Limino

Interactions for Dynamic Blending Virtual and Real Environments
in Head-Mounted Displays

by

Qianyi Chen

in collaboration with Hin Kwan (Billy) Kwok

A thesis submitted in partial satisfaction
of the requirements for the degree of

Master of Design
at the
University of California, Berkeley

Fall, 2022

Faculty Director Signature and Date

Associate Director Signature and Date

By Qianyi Chen
In collaboration
with Billy Kwok

UC Berkeley
Master of Design
Thesis 2022

In fulfillment of the requirements for the degree of Masters in Design (MDes) at the University of California, Berkeley.

Limino

Interactions for Dynamic Blending Virtual and Real Environments in Head-Mounted Displays

Content

1. Abstract
2. Prior Art
3. Motivation
4. Approach
5. Process
6. Final Design
7. Discussion
8. Future Work
9. Conclusion
10. Bibliography

0. Acknowledgement

I would like to express my heartfelt thanks to my teammate, Billy Kwok for the invaluable collaboration and support throughout this process. Your expertise, dedication, and commitment to excellence have been a constant source of inspiration and have contributed significantly to the success of our project.

I would also like to extend my sincere gratitude to my professors Eric Paulos, Danika Cooper, Yoon Bahk, Kyle Steinfeld and GSI Eric Rawn in the Design Studio and Debates in Design for their guidance, feedback, and encouragement throughout the final semester. Their expertise and insights have been invaluable in shaping my understanding of design principles and practices.

I am grateful to the domain experts, Bala Thoravi Kumaravel, Bjoern Hartmann, Michael Naimark, Rajan Vaish for providing valuable feedback on my thesis project. Their insights and critiques have helped me to refine my ideas and approach, and have pushed me to think more deeply about my work.

I am grateful to my family and friends for their unwavering support and encouragement during this journey. Without their encouragement and belief in me, I would not have been able to achieve this important milestone.

Finally, I would like to thank the UC Berkeley MDes community for providing a supportive and collaborative environment and have given me some of the most memorable moments of my academic journey.

1. Abstract

In 1998, Raskar et al. envisioned an office of the future with spatial immersive displays which bring remote coworkers to the workplace (Raskar et al.). This early work opened up discussions of the notion of space by demonstrating how to seamlessly integrate virtual elements with the real world. It also indicates reality can be altered and reconstructed using technologies, which leads to the concept of Mixed Reality (MR) and blended reality.

Existing works have been studying how can the physical world be integrated into the virtual world in MR systems, including different aspects of the blending, ranging from usability on different tasks (McGill et al.) to rendering approaches of objects and environments (Budhiraja et al.). However, only a few works shed light on how the blending can be customized. RealityLens (Wang et al.) touched on this issue by studying the placement of a portal of the physical world, and the methods for triggering the portal. But there remain questions on modes of interaction to blend the virtual scene with real world, and how to provide more support on the system level to reduce users' effort in manual control.

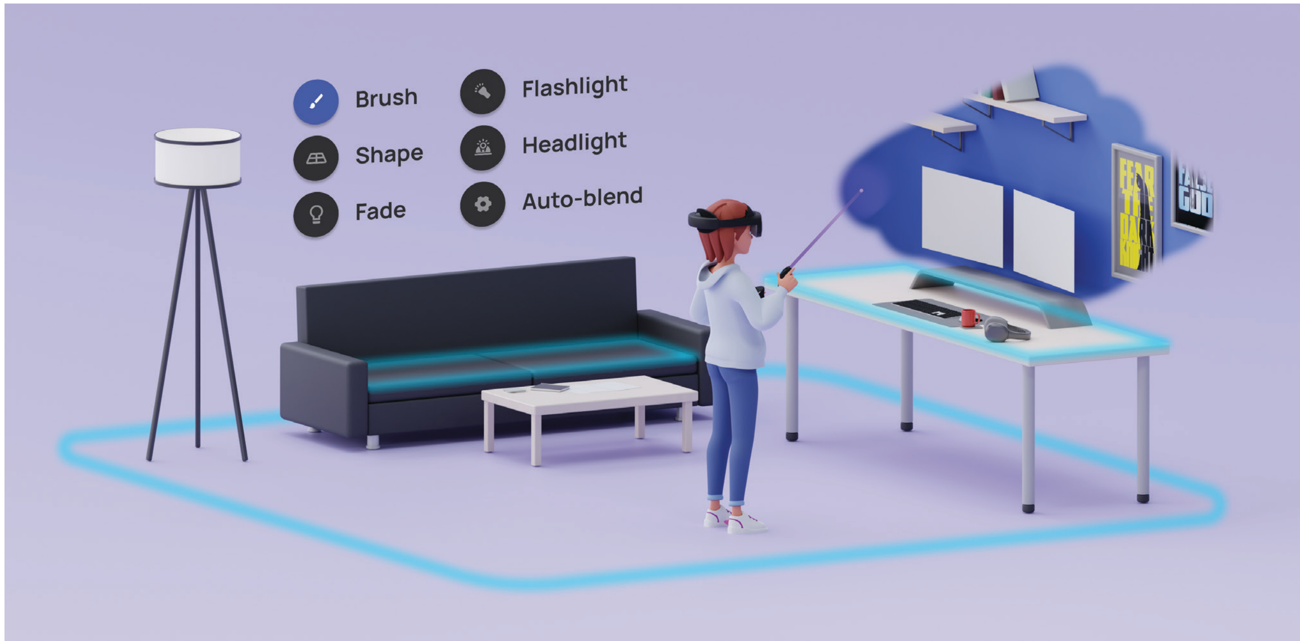
On the other hand, to understand how MR can be used to address the needs in the workspace, such as productivity, the medium of MR itself has pros and cons, such as the advantage of immersion and the drawback of creating an isolated experience. So far only limited works brought blending interactions into this context and study how blending can be used to strike a balance between staying immersed and maintaining context awareness.
Therefore, our work addresses two questions through a series of design explorations :

RQ1: How might we design the interactions for users to customize the passthrough of the real world in a virtual scene?
RQ2: How might we provide support for the customization process to balance the freedom of customization and the cognitive load?

In order to answer RQ1, we design and prototype three categories of interactions: fading, piercing and casting. Each category includes two interactions for defining the passthrough areas. Our second design exploration, aiming at addressing RQ2, studies how to understand the current context of user activity and provide automated adjustments or suggestions for blending options. Finally, we choose three user scenarios to demonstrate the user flow of our prototype. For both design explorations, our approach is a combination of in-situ design and human-centered design. The process typically involves a diverge stage in which we design and prototype a number of interactions, and test them out with users to gather insights for the next interaction.

For the scope of this thesis project, we focus on the use cases of blending for individual tasks and implemented several high-fidelity demos in Unity as proof of concept. Therefore, our work can be extended in future work that explores the technical aspect. For instance, the implementation of the system and more advanced contextual understanding using computer vision or AI. In terms of design, future work can be extended to study more complicated dynamics, such as blending multiple realities in multi-user environments.

By proposing a series of design vocabularies for the blending interaction, we aim at providing inspiration for future studies on how to make blending more customizable and context-aware. This work also opens up new venues for studying mixed-initiative interaction in the context of blending physical and virtual. Ultimately, we hope this work can serve as a provocation for designing future workplaces that seamlessly integrate virtuality with physicality.



