Eye-Perspective View Management
for Optical See-Through Head-Mounted Displays

Gerlinde Emsenhuber
gerlinde.emsenhuber@fh-
salzburg.ac.at
Salzburg University of
Applied Sciences, Austria

Tobias Langlotz
tobias.langlotz@otago.ac.nz
University of Otago
Dunedin, New Zealand

Denis Kalkofen
kalkofen@icg.tugraz.at
Flinders University
Adelaide, Australia

Jonathan Sutton
j.sutton@otago.ac.nz
University of Otago
Dunedin, New Zealand

Markus Tatzgern
markus.tatzgern@fh-salzburg.ac.at
Salzburg University of
Applied Sciences, Austria

ABSTRACT

Optical see-through (OST) head-mounted displays (HMDs) enable users to experience Augmented Reality (AR) support in the form of helpful real-world annotations. Unfortunately, the blend of the environment with virtual augmentations due to semitransparent OST displays often deteriorates the contrast and legibility of annotations. View management algorithms adapt the annotations' layout to improve legibility based on real-world information, typically captured by built-in HMD cameras. However, the camera views are different from the user's view through the OST display which decreases the final layout quality. We present eye-perspective view management that synthesizes high-fidelity renderings of the user's view to optimize annotation placement. Our method significantly improves over traditional camera-based view management in terms of annotation placement and legibility. Eye-perspective optimizations open up opportunities for further research on use cases relying on the user's true view through OST HMDs.

CCS CONCEPTS

• **Computing methodologies → Mixed / augmented reality**; • **Human-centered computing** → *Empirical studies in visualization.*

KEYWORDS

augmented reality, optical see-through, head-mounted displays, label placement, legibility

1 INTRODUCTION

Many Augmented Reality (AR) applications rely on textual or graphical annotations for providing additional cues for users to solve tasks or retrieve additional information about the physical environment. The continuous overlay of those annotations can assist workers with maintenance [20] or assembly [51, 59] tasks. While initially demonstrated mainly in video see-through (VST) displays, we see an increasing trend in utilizing optical see-through (OST) head-mounted display (HMD) technology such as used in Microsoft's Hololenses or Magic Leap's HMDs. All those OST HMDs achieve the graphical overlay using semitransparent displays placed in the user's view. While research has proposed numerous improvements to OST HMDs [24], perceptual issues introduced by semitransparent displays still pose several challenges. In this work, we focus on the poor legibility and decreased contrast of graphical annotations, commonly introduced by the interference between the annotations and the real-world background (Figure 1 (Bottom, Left)). View management combined can improve contrast by adapting the placement [9, 40] and appearance of annotations [14, 22, 25]. However, correct view management requires exact knowledge of how the scene looks to the user when viewed through the OST HMD.

As the user's actual view through the HMD is not available, previous work often relies solely on the built-in HMD camera [9, 17, 40] for view management. However, the offset between the built-in camera and the user's eye leads to incorrect assumptions of the user's view and, thus, label placement, impacting the legibility and contrast of labels against the background (Figure 1 (Left), Figure 2). In addition, most view management techniques optimize layouts for a single camera view but do not consider stereo vision and the natural parallax between the eyes. This approximation often causes a different view for each eye, impacting legibility of text (Figure 2).

In this paper, we introduce eye-perspective view management for OST HMDs that optimizes label layouts from the viewpoint of both user's eyes, thereby supporting stereo vision, avoiding double vision, improving contrast and the overall legibility (Figure 1 (Center, Right)). Our approach to view management is based on eye-perspective rendering (EPR) [10], which reconstructs the user's view through the OST HMD in real-time. EPR is inspired by user-perspective rendering (UPR) for handheld AR [35, 54], where the AR view on a handheld device is distorted so that the device resembles a transparent lens from the user's perspective. As our approach
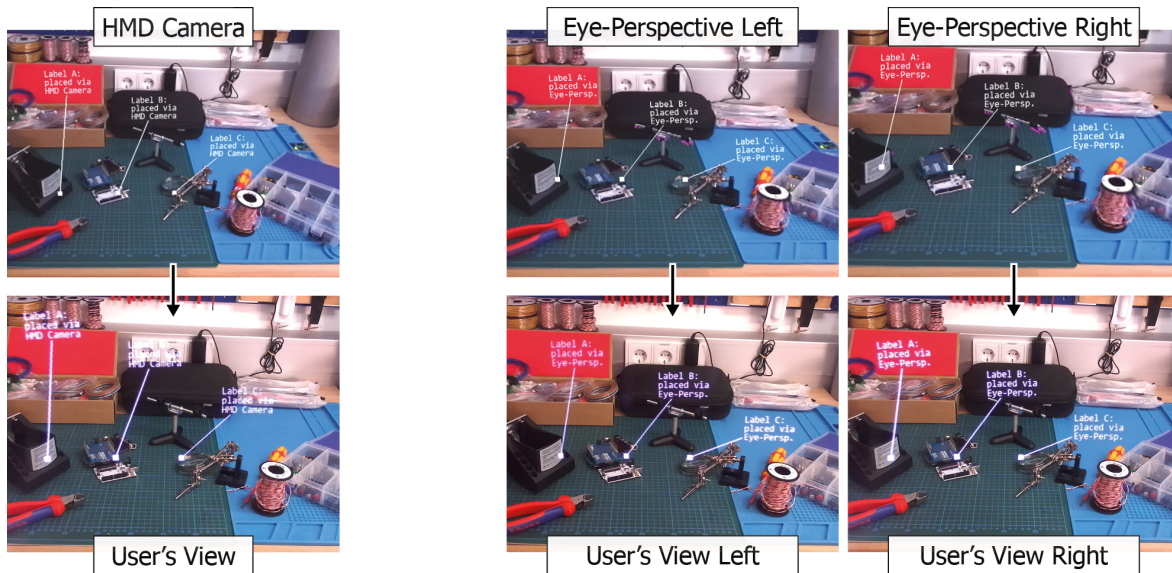
Figure 1: Without knowledge of the user's true view, AR applications do not have information to place augmentations such as labels without them interfering with the background. (Top, Left) Using the camera feed from a built-in head-mounted display camera, a view management algorithm places labels suitable for the respective background areas, assuming this ensures legibility of labels. (Bottom, Left) In the actual view through the display, the user perceives the label in a different location due to the viewpoint offset and parallax between the camera and the user's eyes, which leads to contrast and legibility issues. (Center and Right, Top) Eye-perspective view management synthesizes high-fidelity renderings of the user's view through the display for both eyes in order to adapt label placement in order to avoid visual interference with the background. (Center and Right, Bottom) As eye-perspective view management optimizes the label layout based on the user's actual view through the display, the finally perceived layout is faithful to the calculated layout. Eye-perspective view management also optimizes the layout for both eyes, so that labels share the same uniform background in each eye to legibility issues due to stereo perception.

is mainly software-based, EPR can be implemented on currently available OST HMDs, avoiding hardware modifications that reduce the eye-box and increase the size of the HMD [25, 27, 50].

As the specific challenges in stereoscopic view management for OST HMDs are rarely discussed, we initially perform an exploratory study in which we replicate perceptual issues that have been observed when reading subtitles in stereoscopic Virtual Reality (VR) [38]. In the study, we also evaluate background uniformity behind text as a potential solution for stereoscopic view management to avoid such perceptual issues. We analyze different approaches to implement EPR for real-world scenes of different complexity, which allow for stereoscopic eye-perspective view management and support various HMD hardware configurations. While for simple planar scenes EPR based on an efficient homography of the HMD camera view may be sufficient [54], for more complex scenes 3D proxy geometry of the scene must be created at runtime [3] to avoid visual artifacts in the synthesized views. Based on our analysis, the chosen EPR method for stereoscopic view management strikes a balance between computational performance and reconstruction quality. Finally, we demonstrate the feasibility of our EPR-based, stereoscopic view management in a second user study and compare it against traditional view management relying only on the real-world view of a built-in HMD camera. Our results show

that EPR-based view management significantly improves annotation placement and legibility while compensating for viewpoint and, thus, visible background changes in a mobile AR scenario. In summary, we make the following contributions:

- We demonstrate that stereoscopic perception of annotations in AR HMDs influences their legibility when text is offset from the real-world background and present background uniformity as a potential solution.
- We explore various approaches for synthesizing the user's view and their suitability for view management for OST HMDs. We focus on approaches that fit mobile HMDs and utilize commonly available hardware. Hence, the methods can be integrated into existing HMDs without requiring hardware modifications, thereby, opening up further research into eye-perspective view management and other algorithms and interactions requiring information about the user's view.
- We create a novel eye-perspective view management algorithm for OST HMDs that utilizes scene information from both of the user's eyes for optimization, thereby relying on the user's true view of the real-world scene. The results of our user study show that eye-perspective view management improves the contrast and legibility of annotations compared to view management optimizations based on the integrated camera view of an HMD.

Figure 2: Illustrating the mismatch between HMD camera and user's eyes. We measured the offset between the built-in HMD camera and the user's eye for various commercial AR HMDs and created renderings for a Unity scene illustrating this offset for left and right eye of the user. Labels are placed at 70cm from the HMD, the scene is approximately 2m from the HMD. The background behind the labels changes with each device and eye perspective compared to the HMD camera view. Displacements depend on the location of the HMD camera relative to the user's eyes. Note that, due to the stereo perception of the labels, the label background is different for left and right eye, which can lead to perceptual issues when reading labels.

2 RELATED WORK

In the following, we provide an overview of techniques for improving the rendering of annotations, such as view management algorithms that base their decisions on an analysis of the real-world background that often does not match the user's real view through the display. We then discuss work that focuses on UPR or EPR.

2.1 View Management

A number of approaches have been proposed that use the real-world background to improve AR rendering. Relevant to this work are those that improve the annotations' layout, i.e., view management, and rendering to improve legibility.

Annotation Placement. In early work on view management, Bell et al. [4] optimized label placement in AR and overcame the issue of EPR by simulating their approach in VR. Later work utilized VST AR, where the user perceives the real world via a video feed that can be analyzed and augmented directly as the user's view. Examples include the optimization of annotation placement based on visual features or visual saliency [17, 44], snapping to or aligning with geometric features [26, 39], or the adaptation of label visibility based on a background analysis [17, 28, 40]. For instance, Orlosky et al. [40] utilizes background uniformity and darkness criteria to identify areas where labels would be legible in a video image. However, these approaches have been demonstrated in VST AR as applying

these methods directly to OST HMDs produces erroneous results, because the camera view(s) and, thus, the base for optimization, do not match the user's view of the scene (Figure 3).

