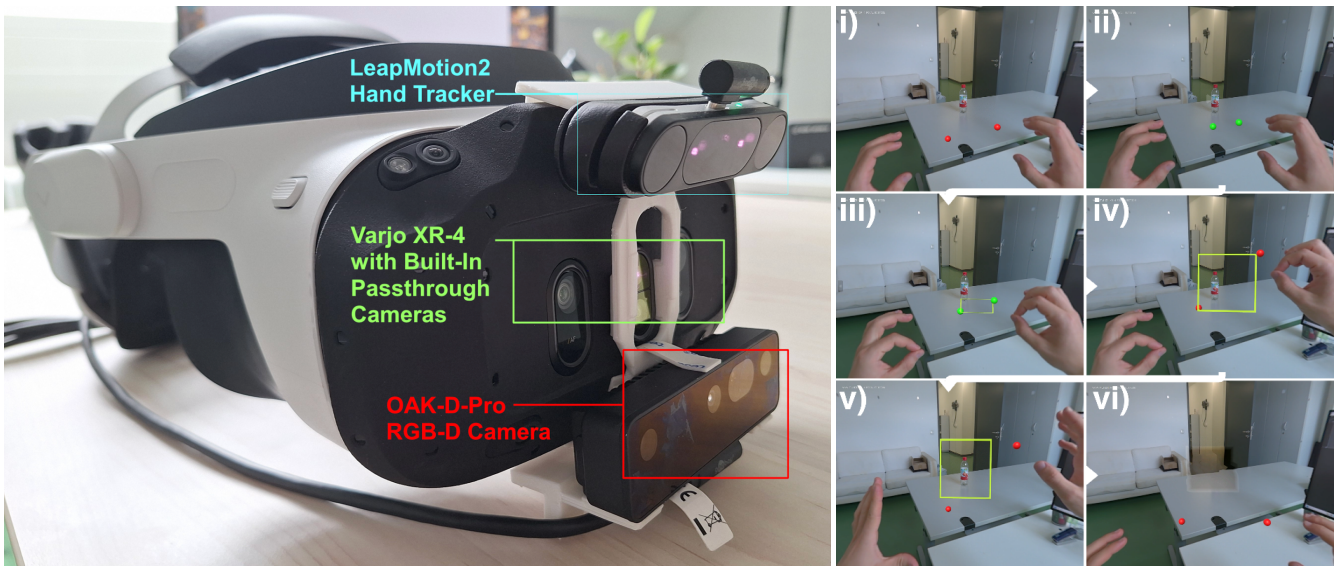
Selective Vision: Designing and Evaluating a Real-Time Diminished Reality System for Mixed Reality Devices

Oliver Hein
University of the Bundeswehr Munich
Munich, Germany
oliver.hein@unibw.de

Robin Piening
LMU Munich
Munich, Germany
robin.piening@campus.lmu.de

Florian Alt
LMU Munich
Munich, Germany
florian.alt@lmu.de



(a) Varjo XR-4 with Leap Motion 2 and OAK-D Pro Camera

(b) Bimanual Pinch Selection Gesture

Figure 1: (a) Hardware setup combining a Varjo XR-4 HMD, Leap Motion 2 hand tracker, and OAK-D Pro RGB-D camera for data input. (b) Bimanual pinch selection gesture: the user defines a rectangular region by pinching with both hands, spanning the selection bounds. When the pinch is released, the selected area will be diminished.

Abstract

Diminished Reality (DR) can visually suppress physical objects and may help users manage attention in mixed-reality environments, yet live, user-controlled DR remains difficult to implement and evaluate. We present *Selective Vision*, a proof-of-concept MR system that combines RGB-D capture, hand-tracking input, inpainting, and 3D mesh updating to let users manually select, diminish, and restore real-world regions. In a study with 28 participants, we examined exploratory and task-oriented use. Participants quickly learned the interaction, used DR to reduce visual clutter, and generally preferred manual control over fully automatic removal. Although blending, alignment, and reconstruction artifacts remained visible, diminished regions were often perceived as acceptable and sometimes less distracting than the original objects. Our findings suggest that future DR systems should prioritize fast, low-effort selection, transparent feedback, reversibility, and user-confirmable suggestions, while improving segmentation, latency, viewpoint robustness, and temporally stable reconstruction.

CCS Concepts

• **Human-centered computing** → **Mixed / augmented reality**; *Empirical studies in HCI*; Gestural input.

Keywords

Diminished Reality, Mixed Reality, Gesture Interactions

ACM Reference Format:
Oliver Hein, Robin Piening, and Florian Alt. 2026. Selective Vision: Designing and Evaluating a Real-Time Diminished Reality System for Mixed Reality Devices. In *Proceedings of the ACM on Human-Computer Interaction (MobileHCI '26)*. ACM, New York, NY, USA, Article MHCI2903, 14 pages.
https://doi.org/10.1145/3821681

Authors' Contact Information: Oliver Hein, University of the Bundeswehr Munich, Munich, Germany, oliver.hein@unibw.de; Robin Piening, LMU Munich, Munich, Germany, robin.piening@campus.lmu.de; Florian Alt, LMU Munich, Munich, Germany, florian.alt@lmu.de.

This work is licensed under a Creative Commons Attribution 4.0 International License.
MobileHCI '26, Swansea, Wales, UK
© 2026 Copyright held by the owner/author(s).
ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/3821681

1 Introduction

Augmented reality (AR) technologies have long promised to enrich users' perception of the physical world by overlaying digital information onto their surroundings. A complementary, but less extensively explored, paradigm is Diminished Reality (DR). Rather than adding visual content, DR aims to remove, hide, or attenuate selected real-world objects from a user's view, often by reconstructing or approximating the background behind them. Such techniques have been proposed to reduce distraction, protect privacy, or simplify visually complex environments. Potential applications include suppressing distracting signage, reducing visual clutter in workspaces, or hiding sensitive objects and bystanders from view. While DR has been studied for more than a decade, many existing systems focus on static images, offline video processing, or mobile AR prototypes with constrained interaction and rendering capabilities. Recent reviews indicate that immersive head-mounted DR systems remain comparatively rare, with only 18 percent of studies employing them in 2024 cite eskandari2024observation. The growing availability of high-resolution mixed-reality headsets creates new opportunities to investigate DR as an interactive, live, and user-controlled experience. However, implementing such systems remains technically challenging. A live DR pipeline must integrate object selection, sensor input, background completion, depth or geometry handling, and low-latency rendering. At the same time, the interaction must remain understandable and fast enough for users to apply DR during ongoing tasks.

In this paper, we present *Selective Vision*, a proof-of-concept diminished reality system for immersive mixed reality. The system is implemented using a Varjo XR-4 headset, an additional RGB-D camera, and hand-tracking input. It allows users to manually select real-world regions, visually diminish them through live inpainting, and restore them when desired. Our system explores manual user control as one possible interaction model for live DR.

To examine how users interact with and perceive this type of system, we conducted a study with 28 participants in exploratory and task-oriented scenarios. Participants were able to learn the interaction quickly and used the system to reduce visual clutter in the environment. They generally preferred manual control over fully automatic removal. At the same time, participants noticed limitations in blending and reconstruction quality, and their responses to the inpainted regions were mixed. These findings suggest that, for some constrained attention-management tasks, diminished regions may be acceptable even when reconstruction is not photorealistic. However, they also highlight important limitations around visual artifacts, selection precision, and technical robustness.

Our work differs from pre-scan-based DR approaches by operating on live sensor input rather than relying on a previously captured model of the environment. This enables a more flexible form of interaction, but also introduces practical constraints related to sensing, latency, viewpoint stability, and computational cost. We therefore treat *Selective Vision* as a proof-of-concept system rather than as a mature deployment-ready solution.

Contribution Statement. This paper makes three contributions: (1) We present *Selective Vision*, a proof-of-concept live diminished reality system for immersive mixed reality that combines RGB-D capture, hand-tracking input, inpainting, and 3D mesh updating. (2) We describe an interaction workflow that allows users to manually select, diminish, and restore real-world regions during live MR use. (3) We report findings from a user study with 28 participants comparing exploratory and task-oriented DR use, providing insights into usability, perceived acceptability, manual control, and the role of visual artifacts in future DR systems.

2 Related Work

2.1 Classification and state of research

Diminished reality, situated opposite to Augmented Reality in Steve Mann's Mediated Reality Continuum [26] can be considered a form of Modulated Reality, and generally achieves the highest grade of perceptual immersion when employed in Mixed Reality HMDs due to their capability for view-encompassing augmentations.

According to Eskandari and Motamedi [7], previous DR research has primarily focused on implementation techniques (50 percent) and application concepts (40 percent), while user experience and perception, particularly in immersive head-mounted displays, have received comparatively less attention. Therefore, we review prior work on technical foundations and investigate use cases, user perceptions, and ethical risks.

2.2 Technical Foundations

Mori et al. [30] describe three core components of diminished reality: defining the area of interest to be removed, obtaining or generating background content to replace it, and compositing the result into the user's view through tracking and blending.

