A Collaborative Multimodal XR Physical Design Environment

Keru Wang
New York University
New York, NY, USA
keru.wang@nyu.edu

Pincun Liu
New York University
New York, NY, USA
pl2253@nyu.edu

Yushen Hu
New York University
New York, NY, USA
yh3856@nyu.edu

Xiaoan Liu
New York University
New York, NY, USA
xl5040@nyu.edu

Zhu Wang
New York University
New York, NY, USA
zhu.wang@nyu.edu

Ken Perlin
New York University
New York, NY, USA
perlin@cs.nyu.edu

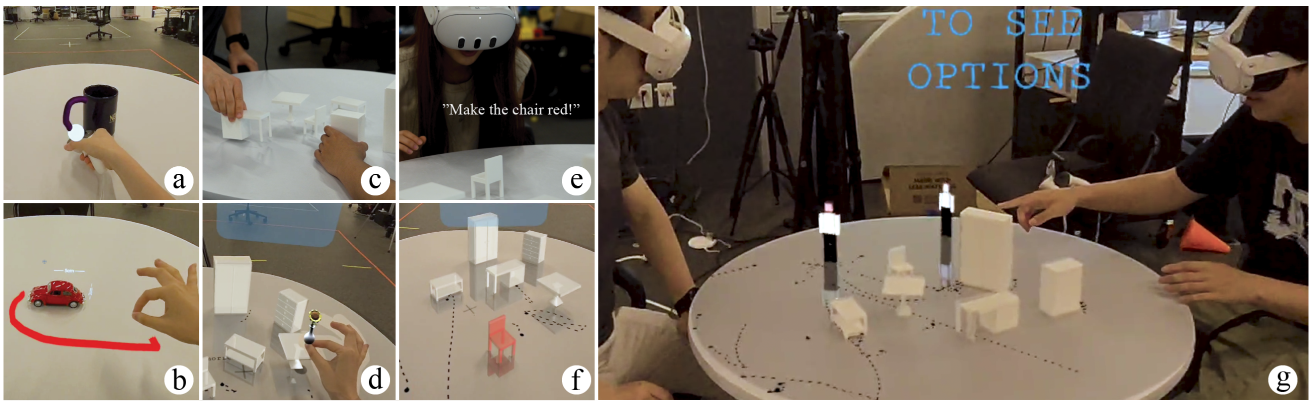

Figure 1: Our collaborative XR system integrates physical and virtual design spaces. It accelerates design iterations via features such as multimodal inputs, real-time physical object tracking, and object-based 3D annotation.

Abstract

Traditional design processes for physical prototypes can be time-consuming and costly due to iterations of prototyping, testing, and refinement. Extended Reality (XR) technology with video passthrough offers unique benefits for alleviating these issues by providing instant visual feedback. We have developed an XR system with multimodal input capability that provides annotations and enables interactive visual modifications by superimposing and aligning visual counterparts to physical objects. This system can help designers to quickly experiment with and visualize a wide range of design options, keep track of design iterations, and explore innovative solutions without the constraints of physical prototyping. As a result, it can significantly speed up the iterative design process, while requiring fewer physical modifications in each iteration.

Keywords

XR Prototype, Multimodal Interaction, Collaborative Experience, Digital Design

ACM ISBN 979-8-4007-1141-1/24/12
https://doi.org/10.1145/3681759.3688914

ACM Reference Format:
Keru Wang, Pincun Liu, Yushen Hu, Xiaoan Liu, Zhu Wang, and Ken Perlin. 2024. A Collaborative Multimodal XR Physical Design Environment. In *SIGGRAPH Asia 2024 XR (SA XR '24), December 03–06, 2024, Tokyo, Japan.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3681759.3688914

1 Introduction

Traditional design processes are often hampered by the separation between a designer's 3D modeling workspace and their physical environment. This leads to inefficiencies and a lack of intuitive interaction. For instance, designing a handle for a physical cup requires multiple steps: modeling the physical cup, transferring it into virtual space, designing the handle on a computer, and then 3D printing the handle to see how it fits with the physical cup. This disconnect can be time-consuming and cumbersome. With advancements in Extended Reality (XR) technologies, such as Meta Quest 3's video passthrough feature, the boundary between virtual and physical worlds is becoming increasingly blurred. Our collocated collaborative XR system is at the forefront of this transformation, merging real and virtual environments to create a seamless design experience. Modern XR platforms, which combine advanced technologies and natural user interfaces such as hand-tracking and speech input alongside stereo displays, provide unique benefits for design, visualization, and interaction.

Our system augments video passthrough visuals to allow users to seamlessly edit their physical prototypes, for instance, as shown in

Figure 1 (**a**), users can add a virtual handle to a mug. Our overhead camera-based tracking system can detect and recognize physical objects, and estimate their 3D pose. Superimposed visuals in XR can therefore always be aligned with the physical objects: as users move and rotate real objects, the system continuously updates the augmented visuals to maintain perfect alignment. Our multimodal user interface allows users to combine speech with freehand or controller-based manipulation for a more intuitive and natural design process. The system supports multi-user collaboration, allowing ideas to be shared in real-time, which creates a dynamic and creative environment for teamwork. In addition, the system helps designers keep track of design evolution by providing object-based annotations (Figure 1 (**b**)) via text, voice memo or sketching that can be associated with any tracked physical items.

