Look over there! Investigating Saliency Modulation for Visual Guidance with Augmented Reality Glasses

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

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

Alexander Plopski*
Graz University of Technology
Graz, Austria
alexander.plopski@icg.tugraz.at

Stefanie Zollmann
University of Otago
Dunedin, New Zealand
stefanie.zollmann@otago.ac.nz

Yuta Itoh†
University of Tokyo
Tokyo, Japan
yuta.itoh@iii.u-tokyo.ac.jp

Holger Regenbrecht
University of Otago
Dunedin, New Zealand
holger.regenbrecht@otago.ac.nz

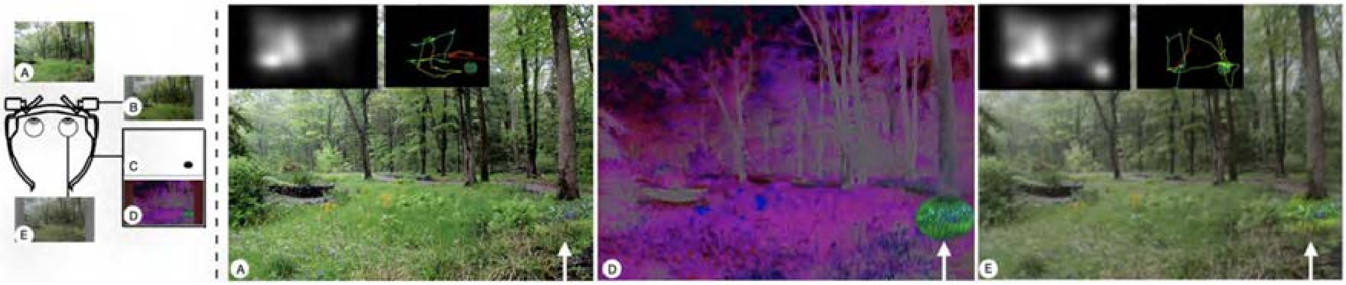


Figure 1: Overview on real-world guidance using saliency modulation in Augmented Reality glasses. (Left) The main steps of our approach that modulate the real world (A) by capturing the scene with an eye-aligned scene camera (B). Applying a mask (C) and a real-time saliency modulation image displayed in the AR glasses (D) allows for changing the perceived scene (E). (Right) (A) Original (un-modulated) scene with insets showing the saliency and example gaze path of study participant. (D) Overlay displayed in the optical see-through AR glasses. (E) Resulting scene when seen through the AR glasses with saliency modulation applied. Insets in (E) show the saliency and example gaze path of study participant. The white arrow pointing out the emphasised area is for illustration only.

Abstract

Augmented Reality has traditionally been used to display digital overlays in real environments. Many AR applications such as remote collaboration, picking tasks, or navigation require highlighting physical objects for selection or guidance. These highlights use graphical cues such as outlines and arrows. Whilst effective, they greatly contribute to visual clutter, possibly occlude scene elements, and can be problematic for long-term use. Substituting those overlays, we explore saliency modulation to accentuate objects in the real environment to guide the user's gaze. Instead of manipulating video streams, like done in perception and cognition research, we investigate saliency modulation of the real world using optical-see-through head-mounted displays. This is a new challenge, since we do not have full control over the view of the real environment. In this work we provide our specific solution to this challenge, including built prototypes and their evaluation.

CCS Concepts

• **Human-centered computing** → *Ubiquitous and mobile computing systems and tools*; **Human computer interaction (HCI)**; **Mixed / augmented reality**.

Keywords

Augmented Reality, Computational Glasses, Visual Guidance, Saliency, Saliency Modulation, Eye tracking, Gaze, Vision Augmentation, Augmented Human, Mixed Reality

ACM Reference Format:
Jonathan Sutton, Tobias Langlotz, Alexander Plopski, Stefanie Zollmann, Yuta Itoh, and Holger Regenbrecht. 2022. Look over there! Investigating Saliency Modulation for Visual Guidance with Augmented Reality Glasses. In *The 35th Annual ACM Symposium on User Interface Software and Technology (UIST '22), October 29-November 2, 2022, Bend, OR, USA.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3526113.3545633

*Also with University of Otago.
†Also with Tokyo Institute of Technology.

UIST '22, October 29-November 2, 2022, Bend, OR, USA

ACM ISBN 978-1-4503-9320-1/22/10…$15.00
https://doi.org/10.1145/3526113.3545633

1 Introduction

Using optical see-through *Augmented Reality* (AR) glasses for visual guidance applications is an obvious example for the utility of AR technology. Early applications and research prototypes usually depict scenarios with bold graphical elements overlaid on top of the user's view. However, for an effective, efficient, and satisfactory user experience we would benefit from a more subtle overlay even using existing visual cues to guide users' visual attention. Visual saliency, "the distinct subjective perceptual quality which makes some items in the world stand out from their neighbours and immediately grab our attention"[1] is a potential solution and a research topic that has attracted research from various areas—in particular researchers in perception and cognition, creating an understanding and models for visual saliency as a quality for certain regions to stand out or attract more attention compared to others. Large parts of this research are driven by utilising eye-tracking technology that allows investigating this effect. Computer Vision and Neuroscience research later tried to compute so-called saliency maps that approximate the visual saliency of a scene or of a given image [23]. This research converged into approaches that were able to even predict human gaze for specific scenes [10, 28]. All this research highlighted the value of visual saliency and validated the general concept.

Based on the established understanding of visual saliency, researchers started to explore approaches that aimed to change the user's perception of images in video footage by modulating the visual saliency of it. The key idea was to modulate parts of the footage to attract the user's attention or reduce the saliency of certain parts, making them stand out less. Most of this research has been implemented using offline image manipulation techniques before being shown to study participants [15, 43, 53, 63].

Researchers have considered AR as an interface technology that could apply saliency modulation to the real world [6, 63]. However, these considerations were mainly conceptual. If realised, it could be used for visual guidance introducing less visual clutter while protecting the actual context. Possible applications are in aiding visual search, acting as a subtle reminder, and would open up many other applications including influencing attention. Unfortunately, existing works come up short in fulfilling the promise of real-world modulation using AR technology. Primarily, because they demonstrate saliency modulation not via an AR device or AR overlay but directly change the appearance of image or video material [3, 63] which is often done offline and displayed on standard monitors. This is far from envisioned application or usage scenarios. Even when using video see-through head-mounted displays (e.g., immersive HMDs equipped with external cameras), something that we have not seen fully implemented yet, one would decouple the user from directly viewing the real world and reducing the fidelity of the world towards the specifications of the displays and cameras used. Whilst this enables complete control over what the user sees and modern devices can match some properties of the human eye (e.g., The Varjo XR-1 can match human foveal vision at the center of it's display), it introduces other issues. For example, the need to completely reproduce all light seen by a user, match many properties of the human eye, and introduces concerns around constrained field of view, reduced social cues (such as eye contact), and safety concerns.

In this work, we go a different, more challenging way by exploring saliency modulation via *optical see-through head-mounted displays* (OSTHMDs). Instead of applying image manipulation to image material shown to the user or decoupling the user from the real world by showing manipulated camera footage, OSTHMDs are conceptually similar to traditional glasses because we can directly see the world in full fidelity but also see a digital overlay shown via the OSTHMD. This potential is highlighted by the industry investments in devices such as Microsoft's Hololens, Magic Leap or Snap's newest Spectacles among others who all follow this concept. Unfortunately, OSTHMDs bring their challenges. Relevant for this project is that we can only add (and not subtract) light to a scene (similar to projectors) and the challenge to precisely align the overlay with the real world.

This paper presents our research that takes inspiration from earlier work utilising OSTHMDs as vision aids [29, 46] and previous work on saliency modulation on images and videos [63]. We present our investigations on using OSTHMDs to modify visual saliency of the physical world (see Figure 1).

In summary, our contributions are a) the overall exploration of visual saliency modulation via OSTHMDs, b) the development of a saliency modulation algorithm considering the specific workings of OSTHMDs and their prototypical implementation in different lab prototypes. Finally, c) as our main contribution we present the result from studies that evaluated the general feasibility and efficacy of our approach using different prototypes. Our insights result from a combination of saliency analysis, eye tracking, and user questionnaires explored in prototypes with a different level of control.

To the best of our knowledge, saliency modulation has never been explored on OSTHMDs before and our work is an important step in advancing saliency or similar modulations of the physical world. It takes them away from studies that applied less constrained image manipulations displayed on a screen, and towards practical systems with many applications in Augmented Reality, Vision Augmentations, Augmented Human, and HCI.

2 Background

Our work involves several different areas of research, such as the general concept of visual saliency, attention modelling, and saliency modulation, which we will briefly introduce in the following sections, focusing on the most relevant works.

Visual Saliency

Early work in cognitive psychology has given evidence of a relationship between the properties of a scene and the attention applied to it. Treismann and Gelade have shown how various features are processed in parallel across the visual field, and that attention is based on these features to process them into complete objects [60]. This feature-based process is commonly referred to as bottom-up saliency. It describes the influence of aspects of a visual scene upon where attention is placed, regardless of conscious influence. The other commonly given aspect of saliency is top-down saliency that describes the influence of conscious effort and goals on where attention is focused on.

[1] http://www.scholarpedia.org/article/Visual_salience

Koch and Ullman proposed a biological model in which the various features being processed in parallel are combined into a singular map that shows the impact of the individual features as a saliency map [26]. Computer vision techniques have been used to model and compute saliency maps to better understand and utilise saliency. Itti et al. proposed one of the first and most well-known maps [23] that builds on the biologically plausible model of Koch and Ullman [26] but since then other saliency maps have been proposed [17, 45, 59]. More recent approaches integrated gaze maps into their models for computing saliency maps with gaze maps showing how much attention will be placed on various areas in images and videos. Examples include SAM-ResNet [10] and MSI-Net [28] for images or Wang et al.'s network for videos [66]. These approaches are tested against datasets of real users' attention, such as CAT2000 [7]. All these models for saliency and visual attention are bottom-up approaches and do not consider conscious influence.

Visual Attention Guidance

Attention guidance has seen a large amount of research due to it having application in many areas such as guiding users through digital information such as webpages [44], aiding order picking in industrial tasks [50], or general in training [52] to only name a few. Many of these techniques for visual guidance use colour adjustments. Azuma et al. demonstrated the use of coloured edges to aid with reading [2]. Nguyen et al. directly replaced colours in desired areas using a graph network [41], whilst Mateescu et al. created a computationally simpler hue shift also to achieve a noticeable colour shift [35]. Changing colours can also be combined with changing other elements such as adjusting the size, the position of elements [44], using graphical elements (e.g., arrows [30]), or changing the structure or shape of elements (e.g., 3D models [25]).