Obtaining Areas of Interest. Areas on objects for diminishing can be chosen beforehand using markers [4, 18, 28, 39, 40], manual selection during study [2, 32, 44], or by object detection [19, 35]

Obtaining Background. Obtained Background can be recorded beforehand (by multiperspective recording [3, 4, 30], by 3D scan of the environment [2, 19, 32] or using images from public sources [24]), realtime recording ([17, 31, 45]) or plausible (generative) inpainting (Algorithmic methods: [1, 12, 13, 39], Generative AI: [5, 6, 9, 11, 14, 16, 22, 23, 36, 37, 41, 48, 51]). It is also feasible to disregard background obtainment, and diminish objects by simply overlaying digital shapes [21], or diminishing visual vectors of a targeted object, e.g. saliency and saturation [2], or potentially replace it with desired template imagery, (e.g. forest), which does not seem to have been done before. The described methods differ primarily in visual realism, informational accuracy, and the (temporal) cost of acquisition, as well as flexibility in adapting to arbitrary scenes. A combination of methods is possible and promising, and has been done in the past [20]

Considerations on Composition. Generating a plausible background for an object itself is, in many cases, insufficient, as the object leaves ambient information in the form of shadows, reflections, etc., which might not be targeted by the user or an automation. Similarly, in the event of changes to the light level or white balance, the applied augmentation might no longer hold up and become visible as such. Kawai et al. [18] applied information from the surrounding area of a marked area to mitigate these issues. Another consideration in head-mounted mixed reality is residual depth information. It might

not be sufficient to simply overlay or inpaint 2D information, as it would be in a picture. Gsaxner et al. [11] designed a technique for "Structure-Aware RGB-D Inpainting", which not only reconstructs the RGB visual data of the selected area but also regenerates a depth mesh from the surrounding depth context.

2.3 Investigated Use Cases

Removal of obstructions. One of the most plausible use cases for diminished reality is indoor renovation, which has been studied by Zhu et al., who used a panoramic camera [49] and generative adversarial networks [50] to reconstruct background information and remove walls and objects in indoor planning. Similarly, DR has been investigated by Inoue et al. [15] for outdoor environmental planning, focusing on improving tracking robustness, and Kido [19] extended the approach with automatic dimming of movable objects and people entering the scene. Mori et al. [31] developed a multiperspective recording technique to remove power tools from a user's hand, and Taylor et al. [45] diminished a robotic arm controlled by the user; however, they found that these approaches offered no significant benefits and increased perceived mental demand. Lindemann and Rigoll [25] reduced opacity on cars to increase drivers' awareness of their surroundings in a virtual scene in a CAVE simulator, with positive results.

Removal of distractions. Lee et al. [21] projected AR objects to diminish the potentially distracting smartphone and investigated the effects of different visual qualities; this reduced distraction despite appearing clearly artificial. Murph et al. [34] investigate DR in a virtual assembly task by removing and gradually reintroducing distractions to help learners adjust to them. Yantaç et. al. [47] explored the idea of using DR to disburden the minds of individuals with autism. McDonald [27] conducted a survey on DR in learning environments, including the effect on people with ADHD.

Privacy. Tabet et. al. [42–44] are developing a framework for mobile diminished reality, which can select, detect and track objects for diminishing in a 3D-scene while implementing safeguards to keep it covered during induced uncertainties. Flores and Belongie [8] and Yagi et al. [46] draw from a set of street-view images to remove pedestrians, granting them more privacy than regular blurring.

2.4 User Perception

A central reference point for our work is Cheng et al.'s 'Towards Understanding Diminished Reality' [2], which conducted the first empirical exploration of how users perceive and apply DR effects in a 3D-immersive scenario. Using VR-based simulations of scanned environments, they compared opacity adjustment against several mediated reality alternatives (e.g., blur, desaturation, scaling) across tasks such as clutter management, search, and privacy protection. Their findings show that users strongly prefer opacity and outline effects, and that acceptability is shaped by contextual factors such as mobility, social presence, and risk of collision. A second experiment demonstrated that participants customized DR to reduce the number of objects by 30 percent, generally avoiding items that might interfere with movement or social interaction. From these insights, Cheng et al. distilled design guidelines, emphasizing the roles of contextual awareness, safety, and user agency in DR.

2.5 Applications Ethics and Risks

In Cheng et al.'s study [2], users themselves could choose which objects to diminish while performing a simple task (including a physical one). Their reports confirm that users strongly prefer a sense of agency, e.g., the ability to change the augmentation's intensity, customize it, or turn it off completely, or at least to have information about its diminishing effects. They also see risks of bad actors hijacking users' devices and manipulating them against their interests by employing diminishing augmentations without their noticing. They also highlight the dystopian outlook of DR being used to diminish people, citing the risk of social exclusion, with a general aversion of participants to diminish other people.

Moffis et al. [29] points out the risk of using the technology in dynamic scenes, especially until the technology has fully matured, as it might accidentally diminish information that could become relevant in the future due to collision risks. They also highlight caution about potentially unresearched effects on the psyche, especially during prolonged use (e.g., reduced situational awareness [33] and cognitive performance), and the risks of use in decision-critical environments, such as nuclear plants and the military.

3 Implementation

In this section, we explain the apparatus (see Figure 1a), the developed interaction techniques (see Figure 1b), and the processing pipeline. Additionally, we discuss our system's performance.

3.1 Hardware Architecture

Head-Mounted Display. To facilitate optimal immersion while simulating the absence of objects and ensuring coverage of augmentations across the entire field of view, a mixed reality-capable HMD was deemed the best solution. At the time of research, the best available device was the Varjo XR-4 headset, offering a 90 Hz refresh rate at greater than 4k resolution per eye, with a 120-degree horizontal field of view. The high 20 MP per-eye pass-through resolution enabled a sharp view of the real world, keeping realism as close to visual see-through AR as possible.

RGB-D Camera Mount. Accessing the headset's pass-through stream was not possible in a sufficient manner; thus, a secondary camera was needed to infer depth data and visual context for inpainting. Multiple RGB-D cameras were considered, including ZED-Mini, Intel RealSense, and OAK-D. The OAK-D was chosen for its good form factor, high resolution, and good dynamic range in depth mapping. The camera was mounted on a self-designed 3D-printed mount, with the RGB sensor placed as close as possible to the headset's pass-through cameras (i.e., the user's perspective) without obstructing the user's field of view.

Hand-Tracking Sensor. Because controllers may not be available or desirable in many real-world DR scenarios, we used hand tracking as the input method, enabling precise selection with a high degree of freedom. After exploring Varjo-supported hand tracking, we found it insufficiently reliable for the selection gestures required in our prototype and therefore decided to use an external Leap Motion 2 device. The hand tracker was mounted on the same 3D-printed structure as the camera (see Figure 1a).

3.2 Interaction Technique

Bimanual Pinch Selection. Users created the selection by pinching with both hands, with the right cursor above the left cursor, expanding the rectangle from the center of the field of view, and placing it by releasing the pinch over the target object or area (see Figure 1b). To support this interaction, the system displayed two spherical cursors, each approximately 2 cm in diameter, about 20 cm in front of the user's hands, and a yellow selection box. The cursors changed from red to green when both hands were close enough to the center of the field of view to start a valid selection.

Deletion Gesture. Users could choose between diminishing and restoring by changing the vertical order of the two cursors. If the left cursor was above the right cursor, releasing the pinch removed an existing diminished region, indicated by a blue selection box (see Figure 2). For restoration, the system checks whether the center vertex of an existing diminished mesh lies within the selection box; if so, the entire mesh is deleted, and the original view is restored.

3.3 Processing Pipeline

Upon the user selecting an area for diminishment, the system generates a mask corresponding to the selected region and projects it onto the RGB-D camera view. The mask and RGB image are sent to the inpainting server, while the system collects and smooths depth data during inpainting. After inpainting and depth processing are complete, the system generates a mesh at the selected location and scale. This mesh displays a depth-aware relief of the inpainted region with faded edges for improved blending (see Figure 1b, 3, 4).

3.3.1 Mask Generation. To generate the mask, the system treats the two cursors as opposing corners of a rectangle that remains upright relative to the user's field of view. The projection from the cursor's 3D coordinates to 2D coordinates on the RGB-D camera feed was performed using polar coordinates relative to the center of the user's field of view. A black-and-white mask image with a rectangle in place of the selection would be generated and sent to the inpainting server, along with the OAK-D camera's RGB image.

3.3.2 LaMa Inpainting. For the inpainting itself, the LaMa (Large Mask) Model [41] was used, as it offered a good tradeoff between realism and processing time. More in-depth diffusion models were considered but not used due to much higher processing times (up to several seconds) and the tendency to hallucinate unwanted objects into the inpainted area. The MAT model [22] was considered as well, as its processing times were on par with LaMa, but the replacement was often more chaotic. The LaMa inpainting was, at times, visually low in detail and sometimes resulted in homogeneous regions of low detail, especially for larger selections, but this was deemed preferable because it was less salient and distracting.