Contrast and Color Enhancement. Particularly in the area of contrast and color enhancement, several approaches have considered the effect of blending displayed information with the real-world background [22, 25, 27, 49]. However, most of these works rely on simulations to demonstrate their approach and do not solve the issue of mapping from a camera view to the OST display as seen by the users. An exception to this is the work of Langlotz et al. [25, 27] who modify an OST HMD with semitransparent mirrors reflecting incoming light towards calibrated cameras to enable EPR in order to compensate for the effect of background color blending by applying pixel-wise corrections. While they proved the efficacy of the approach, it requires hardware modifications that substantially add to the size of the HMD.

Stereo view management. There are several examples in the literature on utilizing depth cues and stereo perception for label placement (e.g. [42, 43]). Recently, there has also been work on identifying and overcoming depth issues and perceptual conflicts between text and background when displaying subtitles in VR scenes [38, 48]. When labels are not aligned with scene geometry but shown at a different depth, double vision of the background negatively impacts legibility. Furthermore, stereo view management will perform

Figure 3: Optimization criteria utilizing incorrect scene information. (Left) The HMD camera view and the (Center) user's real view capture a real-world scene from different perspectives. Therefore, optimization algorithms like view management rely on incorrect scene information when utilizing the HMD camera's view to optimize content for the user's view. (Right) Blue, yellow and white overlays illustrate the results of an optimization criterion utilizing HMD camera information, overlaid over the user's view. Blue indicates pixels that are incorrectly classified as valid regions in the HMD camera view. Yellow indicates pixels that are valid in the user's view, but have not been detected in the HMD view. White indicates results that are correctly classified despite the offset, due to visual similarity between HMD camera view and the user's view in the respective area.

worse in VST AR HMDs because the background negatively affects the legibility of labels, with authors reporting a decrease in performance in search tasks [42]. Overcoming this issue requires understanding the scene as viewed by the user which seemed so far unfeasible for OST HMD.

2.2 Eye- and User-Perspective Rendering in AR

UPR or EPR for AR is typically aimed at VST AR displays and only very few considered related issues for OST HMDs.

Video See-Through AR Displays. Video-based UPR approaches attempt to warp or render a view of a physical scene so that it appears as seen by the user of a system, for instance, based on calculating the desired viewport for a user's eye position when looking through a handheld AR display [21, 31]. Samini et al. [45, 46] demonstrated the feasibility of this UPR technique for interaction and search tasks. A variation of this approach was presented by Tomioka et al. [54] who used 3D feature points of the scene to continuously calculate a homography that transforms the video image to the desired view of the user. To avoid continuous updates and the performance overhead of head tracking, Pucihar et al. [56, 57] assumed a fixed location of the user's head relative to the handheld display. Mohr et al. [35] overcame the restricted head position by efficiently updating the head position only when the user's viewpoint relative to the handheld device changed beyond a certain threshold. 2D approaches that warp the camera feed introduce perceivable artifacts if the scene is not mostly planar, or introduce disocclusion artifacts where parts of the scene are not visible from the user's viewpoint. This can be overcome by using image data and proxy geometry from RGBD cameras [47] or a complete 3D reconstruction [3]. Due to the computational overhead of these methods, other works rely on a single depth image and create missing information by image-based rendering (IBR) [1, 2, 47].

Closest to our technical solutions for realizing EPR is the work of Chaurasia et al. [7] who presented a UPR method to create a video see-through AR mode on an Oculus Quest, a mobile VR platform, rendering the information of greyscale stereo cameras from the

view of the user's eyes. A very recent work aimed at providing an improved 3D reconstruction. Instead of using motion vectors for creating the stereo matching [7], the approach used learning-based stereo matching [29]. Unfortunately, two high-end graphics cards (NVidia Titan) are needed for computing the see-through mode, which makes it unlikely that the method will work for mobile AR hardware soon. Finally, the existing approaches have not been applied in OST HMDs or compared against each other in practical scenarios. In this paper, we perform a systematic evaluation of EPR approaches that are based on UPR to determine feasible solutions for synthesizing the user's view through OST HMDs to perform optimizations based on the real-world background.

Optical See-Through AR Displays. To prototype and evaluate methods that rely on the user's viewpoint, previous work has often relied on software simulations of OST displays [22], or cameras looking through the display instead of the user's eyes [23]. A solution that allows for human subject experiments, is the approach of Langlotz et al. [25, 27, 50] which captures the user's view by introducing additional cameras and beam splitter into a prototypical HMD. While demonstrated in a mobile setup [50] their approach adds to the bulk and weight of the HMD, in particular when supporting larger fields of view. We base our EPR on the idea of UPR for mobile devices such as smartphones or tablets.

3 BACKGROUND UNIFORMITY FOR STEREOSCOPIC TEXT ANNOTATIONS

Perceptual issues such as double vision negatively impact the performance during selection tasks in stereo displays [53] and VR [60] when a mouse cursor and 3D geometry are shown at different depths. Previous research in VR has also shown that double vision causes issues with legibility, when the text is not aligned with the 3D geometry of the scene [38, 48]. Depending on the focus distance, either the text or the scene is perceived as doubled. To avoid perceptual issues, typically text or mouse cursors are aligned with the depth of the 3D scene. However, in AR, aligning labels with
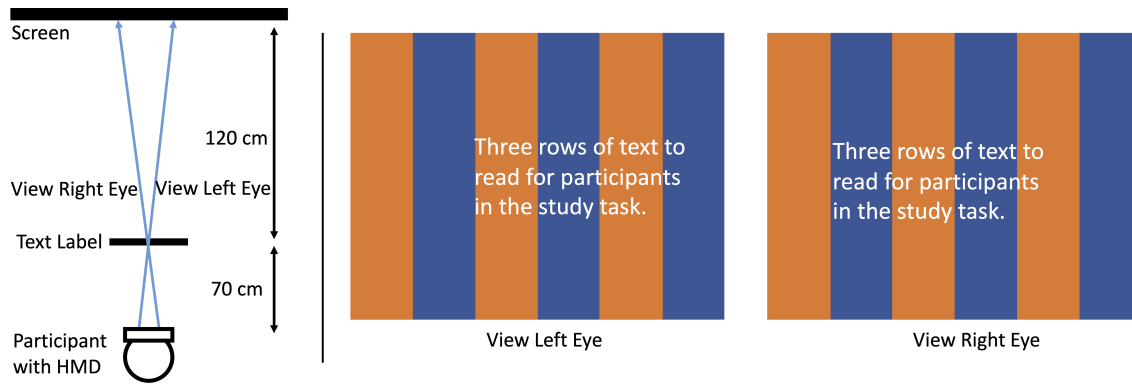
Figure 4: Overview stereo vision study. (Left) A sketch of the apparatus setup and the position of text and user, as well as the distance of the background in the offset condition and uniform condition. (Right) The text label overlaid over the background in the offset condition with striped backgrounds for the left and right eye showing the differing background behind the text leading to double vision. In the aligned condition, the text had the same background for both eyes. Note that for aligned and offset conditions, the width of vertical stripes was adjusted so that the text always had the same background.

the background is not feasible, because labels are also utilized to annotate a real-world context at a certain depth. We performed an exploratory user study to confirm the presence of perceptual issues, such as double vision and poor legibility, when reading stereoscopic AR labels on an OST HMD in front of a real-world background. Furthermore, as an alternative to depth adjustments, we explore the option of using a uniform background to avoid perceptual issues. The results of this study inform the design of our novel eye-perspective view management solution.

Study design. We used a within-subject design where participants read white textual labels shown within an OST HMD. The labels were rendered without a billboard and at a fixed distance of 70 cm from the participants (Figure 4). During the study, the text distance was kept constant, and the background distance varied. The independent variable was *background* with three conditions: (1) vertically striped background placed at a distance *aligned* with the text label, (2) vertically striped background placed at a distance *offset* from the text label by 120 cm, and (3) *uniform* blue background also offset by 120 cm from the label. In the offset condition, the vertical stripes lead to different backgrounds behind the text label, which in turn can cause double vision due to eye vergence on the label in the foreground. The uniform condition represents a potential solution to the encountered perceptual issues.

Participants. For the user study, 12 participants (3 female, $\overline{X}$=30.6 (7.7) years) volunteered. All participants had normal or corrected-to-normal vision and were not affected by color vision deficiency. We determined the dominant eye of participants with the Miles test [33] to align the text with that eye during the tasks, as it is generally used for reading.

Apparatus. Participants were seated in front of a 52 inch screen showing a uniform or striped background depending on the condition. We used a Microsoft Hololens 2 to display white textual labels (Figure 4). The screen could be moved to realize the aligned and offset conditions during the study. Before each condition, users

aligned the text with the background so that its size (2cm) and location was the same between conditions. The text was shown in the Hololens at the highest brightness setting. The uniform condition had a blue background with RGB #00339B, the stripes alternated between blue and orange (RGB #FF6600) (Figure 4 (Right)).

Task. Participants read the text aloud and were instructed to avoid errors. The text consisted of simple three-line sentences, with on average 58 characters and 10 words per text (Figure 4 (Right)).

Data Collection. We measured Task Completion Time (TCT) starting from when the text was shown to users until they finished the reading label. We also measured error as the number of mistakes during reading. Participants filled out a NASA TLX questionnaire after each condition, custom questions using a 7-point scale regarding eye fatigue (Q1), experiencing double vision (Q2), and difficulty reading due to the background (Q3). Participants also stated preference for background pattern (striped, uniform) and preference for offset (offset, aligned).

Procedure. Participants were recruited via public email to a university campus. After filling out an informed consent form and demographics data, participants were seated in front of the apparatus and put the HMD on. Once participants calibrated the setup and performed trial tasks , the first condition started, and they read 30 unique sentences. After finishing a task, participants removed the HMD to fill out questionnaires. After finishing all conditions, participants ranked conditions and the experimenter performed an interview. The experiment lasted approximately 45 minutes. The order of background conditions and text sets was balanced with a mutually orthogonal Latin Square. For analysis, we calculated the mean of TCT and errors for each condition for each participant. With 3 (conditions) x 30 (texts) = 90 repetitions for each participant and task, for 12 participants there were 1060 trials.

Hypothesis. We expected that the offset condition would perform worse than aligned condition where text and object are at the same depth because the offset between text and background would lead
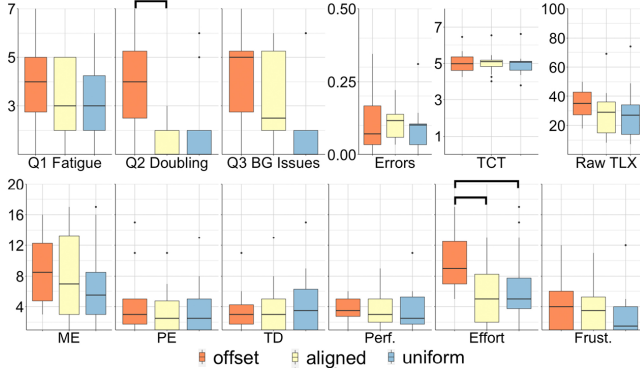
Figure 5: Data stereo vision study. (Top) Questionnaire responses to questions about eye fatigue, double vision, issues with the background while reading, as error, TCT and raw TLX results. (Bottom) The items of the NASA TLX.

to double vision due to eye vergence [38, 48] making it harder to read the text. Furthermore, the alternative approach of using a uniform background for stereoscopically displayed text would perform better than the offset condition.

