This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2025.3616739

See What I Mean? Mobile Eye-Perspective Rendering for Optical See-through Head-mounted Displays

Gerlinde Emsenhuber, Tobias Langlotz, Denis Kalkofen, and Markus Tatzgern

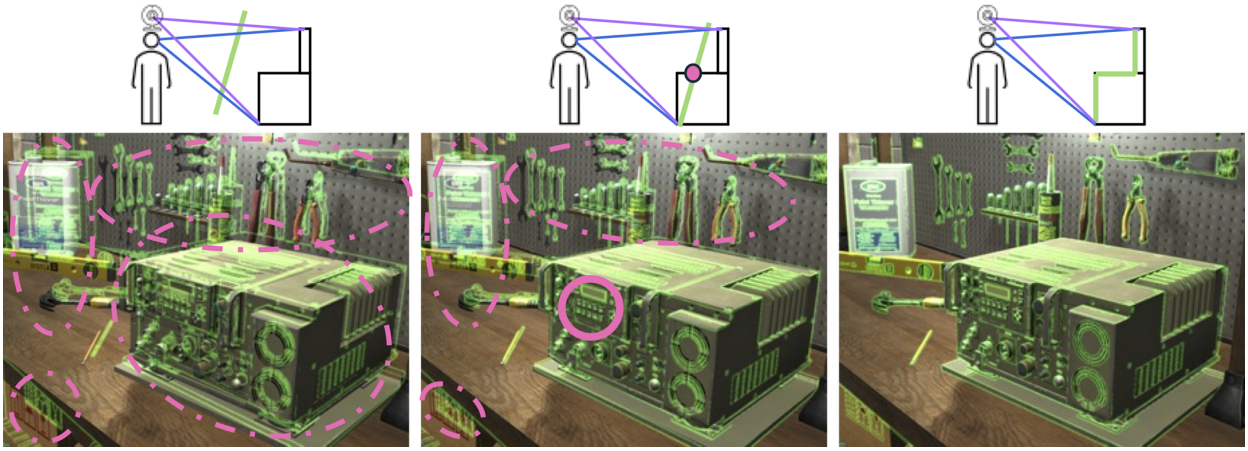

Fig. 1: Edges are extracted from a head-mounted display (HMD) camera view (purple viewport), e.g., to support people with low vision by highlighting scene elements. For accurate alignment, the edges must be shown from the user's eye perspective (blue viewport) on the optical see-through (O ST) display. (Left) A naive Plane-Proxy EPR method projects the camera view onto a fixed-distance proxy plane (green), causing misalignment between visualization and real-world objects (dashed pink areas). (Middle) Our novel Gaze-Proxy EPR approach uses eye-tracking to dynamically adjust the plane to match the depth at the user's gaze point (center pink dot/circle), improving alignment at the focal area but causing misalignment elsewhere. (Right) A Mesh-Proxy EPR technique projects the camera view onto a scene reconstruction (green), ensuring accurate alignment from all perspectives.

Abstract— Image-based scene understanding allows Augmented Reality (AR) systems to provide contextual visual guidance in unprepared, real-world environments. While effective on video see-through (VST) head-mounted displays (HMDs), such methods suffer on optical see-through (OST) HMDs due to misregistration between the world-facing camera and the user's eye perspective. To approximate the user's true eye view, we implement and evaluate three software-based eye-perspective rendering (EPR) techniques on a commercially available, untethered OST HMD (Microsoft HoloLens 2): (1) Plane-Proxy EPR, projecting onto a fixed-distance plane; (2) Mesh-Proxy EPR, using SLAM-based reconstruction for projection; and (3) Gaze-Proxy EPR, a novel eye-tracking-based method that aligns the projection with the user's gaze depth. A user study on real-world tasks underscores the importance of accurate EPR and demonstrates gaze-proxy as a lightweight alternative to geometry-based methods. We release our EPR framework as open source.

Index Terms—Augmented reality, optical see-through, head-mounted displays, eye-perspective rendering, vision augmentation.

1 Introduction

Augmented Reality (AR) combined with optical see-through (OST) head-mounted displays (HMDs) overlays virtual information onto the user's view of the real world. Such systems support assembly and repair tasks [18], assist first responders in hazardous scenarios [7, 33], and aid elderly and visually impaired users [39, 53, 54]. Commonly, AR visualizations highlight task-relevant objects identified through the analysis of images captured by the built-in, world-facing camera of the HMD. Traditionally, these highlighted objects rely on manually pre-registered semantic annotations or automatic detection by specialized computer vision models, restricting support to predefined object types and relationships.

The emergence of context-aware Artificial Intelligence (AI), such as vision language models (VLMs), large language models (LLMs), and open-vocabulary scene graphs [22, 52], reduces the need for extensive prior annotation by enriching captured imagery with semantic context. Image-based scene understanding facilitates a broader adoption of AR in mobile, unprepared environments [14] that can be directly applied to video see-through (VST) HMD setups where the world-facing cameras replace the user's eyes. However, when using an OST device, mapping the image information extracted from a world-camera view to the user's eyes remains challenging due to their different perspectives (Fig. 1(Left)).

Previous work has developed HMD prototypes integrating cameras that capture the user's exact viewpoint by redirecting light through the transparent display using half-silvered mirrors [30, 32]. This setup allows direct usage of the captured image for highlighting object silhouettes [41] or adjusting scene colors to assist individuals with a color vision deficiency (CVD) [32]. However, these hardware modifications

- *Gerlinde Emsenhuber is with Salzburg University of Applied Sciences, Austria and Graz University of Technology, Austria. E-mail: gerlinde.emsenhuber@fh-salzburg.ac.at.*
- *Tobias Langlotz is with Aarhus University, Denmark. E-mail: tobias.langlotz@cs.au.dk.*
- *Denis Kalkofen is with Graz University of Technology, Austria. E-mail: kalkofen@tugraz.at.*
- *Markus Tatzgern is with Salzburg University of Applied Sciences, Austria. E-mail: markus.tatzgern@fh-salzburg.ac.at.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

cannot be applied to existing commercially available HMDs, limiting immediate applicability while also significantly increasing the size of the HMDs.

Software-based alternatives use scene information to reconstruct the user's perspective through eye-perspective rendering (EPR) techniques. The most straightforward method, which we refer to as Plane-Proxy EPR, projects the analyzed world-camera view onto a virtual plane placed at a fixed distance in front of the user [39, 55]. This approach achieves alignment when the virtual plane is aligned in depth with the real-world geometry, but results in visual misalignment when there is an offset (Fig. 1(Left)).

A more accurate method addresses this by projecting the world-camera view onto a 3D mesh reconstructed from depth maps or SLAM data [11, 17, 51]. This method, which we refer to as Mesh-Proxy EPR, significantly improves the alignment [16] between image-based overlays and real-world geometry (Fig. 1(Right)), but requires a full 3D reconstruction, which may be computationally expensive, battery-draining, and not always available in real-time.

We further propose the Gaze-Proxy EPR method, a novel approach that dynamically aligns the projection plane of Plane-Proxy EPR with the scene depth at the user's current gaze target. Unlike Mesh-Proxy EPR, which requires a 3D reconstruction of the environment, Gaze-Proxy EPR instead uses real-time eye tracking to estimate fixation depth, enabling localized alignment without the need for spatial meshes. While Mesh-Proxy EPR aims for global alignment across the entire field of view, Gaze-Proxy EPR prioritizes accuracy in the user's focus region (Fig. 1(Middle)), adapting the projection plane as the gaze shifts. This makes it particularly suitable for mobile OST HMDs, as it relies solely on eye vergence for depth estimation when precise eye tracking is available [36], whereas scene reconstructions may be unavailable, unreliable, or too slow to update (e.g., 1-5 FPS for HoloLens 2).

In this paper, we explore software-based EPR for mobile, untethered OST HMDs by implementing the three aforementioned EPR approaches on a commercially available hardware (HoloLens 2). Previous studies have demonstrated the value of EPR for adapting virtual content to avoid interference with the real-world background [17, 30] and assisting individuals with CVD [32] in tethered, sedentary setups that restrict user mobility and do not allow users to interact with the real world. In contrast, we demonstrate the effectiveness of EPR in a mobile, untethered scenario. We also require users to interact with the augmented real-world scene, thereby, underlining the need for EPR to enable accurate interactions. By providing our EPR framework as open-source software, we facilitate further research beyond controlled laboratory settings and simulations conducted in Virtual Reality (VR) [9, 13], addressing the lack of suitable EPR frameworks for untethered OST HMDs.

In a user study, we compare the three EPR methods in terms of interaction accuracy, usability, and task load. By also collecting qualitative feedback, we provide quantitative and qualitative insights into the use of EPR using an untethered device.