Another method proposed for visual guidance is a subliminal flicker effect or other lighting effects. Bailey et al. first demonstrated this when they varied luminance in the periphery of views to draw attention [3]. It has subsequently been applied to assist search performance [37, 64], storytelling [49, 65], and training [52] on both displays and in virtual reality [13, 49].

Many Augmented Reality techniques use visual guidance techniques. In the context of projector-based AR systems, spotlight systems have been demonstrated [6, 58] while other approaches used techniques such as the AR tunnel using visually overlaid AR arches that act as a tunnel, directing users to a 3D location [4, 5, 50].

Common to all these methods is that they focus on drawing maximum attention which often comes at costs in scene understanding as the techniques often hide or distort other important scene elements.

Saliency modulation also emerged as an approach to guide users' attention or highlight scene elements. Most of the works are screen-based in the sense that they manipulate visual media such as images or videos displayed on standard screens. These existing works looked into different parameters for modulating the saliency of a scene. For example, Kokui et al. and Takimoto et al. applied colour shifts based on saliency maps [27, 57] whilst other approaches looked at modulating the spatial frequency [56] or texture power maps [53]. We have also seen approaches combining colour and intensity modulation using a saliency maps [15, 51], the introduction of subtle blur effects to modulate the visual saliency [16], or the usage of genetic algorithms [43]. More recent approaches combine different parameters affecting visual saliency including, blurring, intensity, saturation, and contrast [55]. Another means to introduce visual guidance could be foveated rendering techniques such as [24, 39] to provide guidance based on focus, similar to that of blur effects. However, this has not yet be demonstrated.

Visual Saliency Modulation for AR

Saliency has been used in AR, initially to place visual information based on saliency so that they do not distract the user from important scene elements [12] but saliency modulation is mainly used for guiding the user. For example, Lu et al. demonstrated how the visual saliency of AR content could be increased to aid users searching for it [31–33]. Another example is the work by Ahn et al. who globally increased the visual saliency to improve the readability of AR screens [1] and to have virtual objects stand out even more which is relatively simple. They do not change the saliency of the real world through AR as aimed at in our work. Some earlier work looked at using Camera-Projector AR systems to adjust saliency [58]. Most notably, the work by Ueda et al. proposed an approach using calibrated projectors and synchronised glasses with focus tunable lenses to guide the users [62]. The idea is that the glasses do a full focal sweep and a part of the scene is lit by the synchronised projector when in focus and other parts when out of focus. While demonstrating impressive visual results, the actual effectiveness as a means of gaze guidance was not demonstrated with users, instead evaluating comfort and usability, and relies on the complex and impractical interplay of external projectors and worn focus tunable lenses.

The most related work to ours is that by Veas et al. [63] and Mendez et al. [38] who looked at achieving subtle visual guidance using saliency modulation on images and videos. However, whilst their general aim does align with ours in modulating the saliency of the physical environment they eventually only demonstrated manipulation of video footage, applicable only to video see-through HMDs and as such putting it's technique closer to prior work in visual attention guidance and not being applicable to current OS-THMDs.

To conclude, there is generally a good understanding of the relevance of bottom-up saliency and its importance for scene understanding. This has been shown with works modulating the visual saliency in images or videos displayed on a screen. Works exploring practical usage of saliency modulation or even saliency modulation of the real world to guide users are mainly in a conceptual stage or use projectors to change the appearance of the real world (spatial AR). In our work, we are targeting the most promising, but also most challenging saliency-based guidance: Saliency modulation of the real-world using optical see-through head-mounted displays. These head-worn displays are often envisioned to be pervasive in the future and are more related to traditional glasses in that users can see their environment through optical glasses, but they can still perceive visual overlays using half-transparent displays.

3 Saliency modulation in Augmented Reality glasses

The core idea of our work is to explore visual guidance by modulating the saliency of the physical environment using OSTHMDs

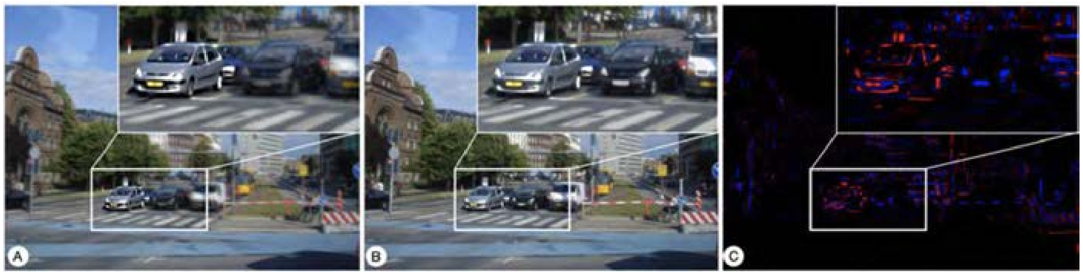


Figure 2: Differences in saliency modulation methods not considering additive only properties of OSTHMDs exemplified by our implementation of the algorithm by Mendez et al. [38] (insets for details): (A) Full colour control allows increasing the saliency of some image elements (e.g., silver car) while decreasing the saliency of other elements (e.g., black car). (B) Modulation using only the additive components fails to demonstrate the full saliency modulation. (C) Visualisation of required additive (red) and subtractive (blue) modulation necessary to highlight the silver car showing the majority of image areas are modulated by subtractions that are not possible in OSTHMDs.

normally used for traditional AR. So far, existing techniques for saliency modulation have only been demonstrated on videos that have been displayed on standard monitors. While providing interesting insight into human perception, this is not applicable for our envisioned real-world application scenarios. More importantly, these approaches would not work on current OSTHMDs because of the differences in how the overlay is displayed to the users (e.g., optically blending virtual overlays, inability to subtract light). In current OSTHMDs, we can only show information by adding light (additive) while interactive state-of-the-art saliency modulation techniques shown on videos (e.g., Veas et al. [63]) always assume full light control (additive and subtractive). The consequence of not being able to subtract light diminishes the effect when using the original approaches (See Fig. 2). Whilst research is being done to create devices capable of subtracting light [20], and commercial devices are in development, current solutions introduce new issues such as heavily reduced light transmission [22].

3.1 Environment modulation in OSTHMDs

Beside the fact that OSTHMDs can only add light to the scene, we also face the issue that we need to precisely modulate the environment. More specifically, we need to firstly capture the scene as perceived by the user to understand the physical environment. Secondly, we need to align the computed modulation with the physical world as seen by the user. Off-the-shelf OSTHMD are not capable of achieving this, as they only integrate off-axis cameras that cannot truly capture the world from the users perspective leading to miss-registration and consequently not allowing us to achieve the desired effect. Solving this required us to create custom OSTHMDs taking inspiration from earlier works on Computational Glasses originally aimed at addressing Colour Vision Deficiency using OSTHMDs [29, 54] which in itself used concepts from Steve Mann's EyeTap [34]. The key is to integrate a scene camera into the OSTHMD, virtually placed at the position of the eye via a beamsplitter (See Fig. 3). The beamsplitter reflects a portion of the scene as seen by the user towards the scene camera. In our work, we use 50/50 beamsplitters; however, depending on the camera used and the sensor sensitivity, other ratios will work as well (e.g., 90/10, requiring more sensitive cameras).

The system is calibrated such that it captures the environment as seen by the user via the scene camera, processes the captured environment image to compute an overlay, which when displayed on the OSTHMD aligns correctly with the physical world. As the scene camera cannot be adjusted accurately enough to meet every individual's eye a further software calibration is required. We use a modified version of the well-known SPAAM approach [61] which we extended to calibrate the scene camera alongside the manual eye-display calibration for each user. We emphasise that this approach does not require spatial tracking, commonly seen in the newest generation of commercial OSTHMDs. In fact spatial tracking would not help much as it tracks the device within a world (with centimetre or millimetre accuracy) while we are able to directly modulate the users view with pixel-precision being dependent only on an observation of the scene and not the user's pose within it. Similarly, the implicit 3D reconstructions commonly created for spatial tracking are too coarse to be of use for a precise modulation of the environment.

3.2 Saliency Modulation via OSTHMDs

As stated before, existing approaches for saliency modulation do not consider OSTHMDs but mainly manipulated video images giving full control of each pixel. Instead, our approach interactively computes a modulation that when displayed in the OSTHMD aligns with the real-world and modulates its saliency (See left Fig. 1).

While even naïve overlays, such as a rendered frame, affect the saliency of a scene, we are looking for a more subtle effect that uses all parameters affecting visual saliency such as colour, orientation, size, motion, and depth [67]. But, because we can only modulate the light entering the eye from the environment, we cannot change some other aspects (e.g., size of an object or motion). Related works usually apply a selection of changes in contrast (blur or sharpening), (de)saturation, or changing the lightness of objects [63]. Whilst some researchers have looked to directly change the hue of an object entirely, we chose not to utilise this effect as it can cause confusions and clutter [36, 41].

Our implementation uses an algorithm combining a set of components based on those described to work in the literature, that can influence saliency whilst considering that they can only be

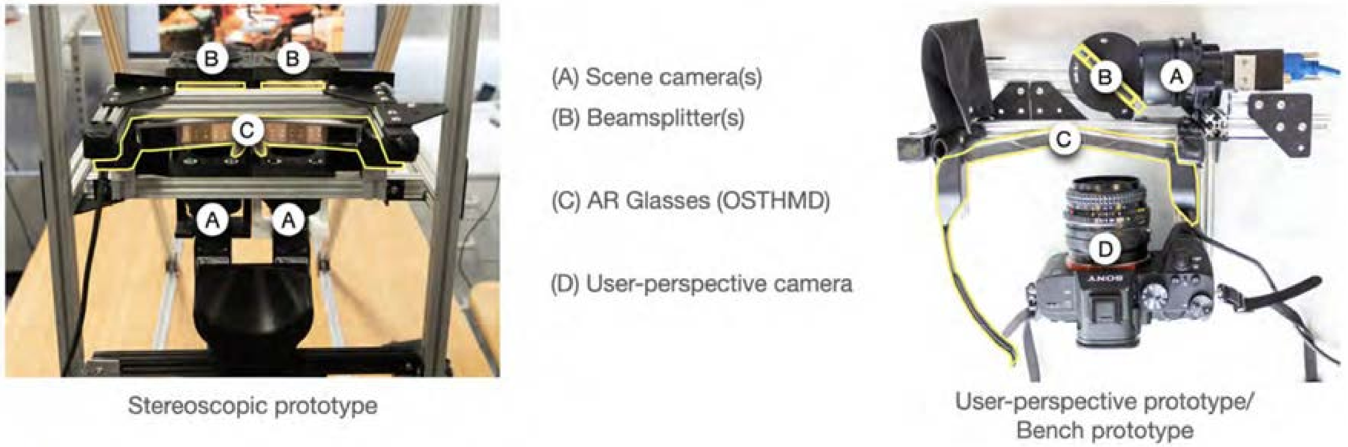


Figure 3: Two of the prototypes developed for this project. (Left) Stereoscopic prototype with the main components highlighted. (Right) Bench prototype (mono) with user perspective camera capturing the image through the AR glasses also showing the beamsplitter and scene camera. This allows for a more controlled study environment.

applied by adding light. In the following, we assume that the reader is familiar with the conceptual properties of colour spaces such as *RGB*, *HSV*, and *Lab*.