Results. If not indicated otherwise, we report numerical values as "mean (sd)". We evaluated the data using a significance level of 0.05. The residuals underlying the data did not fulfill the normality requirement. Therefore, we used non-parametric Friedman tests and post-hoc Wilcoxon signed-rank tests for analysis. We calculate effect size for Wilcoxon signed-rank tests as $r = \frac{Z}{\sqrt{N}}$ [12]. The reported p-values are Holm-Bonferroni corrected. The analysis was performed using the statistics software *R*. We summarise the overall results in Figure 5. Friedman tests revealed significant differences for Q2 regarding double vision ($\chi^2(2)$=4.86,p=0.001) and for effort in the NASA TLX ($\chi^2(2)$=13.15,p=0.001). Wilcoxon signed-rank tests revealed a difference at the significance level for Q2 between offset and aligned condition (Z=2.86,p=.012,r=.58), and for effort between offset and uniform condition (Z=2.91,p=.006,r=.59), and offset and aligned condition (Z=2.63,p=.012,r=.54). 75% of participants preferred the uniform background. Regarding the striped background, 75% of the participants preferred the condition where the screen was aligned with the text label.

Discussion. Our data shows only statistically significant differences with respect to participants perceiving double vision of the background (Q2) in the offset condition, as well as with respect to the effort parameter of the NASA TLX where participants indicated higher effort when reading text in the offset condition. After experiencing all conditions, 58% of participants clearly stated that the uniform background was most comfortable for reading, while the striped background in the offset condition was harder to read due to issues of double vision (75% of participants), and perceived motion of the background (25% of participants) while reading the text. The perceived motion can be explained by the eyes moving along the text during reading, leading to background variations for each eye. While we observed other trends in the data, we refrain from reporting these as a larger sample size or a more

complex task may be required to produce significant results. However, taking into account the preference data and user feedback, the study affirms perceptual issues due to the label offset from the background, similar to previous work from VR [38, 48]. Furthermore, not only aligning content to the background, but also ensuring background uniformity for both eyes appears to be a potential solution to such perceptual issues. These results emphasize the need for eye-perspective view management algorithms that consider the view of each eye for optimizing the view.

4 EYE-PERSPECTIVE VIEW MANAGEMENT

Overcoming the issues of double vision and text readability requires a scene analysis from the user's perspective. More than that, it requires the analysis of the scene as seen by each eye of the user which in turn requires first computing the view of the scene for each eye utilizing EPR. Once done, we must compute the eye-perspective view management from the EPR that gives us the optimal position of the label. In the following, we outline these two main steps.

4.1 Eye-Perspective Rendering

While EPR has not been proposed for OST HMDs, we benefit from the works presenting UPR in the context of handheld VST AR displays. Hence, in the first step, we implement several EPR approaches that are inspired by the current state-of-the-art technology in UPR. To increase the practical applicability of EPR to real-world use cases, we focused our exploration on algorithms that can be implemented on consumer-grade hardware without additional hardware modifications or excessive hardware requirements. Therefore, approaches that require extensive modifications of the HMD [25] or need a dedicated top-end graphic card just for computing the EPR for one eye [29] were not considered.

To facilitate the discussion of EPR, in the following, we define the general EPR system (Figure 6). An EPR system consists of an HMD worn by a user and positioned in the world coordinate system W. Rigidly attached to the HMD is an RGBD sensor that is also used for tracking. T_H is used to transform from world coordinates to the coordinate system of the RGBD sensor and, thus, the head coordinate system H. Our system synthesizes views of the real-world scene for the user's eyes. Hence, the scene must be transformed to the eye coordinate system E_i, where i can be *left* or *right* eye. We achieve this by using the transformation T_{E_i} for each eye to move from H to E_i. For view synthesis, we require input images I_H for EPR captured using the RGBD sensor. For our discussion, we distinguish between the color component of this image $I_{H,RGB}$ and the depth component of the image $I_{H,D}$. The synthesized images for E_i are defined as I_{E_i}.

Based on the UPR literature, we decided to utilize two general approaches to synthesize the user's view I_{E_i} through the display: a homography of the currently captured view $I_{H,RGB}$, and using 3D proxy geometry and visual information of one [47] or more [2] captured views $I_{H,RGBD}$. We implemented three EPR approaches (Figure 6): (1) a homography of the video $I_{H,RGB}$ based on an analysis of the scene that identifies the dominant plane, (2) view synthesis using proxy geometry based on the depth and reprojection of $I_{H,RGBD}$ into the user's view, and (3) view synthesis using the same proxy and IBR for view synthesis. As the semi-transparent display
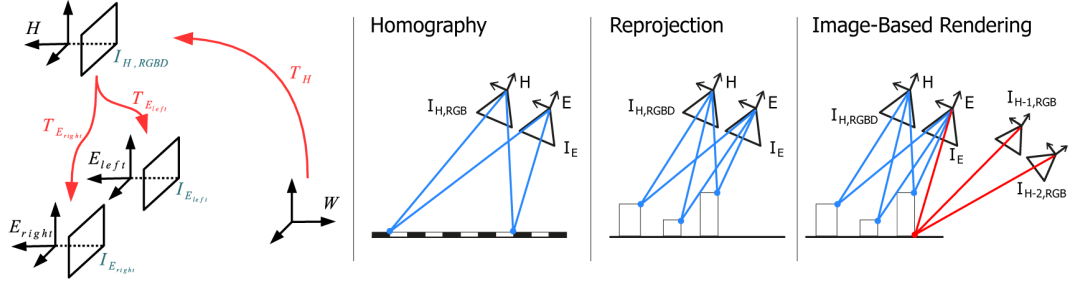
Figure 6: Coordinate systems, transformations and EPR methods. T_H transforms from world to head coordinates of an HMD RGBD camera providing color and depth information $I_{H,RGBD}$. T_{E_1} and T_{E_2} represent the transformation from the RGBD camera to the user's eyes. $I_{E_{left}}$ and $I_{E_{right}}$ are views of the real-world scene synthesized for the user's eyes. The sketches on the right illustrate the setup of the EPR methods homography sufficient for 2D scenes, reprojecton for 3D scenes and image-based rendering also for 3D scenes using recorded views to resolve disocclusions.

material of an OST HMD modifies the scene colors passing through the display [22], the final EPR of each method is modified by a constant factor that depends on the used HMDs. In the following, we discuss details of these approaches.

Homography. An EPR view synthesis can be achieved utilizing 3D information of sparse SLAM-based reconstructions and creating a homography of the captured image $I_{H,RGB}$ [54]. We create a homography of $I_{H,RGB}$ based on 3D points of the largest identified plane in the scene. To ensure temporal coherence, the homography input preferably considers a stable set of points that was visible over multiple frames and, thus has high confidence. An advantage of the homography is that it does not necessarily require a depth camera as SLAM features are sufficient for calculating the homography. A homography method works well in scenes containing mainly planar geometry at close distances, or with more complex scenes when viewed from a distance [5]. As the homography uses the current live video $I_{H,RGB}$, the method can represent changes to the scene structure when planarity or distance constraints are intact.

Reprojection. The reprojection method utilizes depth and color information of the input image $I_{H,RGBD}$ to synthesize the user's view of the real world. The live RGB image is projected onto the depth map and the textured geometry is transformed by T_E into the user's view. The reprojection method is suitable for structured 3D scenes but may suffer from disocclusion artifacts leading to missing depth and color information in the synthesized view I_E. As the method utilizes currently available scene information using I_{RGBD}, it can react to changes in the scene during the interaction. However, view-dependent visual changes cannot be modeled.

Image-based Rendering. The IBR method also transforms the available 3D proxy geometry into the user's view. As depth cameras may have issues with certain materials (e.g., Azure DK), or homogeneously textured areas (e.g., Stereo ZED Mini), a depth inpainting method is used to fill in small holes in the depth map before reprojection, and again after reprojection to fill in holes due to disocclusion artifacts. Our inpainting method is inspired by Schöps et al. [47] but uses an efficient push-pull approach for inpainting [16]. An IBR algorithm [6] fills in missing color information using images that are recorded while the user interacts with the scene. The IBR

method can be applied to complex 3D scenes and allows us to fill in missing information for the user's view with recorded scene information. View-dependent visual changes can also be taken into account when synthesizing views. However, when scene changes occur, the strategy recording $I_{H,RGB}$ views for the IBR must ensure that images are kept up to date.

4.2 View Management

Once we have computed the individual view for each eye using EPR we need to optimize the annotations placements in these views. To do so, we designed a view management method that adapts the layout based on the reliability of EPR information. Our view management is based on state-of-the-art hedgehog labeling [30, 52] that allows us to easily define criteria for view management and provides a temporally coherent layout.

Principle. Labels represent 3D objects and are placed in 3D world space. The layout algorithm projects 3D labels into the user's eyes using the determined eye position to determine a 2D projection of the label. This 2D projection is used to access the information calculated based on the EPR view for a user's eye in order to optimize the label position. The label position is adjusted using a vector field calculated from a set of optimization criteria based on the EPR views, adapting the layout based on information from both of the user's eyes. After the optimization is finished for the current view, the calculated positions are converted back to 3D space via unprojection. In line with Tatzgern et al. [52], the layout algorithm places labels close to their origin, and the label movement is constrained to 3D planes in the scene. Planes are placed at the depth of the label origin which typically corresponds to the 3D position of the annotated object. To ensure temporal coherence of labels, as suggested by Tatzgern et al. [52] and evaluated by Madsen et al. [30], a label layout is frozen after optimization and updated when users change their viewpoint of the scene.

Optimization Criteria. We implemented several criteria that allow us to optimize the label placement. The optimization is based on an analysis of the user's actual view through the HMD generated with EPR. We provide references for criteria that are derived from previous work.