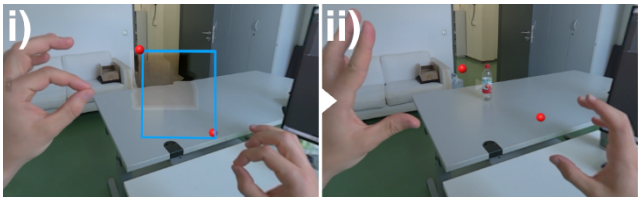


Figure 2: Deletion Gesture: When the selection is released over a diminished area, the system removes it and restores the original view.

3.3.3 Depth-Aware Re-projection. While the OAK's depth stream offered a good dynamic range spanning all possible distances of the setup (except very close), it was noisy and unstable. To work around this, multiple (20) frames were recorded during the processing of the inpainting and refined in 3 steps: 1. averaging over valid values (>0.01, <0.15), 2. still missing non-valid values inferred from the closest valid value 3. smoothed by only considering every pixel divisible by 10, which also reduced processing times. Vertices were shifted up or down depending on the measured depth to account for the slight difference in perspective between the mounted camera and the headset's passthrough cameras. Examples of generated inpaintings and final overlays are shown in Figure 4.

3.4 System Performance & Limitations

A practical limitation of our prototype is that it relied on additional sensing hardware rather than solely on the headset's built-in capabilities. Although the Varjo XR-4 supports hand tracking, the Varjo-supported hand-tracking pipeline was not sufficiently reliable for the selection gestures required in our prototype. We therefore used an externally mounted Leap Motion 2 device to obtain more stable hand input that could be integrated into our custom interaction and processing pipeline. Similarly, access to the headset's native passthrough camera streams and depth information was limited by the available APIs and platform restrictions. As a result, we used an additional RGB-D camera for live scene capture and depth sensing. This configuration enabled us to demonstrate live, user-controlled diminished reality, but it also introduced practical constraints. The externally mounted devices increased the setup's physical complexity and may have affected headset comfort, weight distribution, and perceived encumbrance.

The current implementation is also constrained by processing latency and viewpoint stability. The full processing pipeline, from placing the selection to displaying the smoothed, edge-faded, diminished mesh, took approximately 3 seconds on average. During this time, the system displayed a waiting icon to encourage participants to keep their head still, as head movement could introduce misalignment between the captured RGB-D data, the generated inpainting, and the rendered diminished region (see Figure 3). This constraint limits the current prototype to relatively static scenes and short periods of viewpoint stability. While this was acceptable for our controlled study setup, future mobile or wearable DR systems will require lower latency and more robust handling of head movement, dynamic objects, and changing lighting conditions. Further limitations arise from the selection and reconstruction approach. Object selection was based on rectangular cropping rather than semantic segmentation, which made the interaction simple and fast but reduced precision. As a result, selections could include parts of the surrounding background or fail to align closely with object boundaries. The lack of object-aware segmentation also limited the system's ability to distinguish foreground objects from adjacent surfaces. In addition, the fidelity of the diminished regions was affected by differences in color balance, white-balance changes, slight

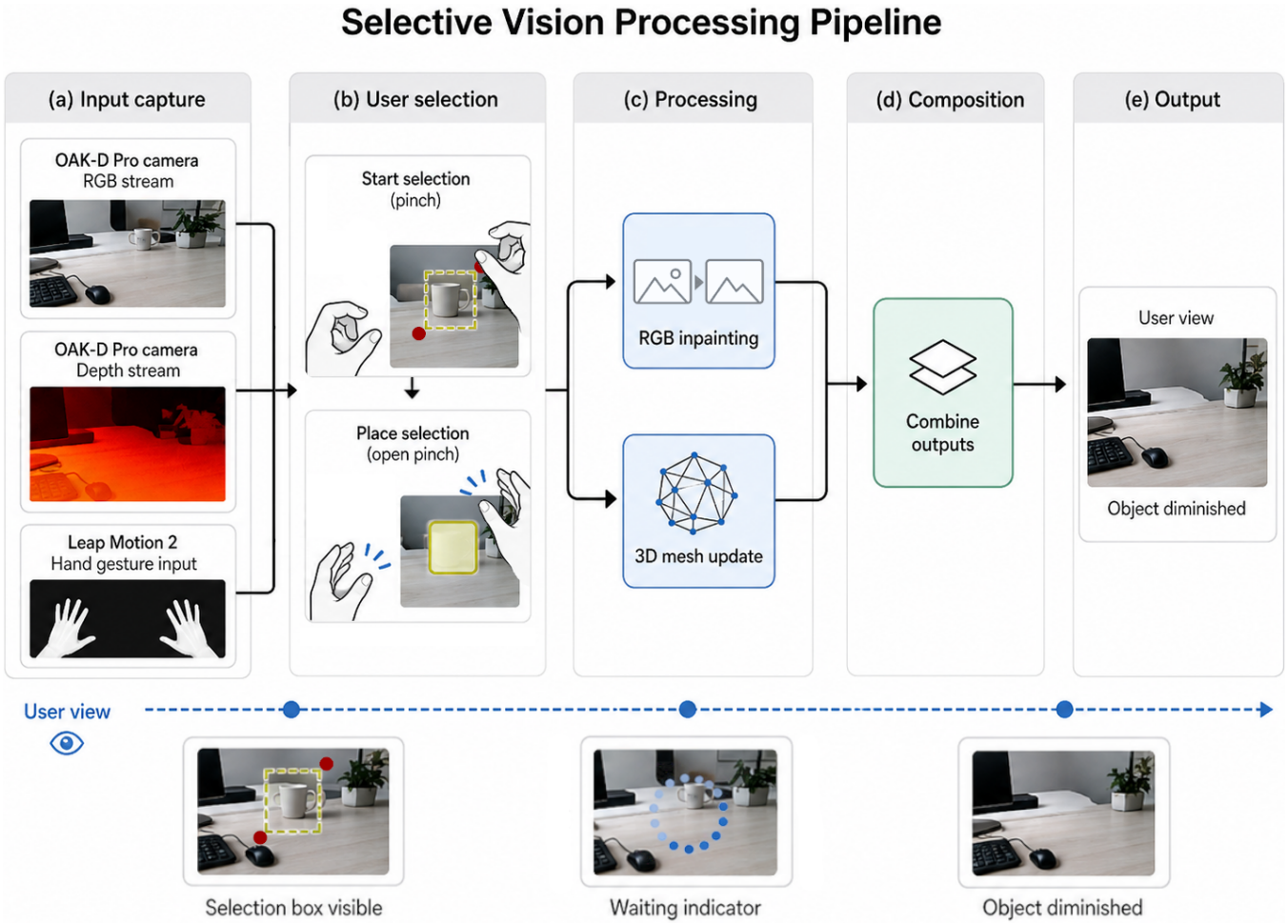


Figure 3: Overview of the processing pipeline: RGB and depth data are captured from the OAK-D Pro, while hand input is captured using Leap Motion. A pinch gesture starts the selection, and opening the pinch places the selected region. The system then performs RGB inpainting and 3D mesh processing in parallel, before combining the two streams into the final diminished view. During processing, the user first sees the selection box and then a waiting indicator; once processing is complete, the selected object is visually diminished.

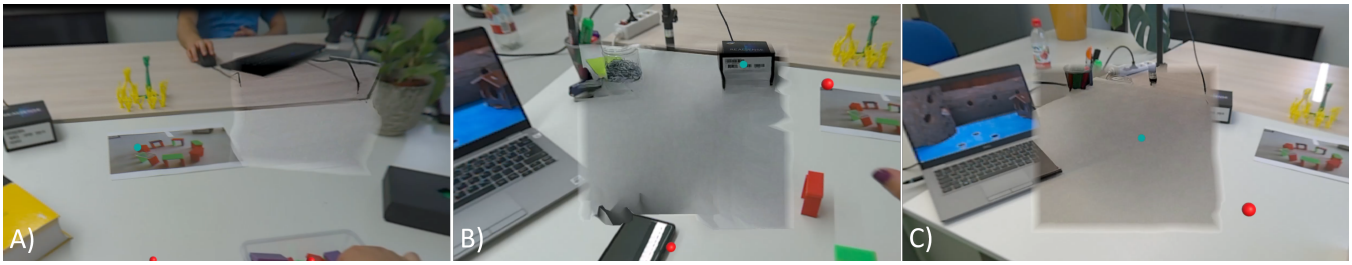


Figure 4: Inpainting examples: A) a visually plausible inpainting, B) color and depth inaccuracies, C) color and alignment inaccuracies.

shifts in the external camera mount, and depth inaccuracies, particularly at close range. The inpainting quality decreased for larger selections and for selections near the border of the OAK-D camera's field of view, where less reliable image and depth information was available.

