Environment for Collaborative Spatial Problem-Solving.. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, 1–5. https://doi.org/10.1145/3613905.3648677

[72] Yvonne Rogers, Kay Connelly, Lenore Tedesco, William Hazlewood, Andrew Kurtz, Robert E. Hall, Josh Hursey, and Tammy Toscos. 2007. Why It's Worth the Hassle: The Value of In-Situ Studies When Designing Ubicomp. In *UbiComp 2007: Ubiquitous Computing*, John Krumm, Gregory D. Abowd, Aruna Seneviratne, and Thomas Strang (Eds.). Springer, Berlin, Heidelberg, 336–353. https://doi.org/10.1007/978-3-540-74853-3_20

[73] Mose Sakashita, Balasaravanan Thoravi Kumaravel, Nicolai Marquardt, and Andrew David Wilson. 2024. SharedNeRF: Leveraging Photorealistic and View-dependent Rendering for Real-time and Remote Collaboration. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–14. https://doi.org/10.1145/3613904.3642945

[74] Arne Seeliger, Raphael P. Weibel, and Stefan Feuerriegel. 2024. Context-Adaptive Visual Cues for Safe Navigation in Augmented Reality Using Machine Learning. *International Journal of Human–Computer Interaction* 40, 3 (Feb. 2024), 761–781. https://doi.org/10.1080/10447318.2022.2122114 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10447318.2022.2122114.

[75] Vishwanath A. Sindagi and Vishal M. Patel. 2017. Generating High-Quality Crowd Density Maps Using Contextual Pyramid CNNs. In *2017 IEEE International Conference on Computer Vision (ICCV)*. 1879–1888. https://doi.org/10.1109/ICCV.2017.206 ISSN: 2380-7504.

[76] Maximilian Speicher, Jingchen Cao, Ao Yu, Haihua Zhang, and Michael Nebeling. 2018. 360Anywhere: Mobile Ad-hoc Collaboration in Any Environment using 360 Video and Augmented Reality. *Proceedings of the ACM on Human-Computer Interaction* 2, EICS (June 2018), 9:1–9:20. https://doi.org/10.1145/3229091

[77] Maximilian Speicher, Brian D. Hall, Ao Yu, Bowen Zhang, Haihua Zhang, Janet Nebeling, and Michael Nebeling. 2018. XD-AR: Challenges and Opportunities in Cross-Device Augmented Reality Application Development. *Proceedings of the ACM on Human-Computer Interaction* 2, EICS (June 2018), 7:1–7:24. https://doi.org/10.1145/3229089

[78] Ivan D. Steiner. 1972. *Group Process and Productivity*. Academic Press, New York, NY.

[79] Richard Stoakley, Matthew J. Conway, and Randy Pausch. 1995. Virtual reality on a WIM: interactive worlds in miniature. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '95)*. ACM Press/Addison-Wesley Publishing Co., USA, 265–272. https://doi.org/10.1145/223904.223938

[80] Xia Su, Han Zhang, Kaiming Cheng, Jaewook Lee, Qiaochu Liu, Wyatt Olson, and Jon E. Froehlich. 2024. RASSAR: Room Accessibility and Safety Scanning in Augmented Reality. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3613904.3642140

[81] Theophilus Teo, Gun A. Lee, Mark Billinghurst, and Matt Adcock. 2019. 360Drops: Mixed Reality Remote Collaboration using 360 Panoramas within the 3D Scene*. In *SIGGRAPH Asia 2019 Emerging Technologies*. ACM, Brisbane QLD Australia, 1–2. https://doi.org/10.1145/3355049.3360517

[82] Balasaravanan Thoravi Kumaravel, Fraser Anderson, George Fitzmaurice, Bjoern Hartmann, and Tovi Grossman. 2019. Loki: Facilitating Remote Instruction of Physical Tasks Using Bi-Directional Mixed-Reality Telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. ACM, New Orleans LA USA, 161–174. https://doi.org/10.1145/3332165.3347872

[83] Balasaravanan Thoravi Kumaravel, Cuong Nguyen, Stephen DiVerdi, and Bjoern Hartmann. 2020. TransceiVR: Bridging Asymmetrical Communication Between VR Users and External Collaborators. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. Association for Computing Machinery, New York, NY, USA, 182–195. https://doi.org/10.1145/3379337.3415827