The earlier work by Veas et al. [63] utilised the method of Mendez et al. [38] which only looked to adjust lightness L then shift opponent colours, *RG*, *BY* to effect conspicuities in *Lab* space thereby simply creating a contrast shift. Unfortunately, our experiments showed that their algorithm strongly relies on darkening image regions that are not of interest and consequently fails to achieve the desired effect in OSTHMDs where this is not possible (see Fig. 2). As our modulation has to be purely additive, we modulated several components in our algorithm. To this end we include saturation increasing and decreasing, a contrast increasing and decreasing, blurring, and sharpening. Each of these components has been used previously for modulating saliency within images shown on normal screens [16, 32, 55, 57], combining them to achieve saliency modulation when reducing environment light per pixel is not possible (e.g., OSTHMDs or Heads-up displays) is novel. We are basically replacing saliency modulations that mainly adjust lightness by utilising other factors that affect visual saliency.

Within our implementation, for each component we define a parameter p used to adjust the degree of modulation. For the saturation component of our algorithm, we set the S component of the colour in *HSV* colour space to *max* for increasing saturation, and *min* for decreasing. For contrast we use a sigmoidal contrast function with $\beta = 10$ and $\alpha = 0.5$ to increase contrast. To decrease it we then use the inverse of the function with the same parameters. To adjust the output of saturation and contrast modulations to the parameter level p we subtract the original value from the modified to get a Δ vector which we then scale by p. To blur the image, we applied a common Gaussian blur with $\sigma = p$, and in order to sharpen we use an sharpening filter scaled by p. We report later on how we established the values for the parameter p adjusting the strength of the modulation. As all components can still produce negative values, which will have no effect within OSTHMDs but will be relevant when simulating the effect on normal screens, we also take Δs for the blur and sharpening components clamp all Δs to ensure they contain no negative values before being added to the original image for simulation or displayed on the OSTHMDs. Overall, our implementation utilises GLSL and the performance is only limited by the camera update rate (41 fps). In fact, first experiments showed sufficient performance even on mobile hardware with a less capable GPU. Therefore we believe it is reasonable to expect that our approach will run sufficiently on future AR glasses with integrated computing units.

3.3 Prototypes

We implemented our approach for real-world saliency modulation using OSTHMDs in several prototypes. These were created for use in our later user studies.

Stereoscopic prototype

In order to enable participants to perceive modulations directly through the Computational Glasses, we built a non-mobile stereoscopic prototype. This prototype utilised an Epson Moverio BT-300 and integrated a 50/50 beamsplitter in front of each eye. Two Point Grey Blackfly cameras were used as scene cameras. We decided for the BT-300 because its OLED display is known to cover almost the entire RGB colour space unlike devices such as the MS Hololens. This prototype was mounted in a stabilising frame and a chin rest was included to enable participants to maintain a comfortable head positioning. We integrated the saliency modulation approach described in the previous section into our stereoscopic prototype and were able to successfully modulate saliency. In order to evaluate the efficacy of our saliency modulation approach, we added a Pupil Labs eye tracker to be able to track the users gaze. While achieving good visual results, upon initial testing we found that due to our specific setup and the nature of using a retrofitted eye tracker the eye tracking accuracy is lower than HMDs with fully integrated eye trackers. We considered using alternative OSTHMDs that already include inbuilt eye tracking, e.g., Hololens 2 or

Magic Leap One. Unfortunately, their integrated displays have a very limited colour space and suffer from low colour correctness and chromatic aberrations[2] which became immediately obvious in our tests. Consequently, using those devices was not an option either, as they would not allow to display correct colours as required for the environment modulation.

User-perspective bench prototype

Our second prototype addresses the identified shortcomings of the earlier prototype, mainly the reduced eye tracking quality (Please note that this is only needed for performing user studies using eye tracking). This prototype again used a modified Epson Moverio BT-300 as an OSTHMD and integrated a 50/50 beamsplitter to reflect a portion of the incoming environment light towards the scene camera (Point Grey Blackfly). However, instead of the user looking directly through the setup, we placed a camera at the position of the user's eye (we used a Sony A7M3) as a user-perspective camera. The output of the camera can either be saved or directly be further processed by accessing its output via HDMI. To address the low quality eye tracking in our first prototype we stream the actual view of the user-perspective camera through the OSTHMD to a VR head-mounted display with integrated eye tracking (HTC Vive Pro Eye) where the camera image is displayed on a plane in the VR scene. VR has often been used in the past to simulate AR interfaces [47] but here we use it to show the output from actual OSTHMD AR glasses instead of a VR simulation of AR. In a similar fashion, user-perspective cameras have been commonly used for evaluating optical devices [9, 18].

The advantage of this approach is that we could utilise the high-quality eye tracking of the used immersive VR HMD while also guaranteeing a good calibration (because we calibrate for the camera) and guaranteeing the same quality of calibration for all participants. This eliminates a large number of possible confounding variables. In particular the calibration of OSTHMDs for actual user's eyes often proves problematic as calibrations are very individual and verification of the quality is hard, introducing at least one confounding variable. Similar prototypes are commonly used for that reason [8, 11, 19, 29].

3.4 Image data-set

For our study, we sourced images of indoor and outdoor scenes from the CAT2000 dataset [7]. We used this dataset as it provides eye-tracking data and saliency maps that we used for identifying areas for saliency modulation and testing. We selected suitable images for our studies by identifying images depicting a distinct object or set of objects of decent size that could be enhanced, which only showed small amounts of initial gaze attraction according to the eye gaze and saliency maps provided.

Although we had eye-tracking data from the CAT2000 dataset, we also took data from 5 people looking at each image in our particular study environment and used this to inform our final selection of areas to emphasise. Final areas were selected as areas that 1-2 people looked at but not more so that we could expect some fixations to analyse in both conditions but still had room to see if we could increase the attention paid to an area. Finally, we also selected a few images that we considered challenging because they are generally highly salient throughout, with many very salient image areas. We used these as almost worst-case scenarios as we would need to detract from very salient image areas.

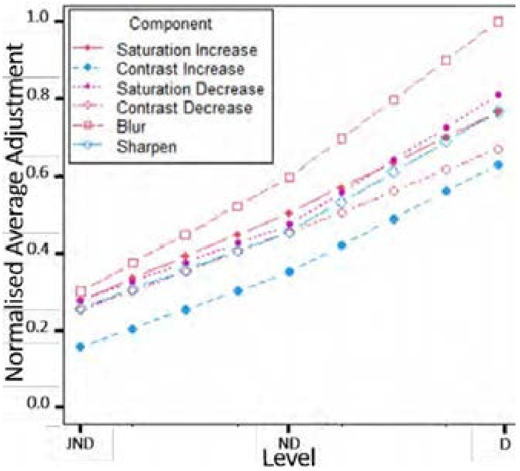


Figure 4: Selection of the appropriate modification level for the primary study. Average selected adjustment levels participants rated as just noticeable (JND), notable (ND), and distracting (D). We define several steps between each rated level and interpolate the corresponding adjustment level.

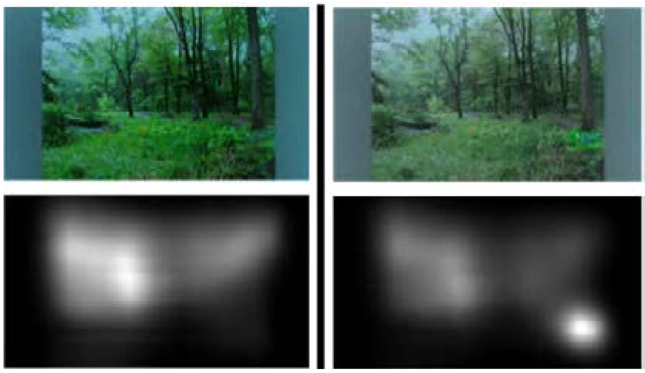


Figure 5: Example of the appropriate modification level for the primary study. Appearance and saliency of an image (left) and results (right) that provide a good balance between colour and saliency change that were selected for the primary study.

4 Selecting modulation levels

Validating saliency modulation is traditionally done by analysing gaze data. However, as for most other existing approaches of saliency modulation not aimed at OSTHMDs [38, 63], we needed to first identify a suitable level of saliency modulation so we explored the parameter space of our algorithm and looked at the effect according to study participants.

[2] https://kguttag.com/2020/07/08/hololens-2-display-evaluation-part-2-comparison-to-hololens-1/

In the following, we describe this study with the main purpose of finding suitable parameters. For this study, we look at normalising the effect of each individual component affecting saliency (i.e. saturation, contrast, blur and sharpening) as each could work drastically differently. We choose to do this based on users' responses and so devised a study where participants set levels for each component, looking for the adjustment levels where the differences first became noticeable (Just Noticeable Difference, JND), once it was having a notable effect (Notable Difference, ND) and the point at which it became distracting to the user (Distracting Difference, D). This provides adjustment levels for our algorithm at which each component would start to have an effect, where the effect was clearly working and where it was over-tuned and causing a detrimental effect. We then linearly combined these results to create one modulation parameter. One could argue that this linear combination is not representing the complex interactions between the components. However, building better models to describe the complex interactions is a research topic on its own and is beyond the scope of this work.

Design: We designed an experiment to test the effect of manipulating each component (i.e. saturation, contrast, blur, and sharpening) within a certain min-max range [0-1]. Our goal was to identify a parameter range for each component representing three different levels: Just noticeable, Notable, and Distracting. We did this for each component separately. The study design was approved within the regulations of the human ethics committee of the University of Otago.

Apparatus: For this study, we seated users in front of a monitor without any of our protoypes, where the image modified by the parameter was shown. We placed a dial, that was used to adjust the level of parameters, and a numpad in front of them. The number pad had labels placed on the relevant keys for running the study (reset, set, none) and the rest of the keys were covered with a single cover to prevent their use.

Procedure: After signing a consent form and completing a demographic questionnaire collecting information on age, gender, and vision impairments (colour and refractive), each participant was seated in front of the monitor and was informed about the study procedure. They were also given the instructions and relevant definitions as text which they were asked to read. Once the participant understood the procedure and had no questions the study was started.