Taken together, these constraints mean that our prototype should be understood as a proof-of-concept system rather than a deployment-ready DR solution. In favorable cases, the system produced a convincing approximation of object diminishment and allowed participants to experience how live, user-controlled DR could function in immersive mixed reality. However, the observed limitations also indicate that future systems will need more robust sensing, direct access to headset camera streams, lower-latency processing, more precise yet low-effort selection methods, semantic or instance-level segmentation, and more temporally stable inpainting and depth reconstruction.

4 User Study

Our study was inspired by Cheng et al.'s prior investigation of diminished reality in a pre-scanned virtual environment [2]. In their study, participants were seated at a table surrounded by clutter and potential targets for diminishing. Their evaluation included two experiments: one comparing different forms of diminishment and one examining the use of DR to reduce distraction in attention-focused and building tasks. We adopt a comparable two-part structure, consisting of an exploratory scenario in which participants could freely try the technique and a task-oriented block-assembly scenario in which DR could be used to reduce distractions.

4.1 Research Questions

The main question in our study is to evaluate the usability and acceptance of the technique to deliver a proof of concept for the prototype by gauging how well the interaction mechanism, 3D inpainting, and the overall concept work for users in an exploratory and task-focused scenario. We also want to gather insights into user preferences regarding the selection of objects and areas for diminishing, as well as the perceived acceptability and subjective responses to diminishing different types of objects. Finally, we compare our findings with Cheng et al.'s prior VR-based study to examine how their insights translate to a live MR setup. Therefore, we formulated the following research questions:

RQ1. Usability and Acceptance: How usable, intuitive, and acceptable is real-time diminished reality (DR) interaction in a mixed reality (MR) environment?

RQ2. User Preferences and Subjective Experiences: What are users' preferences for object selection and their subjective responses to diminishing different types of objects in both free-exploration and task-oriented scenarios?

RQ3. Comparison to Prior Work: How do our findings compare to those reported in previous VR-based DR studies (e.g., Cheng et al.), and do they replicate or extend these insights in a more realistic MR setting?

4.2 Experimental Design

4.2.1 Tasks. After completing the demographic and expertise questionnaire, participants received an introduction to the prototype and its interaction workflow. The experimenter explained how objects or regions could be selected, diminished, and restored. Participants also watched a short video demonstrating the placement-and-removal interaction. Before starting the tasks, participants were informed about the prototype's current technical limitations, including that the system produced the best results for medium-sized selections near the center of the camera view and that they should keep their heads as still as possible during processing. After confirming that they understood the interaction, participants put on the headset and were seated in front of the experimental setup.

Exploratory Use. In the exploratory phase, participants were asked to freely try the diminished reality interaction in a controlled scene containing several physical objects (see Figure 10). They were not instructed to remove specific target objects. Instead, they could choose for themselves which objects or regions to diminish or restore, allowing us to observe how they appropriated the technique without a predefined task goal. Participants were asked to indicate when they felt they had finished exploring the system. Afterward, they completed the UEQ-S questionnaire.

Block-Assembly Task. In the second task, participants assembled a physical block structure based on a reference photo. Before and during the assembly task, they could use the diminished reality interaction to remove objects or regions from view whenever they believed this would help them complete the task, for example, by reducing visual clutter or clearing parts of the visual scene (see Figure 11). Participants were also told that they could activate or deactivate the selection controls by saying the voice commands "activate" and "deactivate". This used a Wizard-of-Oz approach: the experimenter manually enabled or disabled the controls in response to the participant's command.

After completing the block structure, participants removed the headset and completed a second UEQ-S questionnaire. They then filled out a summative experience questionnaire reflecting on the overall experience across both phases.

4.2.2 Scene Setup. The setup happened in a medium-sized office room (ca. 3x4 meters), and users were seated at a large office desk. Notable objects on the table were: 3 small boxes, 1 bright yellow dictionary, office utensils, a small and large plant, a bottle, a large box with cables, a collection of 3D-printed tree structures, and a laptop in the user's periphery playing a video with mice running around (including chirping sounds). Other notable targets in the room were the door (deliberately left open), a couch with some clutter on it, a VR base station, 2 paintings (1 calming nature, 1 high-saliency optical-illusion artwork), and the presenter.

4.3 Participants

We recruited 28 participants through a university mailing list and the research institute's Slack channel. Fourteen participants identified as female, twelve as male, and two as diverse. Most participants were between 18 and 27 years old (n = 17) or between 28 and 37 years old (n = 10), and one participant was older than 57 years. All participants rated themselves as moderately to highly experienced with technology, selecting 3 or higher on a 5-point Likert scale. However, this general technology experience did not align with prior VR or AR experience: five participants had never used VR, thirteen had tried it once or a few times, and ten reported using VR at least occasionally.

4.4 Questionnaires & Measures

4.4.1 Subjective Scales (UEQ-S, Experience Report). The UEQ-S questionnaire measures user experience in terms of a product / prototype's attractiveness and usability. The Experience Report is designed to gather insights into the overall experience, featuring questions on the reception of the interaction mechanic, 3D inpainting, the user's preferences regarding object selection, as well as perceived acceptability, comfort, and distractiveness of the technique, and general preferences about possible usage scenarios. Responses are either given in text or on a 5-point Likert scale.

4.4.2 User Selections – Targeted Positions. For each placed diminishing selection, the position and depth of the mesh's center vertex were logged, thereby marking selected objects and enabling the inference of selection hotspots during reconstruction.

4.4.3 Gaze Logging. During the study, we recorded participants' gaze data and used it to determine how frequently and for how long they focused on diminished areas or objects.

Additional technical measures taken during the study include the time required for inpainting, the distance to the selected areas, and the time users spend looking at the inpainted areas.

4.5 Procedure

The study is structured into three parts: an introduction, followed by two main phases, 'Exploratory' and 'Task'.

First, participants are guided through a consent form and a short questionnaire to collect demographic information and assess their prior experience with mixed reality and related technologies. Next, the study conductor explains the prototype's functions and shows a short video demonstrating the bimanual pinch selection. After addressing any participant questions, the HMD, which is cleaned between sessions, is provided. Eye tracking and headset position are calibrated for optimal fit, after which the Unity scene is launched, and the 'Exploratory' phase begins.

During the 'Exploratory' phase, participants can freely interact with the prototype using a predefined set of objects. They are instructed: "Please try the technique however you like and note the impression it gives you. You can stop at any time." Participants may ask questions or request assistance if any issues occur. The conductor documents verbal comments, and once finished, participants complete a UEQ-S questionnaire [38] to report their experience.

In the 'Task' Phase, the prototype is evaluated in a goal-oriented context. The task replicates Cheng's study [2], in which participants are shown a photo of a shape and asked to recreate it as accurately as possible using a set of building blocks. Participants are provided with the photo, the box of building blocks, and the option to adjust the scene with diminishing augmentations to their preference. They may activate or deactivate controls (such as cursor visualizations, but not previously placed augmentations) via verbal commands, which the conductor executes using a Wizard-of-Oz approach [10]. The technique can be used freely during the task.

The goal of this phase is not to measure task performance but to gather insights into participants' preferences regarding diminished reality in a task-driven setting versus a free-exploration scenario. After the task, participants again complete the UEQ-S questionnaire and provide additional feedback through a more detailed concluding questionnaire on their overall impressions and experience.

4.6 User Study Limitations

The user study was designed as an exploratory evaluation of a proof-of-concept DR prototype rather than a controlled performance comparison. Several limitations should therefore be considered when interpreting the results. First, participants were recruited through a university mailing list and the research institute's Slack channel. This convenience sample likely included participants with relatively high levels of education, technology affinity, and openness to experimental systems, as reflected in their self-reported technology experience. While appropriate for an early-stage evaluation, this limits the generalizability of our findings to broader populations.

Second, the study took place in a controlled office environment with a predefined arrangement of objects. This setup does not fully reflect everyday environments, where lighting, object motion, social context, safety constraints, and task demands may vary substantially. Participants were also seated throughout the study, which limits conclusions about mobile DR use while walking, collaborating, or interacting with dynamic surroundings.

Third, the two study phases differed in task framing, order, and context. Because all participants completed the exploratory phase before the block-assembly task, differences between the two UEQ-S measurements may reflect learning effects, increased familiarity, or changed expectations rather than only differences between free exploration and task-oriented use. The UEQ-S results should therefore be interpreted as descriptive indicators of user experience in the two scenarios, not as a controlled comparison.

Several procedural aspects may also have shaped participant behavior. Participants were told that the prototype worked best for medium-sized selections near the center of the camera view and that they should keep their head still during processing. These instructions were necessary for stable use, but may have encouraged participants to adapt to the system's constraints. Similarly, activation and deactivation of the controls were implemented with a Wizard-of-Oz approach, so the voice-control component itself was not technically evaluated.

We also did not measure objective task performance in the block-assembly task. This was intentional: the task was used to observe how participants appropriated DR in a goal-oriented setting rather than to test whether the prototype improved performance. Future work could examine whether DR can measurably support attention management, productivity, or error reduction in specific tasks.