Limino overview

2

Prior Art

Since our work focuses on using MR for blending physical and virtual worlds in the context of work, the literature we reviewed covered two parts:

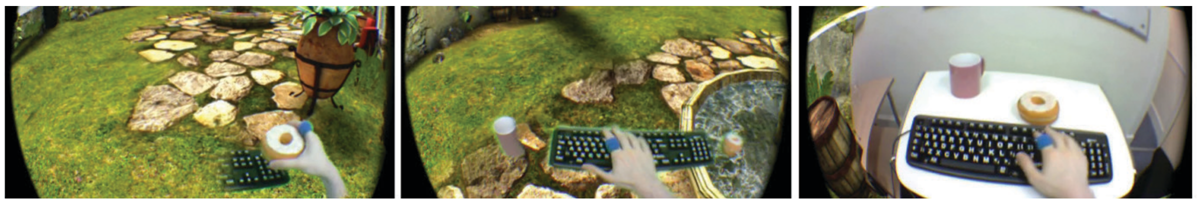
(1) What is the evolution of MR, and what approaches have been studied in terms of blending physical and virtual worlds;
(2) What are pros and cons of MR in supporting work and productivity, and what are the gaps that can be filled by our work

2.1 Mixed Reality and Blended Reality

There are several definitions of Mixed Reality (MR). According to Speicher et al.'s research (Speicher et al.), the most popular source is Milgram et.al's Reality-Virtuality Continuum (Milgram et al.) developed in the 1990s. This continuum positions the real world and the virtual world on the two sides of an axis, defining MR as the interval in between. Virtual Reality (VR), according to this definition, is the right extrema of the axis but not MR. MR includes both Augmented Reality (AR) and Augmented Virtuality (AV). The former refers to an environment where virtual elements are placed on top of the environment which is mostly the real world, and the latter is where a reality view is embedded in a virtual environment.

Compared to AR which has had a plethora of studies since the last century, there is fewer works exploring AV. Some early work includes a system built by Metzger (Metzger) that allows the overlay of images of real-world images on top of a virtual scene and vice versa. It also provides an option to make images transparent and allow the user to see the real world through the virtual world image.

Definition of Mixed Reality adopted from Milgram et al.



The figure all the blending options from McGill et al. Description of the original figure: “Left: Minimal blending (reality around user’s hands). Middle: Partial blending (all interactive objects). Right: Full blending (all of reality)” (McGill et al.).

The figure of all the different renderings from Budhiraja et al. Description of the original figure: “(a) Object & Hand, (b) Object, Hand & Edges, (c) Real World Windowed, (d) Physical Picture In Picture” (Budhiraja et al.).

While AR, AV, and VR are often considered discrete points on the Reality-Virtuality Continuum and studied separately, the transition between AR, AV, and VR has been getting the attention of researchers. The concept of Transitional Interfaces (TIs) (Carvalho et al.) was termed as an approach to seamlessly integrate systems along the Reality to Virtual Reality continuum. However, there is currently no theoretical framework for the tasks and needs related to Transitional Interfaces.

Another set of questions that comes in parallel with the transitional interface is blending reality and virtuality. In a single-user experience, it refers to the amount of reality incorporated into the virtual scene. Studies have also been investigating how different levels of blending influence peripheral interaction and the usability of keyboard typing (McGill et al.). In this study, the researchers define three types of blending based on the portion of reality. The blending is either object-based which only shows all the interactive physical objects or area-based which defines a cutout range around the targeted objects and shows both the object and the area around (McGill et al.). Their study results suggest the necessity of embedding a view of reality under the condition of peripheral interaction, and the user’s preference to have control over the blending levels.

Similarly, Budhiraja et al. study the rendering approach for objects in mixed reality when trying to selectively incorporate the view of the physical world into the virtual scene (Budhiraja et al.). They invested four rendering approaches to integrate the hand in the physical world into the virtual scene, ranging from only showing the objects (the hand and the coffee mug in the hand), outlines for showing the context around the hand, or showing a rectangle real-world window. Through an empirical evaluation, the participants of the study showed strong support for showing both the object and context. Also, an increased workload was observed when participants experienced a mismatch between the actual physical object and the presented image with the same object in a smaller size.

Both works show the importance and usefulness of incorporating a view of the physical world into the virtual environment. Meanwhile, the levels of blending and rendering approaches of the objects in the scene may lead to different user experiences. The context for blending also matters, ranging from typing tasks to maintaining awareness of the bystanders, which opens new avenues for future study.

In addition to adding elements to the environment, the concept of Diminished Reality (DR) explored the effects of removing objects from the scene (Mann). Researchers find that users choose opacity adjustment as the primary way to change object appearance (Cheng et al.), while also emphasizing the need to maintain contextual awareness.

These findings introduce an interesting design space for our work, which is to examine how we can give the user control of the blending parameters, e.g. how they can decide the range for minimal blending and whether that can be automatically changed based on certain conditions. Besides, there is still room for scrutinizing the user needs and contexts associated with the blending of reality and the virtual scene. For example, the need for immersion can vary under tasks with different cognitive loads. The user may want to focus on the current task or maintain awareness of the surroundings. How can we enable the user to switch easily between different modes is another question we can work on.

2.2 XR Productivity Tools for Work and Collaboration

Researchers have long been studying how MR technology can be used to support work and collaboration. One advantage of MR is the ability to augment the physical environment and present virtual elements, such as detailed parameters of an object, or instructions for a task. Compared to other 2D devices, MR systems can utilize the 3D space to display the information spatially. Similarly, many current VR applications such as Infinite Office and Spatial adopt this idea as a way to increase productivity. However, too much information may also increase the cognitive load of the user. Therefore, researchers also looked into how to make a context-aware display of information (Lindlbauer et al.). Some other use cases of MR in individual tasks include cluster management (Cheng et al.) and document management (Iwai and Sato), etc.

MR also has the advantage of manipulating the dimension of time and space. The former enables asynchronous communication between co-workers with richer information compared to text or images. For example, Fender et al. (Fender and Holz) used Microsoft Kinect to capture the real-time point cloud of a team member when the co-worker is focused on the current work and allows the user to play the video back later. In RemixReality (Lindlbauer and Wilson), the user can pause time and inspect the scene based on their needs.

Compared to MR in which users often maintain awareness of the surroundings, the experience in VR is more isolated and immersive. This characteristic of VR can be used to create a working environment that helps the user concentrate on the work (Ruvimova et al.; Mark et al.), which aligns with the research finding that blocking distracting information can bring positive effects on productivity. On the other hand, the downside of immersion is the difficulty of sharing the VR experience with external users. George et al. found out that users in an immersive set-up show fear of disengagement from the real world (George et al.). Therefore, how to balance interruptions and immersion for such systems remains a question.

Another pain point of VR workspace is maintaining awareness of the physical environment. For individual tasks, it's difficult to interact with peripherals and objects such as keyboards or coffee mugs. Some researchers (Endo et al.) came up with a hardware solution that adds modules to the VR headset as a way to extend the peripheral views. However, the augmented version of the headset is not ergonomically friendly to the user, and it requires manual adjustment of each module when switching the views. With the development of consumer-level hardware, such as the passthrough feature of the Meta Quest and Meta Quest Pro, a user can switch between VR and AR via passthrough without taking off the headset. The remaining problem space for this issue then becomes a design question: how can the interactions of switching between the VR/AR environment be designed, and what user needs are related to it?

In summary, MR has its own advantages in supporting work and collaboration. Prior work in MR productivity tools for work and collaboration reveals the potential of this medium and also points out the remaining deficiencies, which brings opportunities for future work to study the relevant use cases and design space.