In summary, we make the following contributions:

- We introduce a novel Gaze-Proxy EPR method for OST HMDs that uses eye tracking to align the projection plane with the scene depth at the user's current focus, providing localized accuracy with minimal computational overhead.
- We implement and compare all three EPR methods on a state-of-the-art, untethered OST HMD (HoloLens 2), providing quantitative and qualitative insights from the first exploration of EPR effects during interactive tasks involving real-world environments.
- Our results clearly demonstrate the need for accurate EPR for image-based analysis, highlighting the impact of misalignments between the world camera and the user's eyes to raise awareness of this challenge.
- We release all three EPR implementations and the study framework as open source, facilitating future research and development of contextual, image-based AR support systems for mobile, untethered use cases.

Please note that the examples used to illustrate EPR techniques in this paper are generated within synthetic scenes. Capturing EPR-aligned imagery via a camera recording through an HMD is inherently limited, as it does not accurately reflect the user's true eye perspective. Consistent with prior work [17, 45], we use simulated views rendered from the user's viewpoint using Unity 2022.3 LTS to visualize algorithmic behavior and alignment artifacts in the paper. In the supplemental material, we also include a video of our best effort of capturing EPR as seen through the HoloLens 2 display. Furthermore, we conduct our experimental user evaluations, including both qualitative and quantitative measurements, in a real-world scenario using a physical, untethered OST HMD, allowing us to assess EPR performance and accuracy under realistic conditions.

2 Related Work

In the following, we provide related work on previous solutions to EPR. Subsequently, we discuss use cases that benefit from knowledge of the correct eye-perspective of users through an OST HMD.

2.1 User- and Eye Perspective Rendering in AR

Perspective-correct rendering for AR is an issue that is not unique to HMDs. It was brought up first in the context of handheld AR interfaces [40] such as mobile phones or tablets. The AR view of handheld devices uses the integrated device camera view for augmentations, where the viewpoint of the user does not match the camera viewpoint requiring the users to mentally map between the different views. To solve this issue, user-perspective rendering [5, 6] has been proposed that synthesizes a view of the scene based on the user's eye position, transforming the VST mobile device into a transparent window onto the scene.

VST HMDs are also affected by the offset between camera position that captures the view of the scene, and the user's eye position. While nowadays most VST HMD place the cameras roughly in front of the user's eye, even this small positional offset combined with the individual differences in eye positions are sufficient to lead to a noticeable difference [23] and negatively impact depth perception [4]. To overcome the positional offset, several approaches have been proposed [11, 28, 51] that either compute a 3D reconstruction [11] relying on video encoders integrated into the HMD to track features between the frames, more recent approaches rely on view generation techniques such as neural rendering [51] or light-field rendering [28]. Unfortunately, neural and light-field rendering techniques require the computational power of a stationary computer (e.g. a NVIDIA Titan RTX for each eye [51]), and, thus, are not applicable to mobile, untethered HMD hardware.

To capture the user's view through an OST HMD, i.e., the light rays right in front of the user's eyes, prior work explored hardware modifications, such as putting a large sensor with lenslets for capturing incoming light rays right in front of the user's eyes [28]. However, implementing such an approach for OST HMDs would block the actual view of the user. Instead, Langlotz et al. [30] proposed to place beam splitters in the optical path of the human eye, where they reflect a fraction of the light towards the camera to precisely change the appearance of the environment e.g. to address color-vision deficiency [44] or real-world saliency modulation [45].

Unfortunately, beamsplitters require careful calibration but, more importantly, increase the HMD's size and require extensive hardware modification. To address these shortcomings, recent work started to explore EPR for OST HMDs, a software-based method that does not require hardware modifications, and demonstrated the effectiveness of the precise knowledge of the user's view through the displays for view management such as label placement [16, 17]. We are exploring and extending previous approaches to software-based EPR for supporting mobile, untethered OST HMDs.

2.2 Applications of Eye-perspective Rendering

AR applications embed virtual information into the user's view. While applications typically rely on 3D pose information for placing the content (e.g., application windows), more advanced placement algorithms perform image-based analysis on the device's camera feed [10]

for scene analysis, to adapt content placement to avoid visual interference [16, 17] or object occlusions [26]. For example, recent algorithms using machine learning techniques for AI object detection [20, 27, 34, 56] detect and identify individual objects in the scene and provide this information to the user via an AR interface [14]. Most of these algorithms work directly on 2D images of the scene, which usually come from a camera attached to the HMD. However, mapping the extracted information into the user's view through an OST HMD leads to inaccuracies and potential ambiguity due to the difference between the camera perspective and the user perspective (Fig. 2(Top)(Left)).

Augmented Vision is another growing area in AR, aiming to support individuals with visual impairments or working in low vision environments (e.g., dust, smoke). It commonly relies on 2D imagery to identify scene features that are modulated and enhanced in the AR view [31]. For instance, vision support systems for individuals with CVD identify critical colors and modulate them via overlays on the OST display — changing color hue and brightness, using patterns that help distinguish them from otherwise similar colors [2, 44], or framing critically colored objects with bounding boxes and additional textual description [55]. Low vision support systems extract edges from camera images using image-based edge detection to emphasize structures in the user's view [25, 31, 39]. However, when naively overlaying in the user's view, the difference in camera perspective and user's eye perspective leads to inaccuracies as the supporting augmentations do not match the real-world scene (Fig. 1(Left), Fig. 2(Top)(Middle, Right)).

Aside from practical implications due to the mismatch between camera and user view impacting the usability of the described use cases, the misalignment of AR overlays with the real-world scene can cause physiological issues such as nausea and cybersickness [12]. Recent studies suggest that a 15 mm offset between the camera and the eye is the threshold at which offsets are becoming noticeable [23]. Even with modern, miniaturized cameras and display optics, achieving a camera placement this close to the human eye is hardly feasible, potentially leading to noticeable negative effects when showing information computed in the camera view.

3 System Overview

In the following, we present implementation details on three EPR techniques included in our open source framework, as well as the image-based analysis approach for our experiment, before illustrating the EPR techniques for selected use cases.

3.1 Eye-Perspective Rendering

Our EPR methods apply a texture projection technique to re-render the world-camera view from the user's perspective using 3D proxy geometry [16, 17]. Each point on the proxy geometry is first projected into the world-camera view to retrieve the image data available to the AR system. The same point is then reprojected into the user's left and right eye views as EPR. When the proxy geometry accurately corresponds to the real-world scene geometry, both the world camera and the user perceive consistent visual information. This alignment is illustrated in Fig. 1(Right), where a mesh closely matches the scene geometry, and in Fig. 1(Middle), where a planar proxy is aligned with the user's gaze target. Correct alignment of EPR views for both eyes is essential for achieving proper stereo perception of image-based augmentations, such as object highlighting (Fig. 2(Left)). Misalignment between augmentations and the physical scene can introduce perceptual conflicts, potentially degrading the user experience [39, 55].

Our system is the first to implement multiple EPR approaches natively on a state-of-the-art OST HMD. We use the HoloLens 2 with the Research Mode API to access sensor streams, and Unity with the Mixed Reality Toolkit (MRTK) 2.8.3 for rendering. While developed for HoloLens 2, the EPR techniques are generalizable to other platforms, including Magic Leap 2 and Snap AR glasses.

We implement three methods that share the same projection technique but differ in their choice of proxy geometry: a fixed plane for Plane-Proxy EPR, a gaze-aligned plane for Gaze-Proxy EPR, and a mesh approximating the scene geometry for Mesh-Proxy EPR.

Plane-Proxy EPR. In this method, the world-camera view is projected onto a virtual 3D proxy plane placed at a fixed distance in front of the user, parallel to the image plane of the dominant eye. The distance can be adjusted manually, to match a typical working distance. However, due to continuous head motion in mobile use, visual augmentations based on the EPR view quickly become misaligned with the real-world scene (Fig. 1(Left), Fig. 2(Top)), negatively affecting the user experience. While this method is prone to misalignment, it may still be viable for viewing distant scene geometry, where the parallax between the camera and the user's eyes becomes negligible [8]. Plane-Proxy EPR is the simplest of the three approaches, requiring no pose tracking, eye tracking, or scene information, and introduces no computational overhead.

Mesh-Proxy EPR. Instead of a planar proxy, this method uses a spatial mesh reconstruction provided by the HMD as a 3D proxy geometry, as proposed by Emsenhuber et al. [17]. The EPR view is generated by projecting the world-camera image onto the mesh and re-rendering it from the user's eye position. Since the proxy geometry closely matches the physical scene, Mesh-Proxy EPR enables more accurate EPR views in which image-based augmentations better align with the real world (Fig. 1(Right) and Fig. 2(Bottom)).

The method's effectiveness depends on the quality of the mesh. As discussed by Weinmann et al. [49], the HoloLens 2's spatial mesh adequately captures basic indoor architecture, but lacks the precision needed to represent finer scene details, potentially leading to visible mismatches between augmentations and physical objects. The HoloLens 2's spatial mapping accuracy can vary depending on environmental conditions and calibration. Under optimal conditions and with proper ground control, studies have demonstrated that the device can achieve spatial mapping accuracy of 1-2 cm [46]. The mesh available through Unity is less precise due to system-level performance constraints. The mesh resolution is governed by the maximum triangles per cubic meter parameter, with a default value of about 500, which is insufficient for capturing finer scene features. In our implementation, we set the parameter controlling this value to "Fine", which corresponds to 2000-3000 triangles per cubic meter, resulting in a runtime rendering performance of approximately 28 fps for stereo EPR views.