Finally, the gaze and selection logs provide useful behavioral context, but have interpretive limitations. Looking less at a diminished region does not necessarily prove that the removed object was distracting or that the inpainted region reduced distraction; it may also indicate that the area was not task-relevant. Gaze data in MR can also be affected by calibration accuracy, headset fit, and visual artifacts. We therefore treat gaze measurements as complementary indicators rather than definitive evidence of attentional effects.

Our findings show how participants experienced live, user-controlled DR in a constrained proof-of-concept setting, but they do not yet demonstrate general effectiveness across everyday environments or user groups.

5 Results

5.1 RQ1: Usability and Acceptance

5.1.1 UEQ-S. UEQ-S ratings exceeded average benchmarks for both Pragmatic (1.18) and Hedonic Quality (1.6) (see Figure 5). Participants slightly preferred the Exploratory scenario over the Task scenario (Pragmatic: 1.19 vs. 1.16; Hedonic: 1.74 vs. 1.46; Overall: 1.47 vs. 1.31). However, Wilcoxon signed-rank tests found no significant differences across scenarios for Pragmatic (p = .74), Hedonic (p = .41), or Overall Attractiveness (p = .66).

5.1.2 Interaction. The bimanual pinch-selection was rated positively: 20 of 28 participants rated it 5/5 for ease of learning (*Mean* = 4.57), and 13 rated it 5/5 for intuitiveness (*Mean* = 4.29). UEQ-S items on difficulty (S-2) and clarity (S-4) were likewise strong. Precision (3.39) and reliability (3.46) were lower but expected, given known limitations. Participants generally agreed they could accomplish intended actions (*Mean* = 3.74).

User suggestions focused on improved hand tracking, reduced input lag, freeform drawing, or object segmentation for finer control, clearer selection colors, and an always-visible tutorial. Some were distracted by the box projection enlarging at the end of a selection. These suggestions indicate that participants not only evaluated the visual output of DR, but also the reliability, legibility, and recoverability of the interaction workflow.

5.1.3 Inpainting. Participants rated blending quality as mediocre (*Mean* = 2.54), but overall plausibility slightly above neutral (3.36) and unobtrusiveness moderate (3.21) (see Figure 6). Despite imperfections, users agreed the technique effectively removed objects (*Mean* = 3.61).

Common issues included poor color adaptation (3), low saturation (2), alignment errors (3), and blockiness (1). Users desired higher fidelity (5), and some found limited blending quality more distracting than the original objects (2). Inpainting performed better on medium-sized, mid-distance items (1–3 meters), and worse on close objects due to depth inaccuracies from the OAK-D Pro. This suggests that current DR prototypes may be acceptable for reducing visual salience, but that artifact severity remains a central factor for trust and perceived usefulness.

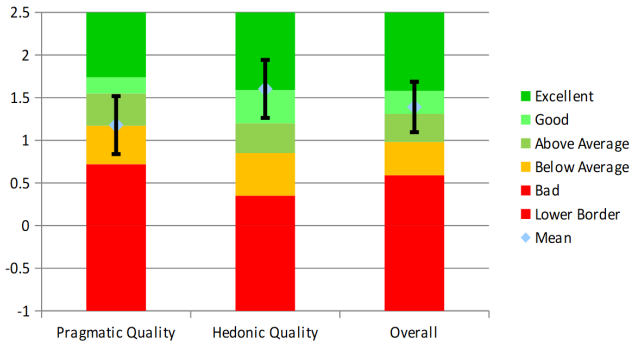


Figure 5: UEQ-S: general results for both tasks combined.

5.2 RQ2: User Preferences and Subjective Experiences

5.2.1 Selection Preferences. Users preferred selecting objects over areas (*Mean* = 1.75; 1 = object-only) (see Figure 7). Objects cited as producing positive effects when diminished included: psychedelic painting (calmer view), laptop/video (reduced peripheral distraction), yellow book (interfered with building), the researcher (less social pressure), bottle and clutter items (cleaner workspace). Objects considered negative to diminish were usually technically difficult: laptop/video (limited blending, dark overlay), close or large objects, small cluttered items, and edge-adjacent objects. Some users disliked removing objects coherent with the scene (e.g., pens), items at risk of collision, or items blocking the task view. One noted superficial removal of 3D objects. Objects users avoided removing included the door (to not feel trapped), items beyond 3 meters, themselves, and task-related materials (instructions and blocks).

5.2.2 Gaze Behavior. Inpainting meshes attracted little attention: 9.1% gaze time in the exploratory scenario and 2.7% during the task, suggesting overlays were not strongly distracting.

Because gaze behavior alone cannot determine whether the original objects were distracting before diminishment, we interpret these values as descriptive indicators of attention rather than as direct evidence of distraction reduction.

5.2.3 Subjective Response. Most participants felt comfortable using the technique (*Mean* = 4). Discomfort stemmed from the headset (3) or technical imperfections (3), not from the concept of DR itself—though overlays were sometimes described as "trippy" or "confusing." Distractiveness ratings leaned slightly toward not distracting (*Mean* = 2.54). Non-distracted users cited interface simplicity; distracted users mentioned cursor interference during the task (4), the added decision load of selecting objects (3), and limited blending (2). Overall perceived impact was near neutral but slightly positive (*Mean* = 3.39), driven largely by novelty (5) and improved focus during the task (3).

5.3 RQ3: Comparison to Prior Work

5.3.1 Potential Use Cases. Participants could imagine using the technique in everyday life if implemented in lighter hardware (*Mean* = 3.10). Most commonly, they would use it to remove distractions while working (12), in high-distraction environments (cafeterias, coworking spaces), or to reduce peripheral movement. Other ideas included removing signs of civilization (2), ad-blocking (1), hiding mess (1), R&D (1), gaming (1), or virtual meetings (1).

A modified version was considered similarly useful (*Mean* = 3.32). Desired improvements included better blending (4), higher inpainting quality (5), automatic suggestions (4), see-through walls (1), moving-object tracking (2), recoloring (1), collision warnings (1), and beautification of scenes (1). Some participants explicitly opposed fully automatic selection (2), preferring optional suggestions that require confirmation. With modifications, participants envisioned use in shared spaces, public transit, architectural planning, and the replacement—not removal—of objects.

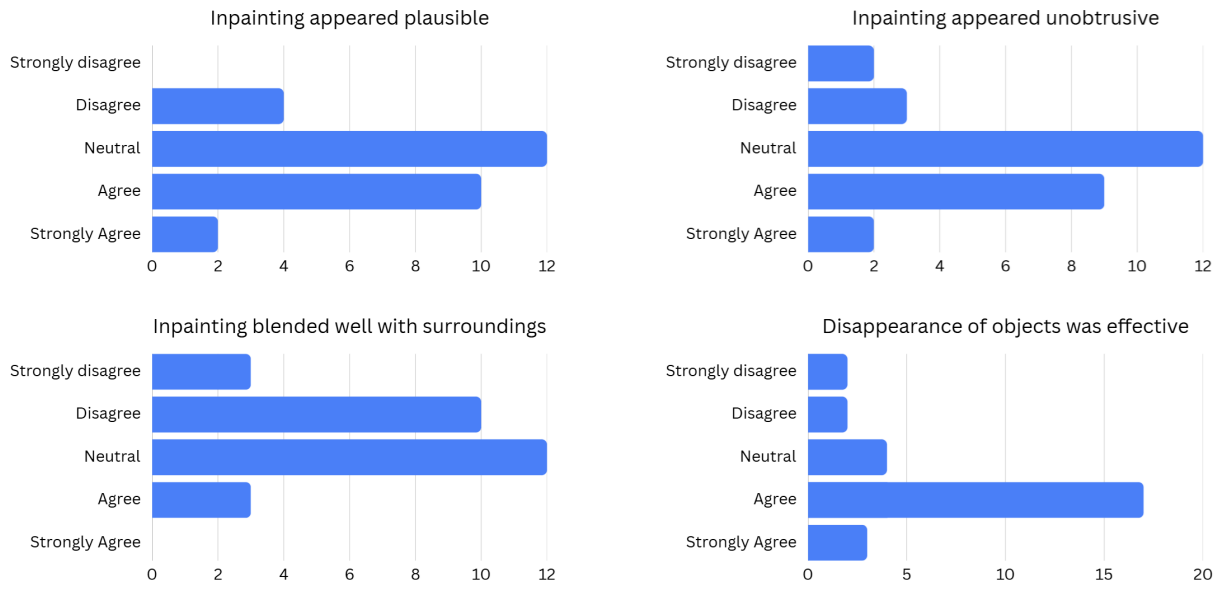


Figure 6: Inpainting ratings

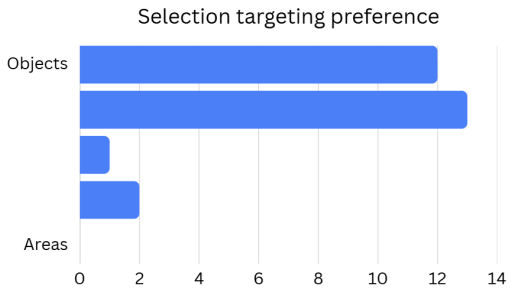


Figure 7: Selection preferences: objects vs areas

5.3.2 Object Placements. During exploration, participants placed 653 meshes (Figure 10); during the task, 139 meshes (Figure 11). Nine participants did not remove peripheral objects during the task.