3. Motivation

As an MDes thesis project, this project presents the cumulative result of my study at MDes. On one hand, I have been engaging in the exploration of emerging technologies throughout the three semesters, with a focus on Virtual Reality and Mixed Reality. On the other hand, by studying design theories from the core courses, I have developed a critical lens on emerging technology and the role of design. This project allows me to combine these two critical disciplines I have picked up during my study and apply my interaction design skills to MR to critically examine the use of such a technology in the near future.

This project comes at the perfect timing when the technical difficulty of building an MR system has started to become lower for designers and the problem space still has much room to explore. The advances and upgrades in hardware lower the barrier for designers to prototype experiences that blend the realms of physical and virtual. For example, the recent release of Meta Quest Pro and MR technology also has the advantage of manipulating the dimension of time and space. The former enables asynchronous communication between co-workers that go beyond the information capacity of text or images, especially in terms of emotive features and many subtleties of human interactions. For example, Fender et al. used Kinect to capture the real-time point cloud of a team member when the co-worker is focused on the current work and allows the user to play the video back later (Fender and Holz). In RemixReality, the user can pause time and inspect the scene based on their needs (Lindlbauer and Wilson).

Regarding the target audience, although the mainstream workflow for information workers still heavily relies on screen-based interactions on laptops, the general public's acceptance of using new technologies for work has been growing since the global pandemic. More and more people got to try out Horizon Workrooms, Microsoft Mesh, and Mozzila hubs for meetings and collaboration. This project can potentially help them reimagine how their needs in work and collaboration can be addressed in an environment where virtual and physical worlds are blended instead of separated.

The project chooses MR as a specific technology to address user needs in the workplace, but it does not come with the presumption that MR is the best solution for work in the future. Instead, this project holds a critical perspective on the technology itself. Through the literature review, we recognize the advantages of MR such as being able to make temporal and spatial modifications of the environment. At the same time, we also point out the limitations of VR such as its isolated nature that leads to difficulty in sharing in-device content and interacting with objects in the physical world. Therefore, we position this project as an attempt to investigate the possible use cases for such a technology, while also critically examining the limitations of the medium.

Overall, we'd like to use this work as a provocation for the design of future MR interactions. We believe that future 3D interfaces and interactions should utilize the additional dimension which taps more potential than the 2D flat screen. While many existing works focus on building systems or solving hardware issues, this project takes a different approach by aiming to establish a series of design vocabulary for the interactions. By presenting concepts around how to enable a user to incorporate the real world into virtual scenes in various workplace settings, we inspire new paradigms of work in an augmented environment that blurs the boundaries between real and virtual.

4 Approach

4.1 Design Approach

In-situ Investigation

This project consists of two parts: design and implementation. Our approach is adapted from the Double Diamond model. Designing for MR experience is challenging since it's hard for people to imagine the experience before they actually test out the project. Therefore, we adjust the original double diamond model and combine it with the idea of in-situ design, which means working with the users "on-site" to define the iterations of the work.

Our project has the diverge and converge stages included in the Double Diamond model, but instead of defining user needs at the very beginning, we started with exploring possible interaction designs, implemented demos for users to test, and then developed the next iteration based on the user feedback.

Our approach is different from the traditional human-centered design since we already define the medium to be Mixed Reality and the focus of interaction to be blending physical and virtual environments. We started with design explorations and then tie the use cases back to the interaction based on the in-situ investigation. In some cases, the user feedback also informs the focus of the next design exploration. Therefore, unlike human-centered design in which the process is usually linear from defining the problem to developing the solution, our project follows a zigzag route that goes back and forth on iterating the design.

4.2 Implementation Approach

Rapid Prototyping

Our thesis project adopts the rapid prototyping approach for implementing our ideas. Since the primary goal of prototyping is to demonstrate a proof of concept, we choose a group of techniques and software that allow us to build prototypes with different fidelities, including ShapesXR for low-fidelity non-interactive visual prototypes, SparkAR for mid-fidelity interactive prototypes, and Unity Game Engine for high-fidelity functional prototypes.

We also evaluate the tradeoff between time cost and functionality when building the prototypes. For example, some of our proposed interactions require computer vision to be fully functional. However, implementing object recognition using computer vision itself is very technically challenging and time-consuming. Therefore, we choose other workarounds to build a proof of concept, while pointing out that computer vision can be used in future attempts of implementing such a system. Such a rapid prototyping approach allows us to spend less time dealing with technical hassles, focus more on mocking up the design, and gather user feedback for iterations.

5. Process

5.1 Design

Research, Ideation, Interviews, Interactions Prototype, and Case Studies

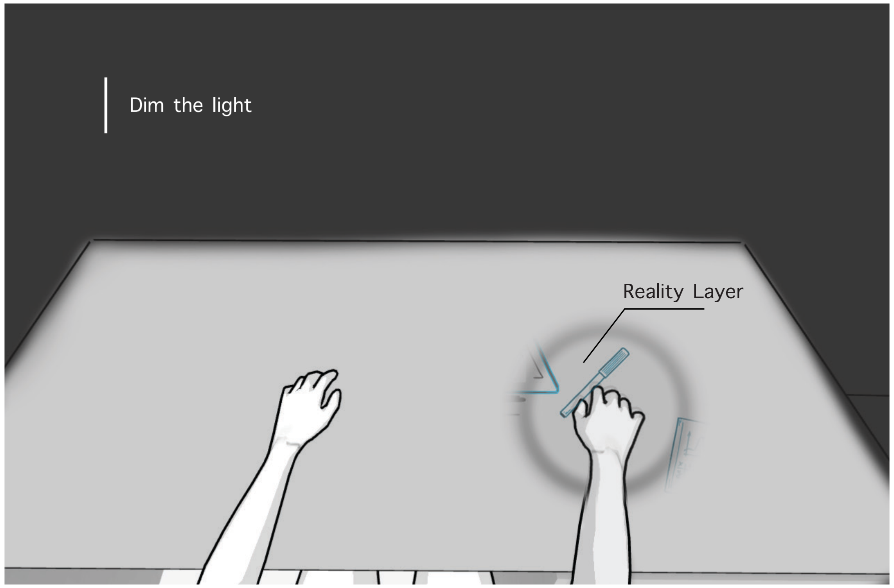
In the initial research stage, we conducted a literature review to understand the gap to be filled in existing works. For the literature review, we specifically looked at papers related to integrating the physical environment into the virtual world. We identified the gap in current literature about balancing the immersion of VR and its isolating nature.

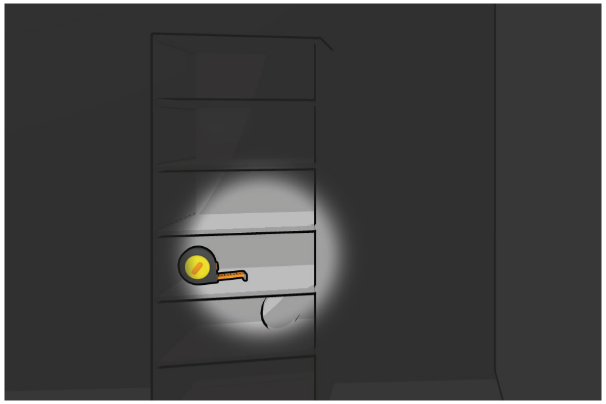
We also conducted exploratory user interviews with people who have experience as remote workers in the field of design, engineering, etc. Our interview findings further demonstrated that the need for immersion varies among individuals under different scenarios. For example, the focus mode is preferred with intense work while maintaining awareness of the surroundings when dealing with less urgent work on a daily routine. This points out the need for having a system or toolkit that allows the user to change between different immersion levels and customize the environment based on their need, which does not exist in the current market.