EPR Views   Optimization Criteria for View Management   Final Layout
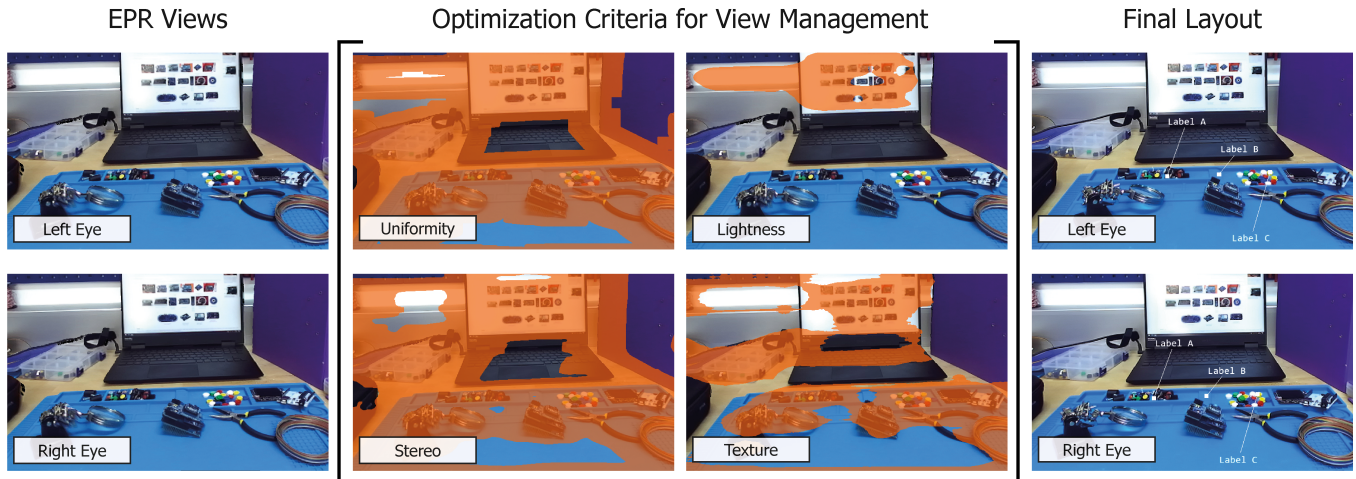

Figure 7: Eye-perspective view management. Our view management utilizes synthesized EPR views for left and right eyes of a user to optimize label placement and improve legibility. Our method relies on four optimization criteria to control label placement: (1) background color uniformity, (2) lightness contrast with respect to the background color, (3) text legibility based on a frequency analysis using Gabor filters and (4) stereo vision uniformity ensuring that the scene background is visually the same for left and right eye. Colored pixels are areas that have been excluded as valid placement regions for labels based on the respective criterion. By combining all criteria valid placement regions are filtered. A vector field moves labels towards the closest valid region.

- The *Background Uniformity* criterion causes labels to be placed in areas with a uniform background color so that a label has a homogenous background. The criterion calculates the standard deviation of the RGB values over the label area [40].

- The *Lightness Contrast* criterion causes labels to be placed in areas, where the text color has high contrast against the background color determined from the EPR view. The criterion evaluates the average lightness over the label area and aim for a recommended lightness difference of 27 [22, 61].

- To ensure legibility, the *Texture Contrast* criterion causes labels to be placed in areas with low-frequency backgrounds based on a Gabor filter analysis [15, 28]. For a candidate position, the criterion evaluates the percentage of textured pixels over the label area.

- To optimize for the different views of each eye, the *Stereo Vision Uniformity* criterion optimizes label placement towards areas where the color of the EPR view is similar between both eyes of the user. Similarity is calculated as the absolute difference between the colors of both views.

- Similarly, the *Disocclusion Errors* criterion prevents label placement in areas, where EPR provides no reliable information about the real-world background due to disocclusion errors. The criterion evaluates the percentage of disoccluded pixels over the label area.

Note that when looking for label placement candidates, our algorithm considers the entire area covered by a label instead of just individual pixels of the projected label position. Summed area tables are used to efficiently evaluate entire areas with respect to our criteria. For each criterion a threshold is used to remove candidate pixels that do not fulfill the criterion. See Figure 7 for a visualization of the results and an overview of the algorithm.

4.3 Implementation

All optimization criteria are realized by analyzing the EPR views of the user's eyes in order to identify good label placement candidates. Each pixel of the EPR view is a potential label position and represents the projected center of the label. Good placement candidates are pixels where the optimization criteria are met for all surrounding pixels that are covered by the respective label. Depending on its lightness, each label has its own set of placement candidates. To avoid overlapping labels the area covered by other labels is excluded from the set of possible label placement candidates. In our current implementation, we use a greedy strategy processing labels sequentially. After calculating the set of placement candidates for a label, we compute a vector field storing the direction towards the nearest candidate for each pixel. This vector field is used to guide labels towards positions that meet our optimization criteria.

Label placement candidates are initially computed in the EPR view of the dominant eye. However, to avoid stereo vision conflicts between the label content and the background, the Stereo Vision Uniformity criterion utilizes EPR views of both eyes. For this purpose, we project each candidate position of the dominant eye into the non-dominant eye. Then, we evaluate if the color of the corresponding candidate position is similar to the dominant eye. The algorithm only considers pixels of the EPR view of the dominant eye as good candidates where the absolute difference between the colors of the two EPR view pixels falls below a certain threshold. This straightforward stereo optimization strategy supplements the other optimization criteria that already take care of

Figure 8: System configurations. The setup used for evaluating EPR techniques and their impact in view management for OST HMDs. (Left) We use the Stereo ZED Mini to capture $I_{H,RGBD}$ representing a typical camera of an OST HMD. A second Stereo ZED Mini placed at the position of the user's eyes represents the user's view and captures ground truth images I_{EG_i} later used for comparison against the synthesized views I_{E_i}. (Right) The setup with the Project North Star OST HMD and one of the Stereo ZEDs attached directly to the HMD. The HMD was used in the second user study by study participants. When capturing views through the display, the device was attached to the capturing rig on the left where the HMD camera replaced one of the capturing rig.

identifying uniform and untextured areas for label placement, and, thus, refines the label position to avoid stereo conflicts.

5 EVALUATION

In the following, we provide a detailed evaluation of our novel eye-perspective view management. We initially describe the results from a technical evaluation of the three EPR methods. The results of the technical evaluation inform our choice of EPR method for a user study investigating the effect of eye-perspective rendering on label layout quality, contrast, legibility, and perceived double vision of the background when reading labels.

5.1 Eye Perspective Rendering

We performed an evaluation to explore the feasibility of the three EPR methods as well as their limitations for our eye-perspective view management solution.

Apparatus. We created a dedicated capturing rig (Figure 8 (Left)) consisting of two Stereo ZED Mini cameras. The top camera represented the built-in HMD camera capturing images $I_{H,RGBD}$ and providing tracking poses using Stereo ZED's native tracking, while the bottom camera captured data representing the user's eyes E_i providing the ground truth I_{EG} for the synthesized EPR views I_E. The spatial arrangement of the cameras represented a typical HMD configuration. We based this configuration on the North Star HMD[1] that we also used in the subsequent user study. We calibrated the intrinsic and extrinsic camera using multi-camera calibration[2]. The setup is connected to a standard PC. The implemented EPR methods run in real-time on an Intel Core i9-10900K, with 3.7 GHz, and an

NVIDIA GeForce RTX 2080 Ti, 11GB. Note that similar approaches have been shown to also run on mobile devices [47, 54].

Scenes. We used two AR scenes in our analysis, which vary in structural complexity in order to explore the methods for various target scenarios Figure 9. We applied our approach to a simple scene consisting mainly of 2D planes (2D scene) and a more complex scene containing depth disparities (3D scene).

Data Collection. Analogously to Figure 3, we compared EPR views generated with the three render methods against a ground truth captured by a stereo camera representing the user's eyes by calculating per-pixel quality differences of an optimization criterion utilizing ground truth or EPR information. For this evaluation, we utilized the quality criterion of background uniformity. Exemplary results are shown in Figure 9.

Discussion. As expected, the homography-based method provided reliable information only for mostly planar scenes for which an accurate plane can be detected. Naturally, any 3D geometry is distorted as seen in the color difference images leading to color and quality mismatches compared to the ground truth. A clear advantage of homography methods is that they can be realized with limited hardware resources as depth cameras are not necessarily required. While the planarity requirement is a clear limitation of the technique, homographies are feasible for use cases where content is applied to planar surfaces such as walls, e.g., when placing virtual windows, or also for distant scenes.

The reprojection method using only RGBD information of the user's current viewpoint generally provided reliable information for view management for 2D and 3D scenes. Note that while we used a stereo-based RGBD camera in our setup, we relied only on the RGB image of one camera as typical HMDs contain only one color camera. Missing information after reprojecting data into the user's view can be clearly identified and handled by the view management algorithm (see cyan pixels). Alternatively, an IBR-based inpainting strategy can be used to fill in missing information.

We noticed that the IBR-based method suffered from ghosting artifacts, which is a common issue of real-time IBR methods and has also been encountered by previous work [11]. These ghosting artifacts come from imprecise geometric proxies due to inaccuracies in the depth map as well as pose inaccuracies due to the incremental nature of SLAM pose tracking and its continuous pose optimizations (e.g., loop closure). To solve these issues, previous work, for instance, utilized precise poses and geometry from dedicated reconstruction algorithms that do not work in real-time [32] or utilized computationally complex algorithms that cannot be deployed to a mobile platform [58].

To strike a balance between reconstruction quality, flexibility regarding the complexity of scene structures, computational complexity, as well as performance, we decided to base our eye-perspective view management algorithm on the reprojection EPR method. This straightforward method allowed us to generate views of even complex 3D structures in real-time. However, a view management algorithm has to consider missing information due to disocclusions.