Most meshes target objects close to the user, especially items on the table. In the exploratory scenario, common distant targets were the couch area, wall paintings (notably the psychedelic artwork), messy boxes, the bottle, the VR lighthouse, and the study conductor. In the task scenario, table objects and the presenter were targeted.

Participants rarely removed the instruction image, block reservoir, or their active build. Average mesh distance in exploration was 1.66 meters ($Median = 1.25$, $\sigma = .90$, $n = 612$; outliers >4 removed).

In the task scenario, it was 1.16 meters ($Median = 0.91$, $\sigma = .66$, $n = 135$).

6 Discussion

We investigated how participants experienced live, user-controlled diminished reality in an immersive mixed-reality setting. Rather than treating DR only as a problem of visual removal or background reconstruction, our findings show that live DR also raises central HCI questions about interaction, control, feedback, trust, and acceptable imperfection. In the following, we discuss what our results imply for future DR systems and how they relate to prior work on diminished reality, interaction design, and user control.

6.1 From Visual Removal to Interactive Diminishment

A central insight from our study is that live diminished reality should be understood not only as a visual computing challenge but also as an interaction problem. In image- or video-based DR, the main goal is often to reconstruct the hidden background as plausibly as possible. In live MR, however, users also need to decide what to diminish, initiate the removal, wait for processing, interpret the result, and restore the original view if needed. The quality of the experience, therefore, depends not only on the visual result but also on the interaction workflow around it.

This distinction helps explain why participants' evaluations were not solely determined by inpainting quality. Participants noticed visual artifacts, but their overall assessment also depended on whether the interaction felt understandable, controllable, and useful in the moment. This suggests that live DR systems should not be evaluated only by reconstruction realism. They should also be assessed by whether they support users in managing attention, reducing clutter, and maintaining confidence in what has been visually altered.

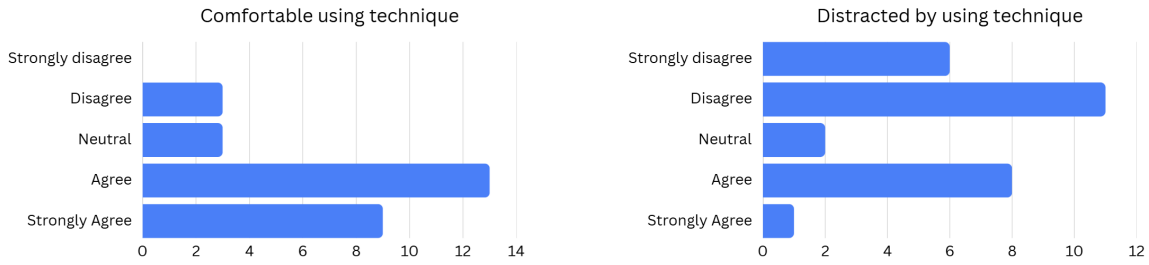


Figure 8: Perceived acceptability: comfort and distractiveness

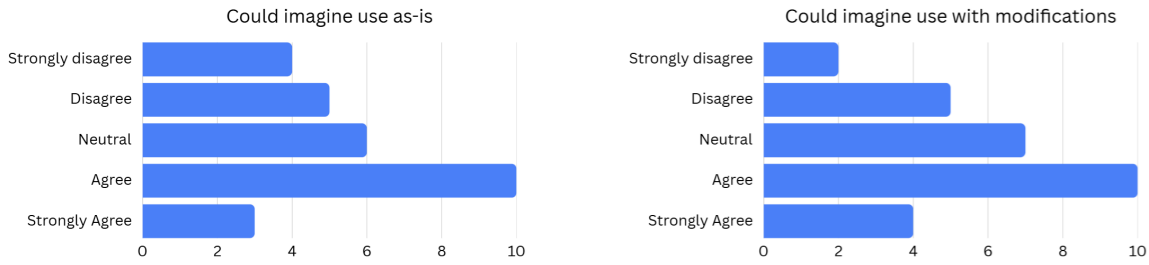


Figure 9: Use cases: as-is vs with modifications



Figure 10: Diminished area placements during the 'Exploratory' phase (marked with yellow dots)

Our work, therefore, contributes beyond the specific prototype implementation by showing how diminished reality can become an active, user-directed form of perception management in MR. Compared with pre-scanned or VR-based DR approaches such as Cheng et al. [2], our system explored DR on live sensor input in a physical environment. This made limitations such as latency, viewpoint dependence, and imperfect reconstruction more apparent, but it also allowed us to study how users respond to DR as an interactive experience rather than as a preconfigured visual effect.

6.2 The Role of Manual Control in DR

Participants' preference for manual control indicates that object removal in DR is not a purely technical decision. Removing an object also removes information from the user's environment, and the relevance of that information depends strongly on context. An object that is distracting in one moment may become important in another, for example, when it is needed for action, navigation, safety, or social awareness. This helps explain why participants were skeptical of fully automatic removal, even when they were open to automatic suggestions.

This finding aligns with prior work emphasizing user agency in DR [2]. However, our results extend this discussion to the use



Figure 11: Diminished area placements during the 'Task' phase (marked with yellow dots)

of live MR, where the system acts on the user's immediate physical surroundings. In such settings, manual confirmation appears especially important because incorrect removal may affect trust, situational awareness, or safety. Participants' openness to system suggestions suggests a promising middle ground: future DR systems could automatically detect potentially distracting, sensitive, or persistent objects, but leave the final decision to the user.

The practical implication is that future DR systems should avoid treating automation as a binary choice between fully manual and fully automatic operation. Instead, they should support adjustable levels of control. For example, users might manually remove objects during exploratory use, accept system-generated suggestions during repetitive tasks, or define persistent rules for static objects in familiar environments. This preserves user control while still reducing the effort of repeated selection.

6.3 Acceptability Beyond Photorealistic Reconstruction

The study suggests that photorealistic reconstruction may not always be necessary for DR to be perceived as useful. Participants noticed blending, alignment, and inpainting artifacts, especially for larger selections, selections near the camera border, or visually complex backgrounds. Nevertheless, diminished regions were often still considered acceptable when they reduced the salience of the original object or made the scene feel less cluttered. This distinction is important: in attention-management scenarios, the goal may not be perfect invisibility but a sufficient reduction in visual salience.

This finding should be interpreted cautiously. Visible artifacts can still undermine trust, realism, and comfort, and they may be unacceptable in safety-critical or accuracy-dependent contexts. For example, DR should not be used where users need precise information about tools, controls, pathways, hazards, or other people. However, for constrained tasks such as reducing visual clutter at a desk, participants' responses suggest that a diminished region can be useful even when the reconstruction is visibly imperfect.

The gaze data provide additional, albeit limited, support for this interpretation. When participants looked less at diminished regions, this may indicate that the removed object or the inpainted area no longer attracted attention. However, this pattern does not prove that the original object was distracting or that the diminished region successfully reduced distraction. It may also mean that the area was not task-relevant, that participants had already inspected it, or that they avoided looking at artifacts. We therefore treat gaze behavior as a complementary indicator of attention, not as definitive evidence of distraction reduction.

The UEQ-S results should be interpreted in a similar way. As participants always completed the exploratory phase before the task phase, differences between the two measurements cannot be attributed solely to the scenario type. They may also reflect learning effects, increased familiarity, fatigue, or changed expectations after experiencing the system's limitations. We therefore use the UEQ-S comparison descriptively, to characterize how the prototype was experienced across two use contexts, rather than as a controlled comparison between exploratory and task-oriented DR.

Severe inpainting inconsistencies were not systematically coded as a separate event category in our study. We therefore cannot report a reliable frequency of severe artifacts. Qualitatively, however, participants and experimenters observed that stronger inconsistencies occurred particularly for large selections, selections close to the edge of the camera view, close-range objects with noisy depth estimates, and backgrounds with complex textures or lighting changes. Future evaluations should explicitly log artifact severity and relate it to selection size, distance, background complexity, and acceptance.

6.4 Interaction Requirements for Real-Time DR

Participants' feedback highlights several interaction requirements for live DR. First, selection must be fast and low effort. The current rectangular selection was easy to understand and allowed participants to quickly mark regions for diminishment. At the same time, it lacked object-level precision. This tradeoff suggests that future systems should combine fast coarse selection with optional refinement, for example, through segmentation-assisted selection, freeform correction, or centered-object selection.