Based on the findings from the literature and interviews, we brainstormed a list of possible interactions for blending, which is our first design exploration. We grouped the interactions into three categories: fading, piercing, and casting which allows various options to define the passthrough area and adjustment. Fading focuses on the manipulation of opacity while piercing and casting directly display the passthrough in areas selected by the user. The difference between the two is how the passthrough areas are updated, which we will cover more details in the next section.

We implemented each interaction in Unity and tested it with users for feedback. During the user testing, we asked users to think out loud about what they would like to use the interactions, and why they think a certain area should be marked as virtual or passthrough view.

With the gathered insights, we move forward to study how can the experience of customizing the passthrough be improved. Since all of the interactions in the first exploration are manual control, we investigate what level of automation can be provided to facilitate the process. Lastly, we studied how blending interactions can be toggled based on the different contexts through case studies.





Initial concept sketches of the interactions

5.2 Prototype

Implementing the Prototype Demo with Different Fidelities

In the early stage of technical investigation, we explored prototyping tools for low-fi and mid-fi, including ShapesXR, VRCeption, and SparkAR. ShapesXR and VRCeption as low-fi prototyping tools have both pros and cons for our exploration. VRCeption provides two pre-build scenes in VR and a photorealistic-style room, which makes it easy for users to start prototyping. However the available assets are very limited, and users in general don't have much room for customizing the interactions. The advantage of ShapesXR is the support for mixed reality prototyping kit and the integration with Oculus passthrough. However, similar to VRCeption, the prototype built in ShapesXR is also static, meaning the interaction can not be demonstrated within the tool.

Overall, the low-fi technical explorations allow us to form an initial understanding of how a blended environment may look from a visual perspective. In order to test out the interactivity, we moved to the next stage of mid-fi prototyping using SparkAR. Among the list of interactions we've brainstormed, we pick the "torch" metaphor for the demo. The idea is a passthrough based on the head position. In the SparkAR demo, we were able to demonstrate an interactive scene with a reality torch, which casts a shadow on digital surfaces and reveals reality.

The limitation of SparkAR is on the device side. Since it is based on mobile devices, the user interactions could be different when moving from mobile phone to headset and controller. Therefore, we moved forward to build a high-fidelity Unity prototype using Oculus Quest Pro and its Color Passthrough feature. We started with prototyping each basic interaction, and also tested for building context understanding. Afterward, we combine context awareness and the interaction, using context awareness as a way to trigger different blending options. More details regarding the implementation will be covered in the section below.

Mid-fi prototye of the headlight interaction using SparkAR

6 Final Design

Our final design is demonstrated through a high-fidelity MR application prototype running on Meta Quest 2 and Quest Pro. The core of the application is an interactive MR workspace loosely resembling the room where the users are physically situated. In addition to completely virtual content, users can map out real-world objects upfront, such as desks, couches, walls, doors, and windows. These mappings will be used to generate their digital twins. Users can also adjust the blending of the physical and virtual environments dynamically through a set of manual and automated interactions. We devised two design explorations to analyze the use cases and trade-offs of these interactions.

6.1 Design Exploration 1

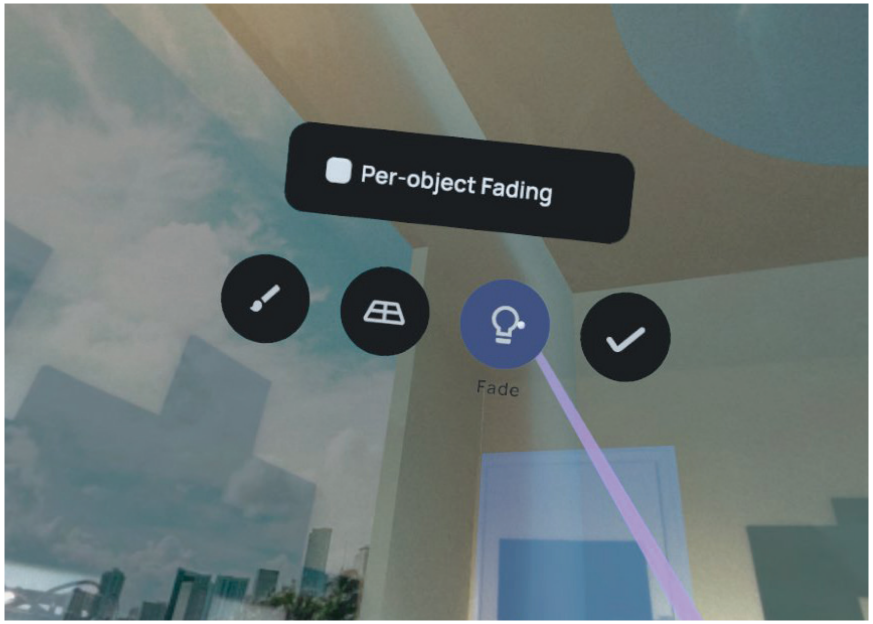
Blending Interactions

In the first design exploration, we focused on designing interactions that unveil the physical world in the virtual environment. These interactions allow users to customize the placement, size, and opacity of the passthrough area. We designed several blending interactions across different levels of usage patterns, based on the alteration of the environment, the blending options are categorized as fading, piercing, and casting.

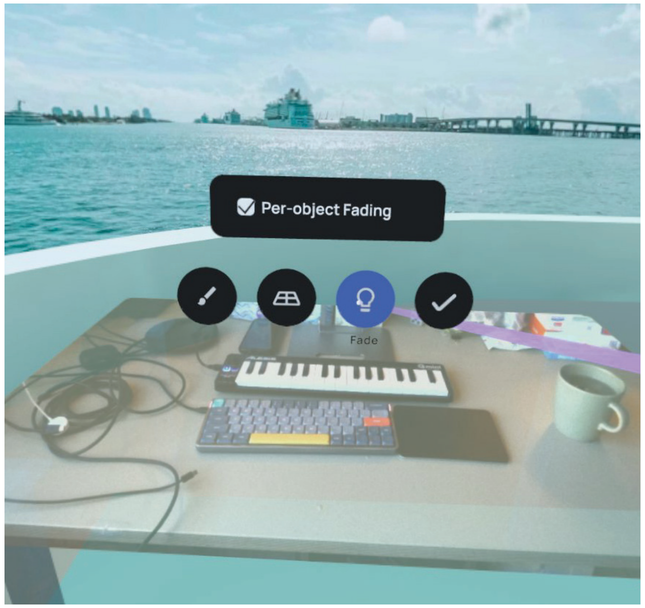
6.1.1 Fading

With fading intersections, users can adjust the opacity of virtual content to reveal the background passthrough behind it. Some objects are completely virtual, such as decorative plants and virtual barriers. Some are digital twins generated from the spatial anchors of their corresponding real-world objects, such as desks, couches, and walls. Their corresponding editing tools are only available in the Edit Mode to avoid unintended changes. In the Edit Mode, all objects are unlocked and subject to changes. The virtual environment also becomes semi-transparent to help users estimate the cutout locations. Since the entire virtual layer can also be considered a global object, we came up with the two fading interactions below.

Global Fading decreases the opacity of all virtual content.
Object Fading decreases the opacity of particular virtual objects or digital twins.