Various approaches for independent and collaborative physical reality augmentation in XR have been explored in prior work[Kari et al. 2021; Wang et al. 2022; Yue et al. 2017]. For example, SceneCtrl focuses on scanning and augmenting static physical scenes but lacks dynamic object recognition[Yue et al. 2017]. Therefore, it is hard for users to directly interact with physical objects. TransforMR uses an iPad to capture and modify video streams in real-time, but does not support user interaction or 3D integration with the physical world. In contrast, our system focuses on providing interactivity [Kari et al. 2021]. It recognizes and tracks objects dynamically and supports active user engagement via multimodal input. Designers can interact with both physical objects and virtual objects/annotations in a unified way.

Our system explores the next generation of creative XR design workflow, offering a glimpse into a more immersive, collaborative, interactive, and enjoyable design process.

2 System Implementation

The system works with modern video passthrough XR headsets. In our experiments, we used Meta Quest 3. We use the headset's built-in controller tracking, hand tracking and speech input.

Object Tracking. To achieve real-time object tracking, an external camera is mounted above the workspace, and this information is calibrated with the corresponding virtual XR scene. Tracking data is continually streamed to our WebXR code base, allowing real-time updates and interactions between physical objects and the virtual environment. This allows designers to seamlessly coordinate interaction between physical objects and virtual objects, making the design process both more efficient and more intuitive.

Network. Our system supports a collaborative experience through a Node.js server network. This network implements a shared blackboard model consisting of a set of state variable. For every state variable, the same value is maintained across all clients. When a client changes the value of a state variable, a update message is immediately sent from that client to the server, which then relays the message to all other clients, thereby ensuring that state remains synchronized between all participants. The values of all state variables are periodically saved onto the server as a JSON file. Whenever a new client joins, these values are sent to that client, bringing it up to date with the ongoing XR session. This supports seamless collaboration by ensuring that all users always have the same state information, and can therefore interact together with the virtual and physical elements in real-time.

Multimodal User Interface. Our system features a multimodal user interface that combines speech and gesture input for a more natural interaction experience. Speech input is streamed from the headset to WebSpeechAPI, allowing users to issue voice commands to interact with the system. We use built-in gesture detection, and developed our own pinch-based interaction interface for virtual content creation and manipulation.

3 User Experience

Our system facilitates accurate and immersive visualization of design ideas, and helps to bridge the gap between concept and reality.

XR can benefit perception during the design process. Users can use their visual, tactile and auditory senses to interact with design prototypes. This enhances spatial awareness by improving users' perception of scale and spatial relationships. As a result, XR has the potential to transform the design process. Users can immersively create, edit and experience virtual prototypes as if the virtual elements were physically present, which can help to acheive a more comprehensive understanding during each design iteration.

We are excited to showcase the capabilities of our system at Siggraph Asia through a playful home planning scenario. Participants will gather around a table with 3D-printed miniature furniture tracked by our system. They can physically move the furniture (Figure 1 (**c**)) and also use speech and gestures both to add virtual objects and to alter the appearance of physical objects. Procedural virtual creatures will interact with their designs, providing playful feedback and further immersing participants in the experience.

The combination of direct physical manipulation with speech and gesture leads to an enhanced design process. For example, a user can say "make the chair red" (Figure 1 (**e**)), and the system will superimpose a virtual red color on the 3D-printed miniature chair (Figure 1 (**f**)), thereby transforming its appearance in XR. Users can also add a virtual object, such as a vase, to adorn a real object with virtual elements (Figure 1 (**d**)). Procedural virtual creatures will interact with the design by moving around in the space and taking an interest in the designs and in their creators (Figure 1 (**g**)), while adding playful feedback to the experience.

Designers can record text notes, voice memos or sketches during each design iteration, and then revisit these annotations between iterations. This allows them to understand their rationales, insights, and alternative options each time they revisit the design process. For collaborative work, these annotations can help reduce the risk of misinterpretation and can help team members understand the respective pros and cons of different approaches.

References

Mohamed Kari, Tobias Grosse-Puppendahl, Luis Falconeri Coelho, Andreas Rene Fender, David Bethge, Reinhard Schütte, and Christian Holz. 2021. Transformr: Pose-aware object substitution for composing alternate mixed realities. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 69–79.

Keru Wang, Zhu Wang, Karl Rosenberg, Zhenyi He, Dong Woo Yoo, Un Joo Christopher, and Ken Perlin. 2022. Mixed Reality Collaboration for Complementary Working Styles. In *ACM SIGGRAPH 2022 Immersive Pavilion*. 1–2.

Ya-Ting Yue, Yong-Liang Yang, Gang Ren, and Wenping Wang. 2017. SceneCtrl: Mixed reality enhancement via efficient scene editing. In *Proceedings of the 30th annual ACM symposium on user interface software and technology*. 427–436.