Alternatively, the Mesh-Proxy EPR method could employ depth maps generated via image-based depth estimation [43] or a built-in depth sensor. However, current mobile hardware may not support real-time image-based depth estimation at sufficient quality, and depth sensors often produce incomplete data with missing regions that require computationally intensive inpainting [21]. Future HMDs may implement such features, when the issue of EPR becomes more recognized for OST.

Gaze-Proxy EPR. This approach builds on the same plane-projection technique as Plane-Proxy EPR, but dynamically positions the projection plane at the 3D point in the scene the user is currently focusing on. In principle, this point can be determined by computing the eye vergence using a built-in eye tracker. However, since the HoloLens 2 only provides a single gaze ray [48], and prior work has shown that vergence estimation degrades with increasing viewing distance [36], we estimate the gaze target by intersecting the eye-tracked gaze ray with the spatial mesh used in Mesh-Proxy EPR. This enables us to position the projection plane at the user's current focus and to generate EPR views that are accurate within the user's focal area (Fig. 1(Middle)). Misalignments outside this region are less likely to impact the user experience, as the visualization updates when the user's gaze shifts.

Although our method currently relies on the spatial mesh to estimate depth, it demonstrates the feasibility of the Gaze-Proxy EPR approach and its potential for lightweight, scene-agnostic rendering, assuming sufficiently accurate vergence tracking in future HMDs. Given that vergence estimation is more reliable for nearby objects, this approach is particularly suitable for close-range tasks, where EPR accuracy is most critical. For distant objects, parallax between the world camera and the user's eyes becomes negligible, reducing the need for dynamic alignment [8].

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2025.3616739

Fig. 2: Use Cases of Image-based Analysis. (Top) The Plane-Proxy EPR method frequently results in misalignments between overlays and the real-world scene due to the fixed distance of the projection plane. (Bottom) The Mesh-Proxy EPR method accurately aligns extracted image information with the real-world scene. Here, we demonstrate the effect for (Left) highlighting outlines of the next picking tray for a maintenance task, (Middle) enhancing scene visibility and text legibility through outline overlays for low vision support, and (Right) applying Daltonization to shift color spaces for individuals with CVD. The top images show the original colors and the colors perceived by an individual with CVD.

3.2 Image Processing

Our framework provides implementations of basic image-based scene analysis routines, which we use to demonstrate the necessity and impact of EPR-aware image analysis in OST AR, both in functional demonstrations in simulated scenes and in our user study using the HoloLens 2. Implementing mobile versions of the image operations enables both quantitative measurements and qualitative feedback on the effects of EPR in real-world scenarios. For our mobile implementation, we avoid computationally intensive object recognition models and instead employ lightweight image processing techniques that allow for real-time manipulation and highlighting of image regions. While we do not currently integrate advanced models, doing so is possible either by executing them on-device if hardware permits, or by streaming world-camera images to an external machine for processing, as shown in previous work [3].

All image processing is performed directly on the RGB image captured by the world-facing camera. The resulting annotations or visual transformations are then projected onto the proxy geometry to create the EPR view. Processing the original camera image rather than the reconstructed EPR view avoids artifacts introduced by mesh inaccuracies in Mesh-Proxy EPR. Limited mesh resolution can lead to distorted or low-polygon silhouettes, particularly for small or detailed objects, which in turn degrade the quality of downstream image analysis. By operating on the artifact-free camera image, we preserve processing accuracy.

Our framework implements the following image-based operations as basis for our EPR use case demonstrations and user study:

- **Edge Overlays:** Canny edge detection is applied to the scene to simulate visual support systems for low vision [39].
- **Hue-Based Segmentation:** Objects of interest are highlighted by segmenting regions based on hue and either coloring them or outlining their contours.
- **Daltonization:** The input image is adjusted using color transformations to improve perception for users with CVDs [32].

3.3 Use Cases of Image-based Analysis

In the following, we describe a selection of use cases that benefit from a precise registration of image-based analysis results utilizing an EPR method such as Mesh-Proxy EPR or Gaze-Proxy EPR that aligns results with the viewpoint of the user and the scene.

Guidance in Manual Tasks. A common use case for AR is assisting users in manual tasks, where virtual augmentations indicate the next element to interact with, such as a part [18], a tool [29] or a control element of a machine [35]. AI-driven, image-based guidance in unstructured scenes is well-suited for such scenarios, where objects of interest are highlighted using their silhouette.

Highlighting objects with artificial silhouettes, rather than bounding boxes, can convey their shape more clearly and helps distinguish them from similar items. Furthermore, in cluttered environments, bounding boxes may enclose multiple similar objects, causing confusion. Fig. 2(Top, Left) illustrates a silhouette highlighting a tray during a sequence of maintenance steps using the naive Plane-Proxy EPR method, where misregistration results in ambiguity. In contrast, Fig. 2(Bottom, Left) demonstrates the precise Mesh-Proxy EPR approach, where the registered silhouette clearly directs attention to the correct object.

Low-vision Support. According to the WHO, 2.2 billion people are suffering from a vision impairment [42], which can severely impact the quality of life, particularly by reducing mobility. Edge overlays can improve perception for people with low vision [39], enabling them to detect obstacles more easily, read text, discern image details, and interpret facial expressions. Beyond assisting individuals with low vision, edge overlays can enhance scene perception in environments with limited visibility, such as dusty work settings or smoke-filled areas in firefighter scenarios.

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2025.3616739

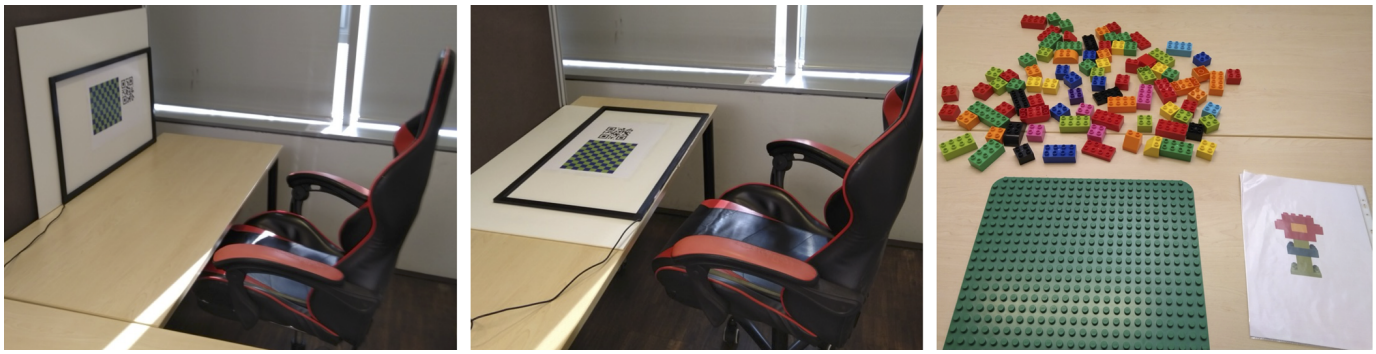

Fig. 3: Study Setup. (Left) The wall-mounted checkerboard (WALL), where users were looking at the targets at a 90° angle, and (Middle) the checkerboard lying on the table (TABLE), where users were looking at an approximately 50° degree angle. (Right) The Duplo playground for the second part of the study.

Simple edge overlays are computationally efficient with image-based analysis. However, previous work has shown that users experience depth perception issues when edge overlays are misaligned with scene geometry and displayed on a fixed-depth plane [39] (Fig. 2(Top, Middle)). Mesh-Proxy EPR allows for precise edge overlays at the correct scene depth, thereby avoiding depth perception conflicts (Fig. 2(Bottom, Middle)).

Note that the HoloLens 2 may not be suitable for people having conditions causing low vision due to the fixed focus plane of the display at 2 m. A retinal projection HMD [1] enables focused augmentations as light is directly focused onto the retina.

Compensating Color Vision Deficiencies. CVD is a condition in which affected individuals have a reduced ability to distinguish between colors, most commonly between red and green. A common approach to CVD compensation is Daltonization, which adjusts colors to reduce combinations that are difficult to differentiate. OST HMDs offer an ideal assistive solution for individuals with CVD, allowing a direct modification of the user's view.

Daltonization can be implemented by segmenting problematic colors in the world-facing camera image and overlaying a mask that shifts scene colors to improve color differentiation [2, 32]. Alternatively, Zhu et al. [55] analyze the camera feed for problematic colors and annotate objects with textual labels and tightly fitting bounding boxes. However, their system renders the output on a fixed-depth plane (1 m), resulting in depth perception issues and misaligned bounding boxes. Using the same system for Daltonization leads to the same issues (Fig. 2(Top)(Right)).