Global fading

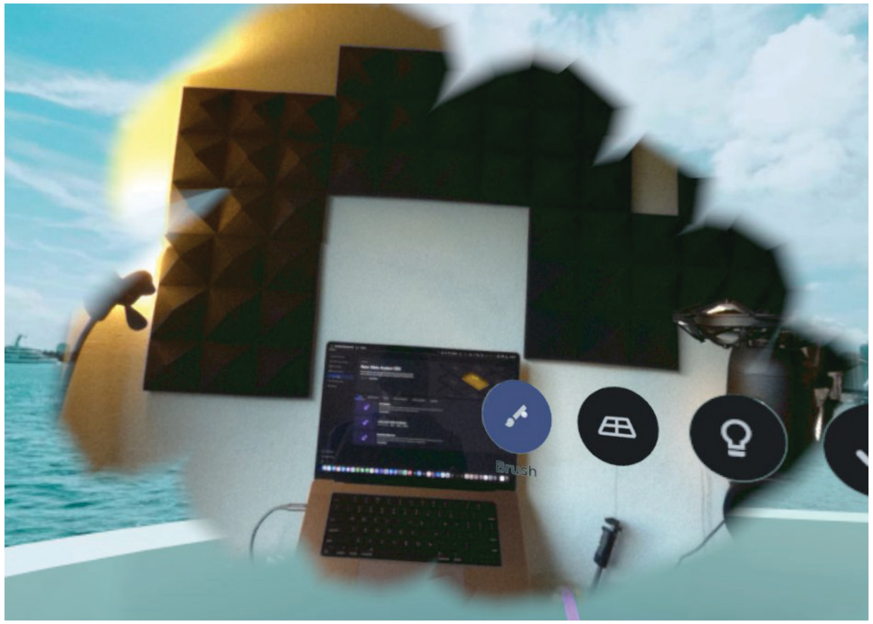


Obejct fading

6.1.2 Piercing

With piercing interactions, users can create, update and remove passthrough cutouts through controllers and a popup interface. These cutouts occlude the underneath content with a slice of the live camera stream of the physical world. Similar to fading interactions, the adjustment to these cutouts is only available in Edit Mode. We implemented the two types of cutouts below. The size of shape and brush can be adjusted.

> Passthrough Shape displays the passthrough image on a surface created by the projection from the controllers.
> Passthrough Brush uses one controller to paint a stroke of reality on top of the virtual environment.



Passthrough brush

Passthrough shape

6.1.3 Casting

With casting interactions, users can cast a passthrough shadow onto the environment as if they are using a searchlight. The searchlight can be attached to different body parts. In our prototype, we chose to track the movement of the hands and head, which can be conveniently mapped to the movement of their corresponding hardware components. Due to the spontaneous nature of the interactions, they are not restricted to the edit mode.

> Flashlight casts passthrough shadows by tracking the hand (controller) movement.
> Headlight casts passthrough shadows by tracking the head (headset) movement.

Flashlight

Headlight

6.2 Design Exploration 2

Context Awareness

6.2.1 Activity Awareness & Environment Awareness

In the first design exploration, all of the interactions are initiated by the user. The manual control of all the passthrough areas results in a high cognitive load. Therefore, we would like to understand what assistance can be provided on the system level in order to lower the effort of manually controlling the blending. For example, whether some interactions can be automatically toggled when some event happens.

Inspired by the findings of related works (McGill et al.), an approach to achieve automation is to associate the trigger of passthrough with events. In order to achieve this, the system should also be able to understand the changing context and suggest when to toggle on the passthrough, and even which passthrough option is a better fit. Therefore, we did a second design exploration on how to add context awareness to the blending interactions.

We identified two types of context awareness that are relevant to the initiation and adjustment of blending: activity and environmental awareness.

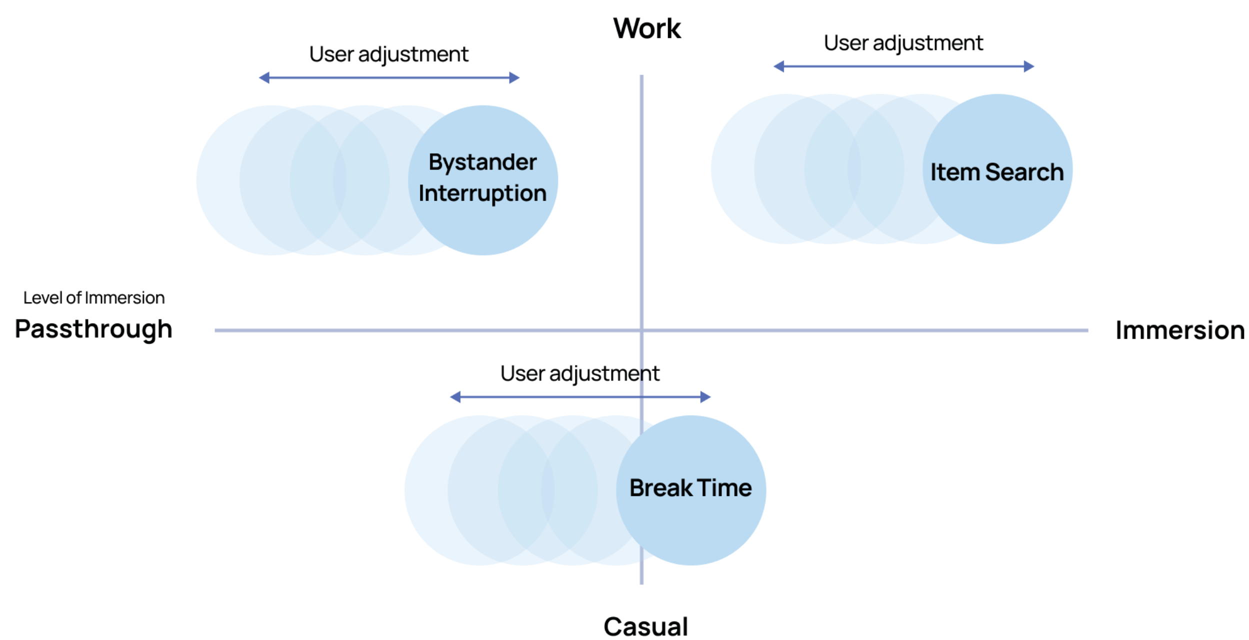
Activity Awareness is the awareness of the current user activity. It captures the active change from the user side when the user switches to a different task, which indicates a potential need for adjusting the passthrough.

Environmental Awareness is the awareness of changes happening in the current space. It captures the changes happening outside the HMD. Due to the isolated nature of VR experience, it is often difficult for users to be aware of and understand what is happening outside the HMD. Leveraging passthrough can be a potential solution for this issue.

Both types of awareness are extrinsic contextual information that can be collected through sensing devices such as cameras or the HMD. With this information, we can detect the current state of the user and the environment, therefore suggesting how the blending can be adjusted. We will talk about more details of how we implement the detection of the context in the technical implementation section.