For the actual study the participant was shown a random image with a random component selected. They were asked to adjust the dial until the effect was just noticeable then press the 'set' key, or the 'none' key if they could not find a value that they believed met the definition. The image was then removed and they were asked to reset the dial. This was then repeated for the notable and distracting levels. The user could reset the values for all levels on that image at any stage. Each participant was asked to set values for each component twice.

Participants: For this study we recruited 10 participants (6 male, 4 female, 20-51 years old $\overline{x} = 29.5$). We excluded participants having colour vision deficiency or vision not corrected to normal.

Outcome: This study allowed us to empirically determine the meaningful range of the modulation parameter p under which the effects of each component can be compared/normalised, and we can expect behaviour to be similar so we can use it to select a singular, combined level across all parameters. To define our final, singular modulation level p , we took the mean JND and ND for each parameter (See Fig. 4) and linearly interpolated between the levels. We used 5 steps. Using our bench prototype, we took images of each of our test images and simulated the levels of modulation on them. We then looked at the changes in the saliency maps and colour shifts and selected the level of our parameters providing a strong response in the corresponding saliency maps while maintaining minimal noticeability (See Fig. 5) and used them in our follow-up studies.

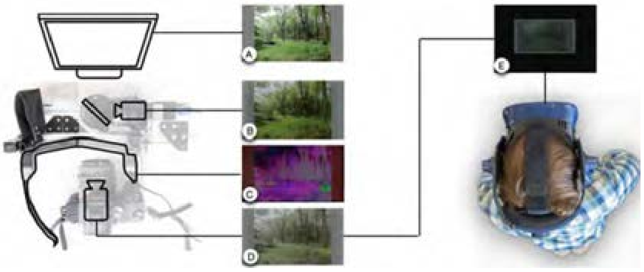


Figure 6: Study apparatus: (A) Scene displayed on a screen representing the real environment is captured by the scene camera (B). The computed saliency modulation (C) is then displayed on the AR glasses. The combined view (D) captured by the user-perspective camera (representing the human eye) is finally displayed in the virtual environment (E).

Parameter Levels Transfer to OSTHMD: While we estimated our parameters on a simulation, this provided us with a saliency map for each degree of noticeability. Meaning, when applied on a different display in the same additive manner, we expect that similar saliency maps will be rated on a similar level of noticeability, allowing scaling of the displayed output to match the desired results. Due to health and safety restrictions for COVID-19 we were limited in our ability to run studies. As such, rather than run further studies to set levels in OSTHMD we choose to utilise these values for our OSTHMD settings. In order to ensure that our values could be as accurate as possible on the OSTHMD we compared the saliency maps from the simulated images on the screen to images taken through the bench prototype. We looked to scale the output on the OSTHMD due to the different contrast produced, until we found the closest output.

5 Main Efficacy User Study

After empirically identifying the main parameter for our saliency modulation, in this main study, we aimed to investigate the effect when modulating the saliency of the environment by way of AR glasses. Similarly to previous studies investigating saliency modulation of images[16, 36, 40, 63], we were interested in recording and analysing gaze data. Here, we were in particular interested in the differences between the gaze data of participants looking at the original images and the gaze data of participants looking at the images when saliency modulated by our AR glasses as this provides insights about the effect of gaze redirection.

Design: We designed a within-subject study to investigate the effect of saliency modulation using our approach. Participants observed views of the images from the image data-set in each of two initial conditions (*unmodulated* and *saliency modulation via AR glasses*) in randomised order.

For each image, we collected the user gaze data and asked participants to rate the image's *Naturalness*, *Obtrusion*, and *Quality* on a Likert-like scale from 1-7. In these conditions, we evaluated the *time until the first fixation* on a target area, the *explored area of the image*, and the *number of participants who fixated* on said area, and the answers to the questionnaires.

After participants viewed and rated all images in both conditions, they were again shown views of the images in a third condition (condition 3: "circles"). In this condition, the images had a circle overlay displayed on the AR glasses around the highlighted area, as an example of traditional guidance in AR. We included this condition to explore how our modulation affects the user's gaze behaviour compared to a traditional AR overlay (showing a circle to highlight an area of interest). In our preliminary tests, we observed that this condition had a strong anchoring effect on participants, affecting their gaze patterns whenever they were exposed to the scene again. We thus opted to show this condition last instead of fully randomising the order in which the participants experienced the three conditions to avoid biasing the results. In this condition, we only collected the participants' gaze data and evaluated the *explored area of the image*.

Our independent variable was the modulation state of the image with the three conditions: 1) "Unmodulated", 2) "Saliency modulation via AR glasses", and 3) "Circles". See Fig. 7 for examples of each condition.

When asking about the naturalness and obtrusion we provided definitions for each word that steered participants towards the requisite measures. This was due to the variability of the definition of such words and to prevent participants from taking drastically different views. We, however, left quality undefined to prevent biasing towards certain aspects of the images, preventing reporting on others. Our provided definition for naturalness was: "having undergone little or no processing" and for obtrusion was: "noticeable or prominent in an unwelcome or unwanted way". The study design was approved within the regulations of the human ethics committee of the University of Otago.

Apparatus: Given the lower accuracy of the external eye tracker in the stereoscopic prototype we opted for our first study to use our bench prototype showing the actual view through our OSTHMD prototype in VR (See Fig. 6). This enables us to utilise high-accuracy eye-tracking whilst providing a more controlled environment where we can overcome the confounding variables such as the eye display calibration quality.

For the VR environment, we created an unlit virtual room with black walls into which the user was placed. They then had a virtual screen placed in front of them that covered a 40^o angle. The screen was always placed directly in front of the user and maintained its visual position relative to the HMD's location throughout the study described later. On that screen, we showed the camera feed as captured through the bench prototype. Thus, this system combined the visual results from an OSTHMD with the quality of the integrated eye tracking from an off-the-shelf VR system commonly used in research (See Fig. 6). The VR environment provides a completely controlled setting where we can ensure all participants are exposed to the same conditions increasing internal validity.

Procedure: Given the context of a global pandemic, we had to take extra precautions. Health and Safety procedures for the study were following institutional and governmental guidelines for COVID-19 safety. As such a distance of >2m was maintained between participants and operator, participants were screened for symptoms, and sterilisation of equipment was used. Before entering the study, each participant completed the screening/contact tracing form, read the supplied information sheet, and signed a consent form. We also asked them to provide their information in a demographic survey collecting information on age, gender, vision impairments, and previous experience with VR (e.g., issues with simulator or motion sickness).

Once completed, we introduced the participants to the use of the HMD, provided an overview of the study procedure and questionnaires, and provide definitions for naturalness and obtrusion. After the participants put on the HMD, we calibrated the integrated eye tracker using the supplied Vive SRanipal calibration. We verified the calibration to be at least within 1.5^o average angular accuracy but often saw values $< 1^o$ across the measured area. This calibration was repeated after every 10 images throughout the study in case participants invalidated the calibration (e.g., by moving the HMD).

During the actual study, each participant was shown 3 different views (one for each condition) of each of 10 images for 5 seconds each. The images were shown in a randomised order and conditions in a semi-randomised order, as detailed in the design. No participant saw the same image in two conditions consecutively. Before showing a new image, we displayed a black screen with a white cross in the centre. Participants were instructed to look at the cross when it appeared. This was done to centre their gaze in the screen for each image. After each image we asked the participants to rate it on the Likert-like scale for naturalness, obtrusion, and quality. Answers were captured by the study conductor. We also asked for general feedback. Participants were gifted vouchers worth approximately $13 (USD) for their time.

Hypotheses: We were primarily interested in the effect saliency modulation has on gaze patterns and on the subjective evaluations of the images, so we formulated three hypotheses:

- H1: Real-world saliency modulation via AR glasses alters gaze pattern when compared to an unmodified scene with respect to time until the first fixation on a target area and number of participants who fixated.
- H2: Real-world saliency modulation via AR glasses is not rated as less natural, more obtrusive, or lower quality compared to unmodulated scenes.
- H3: Real-world saliency modulation via AR glasses is less visually distracting when compared to augmenting geometrical primitives (circles).

Participants: We recruited 20 participants (10 female, 10 male, age ranging from 19 to 47, $\overline{x}$ = 25.2) from students at the university. All recruited participants completed the study according to the procedure above. All participants had normal vision or corrected to normal via contact lenses. Eye tracking was verified to an $\overline{x} = 0.8^o$ and $\theta = 0.36^o$.

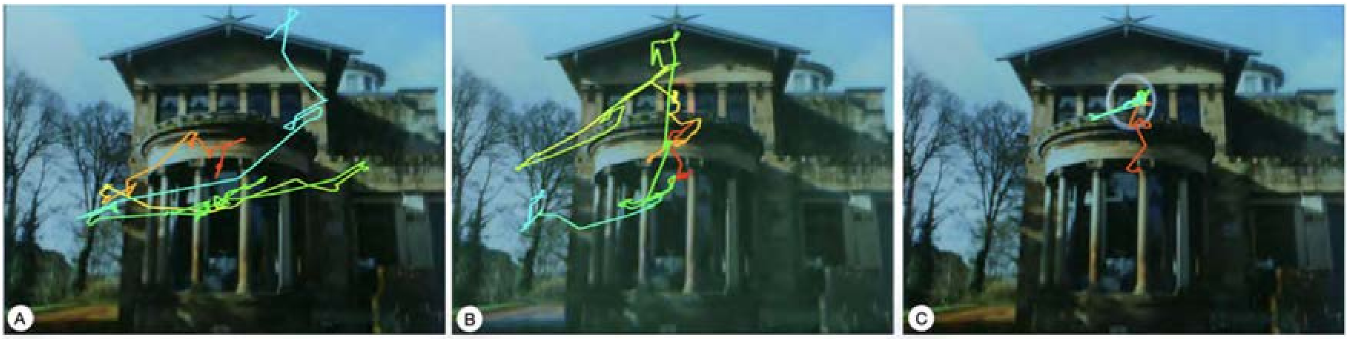


Figure 7: An image from our dataset in both primary conditions; unmodulated (A), and Saliency modulation via AR glasses (B), as well as the secondary condition of circles (C). Overlaid is gaze data from a participant time-coded from red (start) to blue (end).