Second, users need clear feedback throughout the process. As live DR is not instantaneous, participants had to understand when a region was selected, when processing was ongoing, and when the result was ready. The waiting indicator helped communicate this, but it also revealed the importance of reducing latency. A delay of several seconds may be acceptable for exploratory use, but it limits use during dynamic tasks, movement, or time-critical situations.

Third, reversibility is essential. Diminishing an object is not equivalent to hiding a digital overlay; it changes the user's view of the physical world. Easy restoration helps users recover information, correct mistakes, and maintain trust in the system. This is important for objects whose relevance may change over time, such as tools, doors, documents, displays, or people entering the scene.

Fourth, viewpoint stability remains a major constraint on interactions. Because the current prototype relies on single-view capture and projection, head movement during processing can reveal misalignment and reduce the plausibility of the diminished region. This shows that future DR systems need more robust spatial reconstruction, temporally stable inpainting, and better handling of user movement. Otherwise, users may adapt their behavior to the system's limitations rather than using DR naturally.

6.5 Design Implications

Support user-confirmable DR rather than fully automatic removal. Future DR systems should preserve user control over what is diminished. Automatic detection may be useful for suggesting distracting or persistent objects, but removal should remain confirmable, especially in shared, safety-critical, or socially sensitive environments.

Prioritize acceptable reduction of salience over perfect reconstruction for attention-management tasks. For some use cases, such as reducing clutter or suppressing distractions, users may benefit from DR even when the result is not photorealistic. Future systems should therefore consider perceptual acceptability, stability, and reduced visual salience as design goals alongside reconstruction realism.

Combine fast coarse selection with object-aware refinement. Rectangular cropping enabled quick interaction, but it also reduced precision and sometimes included unwanted background regions. Future systems should integrate semantic or instance-level segmentation to help selections better follow object boundaries while preserving the speed and simplicity of manual selection.

Design for latency and uncertainty. As live DR involves sensing, processing, and rendering delays, systems should make processing states visible. Waiting indicators, progress feedback, preview modes, and gradual refinement help users understand what the system is doing and avoid interpreting temporary artifacts as errors.

Keep safety-critical, social, and navigational information visible. Participants' hesitations around people, doors, pathways, and relevant task objects suggest that DR systems need safeguards. People, hazards, routes, interactive controls, and contextually important information should either remain visible by default or require stronger confirmation before being diminished.

6.6 Limitations

Several limitations shape the interpretation of our findings. On the technical side, the prototype relied on additional sensing hardware, constrained viewpoint stability, rectangular cropping, and visible inpainting artifacts. The processing delay also required participants to keep their heads still, which limits ecological validity for mobile or dynamic use. These limitations mean that our results should not be interpreted as evidence that the current system is ready for everyday deployment.

On the methodological side, the study was conducted in a controlled office environment with seated participants and a predefined arrangement of objects. This allowed us to observe interactions in comparable scenes, but it did not capture the variability of everyday environments, such as changing lighting, moving people, outdoor spaces, or safety-critical tasks. In addition, participants were recruited through university and research institute channels, which likely biased the sample toward tech-savvy users.

Finally, several measures have interpretive limitations. The UEQ-S comparison between scenarios is descriptive rather than a controlled test, because the order of phases was fixed. Gaze behavior can indicate where participants looked, but it cannot by itself establish whether diminished regions reduced distraction. Similarly, qualitative feedback about artifacts provides insight into perceived acceptability, but future studies should quantify artifact severity more systematically.

6.7 Future Work

Future work should address both technical and interaction challenges. Technically, live DR systems need lower latency, better temporal consistency, improved depth reconstruction, and more robust handling of viewpoint changes. Direct access to headset camera streams and depth sensors would reduce the need for external hardware and could improve alignment, comfort, and deployment feasibility. Inpainting methods should also be evaluated not only for visual realism, but for stability, perceptual salience, and acceptability during interaction.

Interaction research should compare manual, semi-automatic, and automatic DR workflows in more detail. In particular, confirmable system suggestions may offer a useful compromise between user control and interaction effort. Future systems should also investigate segmentation-assisted selection, restoration techniques, persistent object rules, and multimodal input alternatives such as gaze, voice, or adaptive controllers.

Finally, future evaluations should move beyond short, seated laboratory sessions. Longitudinal studies in more realistic environments are needed to understand how users appropriate DR over time, when they trust it, when they reject it, and how it affects attention, comfort, safety, and social interaction. Such work will be necessary to determine where diminished reality is genuinely useful and where visual removal should be avoided.

7 Conclusion

We presented *Selective Vision*, a proof-of-concept diminished reality system for immersive mixed reality that allows users to manually select, diminish, and restore real-world regions using live RGB-D input, hand tracking, inpainting, and 3D mesh updating. Through

a study with 28 participants, we examined how users experienced the system in exploratory and task-oriented scenarios. Participants quickly learned the interaction, used diminished reality to reduce visual clutter, and generally preferred manual control over fully automatic removal.

Our findings suggest that live DR should be understood not only as a reconstruction problem, but also as an interaction design challenge. Although visual artifacts remained visible, participants often perceived diminished regions as acceptable when they reduced the salience of distracting objects. This indicates that future DR systems for attention-management scenarios may not always require photorealistic reconstruction, but they do require fast selection, clear feedback, reversibility, and user-confirmable control.

At the same time, our prototype highlights important technical and methodological challenges. Future systems will need lower latency, more robust viewpoint handling, improved temporal stability, direct access to headset camera and depth streams, and object-aware segmentation to support more precise yet low-effort selection. Future evaluations should investigate live DR in more diverse, mobile, and long-term settings. Overall, our work contributes an early empirical step toward understanding how user-controlled diminished reality can be designed, experienced, and evaluated in live mixed-reality environments.

AI Disclosure Statement

During the preparation of this work, the authors used OpenAI's GPT-5.5 and Grammarly for rephrasing, grammar and style editing. Parts of Figure 3 were generated using GPT-5.5. All content was reviewed and edited by the authors, who take full responsibility for the final publication.

References

[1] Connelly Barnes, Eli Shechtman, Adam Finkelstein, and Dan B Goldman. 2009. PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Trans. Graph.* 28, 3 (2009), 24.

[2] Yi Fei Cheng, Hang Yin, Yukang Yan, Jan Gugenheimer, and David Lindlbauer. 2022. Towards Understanding Diminished Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 549, 16 pages. doi:10.1145/3491102.3517452

[3] Francesco Cosco, Carlos Garre, Fabio Bruno, Maurizio Muzzupappa, and Miguel A Otaduy. 2012. Visuo-haptic mixed reality with unobstructed tool-hand integration. *IEEE transactions on visualization and computer graphics* 19, 1 (2012), 159–172.

[4] Francesco I. Cosco, Carlos Garre, Fabio Bruno, Maurizio Muzzupappa, and Miguel A. Otaduy. 2009. Augmented touch without visual obtrusion. In *2009 8th IEEE International Symposium on Mixed and Augmented Reality*. IEEE, Washington, DC, 99–102. doi:10.1109/ISMAR.2009.5336492

[5] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A Bharath. 2018. Generative adversarial networks: An overview. *IEEE signal processing magazine* 35, 1 (2018), 53–65.

[6] Qiaole Dong, Chenjie Cao, and Yanwei Fu. 2022. Incremental Transformer Structure Enhanced Image Inpainting With Masking Positional Encoding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Washington, DC, 11358–11368.

[7] Roghieh Eskandari and Ali Motamedi. 2024. Observation-based diminished reality: a systematic literature review. *Virtual Reality* 29, 1 (2024), 7.

[8] Arturo Flores and Serge Belongie. 2010. Removing pedestrians from google street view images. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*. IEEE, Washington, DC, 53–58.

[9] Vasileios Gkitsas, Vladimiros Sterzentsenko, Nikolaos Zioulis, Georgios Albanis, and Dimitrios Zarpalas. 2021. Panodr: Spherical panorama diminished reality for indoor scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Washington, DC, 3716–3726.

[10] Paul Green and Lisa Wei-Haas. 1985. The rapid development of user interfaces: Experience with the Wizard of Oz method. In *Proceedings of the human factors society annual meeting*, Vol. 29. SAGE Publications Sage CA, Los Angeles, CA, 470–474.

[11] Christina Gsaxner, Shohei Mori, Dieter Schmalstieg, Jan Egger, Gerhard Paar, Werner Bailer, and Denis Kalkofen. 2024. DeepDR: Deep Structure-Aware RGB-D Inpainting for Diminished Reality. In *2024 International Conference on 3D Vision (3DV)*. IEEE, Washington, DC, 750–760.

[12] Jan Herling and Wolfgang Broll. 2010. Advanced self-contained object removal for realizing real-time diminished reality in unconstrained environments. In *2010 ieee international symposium on mixed and augmented reality*. IEEE, Washington, DC, 207–212.

[13] Jan Herling and Wolfgang Broll. 2012. Pixmix: A real-time approach to high-quality diminished reality. In *2012 ieee international symposium on mixed and augmented reality (ismar)*. IEEE, Washington, DC, 141–150.