Both Daltonization and bounding geometry benefit from precise EPR methods such as Mesh-Proxy EPR (Fig. 2(Bottom)(Right)), which allow masks or overlays to align with real-world scene geometry at the correct depth, avoiding perceptual inconsistencies.

4 Experiment

The institutional ethics committee of Salzburg University of Applied Sciences approved this study. We conducted a mixed-method, within-subject user study to compare the three EPR approaches in a real-world setting using an untethered, mobile OST HMD (HoloLens 2). The first part of the study focused on quantifiable performance metrics, requiring participants to interact with the scene based on image analysis from the world camera. In the second part, we gathered qualitative feedback as participants walked through a scene simulating a low vision support system with overlaid edge enhancements.

4.1 Quantitative Evaluation

Participants were asked to identify and touch the correct square on a checkerboard, guided solely by a virtual highlight derived from world camera analysis. The study included two independent variables: **Orientation** and **EPR Method**. **Orientation** had two conditions: WALL, with the checkerboard mounted on a wall and viewed head-on, and TABLE, with the checkerboard placed flat on a table and viewed at an angle. These conditions allowed us to examine the impact of viewing angle on EPR accuracy. **EPR Method** had three conditions representing our EPR implementations: PLANE, GAZE, and MESH. PLANE applied a 20 cm offset from the scene geometry, simulating a misaligned proxy plane that has been calibrated for a different workspace. Participants were allowed to compensate for this offset by moving their heads.

Participants. We recruited 18 participants (8 female, $\overline{X}$=33 (8) years). On a scale from one (worst) to seven (best), the mean of self-rated AR experience was 2.5. All participants had normal or corrected-to-normal vision.

Task. During each trial, participants interacted with a checkerboard consisting of 11×9 squares, each measuring 2 cm. An algorithm randomly selected a square, highlighting it with a white outline. Participants were instructed to identify and touch the center of the matching real square, then return their hand to a starting position on their legs. The task was repeated 30 times per EPR method, with the first 5 repetitions discarded to reduce learning effects.

Apparatus. All experiments were conducted with participants seated on an adjustable chair to ensure a consistent viewing angle across different body heights. Participants adjusted their chair to comfortably align their eyes with the middle row of the checkerboard in WALL, positioned at a height of 103 cm (Fig. 3(Left)). For TABLE, the chair height remained unchanged, with the table set at 75 cm, resulting in an approximate viewing angle of 50° (Fig. 3(Middle)). The checkerboard was centered relative to the participant's dominant hand. To support square detection and highlighting, we developed a custom AR application using Unity, the MRTK, and our framework's image analysis functions. For Mesh-Proxy EPR and Gaze-Proxy EPR, we used the built-in HoloLens 2 eye calibration tool to obtain accurate eye position data necessary for computing EPR views. Touch interactions were recorded using a 70.6×39.7 cm infrared touch frame (GreenTouch) with 2 mm precision, connected to a PC. A PC-side application logged touch event timestamps and communicated with the HoloLens 2 via TCP. Upon receiving a touch event, the AR app selected and highlighted a new square, pausing for 3 seconds to allow the participant to return their hand to the resting position. Participants were free to move their upper bodies and heads, which occasionally caused the highlighted square to leave the HoloLens 2's field of view. To determine whether the checkerboard remained within the participant's view, we placed a QR code next to the grid and tracked head position and orientation. When the grid was not visible, the application paused and displayed directional arrows to guide participants back to the correct viewing position. The study was conducted in a room with controlled lighting (LUPO Superpanel Dual Color 60) to ensure consistent conditions across sessions. Ambient light was measured at 317 lux at a color temperature of 3220 Kelvin using a Mavospec Base spectrometer.

Data Collection. We measured task completion time (TCT) as the

[1] https://retissa.biz/en/

	EPR Method	MESH vs PLANE	MESH vs GAZE	PLANE vs GAZE
Correctness	$F(2,85)=129.55$,p<.001	t(84)=14.49,p<.001	-	t(84)=13.32,p<.001
Task Completion Time	$F(2,85)=41.9$,p<.001	t(85)=8.79,p<.001	t(84)=2.19,p=0.032	t(85)=6.6,p<.001
SEQ	$F(2,85)=33.7$,p<.001	t(84)=7.58,p<.001	-	t(85)=6.52,p<.001
TLX Overall	$F(2,85)=61.4$,p<.001	t(85)=9.97,p<.001	-	t(85)=9.17,p<.001
Mental Demand	$F(2,85)=58.3$,p<.001	t(84)=9.55,p<.001	-	t(84)=9.13,p<.001
Physical Demand	-	-	-	-
Temporal Demand	$F(2,85)=3.8$,p=0.026	t(85)=2.72,p=0.024	-	-
Performance	$F(2,85)=83.7$,p<.001	t(85)=10.71,p<.001	-	t(85)=11.64,p<.001
Effort	$F(2,85)=35.1$,p<.001	t(84)=7.56,p<.001	-	t(85)=6.9,p<.001
Frustration	$F(2,85)=14.01$,p<.001	t(85)=4.94,p<.001	-	t(85)=4.13,p<.001

Table 1: Statistically significant results of ART and ART contrasts for main effects EPR method.

time between highlighting a new square on the HoloLens 2 and the participant touching the corresponding square, as detected by the touch frame. Accuracy was defined as the percentage of correctly touched squares. Task load was assessed using the NASA TLX [24], and usability via the Single Ease Question (SEQ). Participants also ranked the EPR methods by preference. To estimate alignment accuracy between the EPR view and the real environment, participants viewed a second checkerboard mounted on a wall, with 1 cm squares overlaid using Canny edge detection and white silhouettes. Positioned 75 cm from the wall, participants estimated the offset between the real and virtual squares in millimeters for Mesh-Proxy EPR and Gaze-Proxy EPR. Plane-Proxy EPR was excluded, as it's accuracy relies on continuous manual alignment.

Procedure. After welcoming participants, they signed an informed consent form and filled out a demographic questionnaire. Their dominant eye was determined using the Miles Test [38] to align the proxy planes of GAZE and PLANE accordingly. Participants then put on the HoloLens 2 and used the built-in eye calibration tool to adjust the device to their eyes.

Participants were seated in the task area and instructed to touch the highlighted square as quickly as possible once it became visible on the HMD, and to move the hand back into its resting position afterwards. Orientation and EPR method were counterbalanced. Half of participants started with TABLE, the other half with WALL. **EPR Method** was balanced using a 3x3 Latin Square Table. After finishing 30 repetitions of an EPR method, participants filled out NASA TLX and SEQ questionnaires, before continuing with the next EPR method. After finishing all EPR conditions, participants ranked the EPR methods and continued with the next Orientation condition.

Participants were allowed to remove the HMD when filling out the questionnaires, or to rest. If the HMD was removed, the eye calibration was performed before the next condition started. After finishing all conditions, participants were asked to estimate the accuracy of the alignment of Mesh-Proxy EPR and Gaze-Proxy EPR using another wall-mounted checkerboard. Participants then continued with the second, qualitative part of the study. We collected 2 (Orientation) x 3 (EPR method) x 25 = 150 data points for each participant, leading to 2700 repetitions over all 18 participants.

Hypotheses.

- **H1.** Due to the alignment of the visualization with the correct depth at the focus point of the user, MESH and GAZE outperform PLANE in **WALL**. We did not expect differences between MESH and GAZE due to correct alignment.
- **H2.** Due to the alignment of the visualization with the correct depth at the focus point of the user, MESH and GAZE outperform PLANE in **TABLE**. We did not expect differences between MESH and GAZE due to correct alignment.
- **H3.** PLANE performs better for WALL than for TABLE as the foreshortening effect of the steeper viewing angle further aggravates the misalignment.
- **H4.** Participant prefer MESH and GAZE over PLANE, as aligning virtual content to the correct depth reduces perceptual depth conflicts.

Results. We used the statistics software *R*, data was evaluated with a significance level of 0.05. The residuals did not fulfill the normality requirement. Therefore, we utilized align-and-rank transform (ART) [50] and follow-up ART contrasts [15] for post-hoc analysis. The reported p-values are Bonferroni-Holm corrected. For each EPR method and Orientation condition, we calculated the mean over all task conditions for each participant. Descriptive statistics are summarized as box plots in Fig. 4, as well as tables in supplemental material. Ranking results are shown in Fig. 5. Statistically significant differences between EPR methods and Orientation conditions are presented in Table 1. In the following, we report on interaction effects between EPR method and Orientation. ART revealed significant differences for correctness ($F(2,85)=19.2$,p<.001), contrasts between WALL*PLANE and WALL*MESH (t(85)=8.7,p<.0001), WALL*GAZE (t(85)=7.7,p<.0001), TABLE*MESH (t(85)=9.4,p<.0001), and TABLE*GAZE (t(85)=9.6,p<.0001), as well as between TABLE*PLANE and WALL*MESH (t(85)=11.2,p<.0001), WALL*GAZE (t(85)=7.7,p<.0001), TABLE*MESH (t(85)=11.9,p<.0001), and TABLE* GAZE (t(85)=12.1,p<.0001). ART revealed significant differences for TCT ($F(2,85)=5.7$,p=.005), contrasts between WALL* PLANE and WALL* MESH (t(85)=8.0,p<.0001), WALL* GAZE (t(85)=7.9,p<.0001), TABLE* MESH (t(85)=6.7,p<.0001), and TABLE* GAZE (t(85)=3.8,p=.002), as well as between TABLE* PLANE and WALL* MESH (t(85)=6.1,p<.0001), WALL* GAZE (t(85)=5.97,p<.0001), and TABLE* MESH (t(85)=4.8,p=.0001). Furthermore, between TABLE* GAZE and WALL* GAZE (t(85)=4.05,p=.0009), WALL* MESH (t(85)=4.2,p=.0006), and TABLE* MESH (t(85)=2.9,p=.027). In terms of **alignment accuracy**, for Mesh-Proxy EPR, participants determined a mean offset of 1.1 mm (sd=0.65 mm), for Gaze-Proxy EPR, a mean offset of 1.3 mm (sd=0.7 mm).