Results: To determine whether participants focused at a target area, we implemented the IV-T fixation detection algorithm as described by Olsen [42]. We identified that a participant fixated at an area of interest when at least one gaze-point associated with a fixation lay within the target area. An example of a participant's gaze data on an unmodified image is shown in Fig. 7(A), compared to the gaze data when modulated via the AR glasses (See Fig. 7(B)). We visually checked all recorded gaze patterns to detect possible errors in the recorded data. We exclude the gaze data of one participant as it exhibited large inconsistencies (e.g., consistent jumps between continuous gaze points). We checked normality of the collected data with the Shapiro-Wilk test and assumed significance at a $\alpha < 0.05$ level. We analysed normally distributed data with a paired one-sided t-test and used a Wilcoxon signed-rank test otherwise.

The results for the remaining 19 participants showed a significantly higher number of detected fixations for the saliency modulated condition than the unmodulated condition using a McNemar test (χ^2 = 25.565, $p < 0.001$) (See Fig. 9(a)). This is supported by our finding that we successfully attracted a higher number of fixations for all but one image, where the number of fixations went down from 4 to 2. This can also be observed in the heatmaps shown in Fig. 8.

We also investigated whether the modulation prompted participants to look at a target faster if they did look at the area of interest. In our first analysis, we grouped by participants. When considering only image observations where participants had an actual fixation (*Cleaned* in Fig. 9), participants fixated at the target area significantly faster in the saliency modulated condition than the unmodulated condition ($t(18) = 3.96, p < 0.001, d = 1.09$; *CI* 0.33-1.078). To compensate the effect of missing fixations in some image observations, Veas et al. assigned the maximum display time (in our cases 5s) for each image when not fixating [63](*All* in Fig. 9). Applying this analysis shows the same result of a significantly faster target fixation when applying saliency modulated ($t(18) = 8.42, p < 0.001, d = 2.3$; *CI* 0.87-1.45) (See Fig. 9(b)).

In our second analysis, we assumed that the time to the first fixation is dominated by the observed image and consequently grouped by image. This evaluation is similar to that of Veas et al. The results violated the normality assumptions. We found that participants fixated onto the target area significantly faster in the modulated condition considering only observations with detected fixations (*Cleaned*) ($t(9) = 3.03, p = 1.4e{-}2, d = 0.81$; *CI* 0.17-1.19). Again, this effect was supported when assigning the maximum time for no detected fixations (*All*) ($t(9) = 4.36, p = 1.8e{-}3, d = 0.86$; *CI* 0.38-1.35)(See Fig. 9(c)).

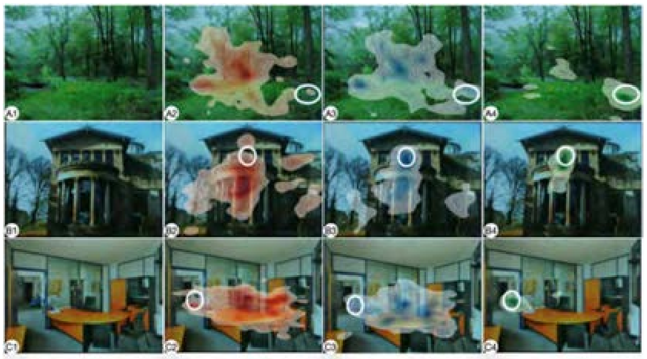
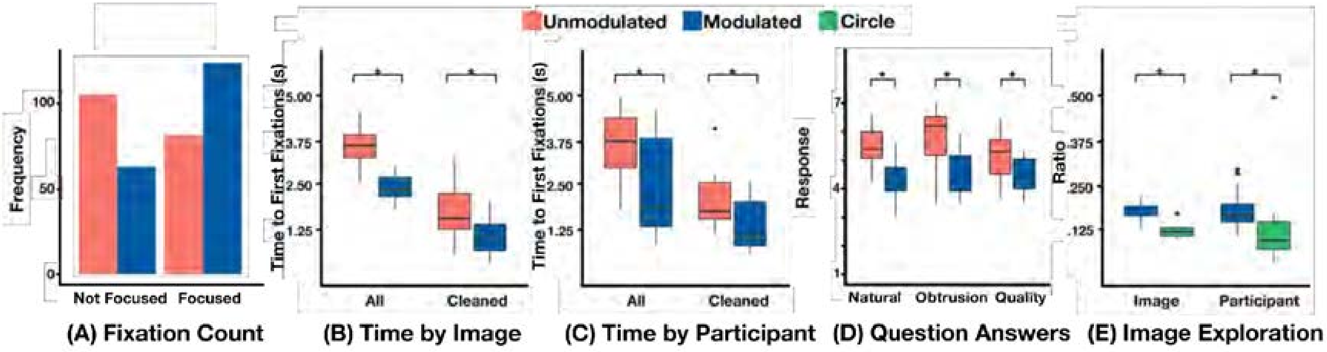


Figure 8: Resulting gaze heatmaps for 3 images. (A1,B1,C1) are the original images, (A2,B2,C2) are the heatmaps in the unmodified condition, (A3,B3,C3) are the heatmaps in the modified condition, (A4,B4,C4) are the heatmaps from the circle condition. Circles indicate the target area.

To investigate if participants perceived a difference in the images based on the Likert-like scales, we evaluated the difference using a paired Wilcoxon signed-rank test on the average scores per user under both the saliency modulated and unmodulated conditions. We found significant differences in all of our metrics; naturalness was significantly reduced ($Z = -3.530$, $p < 0.001$, $r = 0.75$) from a mean of 5.335 ($\sigma = 0.979$) to 4.305 ($\sigma = 0.750$), obtrusion was reduced ($Z = -3.723$, $p < 0.001$, $r = 0.79$) from a mean of 5.78 ($\sigma = 1.008$) to 4.58 ($\sigma = 0.797$), and quality was reduced ($Z = -3.530$, $p < 0.001$, $r = 0.75$) from a mean of 5.15 ($\sigma = 0.754$) to 4.48 ($\sigma = 0.556$) (See Fig. 9(d)). We subsequently evaluated the unmodulated and modulated conditions for each image individually, using paired Wilcoxon tests. Here we found significant differences between the two conditions for the naturalness of $\frac{5}{10}$ images, obtrusion for $\frac{8}{10}$ of the images and quality for $\frac{3}{10}$.



(A) Fixation Count (B) Time by Image (C) Time by Participant (D) Question Answers (E) Image Exploration

Figure 9: Quantitative results of our study for the Unmodulated (red), Modulated (blue), and the Circle (green) conditions. The number of fixations at the target area (A), time to first fixation averaged by image (B) and participant (C), Likert-like scale answers (D), and exploration area averaged by image and participant (E). *Cleaned* includes only observations where participants fixated at the target area and *All* sets the fixation time for the remaining participants to the maximum observation time (5 seconds). Asterisks indicate significant differences ($p < 0.05$) in the comparisons denoted by the bars.

To determine how much saliency modulation affects the participant's exploration of the images when compared against circle overlays, we compared the area covered by the heatmaps generated from the participant gaze data (See Fig. 8). Once again, we compare the covered areas by averaging the data for each participant and each image. When averaging the generated heatmaps for each participant. Participants explored a much larger portion of the image in the saliency modulated condition (M=0.18, SD=0.02) than the circle condition (M=0.12, SD=0.02) ($t(9) = 6.484, p < 0.001, d = 0.61; CI$ 0.036-0.078). We found differences also remained when we average the generated heatmaps for each image ($Z = -3.035, p = 0.002, r = 0.95$) (See Fig. 9(e)).

Discussion: For H1, our results show that saliency modulation via our algorithm on AR glasses not only increases the likelihood that participants look at a target area but also do so faster. We therefore can accept hypothesis H1.

However, we should point out that similar to existing works using saliency modulation (e.g., [63]) saliency modulation does not ensure fixations. Thus, not all participants fixated on the target. This could be due to the inherent saliency of different scenes, the limited time given to users to explore each scene, and top-down influences while exploring an environment. For example, we observed that some participants tended to focus more on the centre of the scene, while others tried to explore as much as possible. For some scenes, the number of fixations remained low, with only 5-6 participants focusing on it, while for others it was as high as 15 and improved by as little as 2 and as much as 10. In only one case, 2 participants fixated less onto the target area. We should also state again that images were shown exactly 5s and we cannot predict if or when participants would have gazed at the target after that. However, we would argue if participants gaze at the target after 5s the relevance for mentioned practical applications is relatively low.

For H2, based on the answers in our questionnaires, we must reject H2 as scores were significantly different. While this was the case over all images, it was not the case for each image which would indicate that there is potential here for further improvements. In particular, we saw a high number of images rated as significantly more obtrusive, whilst only three images were considered to have significantly reduced quality. Naturalness was split evenly. Although we do have a significant difference for all our metrics, all 3 show a mean reduction of only about one step on our scale and whilst our reduction is significant, we do not step past the neutral point. This implies that the differences created, whilst making the images less natural, more obtrusive, and lower quality, we do not expressly make them unnatural, obtrusive, or low quality.

We also should point out that we considered three images as challenging because the masked area was small, or the images were very salient before modulation. However, the results are mixed, and we would argue for more research also considering extreme images, but we see a trend that the amount of saliency modulation required to detract from an already very salient image can lead to artefacts that are considered more obtrusive.

From the feedback of the participants, we noted that they considered our saliency modulation to be part of the static image. In contrast, they considered circles as an overlay or separate to the image, which is an interesting aspect generally supporting the concept of saliency modulation. Participants also noted that whilst the circles readily drew their attention and showed the areas of the image, which they liked for a short viewing of an image, they would not like to have this done constantly.

For H3, our aim for saliency modulation was also to enable a more natural exploration of the scene while directing the user's gaze to the area of interest. Thus, we compared how our modulation affects the gaze behaviour compared to a traditional circle overlay. We found that when presented with our modulation participants explored a much larger portion of the image, thus supporting our hypothesis H3. The presented circle overlay created a very strong anchoring effect, almost gluing the user's gaze to it. As such, although participants were given the instruction to explore the image for all conditions, the overlaid circle significantly hindered their exploration of the scene. These findings support our hypothesis H3. As our modulation did not create a similar anchoring effect, this could also explain why we did not detect fixations on the target areas from some participants. Overall, our saliency modulation

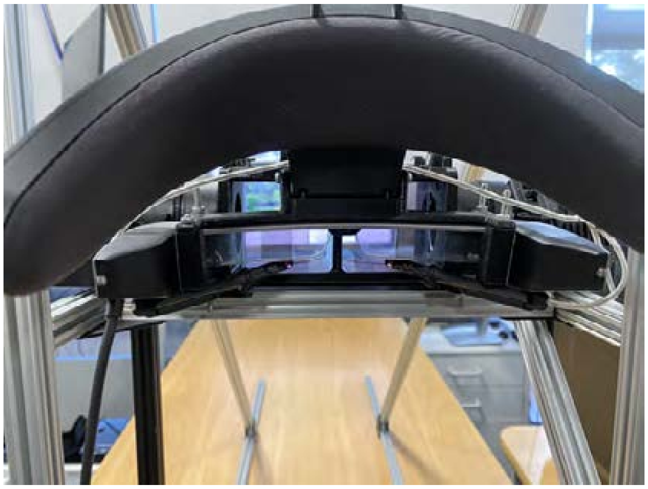


Figure 10: Study apparatus consisting of our stereoscopic prototype with an integrated Pupil-labs eye tracker, with head-band, and chin-rest (not visible).

using AR glasses seems preferable when the goal is to attract the user's attention while allowing natural scene exploration, and a traditional AR overlay is preferable when the user's attention needs to be guided to a critical area.