[84] Huayuan Tian, Gun A. Lee, Huidong Bai, and Mark Billinghurst. 2023. Using Virtual Replicas to Improve Mixed Reality Remote Collaboration. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (May 2023), 2785–2795. https://doi.org/10.1109/TVCG.2023.3247113 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[85] Unity. 2024. Unity: Go Create. https://unity.com/

[86] Unity. 2024. Unity MARS. https://unity.com/products/unity-mars

[87] Rishi Vanukuru, Suibi Che-Chuan Weng, Krithik Ranjan, Torin Hopkins, Amy Banic, Mark D. Gross, and Ellen Yi-Luen Do. 2023. DualStream: Spatially Sharing Selves and Surroundings using Mobile Devices and Augmented Reality. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Sydney, Australia, 138–147. https://doi.org/10.1109/ISMAR59233.2023.00028

[88] Andrés Vargas González, Senglee Koh, Katelynn Kapalo, Robert Sottilare, Patrick Garrity, Mark Billinghurst, and Joseph LaViola. 2019. A Comparison of Desktop and Augmented Reality Scenario Based Training Authoring Tools. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 339–350.

https://doi.org/10.1109/ISMAR.2019.00032 ISSN: 1554-7868.

[89] Eduardo Veas, Raphael Grasset, Ernst Kruijff, and Dieter Schmalstieg. 2012. Extended Overview Techniques for Outdoor Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics* 18, 4 (April 2012), 565–572. https://doi.org/10/fxnspd Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[90] Charles J. Walker. 2010. Experiencing flow: Is doing it together better than doing it alone? *The Journal of Positive Psychology* 5, 1 (Jan. 2010), 3–11. https://doi.org/10.1080/17439760903271116

[91] Cheng Yao Wang, Mose Sakashita, Upol Ehsan, Jingjin Li, and Andrea Stevenson Won. 2020. Again, Together: Socially Reliving Virtual Reality Experiences When Separated. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376642

[92] Ruotong Wang, Lin Qiu, Justin Cranshaw, and Amy X. Zhang. 2024. Meeting Bridges: Designing Information Artifacts that Bridge from Synchronous Meetings to Asynchronous Collaboration. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1 (April 2024), 35:1–35:29. https://doi.org/10.1145/3637312

[93] Zeyu Wang, Cuong Nguyen, Paul Asente, and Julie Dorsey. 2021. DistanciAR: Authoring Site-Specific Augmented Reality Experiences for Remote Environments. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3411764.3445552

[94] Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xinggang Wang. 2024. 4D Gaussian Splatting for Real-Time Dynamic Scene Rendering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*. 20310–20320.

[95] Haijun Xia, Sebastian Herscher, Ken Perlin, and Daniel Wigdor. 2018. Spacetime: Enabling Fluid Individual and Collaborative Editing in Virtual Reality. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. ACM, Berlin Germany, 853–866. https://doi.org/10/gjbvfr

[96] Jacob Young, Tobias Langlotz, Steven Mills, and Holger Regenbrecht. 2020. Mobileportation: Nomadic Telepresence for Mobile Devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 2 (June 2020), 1–16. https://doi.org/10.1145/3397331

[97] Lei Zhang, Ashutosh Agrawal, Steve Oney, and Anhong Guo. 2023. VRGit: A Version Control System for Collaborative Content Creation in Virtual Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3544548.3581136

[98] Stefanie Zollmann, Gerhard Schall, Sebastian Junghanns, and Gerhard Reitmayr. 2012. Comprehensible and Interactive Visualizations of GIS Data in Augmented Reality. In *Advances in Visual Computing (Lecture Notes in Computer Science)*, George Bebis, Richard Boyle, Bahram Parvin, Darko Koracin, Charless Fowlkes, Sen Wang, Min-Hyung Choi, Stephan Mantler, Jürgen Schulze, Daniel Acevedo, Klaus Mueller, and Michael Papka (Eds.). Springer, Berlin, Heidelberg, 675–685. https://doi.org/10.1007/978-3-642-33179-4_64