4.2 Qualitative Feedback

In the second part of the study, we collected qualitative feedback in a use case corresponding to a low vision support system where edges are highlighted to emphasize structures in the environment.

Apparatus. We calculated edge overlays on the HoloLens 2 by applying Canny edge detection to the world-camera video feed. The edges were shown in white color in the field of view of the user using the three EPR methods (PLANE, GAZE, MESH). Similar applications have been proposed, to help people with low vision [25, 39]. However, previous work projected edge enhanced image onto a proxy plane at a fixed distance in front of the user, which corresponds to the PLANE condition. Expecting misalignment between the overlaid edges and the scene, we allowed users to manually adjust the distance of the proxy plane via a menu. Participants had the option to switch between all three EPR methods at any time.

Procedure. Participants were encouraged to walk around the study room while trying out the different EPR methods and conveying their impressions of them. We then asked them to read a Snellen chart on a wall while trying out the three methods. The aim of this was to see if the EPR methods were robust enough the be used for enhancing the reading process by highlighting certain letters or key words, e.g.

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2025.3616739

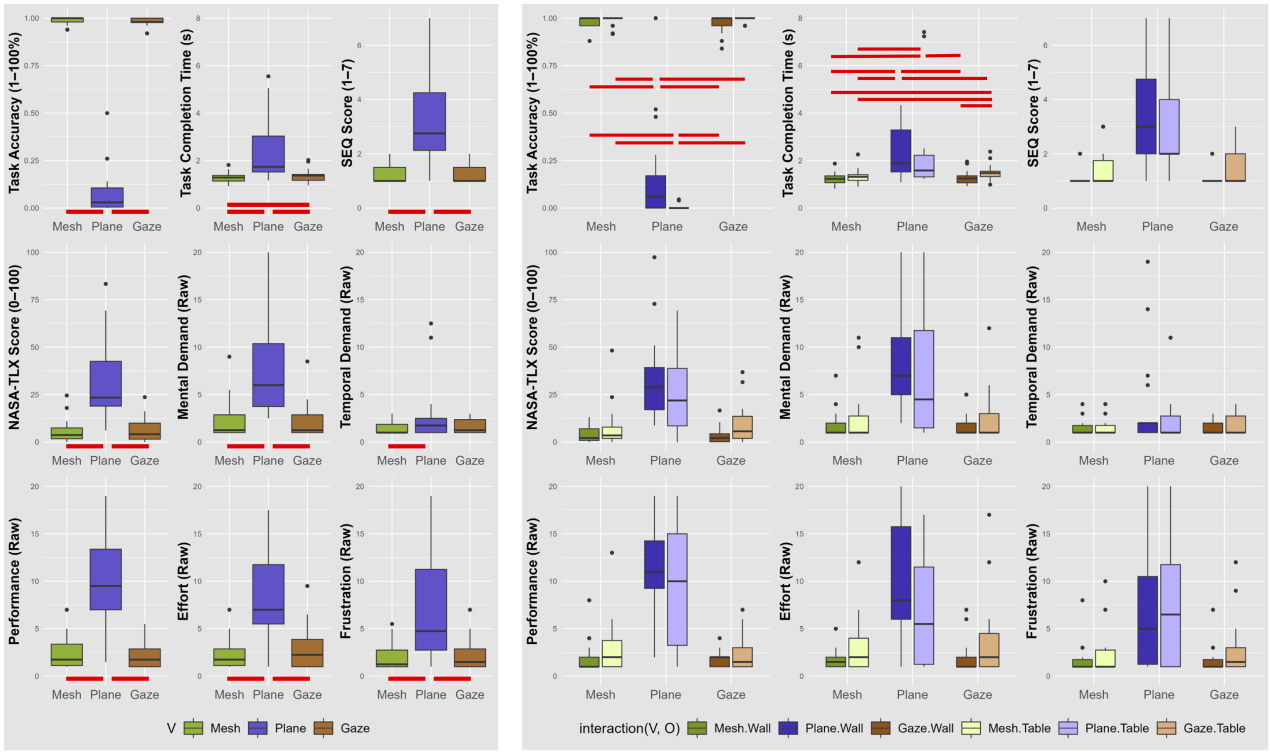

Fig. 4: Box Plots of Overall EPR Methods Conditions and Orientation * EPR Method Conditions. Significant differences are indicated with horizontal lines.

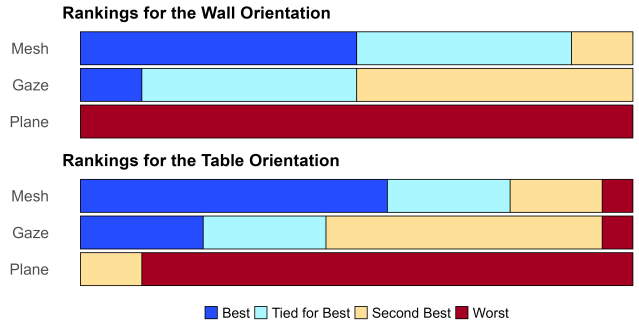

Fig. 5: Ranking of EPR Methods for both Orientations.

as proposed by [47]. Finally, we asked them to sit down at a table where Duplo bricks were laid out (Fig. 3(Right)). The participants were then asked to build three simple shapes out of the available Duplo bricks, one for each EPR method. The shapes consisted of 10 bricks each. Participants used printed images of the shapes as guidance (see supplemental material).

Data Collection. We collected observations and user feedback during this part of the study, while participants worked through the different tasks in the scene. Participants were also encouraged to think-aloud while trying out the EPR methods.

Results. In the following, we present participant feedback and observations, clustered by common themes that emerged.

Use Case: Low Vision. Participants approached the low vision support system in different ways based on their own experience. There was a difference between participants with and without vision issues (e.g., myopia). Seven of the participants wore glasses and one participant mentioned having undergone corrective surgery in the past. These participants were overall more receptive to the idea of using the edges as a reading aide. Four participants mentioned the support of outlines: *"It helps with the clarity"* (P7), *"The contrast is better"* (P16). Participants also opined usage scenarios: *"I think it would be helpful outdoors, for recognizing street signs"* (P8), *"You could use it for street signs"* (P17), *"Could be useful for people with color vision deficiencies"* (P16).

Use Case: Visual Guidance. When participants interacted with the edge overlays of the image-based EPR view analysis in the second part of the study, participants mentioned use cases that related to visual search tasks. Use Cases were mostly related to search tasks: *"It could help me find my glasses"* (P16), *"It could help me find things that I misplaced"* (P17), *"It could be used to point out the right tools"* (P12). Participants pointed out that for a search task more selective highlighting of objects would be beneficial.

Selective Highlighting. During the Duplo building task, only two participants (P8, P4) found the outlines helpful in identifying bricks, which is understandable as the visualization is inspired from low vision support and not guidance. Hence, participants mentioned improvements, such as making the outlines customizable (color, thickness, brightness), offering additional modes (shape-filling instead of outlining), reducing the number of visible outlines, or integrating the outlines into a more intelligent application (offering step-by-step instructions).

Failure of Plane-Proxy EPR. Participants could freely experiment with all three EPR methods. However, all participants eventually focused on either Gaze-Proxy EPR or Mesh-Proxy EPR on their own accord. Despite explanations regarding the functioning of the Plane-Proxy EPR [39, 55] and that it could be aligned manually by setting the distance via an AR button, or by simple head motion, participants were very expressive in pointing out the limitations of the technique: *"The outlines don't fit at all"* (P2), *"The fixed distance method is just nonsense. It's too annoying to adjust it all the time"* (P0), *"Fixed is horrible!"* (P7). When reading from the Snellen chart, participants completely disregarded Plane-Proxy EPR. During the Duplo building task participants further pointed out: *"It's making everything worse"* (P11), *"It is a catastrophe"* (P0), *"I can see the plane, it's blocking my view of the bricks"* (P5), *"It is not helping, it's actively hindering me"* (P7).