6 Direct See-Through Verification

Despite the lower accuracy of the external eye tracker, we also wanted to verify that we could expect to see a similar effect on user's gaze when directly looking through our stereo prototype and modulating the scene saliency. We did so by running a replication of the core component of our earlier study, this time with our stereo prototype that removes the constraints of an user-perspective camera and a VR headset but at the cost of a less controlled study.

Design: We replicated the within-subject study design from our prior study to investigate the effect of saliency modulation. Participants observed views of the images from the image data-set in each of two initial conditions (*unmodulated* and *saliency modulation via AR glasses*) in randomised order. To address Covid restrictions (e.g., need for N95 masks, limited number of participants), and as we only aimed to show general trends from the prior study, we ran a shorter study (e.g., not collected per image questions in this study). This also removed the need for participants to talk throughout, reducing head movement and affecting eyetracking that was already affected by the need to wear N95 masks during the actual study. We consequently also only tested two conditions: 1) *Unmodulated*, and 2) *Saliency modulation via AR glasses*.

Apparatus: We utilised the stereoscopic prototype for this study. Participants heads were secured with a head band adapted from an Oculus Quest and a chin rest was used to support them. We utilised a Pupil-labs integration of their eye tracker for the Epson BT 300 (see Figure 10) and the Pupil-labs pupil service to record all eye gaze data.

Procedure: As per the prior study required precautions were taken but had to be adapted to current Covid guidelines (e.g., mask wearing), and user signed consent forms and filled out demographic forms. The participants were then introduced to the apparatus and seated comfortably in it. Calibration was first completed for the eye-tracker using the pupil labs calibration methods. We verified this was calibrated within 2.5 degrees of accuracy as this the expected reliably achievable accuracy according to Pupil-Labs. They then completed an eye-display calibration for each eye and verified that the calibration was correct for stereo vision. Once completed the participant was shown a white cross to centre their gaze in the image. The participant was then shown each image for five seconds with the cross used to centre their gaze between each. The eye tracker was re-calibrated every 10 images.

Hypotheses: Due to the limitations of our apparatus and the results we could expect to elicit, we did not expect to be able to directly replicate our results, however we expected that our observations would follow similar trends and align with prior results. We hypothesised that:

- H4: Real-world saliency modulation via AR glasses will increase the number of participants who fixate on a target area when compared to an unmodified scene
- H5: Real-world saliency modulation via AR glasses will decrease the time taken by participants to fixate on a target area when compared to an unmodified scene.

As we were not collecting per image questions to mitigate user movement during the study we did not formulate a hypothesis around the impact on scene ratings.

Participants: We recruited 15 participants for our study, however due to instability in the eye tracking caused by different factors (head movement with respect to the eye tracker, mask wearing, different eye tracker) we only received usable data from 8 of these (1 female, 7 male, age ranging from 21 to 36, $\overline{x}$ = 28.2). All participants had normal vision or corrected to normal via contact lenses. Eye tracking was verified to an $\overline{x} = 2.38^o$ and $\theta = 0.28^o$.

Results: We again identified a participant as having fixated on a target area when any gaze point associated with a fixation lay within the target area. When detecting fixations we used the fixation detection provided by the Pupil-Labs player service. To minimise the impact of rotational errors introduced into the gaze data by our apparatus we used the centering point shown before each image as a reference for induced offsets. To enable comparisons to our main user study we utilised the same analysis as used there. We evaluated the number of fixations using a McNemar test and considered the time to fixation on both *cleaned* data where we only included the data points where fixations were recorded, as well as setting instances where fixations where not recorded at the maximum view time *all*. We considered time to first fixation by both participant and image.

Our results show a significantly increased number of fixations in the target area in the modulated condition when compared to the unmodulated condition according to a McNemar test (χ^2 = 6.8182, $p < 0.01$) (See Fig. 11(a)).

Looking at the time to first fixation by image we do find a significant difference when analysing *all* ($t(9)$ = 3.7312, p < 0.005, *CI* 0.33 – 1.365) (See Fig. 11(b)) but see no significant difference in the *cleaned* analysis ($t(9)$ = 1.6036, p = 0.1433, *CI* -0.1563945 – 0.9180341) but

Similarly, looking at it by participant we did not see significant differences in either the *cleaned* ($t(7)$ = 1.1577, p = 0.85, *CI* -0.29

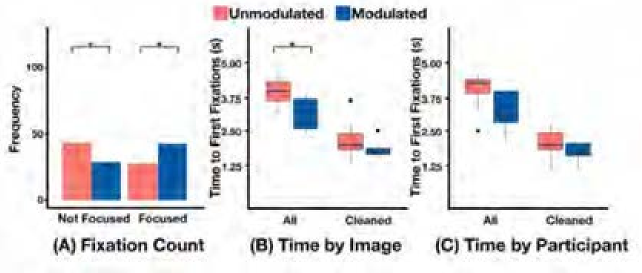


Figure 11: Quantitative results of our study for the Unmodulated (red) and Modulated (blue) conditions. The number of fixations at the target area (A), and time to first fixation averaged by image (B) and participant (C). *Cleaned* includes only observations where participants fixated at the target area and *All* sets the fixation time for the remaining participants to the maximum observation time (5 seconds). Asterisks indicate significant differences ($p < 0.05$) in the comparisons denoted by the bars.

– 0.84) or *all* (V = 32, p = 0.055, CI - 0.012 – 1.65) analyses (See Fig. 11(c)).

Discussion: Overall, this study confirms H1 as, similar to our initial study, we see fixations or see more fixations when the saliency of the scene is modulated via the AR glasses. However, with respect to H2 we see the results as inconclusive. While we see a trend in the data that is even significant in some tests (e.g., significantly faster fixations over all images when using the tests used in the literature [63]) but it is not confirmed in other tests (e.g., not significantly faster fixations using the cleaned data as proposed by us). We think there are various reasons for this: We already acknowledged that the eye tracking setup used for this study is more error prone and we only had usable data from a smaller number of participants. We think that it is even more relevant that, contrary to the well controlled initial study, there was still visual information in the users visual periphery (e.g., lights from the lab) as we did not completely darkened the study environment. When looking into the data we think that this has reduced the difference in time of first fixations requiring more participants and more reliable eye tracking to verify.

In conclusion, we still see an overall agreement with out initial study but also argue that the differences are not as readily demonstrated by less-controlled studies.

7 Discussion and Future Work

In this work, we investigated real-world visual guidance using saliency modulation in optical see-through Augmented Reality glasses. To the best of our knowledge this is the first exploration of practical saliency modulation using AR glasses including the first presenting an actual prototypical implementation and a user study.

Summary: We developed several prototypes that at their core are based on commercially available OSTHMDs but extended them by integrating scene cameras via beamsplitters that capture the world as seen by the user. This is needed to accurately capture the user's perspective of the world, to compute the saliency modulation overlay, and its correct placement. The prototypes include a functional stereoscopic prototype and a bench prototype where users see through the prototype via a camera at the position of the user's eye (user perspective camera). The latter being needed for utilising the integrated eye tracker in the VR HMD that is less error prone and run a study in a controlled study environment reducing confounding variables (e.g., calibration inaccuracy). In addition, we developed a working mobile prototype that we did not further address (see Figure 12 C). The main reason is that there is some perceivable latency (300-500ms), and maintaining a fixed calibrated position on the user's head is challenging. Finally, because the used OSTHMD in the mobile prototype is colour sequential, it is not well suited for capturing results. All our prototypes come with an algorithm for modulating real-world saliency that considers the specific nature of OSTHMDs and their ability only to add light intensities per colour channel, something widely ignored in previous work.

We explored our approach using images from a widely used image database also providing saliency data. The results show that we can guide the user's gaze towards relevant areas with significant effects in time for first fixation and number of actual fixations. These results come from a controlled lab environment with our prototype minimising confounding variables but the results are supported in a second study using the stereoscopic prototype, even though the effect in time to first fixations wasn't as clear there.

Whilst previous research on saliency modulation states that they are able to "imperceptibly" modify image material to change the saliency [63], we did not observe this effect with our approach when modulating the saliency via our AR glasses prototype as questionnaires showed significant differences in perceived image characteristics (e.g., naturalness and unobtrusiveness). After revisiting earlier studies, we would, in general, be careful with targeting the objective of imperceptibility as prior work showed only feeble evidence and critical image material seemed to be not considered. In particular, results suggest that for scenes that are already visually salient, strong saliency modulation is required which is often perceptible. We essentially argue that the line between imperceptible but effective saliency modulation is so thin (much thinner than previously indicated) that it is hard to generalise and, if possible, this imperceptible saliency modulation demands specific knowledge about the scene and parameters tuned for that scene (context-aware AR [14]) which might be hard to realise in interactive non-controlled environments.

Applications: Overall, we believe in the potential of real-world saliency modulation. Not only because we can show a different gaze behaviour, but also because users seemed to look at the targeted area whilst not being overly distracted by it. This was indicated by the gaze analysis, which showed that other image areas have still been explored, something that is relevant when modulating the real world.