6.2.2 Case Study

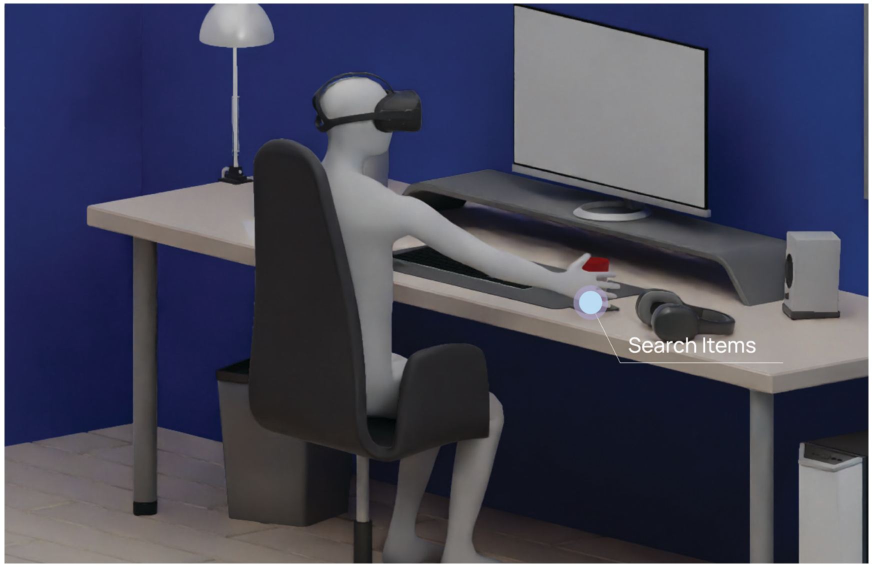
We selected three use cases to study - item searching, break time, and bystander interruption. These use cases cover activities ranging from work to casual entertainment and different levels of immersion. We investigated the input and output for each case from the design perspective.



Mapping of case studies based on immersion level and attention level

(1) Item Searching

A form of activity awareness that assists users in locating items close enough to be reached with a stretched arm but far enough to be outside their peripheral vision. When the system detects the need for item searching, it toggles on the aforementioned headlight interaction. The system understands the user movement by accessing the HMD data, such as the location of the headset and controller.



Scenario demonstration: item searching

Auto trigger of headlight when the user searches for a coffee mug

(2) Break Time

A form of activity awareness that helps users temporarily exit the virtual environment by switching back to reality. The system understands when the user is taking a break from work by detecting the change in the headset position.

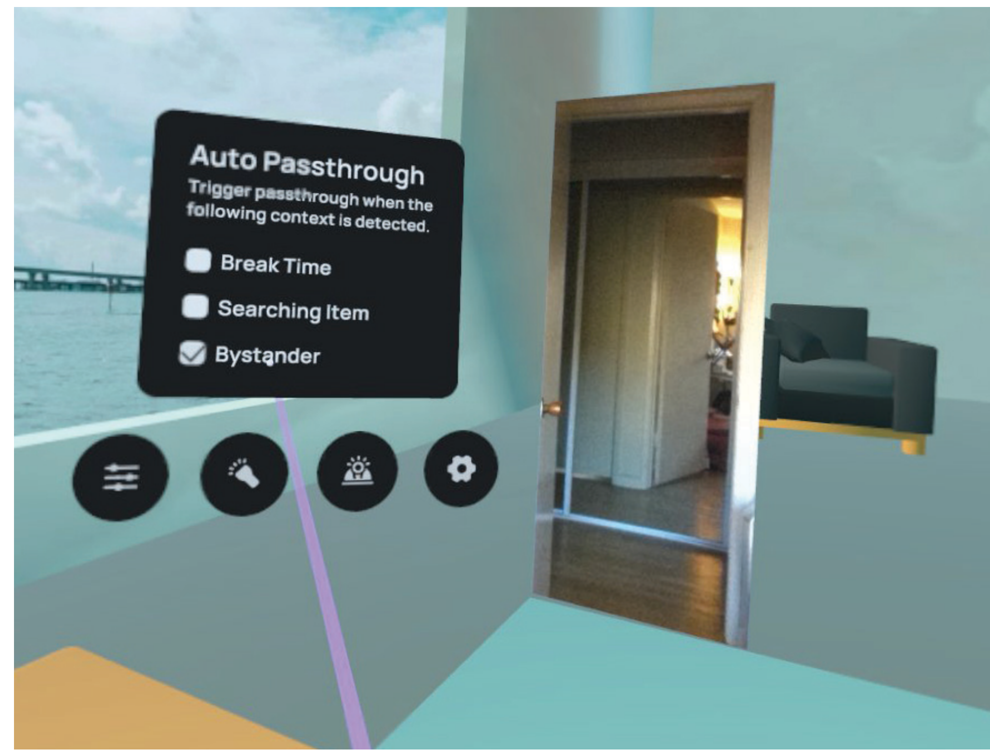
Scenario demonstration: break time

Auto trigger of global fading when the user stands up

(3) Bystander Interruption

A form of environmental awareness that recognizes bystanders entering the predefined activity boundary and fades out the virtual door overlay to reveal the real-world position of the person. Sensing devices ranging from cameras to other IoT systems can be used for detecting changes in the space, such as a non-VR user entering the room. However, this project employs a wizard-of-oz technique that triggers the fading manually.

Scenario demonstration: bystander interruption



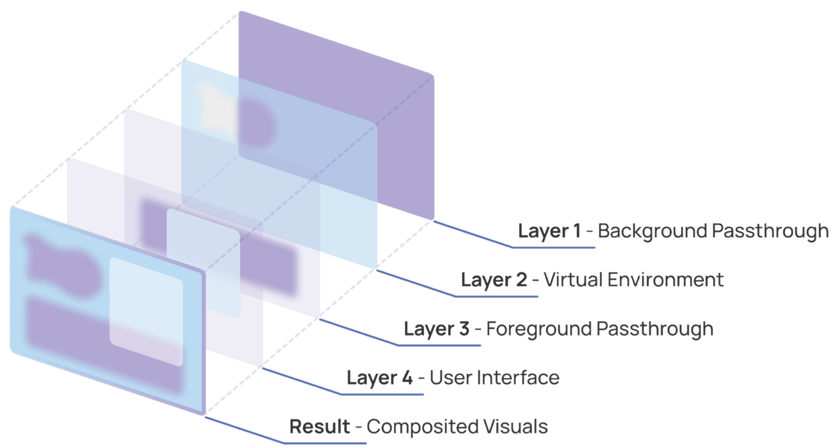
Auto trigger of object (door) fading when the bystander enters the room

6.3 Technical Implementation

The prototype was implemented on Unity using the Oculus XR Plugin and SDK with OpenXR backend. We have chosen Meta Quest 2 and Meta Quest Pro as our testing devices due to their availability and community support, but the experience can be ported to any 6DOF MR headsets that support passthrough, spatial anchors, and controller tracking. The implementation of the system comprises the following four parts.

6.3.1 Compositing Mechanism

The MR workspace is a customizable 3D environment that blends physical and virtual worlds. The intermixing of the two worlds is achieved by compositing different layers together. The position and opacity of the content in each layer contribute to the overall blending of the scene. These layers can be conceptually categorized into four types based on their rendering priority in the depth buffer.



Composition layers of the Limino app

The Background Passthrough layer is the farthest layer, rendering a perspective-mapped live camera feed spanning the entire rendering canvas. Access to individual pixels is prohibited by the Presence Platform SDK to protect privacy. However, the SDK supports a set of predefined image filters, such as edge highlight, as well as changes in saturation, contrast, and brightness.

The Virtual Environment layer consists of virtual and hybrid objects placed on a transparent background. The virtual content in this layer occludes the background passthrough layer, enabling the augmentation and diminishment of real-world objects. However, unoccluded background passthrough pixels are still visible after compositing.

The Foreground Passthrough layer is the same as the background passthrough but rendered on top of the virtual environment. It enables unrestricted control over fine-grained see-through effects, such as soft-edge cut-outs, without modifying the shapes of the objects underneath. Similar to background passthrough, access to individual pixels is also prohibited.