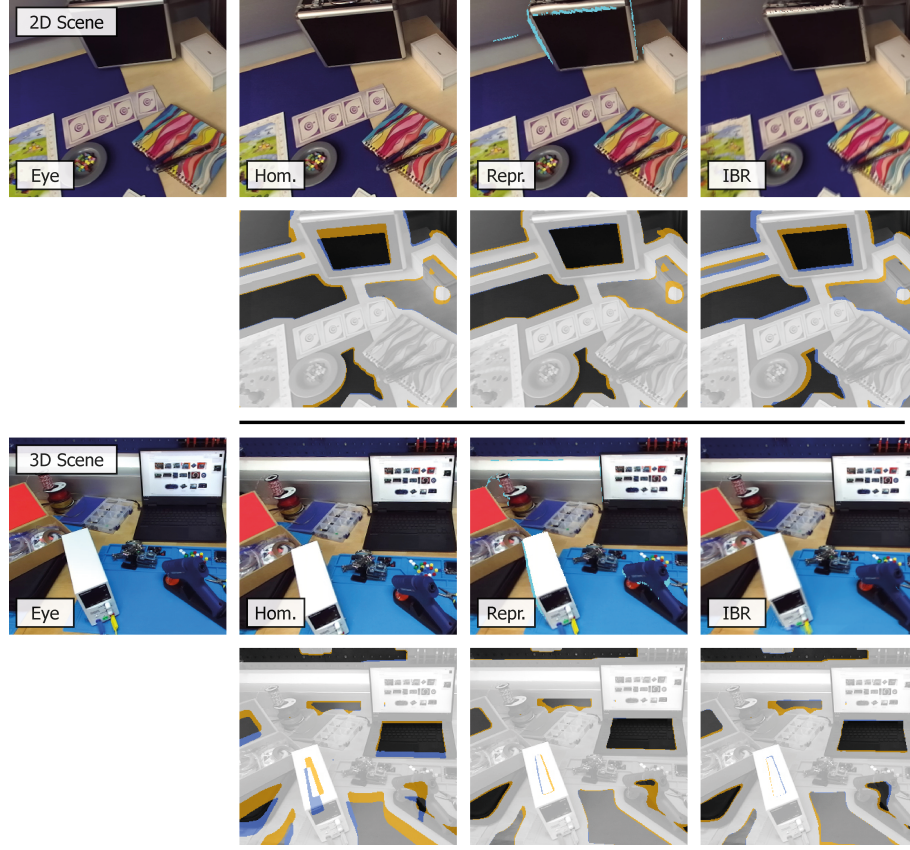
[1]https://developer.leapmotion.com/northstar
[2]https://calib.io

Figure 9: Comparison. We explore the quality of EPR methods by comparing their synthesized views against the ground truth eye-perspective view for (Top) a mainly planar 2D scene and (Bottom) a more complex 3D scene. Annotations: *Eye* for the ground truth, *Hom.* for homography, *Repr.* for the reprojection method, *IBR* for the IBR method. The first row for each scene shows color images of the ground truth, and compared EPR methods. Cyan pixels denote missing information, e.g,. due to disocclusions. The second row visualizes the results of the uniformity optimization criterion calculated on the EPR view and the ground truth. Yellow pixels indicate regions that are valid in the ground truth but have not been detected in the EPR view. Blue indicates pixels are valid pixels in the EPR view but not in the ground truth. White pixels indicate matching valid regions in both views. The results show the limitations of the homography method for non-planar scenes, while IBR matches the ground truth well. However, erroneous poses for the IBR method often lead to ghosting artifacts and thus to an erroneous optimization. The reprojection method is robust, but optimization algorithms must consider missing information due to disocclusions.

5.2 View Management

We designed a within-subject study to evaluate eye-perspective view management. We were interested if the view management would be able to successfully improve contrast, legibility, and avoid interferences with the background when reading labels. The study would also show, if sufficiently precise EPR can be realized for individual users wearing an HMD and changing their viewpoint.

Our independent variable was *view management* with three conditions (Figure 10): (1) a *baseline* condition, where no view management was used to adjust label positions, (2) a condition where view management was based on the information captured by a built-in camera as commonly present in HMDs, and (3) a condition utilizing eye-perspective view management based on *EPR* to optimize label layouts based on the computed user's actual view.

Previous work has evaluated the influence of text contrast and legibility on participants' performance [8, 13]. Therefore, our study focuses on evaluating the ability of eye-perspective view management to enforce the optimization criteria that will result in better label layouts for the user's view through an HMD in terms of contrast, legibility, and avoidance of interferences with the background. The study also explored shortcomings of view management relying only on a built-in HMD camera.

Participants. We recruited 12 participants (3 female, $\overline{X}$=32.3 (7.7) years). On a scale from one to five (best), the mean of self-rated AR experience was 2.1 (sd=1.4, median=1.5). All participants had normal or corrected-to-normal vision and were not affected by color vision deficiency. We determined the dominant eye of participants with the Miles test [33] as input to the view management algorithm.
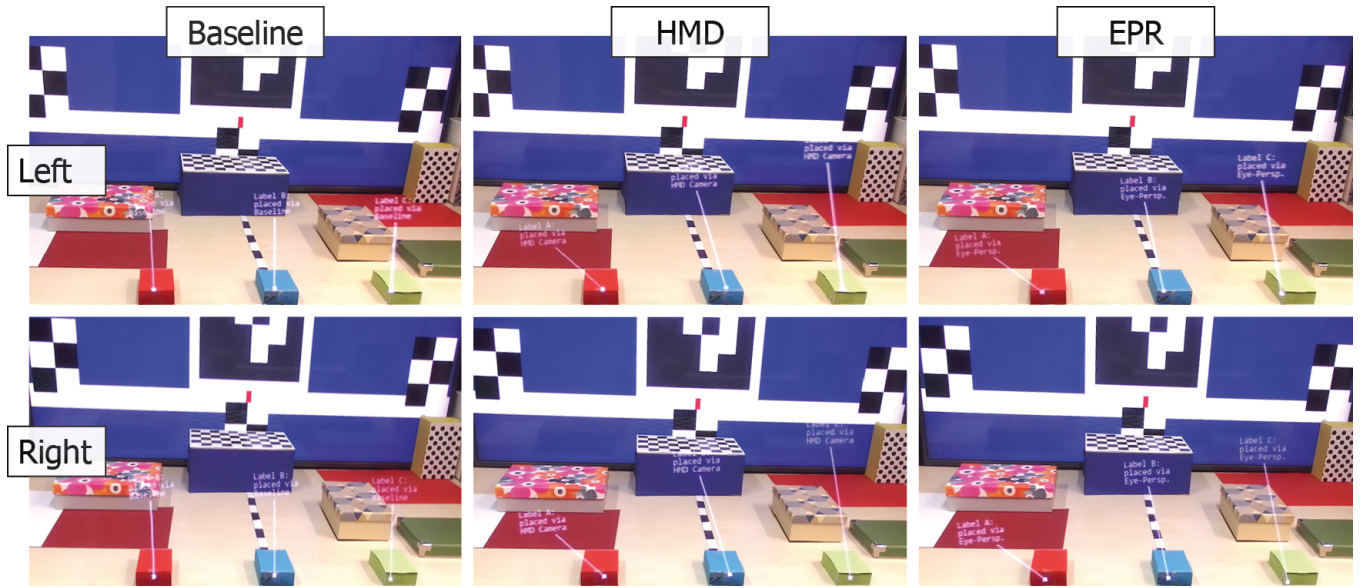
Figure 10: Conditions. The images show left and right views through the HMD displays. The viewing rays of the cameras capturing these images through the HMD were always parallel. (Baseline) In the baseline without view management, the labels are placed offset from their origin at a fixed distance, which leads to frequent interference with the background. (HMD) View management utilizing the view of the HMD camera optimizes the layout based on the camera view. Due to the offset between the user's eye and the camera, the optimization criteria are not always fulfilled, leading to interference with the background. (EPR) Eye-perspective View Management utilizes the synthesized EPR view, which ensures that optimization criteria are fulfilled for both eyes .

Apparatus. Participants performed the experiment in a seated position in front of a table standing in front of a wall. They were seated on a swivel chair that could easily move to change viewpoints during the tasks. The height of the chair was adjusted depending on the height of the users. The scene presented to the participants consisted of various color and texture patches. The back wall of the setup consisted of a 52 inch screen that was used for the initial instructions and example for participants, single-point active alignment method (SPAAM) calibration [55], and background scene elements. When designing the real-world scene for the user study, we took care to use simple geometric shapes in order to improve replicability of the experiment. Earlier studies showed that lightness contrast and not color contrast is the driving issue of perceptual conflicts with the background [61] and thus our labels were all light gray (RGB color: #B9B9B9, lightness 75). Three labels were placed in the scene. The baseline condition did not change the initial label position, while in the other two conditions, the labels were optimized by the respective view management method.

For our study, participants wore a Project North Star HMD with an attached Stereo ZED mini for tracking and RGBD depth data (Figure 8 (Right)). The HMD and cameras were connected to a PC with Intel Core i9-10900K, 3.7 GHz, and an NVIDIA GeForce RTX 2080 Ti, 11GB. To replicate the semi-transparency of the HMD displays affecting the perceived scene colors, the colors of the EPR and HMD view for the respective conditions were modified by subtracting a constant value of RGB #1E0023 that was determined by minimizing the color difference between both views. The environmental

light was set to 628.3 lux at a color temperature of 5604K using a LUPO Superpanel Dual Color 60 and measured via a Mavospec Base Spectrometer. The offset between RGBD camera and the user's eyes was calibrated with SPAAM [19, 55] where participants had to align nine points at two different distances (approx. 70cm and 1m). The quality of the calibration was verified by the participants by overlaying the rendered EPR views using the HMD displays for each eye separately. Participants calibrated until the views were aligned. Furthermore, participants were shown a test label on the HMD at the same distance as the actual labels and asked to read it using both eyes, to ensure a good quality stereo calibration.

Task. Users had to align their viewpoint with a predefined position and orientation and evaluate the quality of the label placement with respect to the optimization criteria of the view management algorithm described in section 4.2. The forced viewpoint changes lead to updates of the layout to reflect the novel viewpoint. Once participants had aligned their viewpoints with the predefined viewpoint, participants were instructed to avoid changing the viewpoint to prevent further changes in the background. Furthermore, during analysis, each label was frozen in place to avoid unwanted layout updates due to involuntary small head motions.

Data Collection. For each label, participants rated the overall legibility and interference of the background using a 5-point rating scale, where 1 was the lowest rating. Participants also analyzed the layout of each label with respect to the utilized optimization criteria:

	EPR	HMD	Baseline
Uniformity	99.7 (1.4), 100	62.4 (33.2), 70	53.9 (34.7), 55
Lightness	99.8 (1.3), 100	82.1 (28.4), 100	89.4 (21.6), 100
Texture	100 (0.5), 100	84.8 (28.9), 100	74.4 (30.8), 87.5
Stereo	99.4 (2.8), 100	82.8 (16.2), 80	84.2 (14.4), 90
Overall Qual.	99.7 (1.4), 100	78.0 (19.0), 84.4	75.5 (18.0), 76.3
Legibility	4.6 (0.7), 5	3.2 (1.4), 3	2.9 (1.5), 3
Background	1.1 (0.2), 1	2.9 (1.6), 3	3.2 (1.5), 3
SEQ	6.6 (0.6), 7	5.4 (1.2), 6	5.1 (1.5), 6
Mental Eff.	1.3 (0.8), 1	2.3 (1.3), 2	2.5 (1.5), 2

Table 1: Descriptive statistics. Data formatted as mean (standard deviation), median. Median is calculated over all labels.