Generally speaking, we see most applications in guiding or highlighting information in the user's context. This was also the main direction of our work. This includes guidance during surgery, where occluding areas can be critical. Similarly, we see applications in guiding and navigation scenarios where introducing additional visual cues might occlude information or introduce visual clutter (see 12 B)). We also see the potential for general AR applications in

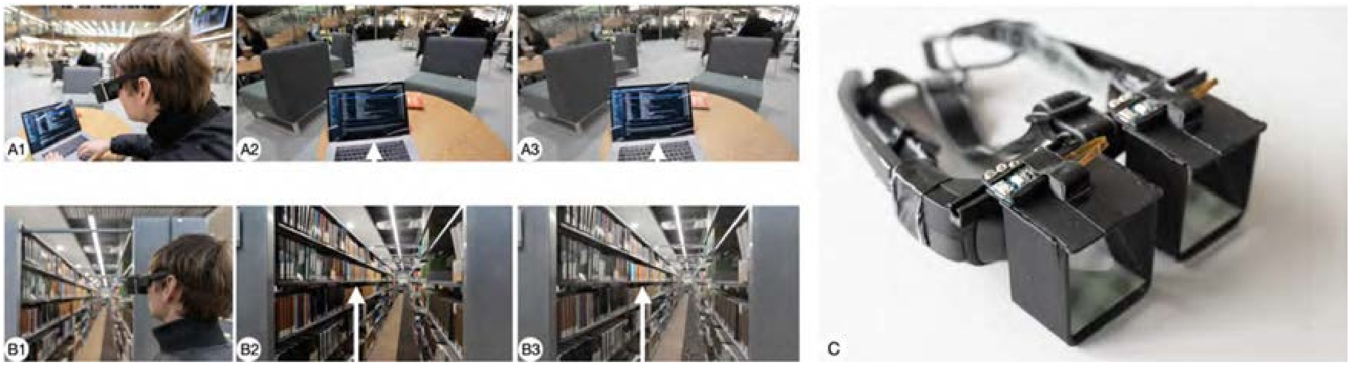


Figure 12: Two conceptual scenarios showing how the concept could be used for focusing on a main task (A) or finding objects (B). Our approach captures the real-world (A2 and B2) and modulates the saliency (A3 and B3) to guide the user's gaze to modulated scene elements (here simulated output of the computer and books on the shelf subtly modulated with our algorithm, as intended for an unobtrusive guide). The white arrow pointing out the emphasised area is for illustration only. (C) Working mobile prototype combining a Lumus DK52 with integrated scene cameras for saliency modulation.

which we can compute the visual saliency of the scene after the AR overlay of new scene elements and correct for it according to the current requirements. Finally, we still think the concept of visual noise cancellation is a strong and interesting concept with applications in many directions (including medical)(see 12 A)). However, it was not necessarily a focus within this work.

Limitations and future work: As the first demonstration of visual guidance using saliency modulation via AR glasses, there are several limitations. The first limitation is that the results are not as readily demonstrated by less-controlled studies as confirming faster fixations in the stereo prototype was inconclusive. We think there are multiple reasons for this but it comes down to having more error sources including a more challenging calibration. However, our approach of mainly relying on bench prototypes is relatively common within the discipline as it allows for the exclusion of several confounding variables (e.g., quality of the individual eye-display calibration). In fact while this could be seen as a limitation relying mainly on bench prototypes is actually a strength: All participants viewed the same content and this allowed us to circumvent potentially confounding variables, like incorrect alignment of the virtual content with the scene, different colour aberrations for each eye, refocusing between the virtual content and the scene, and bad calibration.

Concerning the algorithm, we tried to linearly combine parameters into one parameter, which is used for all image material. However, one could improve the results by optimising the combination and level parameters based on context. To our best knowledge, we are not aware of such an approach, but it seems possible. We also used the masking style from related work (e.g., [63]) but noticed the masks have a very short ramp between the emphasised area and the surroundings and future work could optimise it with likely improved results, particularly in the noticeability of the modulation.

We have also limited ourselves to using an existing image dataset with saliency data. While we chose a wide range of scenes, including some we consider as challenging, the dataset is limited in comparison to the real world. Similarly, we decided to only change the saliency of static images via the AR glasses. This is similar to previous work (e.g., [3, 13]) and not a limitation of our algorithm. In fact, our approach works in real-time and it should be possible to implement it with similar performance on the latest mobile devices. However, using moving scenes as the material would have even further complicated the analysis of the results as finding good scenes with ground truth data (e.g., gaze) is challenging while other cues (dynamic cues such as moving objects) might introduce additional challenges.

There are also limitations to the prototypes used that need to be considered. Alongside the mentioned reasons to use static images, the latency of the current prototypes would have also introduced problems with misalignment when scene content changes. This was largely a problem for our prototypes due to the display system being used. However, for any practical application the motion-to-photon latency would need to be considered and either reduced to a level below human cognition as in active research prototypes [22] or for the impact to be measured and accounted for. Another limitation to note is that for our prototypes the scene cameras are placed on a fixed axis and therefore if the user's eyes deviate from this central axis slight errors will be introduced to the modulations alignment. The design of our main efficacy study avoids this issue by using a static camera, and the limited field of view covered generally minimises the impact of this factor in our prototypes, however consideration of the user's eye position is required for precise modulation, particularly in with a wider field of view display.

As a further consideration, all current OSTHMDs have a fixed focal plane, or two in the case of the Magic Leap. Thus, our modulation mask is "sharp" when focusing on a fixed plane. The displayed images and our focal plane were not at the same distance but also not too far off. In more practical scenarios, not having the modulation mask on the focus plane would cause the mask to be slightly blurred, impacting the precision of the modulations and may introduce slight artefacts in areas where the modulation does not correctly align with the world. This could impact the resulting visual guidance. There are research prototypes that support multiple focal planes. Our approach would benefit from them, but they are far away from being commercially available. Similarly, there is research on OSTHMDs that allows environment modulation by subtracting light intensities [21]. The forthcoming Magic Leap 2 is also given to have a subtractive element using a dimmer layer. This cannot achieve the same pixel resolution as the additive component, and does not create a sharp reduction, therefore does not allow the precise modulations that would be desired for modulating the user's view. Whilst the exact implementation and therefore its full implications, such as on overall light transmittance, are unknown, having a subtractive would enable further scope for modulations. The subtractive elements could be reintroduced into our technique to exploit this and subtler modulations more in line with the prior works [63] may be achievable, although we stand by our prior caution.

Finally, given the small effect size of our results, a greater number of participants beyond the 28 from across our two studies would provide greater strength to our findings.

A successful saliency modulation with optical see-through AR also brings professional and ethical responsibilities for designers, developers, and researchers, amongst others. If our techniques turn out to be highly effective and are used in pervasive AR settings [14], i.e., omnipresent, environmentally adaptive reality augmentation, then careful consideration should be taken in particular regarding health and safety, privacy issues, and produced illusion and belief [48]. We should design our visual modulations in such a way that naturalness and unobtrusiveness can be controlled for the given task, user, and environment. The degree to which we control the perceived difference between reality and virtual reality can lead to unwanted side effects but can also lead to new and meaningful experiences in a host of applications. Visual guidance with saliency modulation can play a major role here.

Overall, our work shows the potential and the practical issues for real-world visual guidance using saliency modulation in AR glasses. This is an important achievement given that prior work has raised the conceptual idea but never actually explored the implementation of saliency modulation via AR glasses or OSTHMDs. Thus, practical issues such as how to capture the environment and modulate it via AR glasses or the limited range for modulating images by only adding light via the AR glasses were not mentioned. Our research findings are thus of relevance for the HCI and AR communities with potential for future work in designing, developing, and comparing novel visual guidance techniques.

Acknowledgments

Tobias Langlotz and Jonathan Sutton are supported by the Marsden Fund Council from Government funding (grant no. MFP-UOO2124).

References

[1] Euijai Ahn, Sungkil Lee, and Gerard Jounghyun Kim. 2018. Real-time adjustment of contrast saliency for improved information visibility in mobile augmented reality. *Virtual Reality* 22, 3 (2018), 245–262.

[2] Kayo Azuma and Hideki Koike. 2018. A Study on Gaze Guidance Using Artificial Color Shifts. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces* (Castiglione della Pescaia, Grosseto, Italy) *(AVI '18)*. Association for Computing Machinery, New York, NY, USA, Article 47, 5 pages. https://doi.org/10.1145/3206505.3206517

[3] Reynold Bailey, Ann McNamara, Nisha Sudarsanam, and Cindy Grimm. 2009. Subtle gaze direction. *ACM Transactions on Graphics* 28, 4 (2009), 1–14. https://doi.org/10.1145/1559755.1559757

[4] Frank Biocca, Charles Owen, Arthur Tang, and Corey Bohil. 2007. Attention Issues in Spatial Information Systems: Directing Mobile Users' Visual Attention Using Augmented Reality. *Journal of Management Information Systems* 23, 4 (2007), 163–184. https://doi.org/10.2753/mis0742-1222230408

[5] Frank Biocca, Arthur Tang, Charles Owen, and Fan Xiao. 2006. Attention Funnel: Omnidirectional 3D Cursor for Mobile Augmented Reality Platforms. In *Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI '06*. Association for Computing Machinery, New York, NY, USA, 1115. https://doi.org/10.1145/1124772.1124939

[6] Thomas Booth, Srinivas Sridharan, Ann McNamara, Cindy Grimm, and Reynold Bailey. 2013. Guiding Attention in Controlled Real-World Environments. In *Proceedings of the ACM Symposium on Applied Perception* (Dublin, Ireland) *(SAP '13)*. Association for Computing Machinery, New York, NY, USA, 75–82. https://doi.org/10.1145/2492494.2492508

[7] Ali Borji and Laurent Itti. 2015. CAT2000: A Large Scale Fixation Dataset for Boosting Saliency Research. *CVPR 2015 workshop on "Future of Datasets"* 0, 0 (2015), 4 pages. arXiv preprint arXiv:1505.03581.

[8] P. Chakravarthula, D. Dunn, K. Akşit, and H. Fuchs. 2018. FocusAR: Auto-focus Augmented Reality Eyeglasses for both Real World and Virtual Imagery. *IEEE TVCG* 24, 11 (2018), 2906–2916. https://doi.org/10.1109/TVCG.2018.2868532

[9] Praneeth Chakravarthula, Ethan Tseng, Tarun Srivastava, Henry Fuchs, and Felix Heide. 2020. Learned Hardware-in-the-loop Phase Retrieval for Holographic Near-Eye Displays. *ACM Transactions on Graphics (TOG)* 39, 6 (2020), 186.

[10] M. Cornia, L. Baraldi, G. Serra, and R. Cucchiara. 2018. Predicting Human Eye Fixations via an LSTM-Based Saliency Attentive Model. *IEEE Transactions on Image Processing* 27 (2018), 5142–5154.

[11] J. L. Gabbard, J. E. Swan, J. Zedlitz, and W. W. Winchester. 2010. More than meets the eye: An engineering study to empirically examine the blending of real and virtual color spaces. In *IEEE VR*. IEEE, Boston, MA, USA, 79–86. https://doi.org/10.1109/VR.2010.5444808

[12] R. Grasset, T. Langlotz, D. Kalkofen, M. Tatzgern, and D. Schmalstieg. 2012. Image-driven view management for augmented reality browsers. In *Mixed and Augmented Reality (ISMAR), 2012 IEEE International Symposium on*. IEEE, Atlanta, GA, USA, 177–186. https://doi.org/10.1109/ISMAR.2012.6402555

[13] Steve Grogorick, Michael Stengel, Elmar Eisemann, and Marcus Magnor. 2017. Subtle Gaze Guidance for Immersive Environments. In *Proceedings of the ACM Symposium on Applied Perception* (Cottbus, Germany) *(SAP '17)*. Association for Computing Machinery, New York, NY, USA, Article 4, 7 pages. https://doi.org/10.1145/3119881.3119890

[14] J. Grubert, T. Langlotz, S. Zollmann, and H. Regenbrecht. 2017. Towards Pervasive Augmented Reality: Context-Awareness in Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics* 23, 6 (2017), 1706–1724.