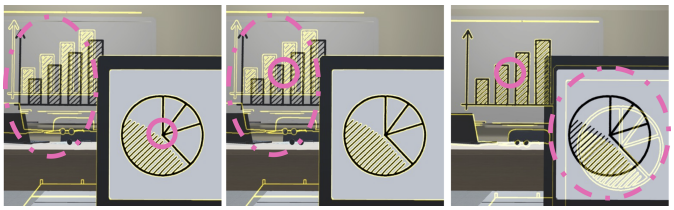

Fig. 6: Adaptation Process of Gaze-Proxy EPR. Adapting the proxy plane to a new depth using eye tracking potentially introduces temporary ambiguities due to system latency. (Left) As the user focuses on the foreground (purple circle), the proxy plane is aligned with the foreground geometry. The edge highlights in the background are misaligned (dashed lines). (Middle) When the user changes their gaze focus to the background (circle), they may perceive a short misalignment of highlights and background (dashed) before the system detects the user's intention, and (Right) adapts the proxy plane to the new depth.

Perceptual Issues. A common criticism was the lack of temporal coherence in edge overlays across all EPR methods, during head motion (P1, P8, P16). The simple Canny edge detection also lead to inconsistencies in outlining differently colored Lego bricks making it harder to distinguish between them for seven participants.

5 Discussion

In the following, we discuss results in relation to our hypotheses, before presenting our insights in more detail.

5.1 Hypotheses

The main effects between the EPR methods were statistically significant so that overall MESH and GAZE outperformed PLANE in all measures. However, there were almost no additional interaction effects for ART for (EPR method * Orientation) aside from accuracy and TCT.

H1. For WALL, **we accept H1**, as MESH and GAZE clearly outperformed PLANE, which displayed the image-based analysis at a fixed distance from the user, both overall (Fig. 4(Left)) and when analyzing interaction effects in terms of TCT and accuracy (Fig. 4(Right)).

H2. For TABLE, **we partially accept H2**, as MESH and GAZE again outperformed PLANE across all measures (Fig. 4(Left)) and in interaction effects for TCT and accuracy (Fig. 4(Right)).

However, TCT for GAZE*TABLE was significantly different from GAZE*WALL, MESH*TABLE, and MESH*WALL. We partly attribute this to a perceivable delay when updating the proxy plane depth, due to a combination of eye tracker latency and the larger misalignment between the proxy plane that is orthogonal to the user's view and the horizontal table surface compared to WALL, which potentially requires users to reorient as the visualization adapts (Fig. 6).

H3. Although PLANE performed better on average in the WALL condition in terms of accuracy, **we reject H3**, as the difference was not statistically significant. Despite participants being allowed to adjust their head position after the task began, only two were able to compensate for the offset caused by the misalignment between the plane and real-world geometry. While we did not observe a significant difference for PLANE across Orientations, we found a significant TCT difference between GAZE*TABLE and GAZE*WALL as described in the discussion of H2.

H4. We accept H4, as participants clearly preferred MESH and GAZE in both Orientation conditions (Fig. 5). Some participants could not distinguish between MESH and GAZE in terms of quality and ranked both equally. This was more common in the WALL condition ($n = 7/18$) than in TABLE ($n = 4/18$), which may indicate that temporary misalignments in EPR results with GAZE due to the plane adaptation process(Fig. 6) were more noticeable in TABLE (see discussion of H2 and H3).

5.2 Insights

Based on the quantitative data, qualitative feedback, and observations, we present the following insights.

Gaze-Proxy EPR as Effective Alternative. Overall, Gaze-Proxy EPR performed comparably to Mesh-Proxy EPR, which depends on detailed mesh reconstruction being available. Participants estimated only minor alignment errors of approximately 1 mm with the real world in both cases. Thus, Gaze-Proxy EPR represents a viable alternative to Mesh-Proxy EPR, requiring no additional scene information such as depth maps or mesh reconstructions when relying on precise eye tracking for vergence calculation [36,48]. Since the aim of Gaze-Proxy EPR is to create localized alignment directly in the area around the user's focus point by approximating scene geometry in that area with a plane, scene depth variations will cause misalignments between the virtual content and the real world. To avoid this, the gaze target should fill out the area in the user's field of view with a uniform depth. In our current implementation we did not enforce a specific size for the gaze target, however participants did not mention noticing such alignment errors during the study, e.g. when interacting with the Duplo bricks. However, future work should determine the minimum size at which the gaze target must fill out the user's visual field to ensure perceptually seamless alignment, potentially on the order of the human foveal region (approximately 1–2° of visual angle) in line with foveated rendering [1].

Gaze-Proxy EPR using Vergence-based Depth Estimates. Although our results show that Gaze-Proxy EPR is viable, our implementation relies on the spatial mesh to estimate gaze-target distance because the HoloLens 2 provides only a single combined gaze ray, preventing direct vergence calculations [48]. Placing the proxy plane via the intersection of the spatial mesh and the gaze ray can introduce errors e.g. when the spatial mesh does not reflect changes in the environment fast enough. Using an eye tracker that directly measures vergence distance avoids these issues as Gaze-Proxy EPR does not rely on a potentially imprecise or slowly updating mesh. Vergence-based depth estimates are more precise at smaller distances and degrade with distance due to angular errors [36] making Gaze-Proxy EPR particularly suitable for close-range tasks. For distant scene geometry, the parallax between the world camera and the user's eyes becomes minimal [8], and continuous EPR updates may not be required at ranges where vergence-based estimates are less precise. Further experiments are needed to fully explore the potential of vergence-based Gaze-Proxy EPR at different distances.

Delayed Eye Tracker Updates. Our experiments show reduced TCT in GAZE*TABLE compared to GAZE*WALL and both MESH conditions. This can be explained by the HoloLens 2's eye-tracking refresh rate (30 Hz) combined with the MRTK gaze-pointer implementation requiring at least 40 gaze samples for stabilization. These factors introduce a noticeable delay whenever the proxy plane is realigned via gaze refocus (Fig. 6). In GAZE*WALL, the proxy plane's angular alignment with the grid meant that gaze shifts toward the highlighted square required minimal depth adjustment, masking the delay. In GAZE*TABLE, the poorer alignment from the viewing angle caused larger depth adjustments, forcing participants to wait for the proxy plane to update before touching the highlighted square. Several participants reported noticing this delay with Gaze-Proxy EPR. Replacing the MRTK gaze pointer with a custom implementation using fewer samples can improve the delay, but risks of reducing pointer stability due sensor jitter and head movements not being smoothed. Further research is needed to fully understand how gaze-based proxy-plane updates affect task performance and to quantify the improvements achievable with higher-refresh-rate eye trackers.

High Accuracy of EPR. Since capturing the user's EPR view is not possible with an OST HMD, we asked participants to estimate the alignment error of edge overlays. While this is a subjective measure, it enhances ecological validity compared to using instrumentation such as a head rest by allowing unrestricted user movement when wearing the HMD. At 75 cm from a wall, participants reported alignment errors of approximately 1 mm for both Mesh-Proxy EPR and Gaze-Proxy EPR, indicating that EPR supports precise image-based vision augmentation.

Alignment accuracy was evaluated using only a single HMD model (HoloLens 2). However, given the low computational overhead of the proposed software-based EPR methods, we are confident in their applicability to a broad range of HMDs. Nonetheless, future investigations should explore EPR performance across a more diverse set of devices,

including wearables, e.g. Snap glasses.

Correct Depth Perception. Most participants using Plane-Proxy EPR reported seeing double overlays, i.e., one virtual square highlight for each eye. This visual offset arises from the misalignment between the user's real-world focus point and the fixed proxy plane. As a result, participants could not focus on both simultaneously. Participants did not report double vision for Mesh-Proxy EPR and Gaze-Proxy EPR, which underlines the importance of aligning image analysis with the correct scene depth.

Disregard Fixed-distance Plane-Proxy EPR. Plane-Proxy EPR performed significantly worse than other techniques, and, like in previous work [39, 55], participants strongly criticized it. Despite being instructed on how to compensate for the fixed offset, either via a software setting or head movement, participants found Plane-Proxy EPR impractical and ineffective for any given tasks. Therefore, we strongly advise researchers and practitioners to avoid fixed-distance Plane-Proxy EPR for image-based vision augmentation on OST HMDs and use depth-aligned methods such as Mesh-Proxy EPR or Gaze-Proxy EPR instead.

Optimize Object Highlights. For the qualitative study, we used Canny edge detection with a fixed highlight color to simulate a low vision support system. Participants noted that the edge overlays and chosen color sometimes obscured object recognition. Future research should explore adaptable color schemes that achieve a good contrast to the scene [19]. Furthermore, due to the precision of EPR, future systems can move beyond coarse bounding boxes [55] and use fine-grained silhouettes or minimal augmentations like points or lines rendered within a small object of interest, potentially reducing clutter introduced by additional augmentations. In procedural tasks (e.g., training, assembly), filtering techniques can further reduce visual clutter of highlights.