[14] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. 2017. Globally and locally consistent image completion. *ACM Transactions on Graphics (ToG)* 36, 4 (2017), 1–14.

[15] KAZUYA Inoue, TOMOHIRO Fukuda, R Cao, and Nobuyoshi Yabuki. 2018. Tracking robustness and green view index estimation of augmented and diminished reality for environmental design. *Proceedings of CAADRIA 2018* 1 (2018), 339–348.

[16] Xuan Ju, Xian Liu, Xintao Wang, Yuxuan Bian, Ying Shan, and Qiang Xu. 2024. Brushnet: A plug-and-play image inpainting model with decomposed dual-branch diffusion. In *European Conference on Computer Vision*. Springer, Cham, 150–168.

[17] Yoshinari Kameda, Taisuke Takemasa, and Yuichi Ohta. 2004. Outdoor see-through vision utilizing surveillance cameras. In *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*. IEEE, Washington, DC, 151–160.

[18] Norihiko Kawai, Masayoshi Yamasaki, Tomokazu Sato, and Naokazu Yokoya. 2012. AR marker hiding based on image inpainting and reflection of illumination changes. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Washington, DC, 293–294.

[19] Daiki Kido, Tomohiro Fukuda, and Nobuyoshi Yabuki. 2020. Diminished reality system with real-time object detection using deep learning for onsite landscape simulation during redevelopment. *Environmental Modelling & Software* 131 (2020), 104759.

[20] Christian Kunert, Tobias Schwandt, and Wolfgang Broll. 2019. An Efficient Diminished Reality Approach Using Real-Time Surface Reconstruction. In *2019 International Conference on Cyberworlds (CW)*. IEEE, Washington, DC, 9–16. doi:10.1109/CW.2019.00010

[21] JangHyeon Lee and Lawrence H. Kim. 2025. DiminishAR: Diminishing Visual Distractions via Holographic AR Displays. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 30, 16 pages. doi:10.1145/3706598.3713415

[22] Wenbo Li, Zhe Lin, Kun Zhou, Lu Qi, Yi Wang, and Jiaya Jia. 2022. Mat: Mask-aware transformer for large hole image inpainting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. IEEE, Washington, DC, 10758–10768.

[23] Zhen Li, Cheng-Ze Lu, Jianhua Qin, Chun-Le Guo, and Ming-Ming Cheng. 2022. Towards an end-to-end framework for flow-guided video inpainting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. IEEE, Washington, DC, 17562–17571.

[24] Zhuwen Li, Yuxi Wang, Jiaming Guo, Loong-Fah Cheong, and Steven ZhiYing Zhou. 2013. Diminished reality using appearance and 3D geometry of internet photo collections. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Washington, DC, 11–19.

[25] Patrick Lindemann and Gerhard Rigoll. 2017. A diminished reality simulation for driver-car interaction with transparent cockpits. In *2017 IEEE Virtual Reality (VR)*. IEEE, Washington, DC, 305–306.

[26] Steve Mann. 2002. Mediated reality with implementations for everyday life. *Presence Connect* 1 (2002), 2002.

[27] Mitchell Dewayne McDonald. 2021. *The Efficacy of Diminished Reality on Learning in Complex Environments*. North Carolina State University, Raleigh, North Carolina.

[28] Siim Meerits and Hideo Saito. 2015. Real-time diminished reality for dynamic scenes. In *2015 IEEE International Symposium on Mixed and Augmented Reality Workshops*. IEEE, Washington, DC, 53–59.

[29] Katlin T Moffis, Yves Valentin, and HeeSun Choi. 2024. Beyond Illusions: Diminished Reality's Potential, Pitfalls, and Ethical Reflections. In *International Conference on Human-Computer Interaction*. Springer, Cham, 34–42.

[30] Shohei Mori, Sei Ikeda, and Hideo Saito. 2017. A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects. *IPSJ Transactions on Computer Vision and Applications* 9 (2017), 1–14.

[31] Shohei Mori, Momoko Maezawa, and Hideo Saito. 2017. A work area visualization by multi-view camera-based diminished reality. *Multimodal Technologies and Interaction* 1, 3 (2017), 18.

[32] Shohei Mori, Fumihisa Shibata, Asako Kimura, and Hideyuki Tamura. 2015. Efficient use of textured 3D model for pre-observation-based diminished reality. In *2015 IEEE international symposium on mixed and augmented Reality workshops*. IEEE, Washington, DC, 32–39.

[33] Imani Murph, M McDonald, K Richardson, M Wilkinson, S Robertson, A Karunakaran, M Gandy Coleman, V Byrne, and AC McLaughlin. 2021. Diminishing reality: potential benefits and risks. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 65. SAGE Publications Sage CA, Los Angeles, CA, 164–168.

[34] I Murph, K Richardson, and AC McLaughlin. 2022. Methods of training to overcome distraction via diminished reality. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 66. SAGE Publications Sage CA, Los Angeles, CA, 1844–1848.

[35] Yoshikatsu Nakajima, Shohei Mori, and Hideo Saito. 2017. Semantic object selection and detection for diminished reality based on slam with viewpoint class. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*. IEEE, Washington, DC, 338–343.

[36] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. IEEE, Washington, DC, 10684–10695.

[37] Andranik Sargsyan, Shant Navasardyan, Xingqian Xu, and Humphrey Shi. 2023. Mi-gan: A simple baseline for image inpainting on mobile devices. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. IEEE, Washington, DC, 7335–7345.

[38] Martin Schrepp, Andreas Hinderks, et al. 2017. Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence* 4, 6 (Dec. 2017), 103–108. doi:10.9781/ijimai.2017.09.001

[39] Sanni Siltanen. 2006. Texture generation over the marker area. In *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*. IEEE, Washington, DC, 253–254.

[40] Sanni Siltanen. 2017. Diminished reality for augmented reality interior design. *The Visual Computer* 33 (2017), 193–208.

[41] Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. 2022. Resolution-robust large mask inpainting with fourier convolutions. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. IEEE, Washington, DC, 2149–2159.

[42] Salam Tabet, Ayman Kayssi, and Imad H Elhajj. 2023. Mobile Diminished Reality for Preserving 3D Visual Privacy. In *2023 International Conference on Intelligent Metaverse Technologies & Applications (iMETA)*. IEEE, Washington, DC, 01–07.

[43] Salam Tabet, Ayman Kayssi, and Imad H Elhajj. 2024. Adaptive Mobile Diminished Reality Framework for 3D Visual Privacy. In *2024 International Conference on Intelligent Systems and Computer Vision (ISCV)*. IEEE, Washington, DC, 1–8.

[44] Salam Tabet, Ayman Kayssi, and Imad H Elhajj. 2024. AI-Enhanced Mobile Diminished Reality for Preserving 3D Visual Privacy. In *2024 2nd International Conference on Intelligent Metaverse Technologies & Applications (iMETA)*. IEEE, Washington, DC, 141–148.

[45] Ada V Taylor, Ayaka Matsumoto, Elizabeth J Carter, Alexander Plopski, and Henny Admoni. 2020. Diminished reality for close quarters robotic telemanipulation. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Washington, DC, 11531–11538.

[46] Kentaro Yagi, Kunihiro Hasegawa, and Hideo Saito. 2017. Diminished reality for privacy protection by hiding pedestrians in motion image sequences using structure from motion. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*. IEEE, Washington, DC, 334–337.

[47] Asim Evren Yantaç, Doga Çorlu, Morten Fjeld, and Andreas Kunz. 2015. Exploring diminished reality (DR) spaces to augment the attention of individuals with autism. In *2015 IEEE International Symposium on Mixed and Augmented Reality Workshops*. IEEE, Washington, DC, 68–73.

[48] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. 2019. Free-form image inpainting with gated convolution. In *Proceedings of the IEEE/CVF international conference on computer vision*. IEEE, Washington, DC, 4471–4480.

[49] Yuehan Zhu, Tomohiro Fukuda, and Nobuyoshi Yabuki. 2019. Synthesizing 360-degree live streaming for an erased background to study renovation using mixed reality. In *Proceedings of the 24th International Conference of the Association for Computed-Aided Architectural Design Research in Asia (CAADRIA), Victoria University of Wellington, Faculty of Architecture & Design, Wellington, New Zealand*. Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong, 15–18.

[50] Yuehan Zhu, Tomohiro Fukuda, and Nobuyoshi Yabuki. 2024. A Mixed Reality Design System for Interior Renovation: Inpainting with 360-Degree Live Streaming and Generative Adversarial Networks after Removal. *Technologies* 12, 1 (2024), 9.

[51] Junhao Zhuang, Yanhong Zeng, Wenran Liu, Chun Yuan, and Kai Chen. 2025. A Task Is Worth One Word: Learning with Task Prompts for High-Quality Versatile Image Inpainting. In *Computer Vision – ECCV 2024*, Aleš Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol (Eds.). Springer Nature Switzerland, Cham, 195–211.