The User Interface layer is always rendered on top of other content regardless of their world positions because it is always in active use when shown, and stays hidden otherwise.

6.3.2 Environment Curation

The virtual environment is generated based on the scene anchors provided by the Presence Platform SDK. These scene anchors capture the world positions and dimensions of the bounding planes and boxes that enclose real-world objects, such as desks, couches, walls, doors, and windows. The users need to define them upfront through the Meta Quest room mapping setting. Digital twins of the mapped real-world objects are then generated based on the scene anchors. In addition to these virtual replicas, completely virtual items, such as decorative plants and soothing scenery, and are also available to support the experience aesthetically or functionally. These virtual and hybrid objects are either designed in Blender by ourselves or obtained from open-source 3D asset platforms. They are imported into Unity as GLTF or FBX files.

6.3.3 Blending Interactions

Passthrough content is assigned with a passthrough material rendered with a special HLSL shader that exposes the underlying camera live stream and creates soft edges along the passthrough boundaries. However, the implementation of blending interactions depends on the types of blending they create.

For fading interactions, we exposed the opacity control of the entire virtual environment layer and its content to the users through a slider interface. For piercing interactions, we implemented the passthrough surfaces by assigning passthrough materials and editable behavior to individual planes. The passthrough stroke implementation is based on an Oculus XR Integration sample with support for brush size and opacity adjustment added by us. For casting interactions, the passthrough shadow is projected based on a light volume model that tracks the user movement. The required tracking data was obtained from the built-in sensors through the Presence Platform SDK.

6.3.4 Context Awareness

The context-aware component in our prototype transforms events, including user activities and environmental changes, into the blending adjustments mentioned above.

(1) Activity Awareness

Activity awareness is based on motion-based, for instance, standing, arms movement, break-time reminders, and calendar meetings. Motion-based event detection was achieved using the built-in sensors and Presence Platform SDK. For example, the application can detect standing by monitoring the vertical position of the headset relative to the floor and arm movement by measuring the distance between the headset and the hands or controllers. While time-based event detection can be implemented relatively easily via third-party calendar integration, it is triggered behind the scene using one of the controller buttons in our prototype, allowing us to focus on the interactions themselves.

(2) Environmental Awareness

For environmental awareness, our prototype relies on an external camera for visual input due to the restricted access to passthrough data on Quest devices. The video stream from the camera is fed into a pose estimation Tensorflow model to predict the presence and movement of bystanders in the environment. The derived data is pushed to the MR application via a local WebSocket server and triggers the corresponding blending changes.

7 Discussion

We observed a number of patterns and issues from the user testing with our prototype, which are grouped into the following themes.

7.1 Desired Immersion Level and the Changing Contexts

Immersion and context awareness are key factors in our design considerations. During our design explorations, we did not hold preassumptions about the relationship between certain passthrough features and the level of immersion. Our approach is identifying passthrough parameters that may be relevant to the immersion level and prototyped a range of interactions that allow the user the adjust these parameters. For example, our blending interactions (fading, piercing, casting) include three dimensions: opacity of virtual elements, the size of passthrough areas, and passthrough duration. All of these can be adjusted by the user, which allows us to hear users' thoughts while they are actually using these interactions in situ.

From this design exploration and user testing, we found that the need for immersion varies based on the ongoing task. For example, some users attempted to minimize the amount of passthrough content when working on tasks that require deep concentration, such as writing formal documents and consuming complex information. However, when performing tasks that allowed distraction, some users preferred to see a larger part of the physical world. Although we did not conduct a quantitative user evaluation on a large scale, the observation and qualitative feedback we got still emphasizes the need for blending adjustments based on the cognitive load of current user activities.

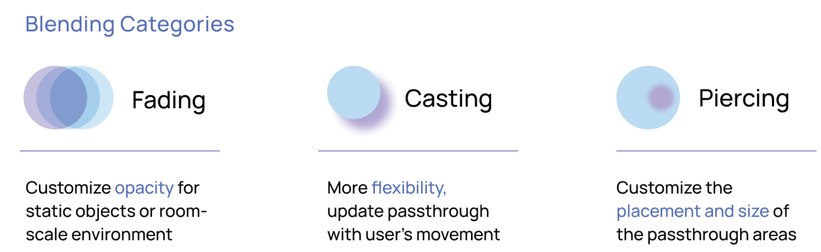
The interaction between the HMD user and non-HMD users in the room is another major design consideration for passthrough. This use case is common among co-residents, parents, and pet owners. These users need to keep an eye on their living space when they are in MR sessions and be prepared for having interactions with people or pets sharing the same physical space with them.

7.2 Fading vs. Piercing vs. Casting

In addition, the first exploration has given us insights into some of the pros and cons of each type of blending interaction.

Fading interaction enables relatively convenient passthrough setup and large-scale modifications of the virtual scene, for example, making the entire desk a passthrough without manually drawing out the desk area, and showing a full-screen passthrough with global fading. However, the downsides of fading are the lower customizability and simple yet required manual configurations.

Piercing interaction, on the other hand, provides users with more precise control over the passthrough content. But the amount of manual configuration required, ranging from setting opacity to drawing out passthrough shapes, steepens the learning curve and increases the cognitive load of the users.



Comparison between three blending categories

Casting interaction, which tracks hand and head movement, requires the least manual effort to define the passthrough configurations. Since the passthrough shadow follows user movement, they can have a dynamically defined view of the real world. Since the area cast in shadow is likely where users want to see in passthrough, and the rest of the environment stays virtual, they can maintain a relatively high sense of immersion. Nonetheless, the dynamic focus of the interaction may become a limitation under specific circumstances. For example, it could be hard for users to see multiple spots of the physical environment or the area outside the reach of their gaze or arm. It may also introduce accessibility issues due to the reliance on body tracking.

Therefore, an important takeaway from the first design exploration is that there is no one-fits-all interaction for blending adjustment and customization. A better approach is to combine manual configuration and dynamic manipulation, allowing users to choose their preferred interactions in different scenarios. Another insight is the potential of using automation to assist the three types of interactions. For instance, in the Passthrough Strokes interaction, the strokes can snap to the outline of the background objects so that users do not need to draw precisely.

7.3 False Positives and Negatives in Context Detection

In the first exploration, we observed users fine-tuning and re-adjusting the blending to find the optimal configuration. This behavior coincides with our initial motivation of reducing manual effort through context-aware automation. However, in the subsequent exploration, we discovered that while context awareness is effective in reducing manual effort, there were instances of false positive and negative detections. These detection errors can be grouped into two categories.

The first type of detection error is technical inaccuracy. For instance, when detecting arm stretching in the item searching use case, the relative position between the hand and the headset can be imprecise due to the limitations of the sensing algorithm and hardware. In this case, the availability of better sensing technologies would affect context detection and the resulting blending experience. Another example of technical inaccuracy is the imprecise detection of bystanders in the activity boundary. The lack of a built-in 360-degree camera and the inability to access passthrough data both contributed to this issue. In particular, the reliance on an external camera increased the detection latency and introduced mapping inaccuracy in the bystander position, which made us fall back to wizard-of-oz method. As a result, better context-aware blending for MR experience would benefit from improved sensing capabilities of HMDs in some use cases.

The second type of detection error is intent mismatches. For instance, arm stretching does not always imply that the users are reaching for an object. Instead, they could be stretching their arms to relax their muscles. Similarly, the end of meeting events on calendars only sometimes signifies that the users prefer to take a break. These mismatches between user intent and detected events could lead to unneeded distraction and increased cognitive load.