(1) background uniformity, (2) lightness contrast, (3) texture contrast, and (4) stereo uniformity. For the analysis, we calculated the overall label quality for each label as the average of the four quality criteria. Participants also rated the legibility of labels (1 to 5, 5 best) and the negative influence of the background on legibility (1 to 5, 5 worst). After rating all labels for a viewpoint and view management condition, participants rated task difficulty using the Single Ease Question (SEQ) and mental effort using the Paas scale [41]. After finishing all three view management conditions for a viewpoint, participants were asked to rank the three strategies according to perceived layout quality. Participants rated the optimization criteria by estimating percentages in steps of 10 where higher percentages meant better quality. Background uniformity was rated by estimating the percentage occupied by the largest uniform area behind the label. Lightness contrast and texture contrast were rated by estimating the percentage of the label background that was not too bright or textured, respectively. Stereo uniformity was evaluated by estimating the percentage of uniform background between both eyes. This analysis was performed for both eyes separately. To avoid unnecessarily extending the user study, participants only performed this analysis for the dominant eye, when the background turned out to be uniform for both eyes.

Procedure. Participants were recruited via university mailing lists and performed the study in a room on campus. After filling out an informed consent form and demographics questionnaire, participants were given a short introduction to the study. The introduction included a tutorial on SPAAM calibration to train participants to perform the procedure by themselves. To give participants an impression of various SPAAM calibration qualities, the experimenter showed them overlays filmed through the HMD that were recorded before the experiment. To make sure, participants understood the task and the involved analysis, they were also shown exemplary labels filmed through the HMD. Participants were then seated, put on the HMD, and calibrated their eye positions using SPAAM. Afterwards, the first view management method placed labels in the scene and participants assumed the first viewpoint. The experimenter queried the participants for each label and noted their answers. After rating all three labels for a viewpoint, the next view management condition was shown. After all view management conditions, participants had to rank the conditions from best to worst based on the layout quality. Afterward, participants assumed the next viewpoint and the procedure was repeated. In total, participants

had to assume three different viewpoints leading to various backgrounds. An experiment lasted for approximately 90 minutes. The order of conditions and view positions was balanced using an orthogonal Latin Square table. We calculated the mean of each rating for each condition for each participant over all the viewpoints. With 3 (conditions) x 3 (labels) x 3 (viewpoints) = 27 repetitions for each participant and task, for 12 participants there were 324 trials.

Hypotheses. We expected that (**H1**) world-camera view management and eye-perspective view management outperform the baseline condition (no view management) because even though optimizations that rely on the integrated HMD camera are not precise, labels are generally pushed towards more suitable areas. We expected that (**H2**) eye-perspective view management will outperform HMD camera view management because the EPR view replicates the user's view and places labels in areas that the view management identified as suitable due to the optimization.

Results. We analyzed the statistics as described in section 3. Results are summarized in Table 1 and Table 2, box plots and ranking data are visualized in Figure 11. Participants ranked eye-perspective view management first in 97% of all cases, participants ranked HMD-based view management second in 69% of all cases, and baseline third in 69% of all cases. In 16% of all cases, participants could not decide to rank HMD or baseline method second or third. However, they ranked eye-perspective view management in the first place.

5.3 Discussion

In the following, we provide a summary of our learnings that can guide the use of eye-perspective view management algorithms.

Eye-Perspective View Management Improves Layouts. Our user study clearly showed that eye-perspective view management using EPR can optimize label layouts for individual users wearing an OST HMD. Based on the statistically significant results of the user study, we partially accept H1 as eye-perspective view management outperformed the baseline layout. In addition, we accept H2 as eye-perspective view management outperformed HMD camera view management in all measured dependent variables. Overall, the average label layout quality, judged by participants was close to 100% for the EPR layout, compared to 78% for the HMD camera layout, and 75.5% for the baseline naive layout. As the view management took care of finding homogeneous areas for placing labels, the EPR-based label layout leads to significantly better legibility of

	Friedman	EPR - Baseline	EPR - HMD	HMD - Baseline
Uniformity	$\chi^2(2)=19.5$,p=.00005	Z=3.06,p=.004,r=.63	Z=3.06,p=.007,r=.63	Z=1.96,p=.05,r=0.4
Lightness	$\chi^2(2)=18.7$,p=.00008	Z=3.03,p=.005,r=.62	Z=3.06,p=.007,r=.63	Z=2.16,p=.03,r=0.44
Texture	$\chi^2(2)=21.8$,p=.00002	Z=3.06,p=.007,r=.63	Z=3.06,p=.004,r=.63	Z=2.87,p=.004,r=0.59
Stereo	$\chi^2(2)=18.4$,p=.0001	Z=3.06,p=.007,r=.63	Z=3.06,p=.004,r=.63	
Overall Qual.	$\chi^2(2)=18.2$,p=.0001	Z=3.06,p=.004,r=.63	Z=3.06,p=.007,r=.63	
Legibility	$\chi^2(2)=18.8$,p=.00008	Z=3.06,p=.004,r=.63	Z=3.06,p=.007,r=.63	
Background	$\chi^2(2)=18.7$,p=.00009	Z=3.06,p=.004,r=.63	Z=3.06,p=.007,r=.63	
SEQ	$\chi^2(2)=16.8$,p=.0002	Z=3.06,p=.007,r=.63	Z=3.0,p=.006,r=.61	
Mental Eff.	$\chi^2(2)=15.7$,p=.0004	Z=2.96,p=.009,r=.6	Z=2.86,p=.009,r=.58	

Table 2: Statistical analysis. Statistically significant results for the collected data analysed with Friedman test and Wilcoxon signed-rank tests for pairwise comparisons, including effect sizes. Pairwise comparisons are Bonferroni-Holm corrected.
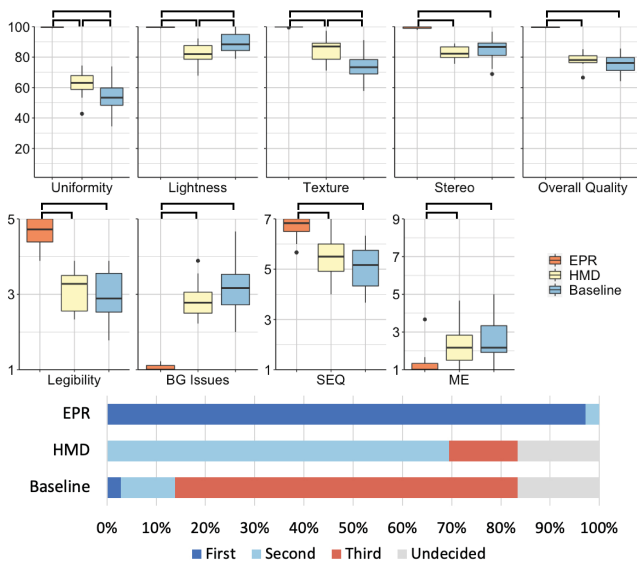

Figure 11: Data eye-perspective view management study. (Top) Answers about label quality: uniformity, lightness contrast, texture contrast, stereo uniformity, and overall label quality. (Middle) Answers regarding label legibility, issues with the background, task difficulty (SEQ), and mental effort (ME). (Bottom) Ranking of view management conditions. Participants could not decide between second and third place for HMD or baseline condition in 16% of all cases.

labels and fewer distractions caused by the background. In 97% of all cases, participants preferred the EPR label layout, emphasizing the impact of layout optimizations from the user's eye perspective.

Camera-Based Optimization Still Beneficial. We only partially accept H1 in terms of the HMD layout outperforming the baseline layout. The HMD layout outperformed the baseline layout in terms of uniformity and texture contrast. We did not find a significant difference in stereo quality for HMD and baseline, which is not surprising as both do not optimize the layout taking into account both eyes of the user. Interestingly, the HMD layout did not outperform the baseline in terms of legibility, background influence,

or overall label quality. The baseline layout was even significantly better in terms of lightness contrast than the HMD layout. This may indicate that layouts based on wrong assumptions of the user's view (e.g., HMD camera view), do not offer benefits compared to a very basic layout that is not optimized. However, more likely, this result can be explained by labels being placed in areas of the scene exhibiting different lightness characteristics. This placement may have adversely affected the overall label quality of the HMD view.

When exploring the lightness contrast issue, we noticed that during the study both HMD and baseline conditions placed labels in low lightness contrast areas. However, the HMD labels were overlapping with areas of comparably higher lightness. Due to the additive nature of the OST HMD display, these areas made it likely impossible for participants to read the labels, thereby, influencing the lightness contrast estimates. To explore this issue, we removed all label data that participants evaluated as having lightness contrast issues (90 of 324 labels), i.e., we filtered the data and kept only labels that were estimated as 100% good lightness contrast labels. While eye-perspective view management still outperformed both HMD and baseline layout, the HMD layout quality was significantly better than the baseline layout quality (HMD: 84.8 (17.6), 82.5; baseline: 78.3 (18.2), 81.25), than the baseline legibility (HMD: 3.6 (1.4), 4; baseline: 3.1 (1.6), 3) and than the baseline background influence (HMD: 2.4 (1.6), 2; baseline: 3.0 (1.6), 3). Note that we refrain from reporting results from statistical testing, as this analysis was not part of the initial hypotheses and should be investigated further.

In summary, optimization relying on the HMD camera may still be more beneficial than having no layout optimization in place. This is backed up by preference data, as the majority of participants (69% of all viewpoints) ranked HMD layout in second place and the baseline layout in third place (69% of all viewpoints).

EPR Replicates the User's Views. Our main study showed that EPR can replicate the user's view through an OST HMD with sufficient precision so that eye-perspective optimization algorithms such as view management can adapt label layouts to improve legibility. More importantly, unlike previous studies utilizing OST displays [14, 27], participants were not forced to use a headrest but were allowed to change their view during the user study. Hence, we demonstrated that EPR can be utilized in mobile real-world scenarios which enables further exploration of effective view management for real-world use cases, such as a localized assembly task, or when

users walk through a scene. EPR can also be applied beyond view management for other use cases for OST HMDs relying on real-world background information, such as content-aware interaction methods [39], or supporting people with visual impairments [34].

Choice of EPR Algorithms. Based on our analysis of EPR algorithms, we decided to utilize the rather straightforward reprojection method as input to the eye-perspective view management. The reprojection method has the advantage that due to the low computational demand, it can run on mobile hardware and utilizes less battery power compared to more elaborate methods [58]. Furthermore, the reprojection method utilizes only the information from the current viewpoint of users and, therefore, can also react to scene changes. Therefore, the reprojection method is generally a good choice to explore the impact of EPR for various OST HMD use cases. However, missing information due to disocclusion artifacts may be a strong limiting factor of this method under certain conditions. Note that EPR is most useful for scenarios where a user is close to scene geometry. EPR may not be necessary when users view distant scenes as the parallax between camera and the user's eyes does not introduce noticeable differences in both views [5].