Need for Stable Image Analysis. Like in prior work [39], participants experienced issues with the temporal stability of edge overlays. Our qualitative feedback scenario used a basic Canny edge implementation, which led to flickering due to constant recomputation. Limited by the computational and battery constraints of the HoloLens 2, we could not deploy more advanced algorithms in this experiment. As mobile hardware improves or shifts to edge/cloud processing, future systems may support more stable, intelligent vision augmentation [3].

6 Conclusion

In this paper, we compared three EPR methods to collect quantitative and qualitative measures for vision augmentation for OST HMDs that need alignment with a real world scene. Our results clearly showed that the commonly used Plane-Proxy EPR approach that places a proxy plane at a fixed distance from the user [25, 39, 55] fails due to severe misalignments of the analyzed scene information and the users view.

Researchers and practitioners should focus on EPR methods that align the augmentations with the corresponding real-world geometry within the field of view of the user instead of the world-camera view. Mesh-Proxy EPR achieves this via a depth map from an RGBD camera or a mesh proxy based on a scene reconstruction, while our novel Gaze-Proxy EPR approach moves the proxy plane to the depth the user currently focuses on. We find that Gaze-Proxy EPR provides an effective alternative to the computationally more complex Mesh-Proxy EPR, and can be realized without scene geometry given a sufficiently precise eye tracker for eye vergence depth detection [37].

We provide our mobile EPR methods and study framework as open source to foster future research for vision augmentations for mobile, untethered OST HMDs in real-world environments. EPR is highly relevant for scenarios that only rely on image-based analysis [10] and utilize advanced AI-based LLMs and VLMs [14], which require that results are rendered from the user's eye perspective.

While we propose a software-based framework for EPR to stimulate further research, long-term, we see the need to integrate EPR natively into OST HMDs either via efficient compute hardware specialized for this task, or novel hardware architectures [30, 32]. Our hope is that if more research explores the effectiveness of EPR for real-world use cases, solution providers will pick up EPR as a standard feature for OST HMD so that future applications can make full use of advanced image analysis models to support users in their everyday tasks.

Supplemental Materials

The open source repository for the EPR framework: https://github.com/DigitalRealitiesLab/MobileEPR

Acknowledgments

This work was supported by a grant from the Austrian Research Promotion Agency (grant no. 877104).

References

[1] R. Albert, A. Patney, D. Luebke, and J. Kim. Latency requirements for foveated rendering in virtual reality. *ACM Transactions on Applied Perception (TAP)*, 14(4):1–13, 2017. 8

[2] B. S. Ananto, R. F. Sari, and R. Harwahyu. Color transformation for color blind compensation on augmented reality system. In *2011 International Conference on User Science and Engineering (i-USEr)*, pp. 129–134, 2011. doi: 10.1109/iUSEr.2011.6150551 3, 5

[3] H. Bahri, D. Krčmařík, and J. Kočí. Accurate object detection system on hololens using yolo algorithm. In *2019 International Conference on Control, Artificial Intelligence, Robotics & Optimization (ICCAIRO)*, pp. 219–224, 2019. doi: 10.1109/ICCAIRO47923.2019.00042 4, 9

[4] J. N. Bailenson, B. Beams, J. Brown, C. DeVeaux, E. Han, A. C. M. Queiroz, R. Ratan, M. Santoso, T. Srirangarajan, Y. Tao, and P. Wang. Seeing the World Through Digital Prisms: Psychological Implications of Passthrough Video Usage in Mixed Reality. *Technology, Mind, and Behavior*, 5(2: Summer 2024), 2024. 2

[5] D. Baričević, T. Höllerer, P. Sen, and M. Turk. User-perspective augmented reality magic lens from gradients. In *ACM Symposium on Virtual Reality Software and Technology*, pp. 87–96. ACM, 2014. doi: 10.1145/2671015.2671027 2

[6] D. Baričević, C. Lee, M. Turk, T. Höllerer, and D. A. Bowman. A handheld ar magic lens with user-perspective rendering. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 197–206, 2012. doi: 10.1109/ISMAR.2012.6402557 2

[7] M. Bhattarai, A. R. Jensen-Curtis, and M. Martínez-Ramón. An embedded deep learning system for augmented reality in firefighting applications. In *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 1224–1230, 2020. doi: 10.1109/ICMLA51294.2020.00193 1

[8] R. A. Borsoi and G. H. Costa. On the Performance and Implementation of Parallax Free Video See-Through Displays. *IEEE Transactions on Visualization and Computer Graphics*, 24(6):2011–2022, 2018. doi: 10.1109/TVCG.2017.2705184 3, 8

[9] A. Burova, J. Mäkelä, J. Hakulinen, T. Keskinen, H. Heinonen, S. Siltanen, and M. Turunen. Utilizing vr and gaze tracking to develop ar solutions for industrial maintenance. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 13 pages, p. 1–13. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3313831.3376405 2

[10] J. Cao, K.-Y. Lam, L.-H. Lee, X. Liu, P. Hui, and X. Su. Mobile augmented reality: User interfaces, frameworks, and intelligence. *ACM Comput. Surv.*, 55(9), article no. 189, 36 pages, Jan. 2023. doi: 10.1145/3557999 2, 9

[11] G. Chaurasia, A. Nieuwoudt, A.-E. Ichim, R. Szeliski, and A. Sorkine-Hornung. Passthrough+: Real-time stereoscopic view synthesis for mobile mixed reality. *Proc. ACM Comput. Graph. Interact. Tech.*, 3(1), article no. 7, 17 pages, May 2020. doi: 10.1145/3384540 2

[12] T. E. Chemaly, M. Goyal, T. Duan, V. Phadnis, S. Khattar, B. Vlaskamp, A. Kulshrestha, E. L. Turner, A. Purohit, G. Neiswander, et al. Mind the gap: Geometry aware passthrough mitigates cybersickness. *arXiv preprint arXiv:2502.11497*, 2025. 3

[13] Y. Cheng, Y. Yan, X. Yi, Y. Shi, and D. Lindlbauer. Semanticadapt: Optimization-based adaptation of mixed reality layouts leveraging virtual-physical semantic connections. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, UIST '21, 16 pages, p. 282–297. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3472749.3474750 2

[14] M. D. Dogan, E. J. Gonzalez, K. Ahuja, R. Du, A. Colaço, J. Lee, M. Gonzalez-Franco, and D. Kim. Augmented object intelligence with xr-objects. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, UIST '24, article no. 19, 15 pages.

Association for Computing Machinery, New York, NY, USA, 2024. doi: 10.1145/3654777.3676379 1, 3, 9

[15] L. A. Elkin, M. Kay, J. J. Higgins, and J. O. Wobbrock. An aligned rank transform procedure for multifactor contrast tests. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, UIST '21, 15 pages, p. 754–768, 2021. doi: 10.1145/3472749.3474784 6

[16] G. Emsenhuber, M. Domhardt, T. Langlotz, D. Kalkofen, and M. Tatzgern. Towards eye-perspective rendering for optical see-through head-mounted displays. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 640–641, 2022. doi: 10.1109/VRW55335.2022.00171 2, 3

[17] G. Emsenhuber, T. Langlotz, D. Kalkofen, J. Sutton, and M. Tatzgern. Eye-perspective view management for optical see-through head-mounted displays. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, article no. 707, 16 pages. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544548.3581059 2, 3

[18] M. Eswaran, A. K. Gulivindala, A. K. Inkulu, and M. Raju Bahubalendruni. Augmented reality-based guidance in product assembly and maintenance/repair perspective: A state of the art review on challenges and opportunities. *Expert Systems with Applications*, 213:118983, 2023. doi: 10.1016/j.eswa.2022.118983 1, 4

[19] J. L. Gabbard, J. E. Swan, D. Hix, Si-Jung Kim, and G. Fitch. Active Text Drawing Styles for Outdoor Augmented Reality: A User-Based Study and Design Implications. In *IEEE Virtual Reality*, pp. 35–42. IEEE, mar 2007. doi: 10.1109/VR.2007.352461 9

[20] Y. Ghasemi, H. Jeong, S. H. Choi, K.-B. Park, and J. Y. Lee. Deep learning-based object detection in augmented reality: A systematic review. *Computers in Industry*, 139:103661, 2022. doi: 10.1016/j.compind.2022.103661 3

[21] C. Gsaxner, S. Mori, D. Schmalstieg, J. Egger, G. Paar, W. Bailer, and D. Kalkofen. Deepdr: Deep structure-aware rgb-d inpainting for diminished reality. In *2024 International Conference on 3D Vision (3DV)*, pp. 750–760, 2024. doi: 10.1109/3DV62453.2024.00037 3

[22] Q. Gu, A. Kuwajerwala, S. Morin, K. M. Jatavallabhula, B. Sen, A. Agarwal, C. Rivera, W. Paul, K. Ellis, R. Chellappa, C. Gan, C. M. de Melo, J. B. Tenenbaum, A. Torralba, F. Shkurti, and L. Paull. Conceptgraphs: Open-vocabulary 3d scene graphs for perception and planning. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5021–5028, 2024. doi: 10.1109/ICRA57147.2024.10610243 1