The two types of detection error showed that hardware limitations and intent mismatch are both important considerations when designing context-aware blending interactions.

8 Future Work

8.1 Beyond Passthrough-based Blending

While our work integrates the view of the physical world using passthrough, which, as the name suggested, directly shows a peak into the physical world, there are other approaches for showing the physical environment. For example, the Space Sense feature in Meta Quest only renders the outline of objects to provide the user with awareness of the physical environment. Future work can explore using our proposed blending interactions and contextual awareness to switch between different visualizations. Besides, future work can also extend our work into non-visual and cross-sensory blending. For example, environment sound can be recognized and mixed into the audio played in the virtual environment.

8.2 Blending for Co-located Users

If we had the time to take a step further, we would also like to study the blending interactions in multi-user scenarios. Shared spatial anchors have recently been made available in the Presence Platform SDK. With this feature, future work may introduce adaptive blending to a shared MR environment. For example, blending may be personalized based on the preference of each user in the same space. The practicality and trade-offs of such multi-user blending experiences could be further explored.

8.3 Improvements in Context Awareness

In our project, we investigated principles for designing context-aware blending and implemented the prototype as a proof of concept. Future work can look deeper into the technical aspect of developing context awareness for HMDs. For example, AI or computer vision can be used to build a more precise understanding of the environment, therefore reducing the false positive rate.

While we design the blending interactions around the use case of work and home office, we can also envision the context awareness being applied to other scenarios, including entertainment and meditation experiences.

9 Conclusion

Our work identifies opportunities to explore interaction design for adaptive blending in MR based on our literature review. With a focus on adaptive blending, we designed, implemented, and tested a set of blending interactions to improve our understanding of their benefits and constraints.

Specifically, we conducted two in-situ explorations to synthesize the relevant design considerations and tradeoffs. The first exploration investigated three types of blending interactions, namely fading, piercing, and casting, each focusing on different aspects of the blending options. We learned about how can passthrough parameters, including passthrough duration, manual customization of passthrough, and opacity be adjusted through interaction. We also studied how user experience, including the need for immersion and cognitive load in customizing passthrough can be associated with the blending experiences.

The second exploration applied context awareness to the blending interactions as a way to support the customization of passthrough. Through three case studies, we demonstrated how the interactions can be used for different levels of immersion and in different contexts ranging from work to casual activities. It also revealed that two types of false positive and negative detections, namely technical inaccuracy and intent mismatches, should be considered when integrating context awareness into the blending interactions. The two explorations together demonstrated the potential of making MR experiences more versatile and personalized through context-aware blending interactions.

With a series of design explorations and proposed design languages in adaptive blending, the project contributes to the overall MR research landscape by categorizing the design space for blending interactions, as well as introducing context awareness as an approach to implement automation for triggering blending. We hope to expand our work in the future and that our work will inspire more explorations in context-aware MR experiences, Mixed initiative MR experiences, and blending interactions.

10. Bibliography

Budhiraja, Pulkit, et al. Where's My Drink? Enabling Peripheral Real World Interactions While Using HMDs. Feb. 2015, arXiv:1502.04744 [cs]. arxiv.org/abs/1502.04744. Accessed 17 Oct. 2022.

Carvalho, Felipe G., et al. "Toward the design of transitional interfaces: an exploratory study on a semi-immersive hybrid user interface." Virtual Reality, vol. 16, no. 4, Nov. 2012, pp. 271–88. https://doi.org/10.1007/s10055-011-0205-y.

Cheng, Yi Fei, et al. "Towards Understanding Diminished Reality." Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, Apr. 2022, pp. 1–16, https://doi.org/10.1145/3491102.3517452.

Endo, Isamu, et al. "ModularHMD: A Reconfigurable Mobile Head-Mounted Display Enabling Ad-Hoc Peripheral Interactions with the Real World." The 34th Annual ACM Symposium on User Interface Software and Technology. UIST '21, Association for Computing Machinery, 2021, pp. 100–117, https://doi.org/10.1145/3472749.3474738.

Fender, Andreas Rene, and Christian Holz. "Causality-preserving Asynchronous Reality." Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, Apr. 2022, pp. 1–15, https://doi.org/10.1145/3491102.3501836.

George, Ceenu, et al. “Fearing Disengagement from the Real World.” 25th ACM Symposium on Virtual Reality Software and Technology. VRST ’19, Association for Computing Machinery, 2019, pp. 1–5, https://doi.org/10.1145/3359996.3364273.

Gruenefeld, Uwe, et al. “VRception: Rapid Prototyping of Cross-Reality Systems in Virtual Reality.” Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI ’22, Association for Computing Machinery, 2022, pp. 1–15, https://doi.org/10.1145/3491102.350182

Iwai, Daisuke, and Kosuke Sato. “Limpid desk: see-through access to disorderly desktop in projection-based mixed reality.” Proceedings of the ACM symposium on Virtual reality software and technology. VRST ’06, Association for Computing Machinery, Nov. 2006, pp. 112–15, https://doi.org/10.1145/1180495.1180519.

Lindlbauer, David, and Andy D. Wilson. “Remixed Reality: Manipulating Space and Time in Augmented Reality.” Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI ’18, Association for Computing Machinery, 2018, pp. 1–13, https://doi.org/10.1145/3173574.3173703.

Lindlbauer, David, et al. “Context-Aware Online Adaptation of Mixed Reality Interfaces.” Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology. UIST ’19, Association for Computing Machinery, Oct. 2019, pp. 147–60, https://doi.org/10.1145/3332165.3347945.

Mann, Steve. “Mediated Reality.” Linux Journal, vol. 1999, 59es, Mar. 1999, 5–es.

Mark, Gloria, et al. “How blocking distractions affects workplace focus and productivity.” Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers. UbiComp ’17, Association for Computing Machinery, 2017, pp. 928–34, https://doi.org/10.1145/3123024.3124558.

McGill, Mark, et al. “A Dose of Reality: Overcoming Usability Challenges in VR Head-Mounted Displays.” Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. CHI ’15, Association for Computing Machinery, 2015, pp. 2143–52, https://doi.org/10.1145/2702123.2702382.

Metzger, Paul J. “Adding reality to the virtual.” Proceedings of IEEE Virtual Reality Annual International Symposium. IEEE, 1993, pp. 7–13.

Milgram, Paul, et al. “Augmented reality: A class of displays on the reality-virtuality continuum.” Telemanipulator and Telepresence Technologies, vol. 2351, Jan. 1994. https://doi.org/10.1117/12.197321.

Raskar, Ramesh, et al. “The office of the future: a unified approach to image-based modeling and spatially immersive displays.” Proceedings of the 25th annual conference on Computer graphics and interactive techniques - SIGGRAPH ’98. ACM P, 1998, pp. 179–88, https://doi.org/10.1145/280814.280861.

Ruvimova, Anastasia, et al. ““Transport Me Away”: Fostering Flow in Open Offices through Virtual Reality.” Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI ’20, Association for Computing Machinery, Apr. 2020, pp. 1–14, https://doi.org/10.1145/3313831.3376724.

Speicher, Maximilian, et al. “What is Mixed Reality?” Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI ’19, Association for Computing Machinery, May 2019, pp. 1–15, https://doi.org/10.1145/3290605.3300767.

Wang, Chiu-Hsuan, et al. “RealityLens: A User Interface for Blending Customized Physical World View into Virtual Reality.” Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology. UIST ’22, Association for Computing Machinery, 2022, https://doi.org/10.1145/3526113.3545686.