Disocclusion Artifacts. Disocclusion artifacts cause missing information in the synthesized eye-perspective view due to the offset between the user's eyes and the camera providing color information of the scene. Disocclusion artifacts occur at the borders of 3D scene geometry and grow larger the closer the view is to the scene [5]. While our eye-perspective view management can handle disocclusion artifacts, it may fail to find valid placement areas for larger disocclusions when 3D geometry is too close to the user, e.g., when users pick up the geometry for closer inspection using their hands. Hence, advanced inpainting methods may be required to reconstruct the occluded scene [36, 37, 58], or algorithms may rely on alternative optimization methods.

Alternative Optimizations. When reliable scene information is not available, e.g., due to disocclusion artifacts, or there is an insufficient number of valid locations, view management algorithms can utilize alternative forms of optimizations. Furthermore, while freezing label layouts [30, 52] may be an appropriate temporal coherence strategy for localized use cases, alternative strategies must be developed for scenarios where the real-world background constantly changes, e.g., while walking through a scene. Previous work has demonstrated that adapting the representation of a label using a uniform billboard color as label background can improve legibility, compensating for background interferences [8, 14]. However, such adaptations change the label design and may alter the meaning of information when using distinct color codes to communicate information such as security-critical details. Alternatively, label positions can be modified by sticking them to real-world scene geometry so that labels are at the same depth as the real-world background, thereby avoiding labels positioned in midair [40]. Such an approach also improves temporal coherence, as the label background will not change during viewpoint changes, and allows an optimization algorithm to utilize the view of the HMD camera as a reliable source of color information because the label background will be the same from the HMD camera's view and the user's eye

view. However, sticking labels to real-world geometry in the background but annotating foreground objects will lead to a perceptual conflict as users have to switch between the depth of the annotated foreground object and the label positioned in the background.

Optimization as Loss Function. In our user evaluation, we did not allow the eye-perspective view management tolerances, which forced labels to be placed in locations where all the constraints are fulfilled. Hence, we compared an optimal solution for label placement layout to the potentially worst outcome using imprecise scene information. HMD-camera-based layout and baseline layout inadvertently lead to layouts, where label quality deteriorated on average to below 80%. However, such a strict interpretation of the optimization criteria may lead to layouts where labels are placed at a large distance from their origin, or to situations where the scene structure does not allow the algorithm to find a layout solution. Instead of strictly enforcing the criteria, they can also be integrated into a loss function for a more lenient optimization that always finds a solution for label placement. Future work should investigate which criteria can be violated to which degree in order to achieve a layout with legible labels. This information can then be utilized to identify weights for the optimization criteria.

6 CONCLUSION

In this paper, we discussed the often overlooked issues that arise from a lack of knowledge of the user's view of the physical scene when using an AR OST HMD. In particular, we demonstrated the advantage of eye-perspective view management in AR to achieve optimal label placements instead of relying on image information from an HMD-integrated camera that does not match the user's view. We have demonstrated that reliable EPR for OST HMDs can come without a large hardware and software overhead and using currently available HMDs. Our approach provides new opportunities for further research on algorithms and methods relying on perceptual information of the user's view through the HMD in real-world conditions and is directly opening up two directions for future work: Firstly, the exploration of fully mobile real-life scenarios such as assembly tasks. This might require adding support for more commonly used OST HMDs such as the Microsoft Hololens 2. Real-world scenarios will also bring additional challenges such as an increased level of user activity as well as temporal coherence strategies for label placement that can be applied to scenarios with frequently changing backgrounds such as walking through a scene. Secondly, future work should consider other areas in AR requiring knowledge of the user's true view. Examples include color harmonization when colors are matched to the real-world background [18], advanced interaction methods relying on a scene analysis [39], or supporting people with visual impairments by emphasizing scene structures [34, 50].

REFERENCES

[1] Domagoj Baričević, Tobias Höllerer, Pradeep Sen, and Matthew Turk. 2014. User-perspective augmented reality magic lens from gradients. In *ACM Symposium on Virtual Reality Software and Technology*. ACM, 87–96. https://doi.org/10.1145/2671015.2671027

[2] Domagoj Baričević, Tobias Höllerer, Pradeep Sen, and Matthew Turk. 2017. User-Perspective AR Magic Lens from Gradient-Based IBR and Semi-Dense Stereo. *IEEE Transactions on Visualization and Computer Graphics* 23, 7 (2017), 1838–1851. https://doi.org/10.1109/TVCG.2016.2559483

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

[4] Blaine Bell, Steven Feiner, and Tobias Höllerer. 2001. View management for virtual and augmented reality. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*. ACM, Orlando, Florida, 101–110. https://doi.org/10.1145/502348.502363

[5] Ricardo Augusto Borsoi and Guilherme Holsbach Costa. 2018. On the Performance and Implementation of Parallax Free Video See-Through Displays. *IEEE Transactions on Visualization and Computer Graphics* 24, 6 (2018), 2011–2022. https://doi.org/10.1109/TVCG.2017.2705184

[6] Chris Buehler, Michael Bosse, Leonard McMillan, Steven Gortler, and Michael Cohen. 2001. Unstructured lumigraph rendering. *Proc. of the 28th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 2001* (2001), 425–432. https://doi.org/10.1145/383259.383309

[7] Gaurav Chaurasia, Arthur Nieuwoudt, Alexandru-Eugen Ichim, Richard Szeliski, and Alexander Sorkine-Hornung. 2020. Passthrough+: Real-Time Stereoscopic View Synthesis for Mobile Mixed Reality. *Proc. ACM Comput. Graph. Interact. Tech.* 3, 1, Article 7 (apr 2020), 17 pages. https://doi.org/10.1145/3384540

[8] Saverio Debernardis, Michele Fiorentino, Michele Gattullo, Giuseppe Monno, and Antonio Emmanuele Uva. 2014. Text Readability in Head-Worn Displays: Color and Style Optimization in Video versus Optical See-Through Devices. *IEEE Transactions on Visualization and Computer Graphics* 20, 1 (2014), 125–139. https://doi.org/10.1109/TVCG.2013.86

[9] John J. Dudley, Jason T. Jacques, and Per Ola Kristensson. 2021. Crowdsourcing Design Guidance for Contextual Adaptation of Text Content in Augmented Reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 731, 14 pages. https://doi.org/10.1145/3411764.3445493

[10] Gerlinde Emsenhuber, Michael Domhardt, Tobias Langlotz, Denis Kalkofen, and Markus Tatzgern. 2022. 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)*. 640–641. https://doi.org/10.1109/VRW55335.2022.00171

[11] Okan Erat, Markus Hoell, Karl Haubenwallner, Christian Pirchheim, and Dieter Schmalstieg. 2019. Real-Time View Planning for Unstructured Lumigraph Modeling. *IEEE Trans. on Visualization and Computer Graphics* 25, 11 (2019), 3063–3072. https://doi.org/10.1109/TVCG.2019.2932237

[12] Catherine O. Fritz, Peter E. Morris, and Jennifer J. Richler. 2012. Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General* 141, 1 (2012), 2–18. https://doi.org/10.1037/a0024338

[13] Joseph L. Gabbard, J. Edward Swan, and Deborah Hix. 2006. The Effects of Text Drawing Styles, Background Textures, and Natural Lighting on Text Legibility in Outdoor Augmented Reality. *Presence* 15, 1 (2006), 16–32. https://doi.org/10.1162/pres.2006.15.1.16

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

[15] Michele Gattullo, Antonio Emmanuele Uva, Dipartimento Meccanica, Politecnico Bari, Politecnico Bari, and Dipartimento Meccanica. 2015. Predicting Text Legibility over Textured Digital Backgrounds for a Monocular Optical See-Through Display. *Presence* 26, 1 (2015), 1–15. https://doi.org/10.1162/PRES

[16] Steven J. Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F. Cohen. 1996. The lumigraph. *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1996* (1996), 43–54. https://doi.org/10.1145/237170.237200

[17] Raphaël Grasset, Tobias Langlotz, Denis Kalkofen, Markus Tatzgern, and Dieter Schmalstieg. 2012. Image-driven view management for augmented reality browsers. In *2012 IEEE Int. Symp. on Mixed and Augmented Reality (ISMAR)*. 177–186. https://doi.org/10.1109/ISMAR.2012.6402555

[18] Lukas Gruber, Denis Kalkofen, and Dieter Schmalstieg. 2010. Color harmonization for Augmented Reality. In *IEEE International Symposium on Mixed and Augmented Reality*. IEEE, 227–228. https://doi.org/10.1109/ISMAR.2010.5643580

[19] Jens Grubert, Yuta Itoh, Kenneth Moser, and J. Edward Swan. 2018. A Survey of Calibration Methods for Optical See-Through Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics* 24, 9 (2018), 2649–2662. https://doi.org/10.1109/TVCG.2017.2754257 arXiv:1709.04299

[20] Steven Henderson and Steven Feiner. 2011. Exploring the benefits of augmented reality documentation for maintenance and repair. *IEEE Transactions on Visualization and Computer Graphics* 17, 10 (2011), 1355–1368. https://doi.org/10.1109/TVCG.2010.245

[21] Alex Hill, Jacob Schiefer, Jeff Wilson, Brian Davidson, Maribeth Gandy, and Blair MacIntyre. 2011. Virtual transparency: Introducing parallax view into video see-through AR. In *IEEE International Symposium on Mixed and Augmented Reality*. 239–240. https://doi.org/10.1109/ISMAR.2011.6092395

[22] Juan David Hincapié-Ramos, Levko Ivanchuk, Srikanth K. Sridharan, and Pourang P. Irani. 2015. SmartColor: Real-Time Color and Contrast Correction for Optical See-Through Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics* 21, 12 (2015), 1336–1348. https://doi.org/10.1109/TVCG.2015.2450745

[23] Yuta Itoh, Maksym Dzitsiuk, Toshiyuki Amano, and Gudrun Klinker. 2015. Semi-Parametric Color Reproduction Method for Optical See-Through Head-Mounted Displays. *IEEE Trans. on Visualization and Computer Graphics* 21, 11 (2015), 1269–1278. https://doi.org/10.1109/TVCG.2015.2459892

[24] Yuta Itoh, Tobias Langlotz, Jonathan Sutton, and Alexander Plopski. 2021. Towards Indistinguishable Augmented Reality: A Survey on Optical See-through Head-Mounted Displays. *ACM Comput. Surv.* 54, 6, Article 120 (July 2021), 36 pages. https://doi.org/10.1145/3453157

[25] Tobias Langlotz, Matthew Cook, and Holger Regenbrecht. 2016. Real-time radiometric compensation for optical see-through head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics* 22, 11 (2016), 2385–2394. https://doi.org/10.1109/TVCG.2016.2593781