[23] P. Guan, E. Penner, J. Hegland, B. Letham, and D. Lanman. Perceptual requirements for world-locked rendering in ar and vr. In *SIGGRAPH Asia 2023 Conference Papers*, SA '23, article no. 35, 10 pages. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3610548.3618134 2, 3

[24] S. G. Hart and L. E. Staveland. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock and N. Meshkati, eds., *Human Mental Workload*, vol. 52 of *Advances in Psychology*, pp. 139–183. North-Holland, 1988. 6

[25] A. D. Hwang and E. Peli. An augmented-reality edge enhancement application for google glass. *Optometry and Vision Science*, 91(8), 2014. 3, 6, 9

[26] J. Jia, S. Elezovikj, H. Fan, S. Yang, J. Liu, W. Guo, C. C. Tan, and H. Ling. Semantic-aware label placement for augmented reality in street view. *Visual Computer*, 37(7):1805–1819, 2021. doi: 10.1007/s00371-020-01939-w 3

[27] S. K. Kim and M. Y. Kim. Deep learning-based calibration method for an augmented reality surgical navigation system without head-mounted optical markers. In *2024 24th International Conference on Control, Automation and Systems (ICCAS)*, pp. 453–458, 2024. doi: 10.23919/ICCAS63016.2024.10773365 3

[28] G. Kuo, E. Penner, S. Moczydlowski, A. Ching, D. Lanman, and N. Matsuda. Perspective-correct vr passthrough without reprojection. In *ACM SIGGRAPH 2023 Conference Proceedings*, SIGGRAPH '23, article no. 15, 9 pages. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3588432.3591534 2

[29] Z.-H. Lai, W. Tao, M. C. Leu, and Z. Yin. Smart augmented reality instructional system for mechanical assembly towards worker-centered intelligent manufacturing. *Journal of Manufacturing Systems*, 55:69–81, 2020. doi: 10.1016/j.jmsy.2020.02.010 4

[30] T. Langlotz, M. Cook, and H. Regenbrecht. Real-time radiometric compensation for optical see-through head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, 22(11):2385–2394, 2016. doi: 10.1109/TVCG.2016.2593781 1, 2, 9

[31] T. Langlotz, J. Sutton, and H. Regenbrecht. A design space for vision augmentations and augmented human perception using digital eyewear. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, article no. 966, 16 pages. Association for Computing Machinery, New York, NY, USA, 2024. doi: 10.1145/3613904.3642380 3

[32] T. Langlotz, J. Sutton, S. Zollmann, Y. Itoh, and H. Regenbrecht. Chromaglasses: Computational glasses for compensating colour blindness. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, 12 pages, p. 1–12. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3173964 1, 2, 4, 5, 9

[33] J. Liu, A. S. Rao, F. Ke, T. Dwyer, B. Tag, and P. D. Haghighi. Arfacilitated safety inspection and fall hazard detection on construction sites. In *2024 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 12–14, 2024. doi: 10.1109/ISMAR-Adjunct64951.2024.00010 1

[34] L. Liu, H. Li, and M. Gruteser. Edge assisted real-time object detection for mobile augmented reality. In *The 25th Annual International Conference on Mobile Computing and Networking*, MobiCom '19, article no. 25, 16 pages. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3300061.3300116 3

[35] L. Lutnyk, K. G. Kim, A. Sarbach, P. Kiefer, R. Häusler, and M. R. and. Context-sensitive augmented reality assistance in the cockpit. *International Journal of Human–Computer Interaction*, 0(0):1–20, 2024. doi: 10.1080/10447318.2024.2440976 4

[36] D. Mardanbegi, T. Langlotz, and H. Gellersen. Resolving target ambiguity in 3d gaze interaction through vor depth estimation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, 12 pages, p. 1–12. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300842 2, 3, 8

[37] Y. Matsuda, F. Shibata, A. Kimura, and H. Tamura. Poster: Creating a user-specific perspective view for mobile mixed reality systems on smartphones. In *IEEE Symposium on 3D User Interface*, pp. 157–158. IEEE, 2013. doi: 10.1109/3DUI.2013.6550226 9

[38] W. R. Miles. Ocular dominance in human adults. *The Journal of General Psychology*, 3(3):412–430, 1930. doi: 10.1080/00221309.1930.9918218 6

[39] H. Min Htike, T. H. Margrain, Y.-K. Lai, and P. Eslambolchilar. Augmented reality glasses as an orientation and mobility aid for people with low vision: a feasibility study of experiences and requirements. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, article no. 729, 15 pages. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445327 1, 2, 3, 4, 5, 6, 7, 9

[40] P. Mohr, M. Tatzgern, J. Grubert, D. Schmalstieg, and D. Kalkofen. Adaptive user perspective rendering for handheld augmented reality. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 176–181, 2017. doi: 10.1109/3DUI.2017.7893336 2

[41] J. Mooser, S. You, and U. Neumann. Real-time object tracking for augmented reality combining graph cuts and optical flow. In *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 145–152, 2007. doi: 10.1109/ISMAR.2007.4538839 1

[42] W. H. Organization. Blindness and visual impairment, 2023. Accessed: 2025-03-24. 4

[43] U. Rajapaksha, F. Sohel, H. Laga, D. Diepeveen, and M. Bennamoun. Deep learning-based depth estimation methods from monocular image and videos: A comprehensive survey. *ACM Comput. Surv.*, 56(12), article no. 315, 51 pages, Oct. 2024. doi: 10.1145/3677327 3

[44] J. Sutton, T. Langlotz, and A. Plopski. Seeing colours: Addressing colour vision deficiency with vision augmentations using computational glasses. *ACM Trans. Comput.-Hum. Interact.*, 29(3), article no. 26, 53 pages, Jan. 2022. doi: 10.1145/3486899 2, 3

[45] J. Sutton, T. Langlotz, A. Plopski, S. Zollmann, Y. Itoh, and H. Regenbrecht. Look over there! investigating saliency modulation for visual guidance with augmented reality glasses. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–15, 2022. 2

[46] S. Teruggi and F. Fassi. Hololens 2 spatial mapping capabilities in vast monumental heritage environments. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVI-2/W1-2022:489–496, 2022. doi: 10.5194/isprs-archives-XLVI-2-W1-2022-489-2022 3

[47] T. Toyama, W. Suzuki, A. Dengel, and K. Kise. User attention oriented augmented reality on documents with document dependent dynamic overlay. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 299–300, 2013. doi: 10.1109/ISMAR.2013.6671814 7

[48] Z. Wang, Y. Zhao, and F. Lu. Gaze-vergence-controlled see-through vision in augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 28(11):3843–3853, 2022. doi: 10.1109/TVCG.2022.3203110 3, 8

[49] M. Weinmann, S. Wursthorn, M. Weinmann, and P. Hübner. Efficient 3D Mapping and Modelling of Indoor Scenes with the Microsoft HoloLens: A Survey. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 89(4):319–333, Aug. 2021. doi: 10.1007/s41064 -021-00163-y 3

[50] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, 4 pages, p. 143–146, 2011. doi: 10.1145/1978942.1978963 6

[51] L. Xiao, S. Nouri, J. Hegland, A. G. Garcia, and D. Lanman. Neural-passthrough: Learned real-time view synthesis for vr. In *ACM SIGGRAPH 2022 Conference Proceedings*, SIGGRAPH '22, article no. 40, 9 pages. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3528233.3530701 2

[52] C. Xu, R. Kumaran, N. Stier, K. Yu, and T. Höllerer. Multimodal 3d fusion and in-situ learning for spatially aware ai. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 485–494, 2024. doi: 10.1109/ISMAR62088.2024.00063 1

[53] O. Younis, W. Al-Nuaimy, F. Rowe, and M. H. Alomari. A smart context-aware hazard attention system to help people with peripheral vision loss. *Sensors*, 19(7), 2019. doi: 10.3390/s19071630 1

[54] X. Yu, G. Yang, S. Jones, and J. Saniie. Ar marker aided obstacle localization system for assisting visually impaired. In *2018 IEEE International Conference on Electro/Information Technology (EIT)*, pp. 0271–0276, 2018. doi: 10.1109/EIT.2018.8500166 1

[55] Z. Zhu, J. Li, Y. Tang, K. Go, M. Toyoura, K. Kashiwagi, I. Fujishiro, and X. Mao. Cc-glasses: Color communication support for people with color vision deficiency using augmented reality and deep learning. In *Proceedings of the Augmented Humans International Conference 2023*, AHs '23, 10 pages, p. 190–199. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3582700.3582707 2, 3, 5, 7, 9

[56] M. Łysakowski, K. Żywanowski, A. Banaszczyk, M. R. Nowicki, P. Skrzypczyński, and S. K. Tadeja. Real-time onboard object detection for augmented reality: Enhancing head-mounted display with yolov8. In *2023 IEEE International Conference on Edge Computing and Communications (EDGE)*, pp. 364–371, 2023. doi: 10.1109/EDGE60047.2023. 00059 3