[26] Tobias Langlotz, Thanh Nguyen, Dieter Schmalstieg, and Raphael Grasset. 2014. Next-Generation Augmented Reality Browsers: Rich, Seamless, and Adaptive. *Proc. IEEE* 102, 2 (2014), 155–169. https://doi.org/10.1109/JPROC.2013.2294255

[27] Tobias Langlotz, Jonathan Sutton, Stefanie Zollmann, Yuta Itoh, and Holger Regenbrecht. 2018. ChromaGlasses: Computational glasses for compensating colour blindness. In *Conf. on Human Factors in Computing Systems*, Vol. 2018-April. 1–12. https://doi.org/10.1145/3173574.3173964

[28] A. Leykin and M. Tuceryan. 2004. Automatic determination of text readability over textured backgrounds for augmented reality systems. In *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*. 224–230. https://doi.org/10.1109/ISMAR.2004.22

[29] L. Lipson, Z. Teed, and J. Deng. 2021. RAFT-Stereo: Multilevel Recurrent Field Transforms for Stereo Matching. In *2021 International Conference on 3D Vision (3DV)*. IEEE Computer Society, Los Alamitos, CA, USA, 218–227. https://doi.org/10.1109/3DV53792.2021.00032

[30] Jacob Boesen Madsen, Markus Tatzqern, Claus B. Madsen, Dieter Schmalstieg, and Denis Kalkofen. 2016. Temporal Coherence Strategies for Augmented Reality Labeling. *IEEE Trans. on Vis. and Computer Graphics* 22, 4 (2016), 1415–1423. https://doi.org/10.1109/TVCG.2016.2518318

[31] Yuki Matsuda, Fumihisa Shibata, Asako Kimura, and Hideyuki Tamura. 2013. Poster: Creating a user-specific perspective view for mobile mixed reality systems on smartphones. In *IEEE Symposium on 3D User Interface*. IEEE, 157–158. https://doi.org/10.1109/3DUI.2013.6550226

[32] Ben Mildenhall, Pratul P. Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. 2019. Local Light Field Fusion: Practical View Synthesis with Prescriptive Sampling Guidelines. *ACM Trans. Graph.* 38, 4, Article 29 (jul 2019), 14 pages. https://doi.org/10.1145/3306346.3322980

[33] Walter R. Miles. 1930. Ocular dominance in human adults. *Journal of General Psychology* 3 (1930), 412–430. Issue 3. https://doi.org/10.1080/00221309.1930.9918218

[34] Hein Min Htike, Tom H. Margrain, Yu-Kun Lai, and Parisa Eslambolchilar. 2021. 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* (Yokohama, Japan) *(CHI '21)*. New York, NY, USA, Article 729, 15 pages. https://doi.org/10.1145/3411764.3445327

[35] P. Mohr, M. Tatzgern, J. Grubert, D. Schmalstieg, and D. Kalkofen. 2017. Adaptive user perspective rendering for Handheld Augmented Reality. In *2017 IEEE Symposium on 3D User Interfaces, 3DUI 2017 - Proceedings*. https://doi.org/10.1109/3DUI.2017.7893336

[36] Shohei Mori, Okan Erat, Wolfgang Broll, Hideo Saito, Dieter Schmalstieg, and Denis Kalkofen. 2020. InpaintFusion: Incremental RGB-D Inpainting for 3D Scenes. *IEEE Transactions on Visualization and Computer Graphics* 26, 10 (2020), 2994–3007. https://doi.org/10.1109/TVCG.2020.3003768

[37] Shohei Mori, Dieter Schmalstieg, and Denis Kalkofen. 2022. Good Keyframes to Inpaint. *IEEE Transactions on Visualization and Computer Graphics* (2022), 1–1. https://doi.org/10.1109/TVCG.2022.3176958

[38] Cuong Nguyen, Stephen DiVerdi, Aaron Hertzmann, and Feng Liu. 2018. Depth Conflict Reduction for Stereo VR Video Interfaces. In *Proc. of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. New York, NY, USA, 1–9. https://doi.org/10.1145/3173574.3173638

[39] Benjamin Nuernberger, Eyal Ofek, Hrvoje Benko, and Andrew D. Wilson. 2016. SnapToReality: Aligning Augmented Reality to the Real World. In *Proc. of the 2016 CHI Conf. on Human Factors in Computing Systems*. New York, New York, USA, 1233–1244. https://doi.org/10.1145/2858036.2858250

[40] Jason Orlosky, Kiyoshi Kiyokawa, and Haruo Takemura. 2013. Dynamic text management for see-through wearable and heads-up display systems. In *Proc. of the 2013 international conference on Intelligent user interfaces (IUI '13)*. ACM, 363–370. https://doi.org/10.1145/2449396.2449443

[41] Fred G. Paas. 1992. Training strategies for attaining transfer of problem solving skills in statistics : a cognitive-load approach. *Journal of Educational Psychology* 84, 4 (1992), 429–434. https://doi.org/10.1037/0022-0663.84.4.429

[42] Stephen D. Peterson, Magnus Axholt, and Stephen R. Ellis. 2008. Comparing Disparity Based Label Segregation in Augmented and Virtual Reality. In *Proc. of the 2008 ACM Symposium on Virtual Reality Software and Technology (VRST '08)*. 285–286. https://doi.org/10.1145/1450579.1450655

[43] Stephen D. Peterson, Magnus Axholt, and Stephen R. Ellis. 2009. Objective and subjective assessment of stereoscopically separated labels in augmented reality. *Computers & Graphics* 33, 1 (2009), 23–33. https://doi.org/10.1016/j.cag.2008.11.006

[44] Edward Rosten, Gerhard Reitmayr, and Tom Drummond. 2005. Real-Time Video Annotations for Augmented Reality. In *Advances in Visual Computing*, George Bebis, Richard Boyle, Darko Koracin, and Bahram Parvin (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 294–302.

[45] Ali Samini and Jarljohan Lundin Palmerius. 2016. A User Study on Touch Interaction for User-Perspective Rendering in Hand-Held Video See-Through Augmented Reality. In *Augmented Reality, Virtual Reality, and Computer Graphics*. Springer International Publishing, 304–317. https://doi.org/10.1007/978-3-319-40651-0

[46] Ali Samini and Karljohan Lundin Palmerius. 2014. A perspective geometry approach to user-perspective rendering in hand-held video see-through augmented reality. In *ACM Symposium on Virtual Reality Software and Technology*. 207–208. https://doi.org/10.1145/2671015.2671127

[47] Thomas Schöps, Martin R. Oswald, Pablo Speciale, Shuoran Yang, and Marc Pollefeys. 2017. Real-Time View Correction for Mobile Devices. *IEEE Transactions on Visualization and Computer Graphics* 23, 11 (2017), 2455–2462. https://doi.org/10.1109/TVCG.2017.2734578

[48] Ludwig Sidenmark, Nicolas Kiefer, and Hans Gellersen. 2019. Subtitles in interactive virtual reality: Using gaze to address depth conflicts. In *Workshop on Emerging Novel Input Devices and Interaction Techniques*.

[49] Srikanth Kirshnamachari Sridharan, Juan David Hincapié-Ramos, David R. Flatla, and Pourang Irani. 2013. Color Correction for Optical See-through Displays Using Display Color Profiles. In *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology* (Singapore) *(VRST '13)*. Association for Computing Machinery, New York, NY, USA, 231–240. https://doi.org/10.1145/2503713.2503716

[50] Jonathan Sutton, Tobias Langlotz, and Alexander Plopski. 2022. Seeing Colours: Addressing Colour Vision Deficiency with Vision Augmentations Using Computational Glasses. *ACM Trans. Comput.-Hum. Interact.* 29, 3, Article 26 (jan 2022), 53 pages. https://doi.org/10.1145/3486899

[51] Tang, Owen, Biocca, and Mou. 2003. Comparative effectiveness of augmented reality in object assembly. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 73–80. https://doi.org/10.1145/642611.642626

[52] Markus Tatzgern, Denis Kalkofen, Raphael Grasset, and Dieter Schmalstieg. 2014. Hedgehog labeling: View management techniques for external labels in 3D space. In *2014 IEEE Virtual Reality (VR)*. 27–32. https://doi.org/10.1109/VR.2014.6802046

[53] Robert J. Teather and Wolfgang Stuerzlinger. 2015. Factors Affecting Mouse-Based 3D Selection in Desktop VR Systems. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction* (Los Angeles, California, USA) *(SUI '15)*. 10–19. https://doi.org/10.1145/2788940.2788946

[54] Makoto Tomioka, Sei Ikeda, and Kosuke Sato. 2013. Approximated user-perspective rendering in tablet-based augmented reality. *2013 IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2013* (2013), 21–28. https://doi.org/10.1109/ISMAR.2013.6671760

[55] M. Tuceryan and N. Navab. 2000. Single point active alignment method (SPAAM) for optical see-through HMD calibration for AR. *Proceedings - IEEE and ACM International Symposium on Augmented Reality, ISAR 2000* 11, 3 (2000), 149–158. https://doi.org/10.1109/ISAR.2000.880938

[56] Klen; Čopič Pucihar, Paul; Coulton, and Jason; Alexander. 2013. Evaluating dual-view perceptual issues in handheld augmented reality : device vs . user perspective rendering. In *International Conference on Multimodal Interaction*. ACM, 381–388. https://doi.org/10.1145/2522848.2522885

[57] Klen Čopič Pucihar, Paul Coulton, and Jason Alexander. 2014. The use of surrounding visual context in handheld AR. In *Conference on Human Factors in Computing Systems*. ACM, 197–206. https://doi.org/10.1145/2556288.2557125

[58] Lei Xiao, Salah Nouri, Joel Hegland, Alberto Garcia Garcia, and Douglas Lanman. 2022. NeuralPassthrough: Learned Real-Time View Synthesis for VR. (2022). https://doi.org/10.48550/ARXIV.2207.02186

[59] Masahiro Yamaguchi, Shohei Mori, Peter Mohr, Markus Tatzgern, Ana Stanescu, Hideo Saito, and Denis Kalkofen. 2020. Video-Annotated Augmented Reality Assembly Tutorials. In *ACM User Interface Software and Technology (UIST '20)*. 1010–1022. https://doi.org/10.1145/3379337.3415819

[60] Qian Zhou, George Fitzmaurice, and Fraser Anderson. 2022. In-Depth Mouse: Integrating Desktop Mouse into Virtual 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 354, 17 pages. https://doi.org/10.1145/3491102.3501884

[61] Silvia Zuffi, Carla Brambilla, Giordano Beretta, and Paolo Scala. 2007. Human Computer Interaction: Legibility and Contrast. In *14th International Conference on Image Analysis and Processing (ICIAP 2007)*. 241–246. https://doi.org/10.1109/ICIAP.2007.4362786