[15] Akira Hagiwara, Akihiro Sugimoto, and Kazuhiko Kawamoto. 2011. Saliency-Based Image Processing for Guiding Visual Attention. In *PETMEI*. Association for Computing Machinery, New York, NY, USA, 1–8.

[16] Hajime Hata, Hideki Koike, and Yoichi Sato. 2016. Visual Guidance with Unnoticed Blur Effect. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (Bari, Italy) *(AVI '16)*. Association for Computing Machinery, New York, NY, USA, 28–35. https://doi.org/10.1145/2909132.2909254

[17] X. Hou and L. Zhang. 2007. Saliency Detection: A Spectral Residual Approach. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, New York, NY, USA, 1–8.

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

[19] Y. Itoh, T. Hamasaki, and M. Sugimoto. 2017. Occlusion Leak Compensation for Optical See-Through Displays Using a Single-Layer Transmissive Spatial Light Modulator. *IEEE TVCG* 23, 11 (2017), 2463–2473. https://doi.org/10.1109/TVCG.2017.2734427

[20] Y. Itoh, T. Langlotz, D. Iwai, K. Kiyokawa, and T. Amano. 2019. Light Attenuation Display: Subtractive See-Through Near-Eye Display via Spatial Color Filtering. *IEEE Transactions on Visualization and Computer Graphics* 25, 5 (May 2019), 1951–1960. https://doi.org/10.1109/TVCG.2019.2899229

[21] Y. Itoh, T. Langlotz, D. Iwai, K. Kiyokawa, and T. Amano. 2019. Light Attenuation Display: Subtractive See-Through Near-Eye Display via Spatial Color Filtering. *IEEE TVCG* 25, 5 (May 2019), 1951–1960. https://doi.org/10.1109/TVCG.2019.2899229

[22] 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 (jul 2021), 36 pages. https://doi.org/10.1145/3453157

[23] L. Itti, C. Koch, and E. Niebur. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20 (Nov 1998), 1254–1259. https://doi.org/10.1109/34.730558

[24] Anton S. Kaplanyan, Anton Sochenov, Thomas Leimkühler, Mikhail Okunev, Todd Goodall, and Gizem Rufo. 2019. DeepFovea: Neural Reconstruction for Foveated Rendering and Video Compression Using Learned Statistics of Natural

Videos. *ACM Trans. Graph.* 38, 6, Article 212 (nov 2019), 13 pages.

[25] Youngmin Kim and Amitabh Varshney. 2008. Persuading Visual Attention Through Geometry. *IEEE Transactions on Visualization and Computer Graphics* 14, 4 (July 2008), 772–782. https://doi.org/10.1109/TVCG.2007.70624

[26] C. Koch and S. Ullman. 1987. Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry. *Matters of Intelligence* 4, 4 (1987), 219—-227.

[27] Tatsuhiko Kokui, Hironori Takimoto, Yasue Mitsukura, Mitsuyoshi Kishihara, and Kensuke Okubo. 2013. Color image modification based on visual saliency for guiding visual attention. In *2013 IEEE RO-MAN*. IEEE, New York, NY, USA, 467–472. https://doi.org/10.1109/ROMAN.2013.6628548

[28] Alexander Kroner, Mario Senden, Kurt Driessens, and Rainer Goebel. 2019. Contextual Encoder-Decoder Network for Visual Saliency Prediction. *CoRR* abs/1902.06634 (2019), 261–270. arXiv:1902.06634

[29] Tobias Langlotz, Jonathan Sutton, Stefanie Zollmann, Yuta Itoh, and Holger Regenbrecht. 2018. ChromaGlasses : Computational Glasses for Compensating Colour Blindness. In *CHI '18 Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–12.

[30] Y.-C. Lin, Y.-J. Chang, H.-N. Hu, H.-T. Cheng, C.-W. Huang, and M. Sun. 2017. Tell me where to look: Investigating ways for assisting focus in 360° video. *Conference on Human Factors in Computing Systems - Proceedings* 2017-May (2017), 2535–2545.

[31] Weiquan Lu, Been-Lirn Henry Duh, and Steven Feiner. 2012. Subtle cueing for visual search in augmented reality. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Atlanta, GA, USA, 161–166.

[32] Weiquan Lu, Henry Been Lirn Duh, Steven Feiner, and Qi Zhao. 2014. Attributes of subtle cues for facilitating visual search in augmented reality. *IEEE Transactions on Visualization and Computer Graphics* 20, 3 (2014), 404–412. https://doi.org/10.1109/TVCG.2013.241

[33] Weiquan Lu, Dan Feng, Steven Feiner, Qi Zhao, and Henry Been-Lirn Duh. 2013. Subtle cueing for visual search in head-tracked head worn displays. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, New York, NY, USA, 271–272. https://doi.org/10.1109/ISMAR.2013.6671800

[34] W Steve G Mann. 2001. Wearable camera system with viewfinder means. US Patent 6,307,526.

[35] Victor Mateescu and Ivan Bajic. 2014. Attention Retargeting by Color Manipulation in Images. *Proceedings of the 1st International Workshop on Perception Inspired Video Processing* 1 (2014), 15–20. https://doi.org/10.1145/2662996.2663009

[36] Victor A. Mateescu and Ivan V. Bajić. 2014. Can Subliminal Flicker Guide Attention in Natural Images?. In *Proceedings of the 1st International Workshop on Perception Inspired Video Processing* (Orlando, Florida, USA) *(PIVP '14)*. Association for Computing Machinery, New York, NY, USA, 33–34. https://doi.org/10.1145/2662996.2663012

[37] Ann McNamara, Reynold Bailey, and Cindy Grimm. 2008. Improving Search Task Performance Using Subtle Gaze Direction. In *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization (APGV '08)*. ACM, New York, NY, USA, 51–56. https://doi.org/10.1145/1394281.1394289

[38] Erick Mendez, Steven Feiner, and Dieter Schmalstieg. 2010. Focus and context in mixed reality by modulating first order salient features. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 6133 LNCS (2010), 232–243. https://doi.org/10.1007/978-3-642-13544-6_22

[39] Xiaoxu Meng, Ruofei Du, Matthias Zwicker, and Amitabh Varshney. 2018. Kernel Foveated Rendering. *Proc. ACM Comput. Graph. Interact. Tech.* 1, 1, Article 5 (jul 2018), 20 pages. https://doi.org/10.1145/3203199

[40] Junpei Miyamoto, Hideki Koike, and Toshiyuki Amano. 2018. Gaze Navigation in the Real World by Changing Visual Appearance of Objects Using Projector-Camera System. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology* (Tokyo, Japan) *(VRST '18)*. Association for Computing Machinery, New York, NY, USA, Article 14, 5 pages. https://doi.org/10.1145/3281505.3281537

[41] Tam V. Nguyen, Bingbing Ni, Hairong Liu, Wei Xia, Jiebo Luo, Mohan Kankanhalli, and Shuicheng Yan. 2013. Image re-attentionizing. *IEEE Transactions on Multimedia* 15, 8 (2013), 1910–1919. https://doi.org/10.1109/TMM.2013.2272919

[42] Anneli Olsen. 2012. The Tobii I-VT fixation filter. *Tobii Technology* 0, 0 (2012), 1–21.

[43] Rajarshi Pal and Dipanjan Roy. 2018. Enhancing Saliency of an Object Using Genetic Algorithm. *Proceedings - 2017 14th Conference on Computer and Robot Vision, CRV 2017* 2018-January (2018), 337–344. https://doi.org/10.1109/CRV.2017.33

[44] Xufang Pang, Ying Cao, Rynson W. H. Lau, and Antoni B. Chan. 2016. Directing user attention via visual flow on web designs. *ACM Transactions on Graphics* 35, 6 (2016), 1–11. https://doi.org/10.1145/2980179.2982422

[45] Derrick Parkhurst, Klinton Law, and Ernst Niebur. 2002. Modeling the role of salience in the allocation of overt visual attention. *Vision Research* 42, 1 (2002), 107 – 123. https://doi.org/10.1016/S0042-6989(01)00250-4

[46] Eli Peli, Gang Luo, Alex Bowers, and Noa Rensing. 2007. Applications of Augmented Vision Head-Mounted Systems in Vision Rehabilitation. *Journal of the Society for Information Display* 15, 12 (2007), 1037–1045. https://doi.org/10.1889/1.2825088 arXiv:NIHMS150003

[47] E. Ragan, C. Wilkes, D. A. Bowman, and T. Hollerer. 2009. Simulation of Augmented Reality Systems in Purely Virtual Environments. In *2009 IEEE VR*. IEEE, New York, NY, USA, 287–288.

[48] H. Regenbrecht, S. Zwanenburg, and T. Langlotz. 5555. Pervasive Augmented Reality - Technology and Ethics. *IEEE Pervasive Computing* 1, 1 (mar 5555), 1–8.

[49] Sylvia Rothe, Felix Althammer, and Mohamed Khamis. 2018. GazeRecall: Using Gaze Direction to Increase Recall of Details in Cinematic Virtual Reality. In *MUM'18*. Association for Computing Machinery, New York, NY, USA, 115–119. https://doi.org/10.1145/3282894.3282903

[50] Björn Schwerdtfeger and Gudrun Klinker. 2008. Supporting order picking with augmented reality. *Proceedings - 7th IEEE International Symposium on Mixed and Augmented Reality 2008, ISMAR 2008* 1, 1 (2008), 91–94. https://doi.org/10.1109/ISMAR.2008.4637331

[51] Tao Shi and Akihiro Sugimoto. 2015. Video Saliency Modulation in the HSI Color Space for Drawing Gaze. *PSIVT* 8333, July 2015 (2015), 206–219. https://doi.org/10.1007/978-3-642-53842-1

[52] Srinivas Sridharan, Ann McNamara, and Cindy Grimm. 2012. Subtle gaze manipulation for improved mammography training. *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA '12* 1, 212 (2012), 75.

[53] Sara L. Su, Frédo Durand, and Maneesh Agrawala. 2005. De-Emphasis of Distracting Image Regions Using Texture Power Maps. In *Proceedings of the 2nd Symposium on Applied Perception in Graphics and Visualization* (A Coroña, Spain) *(APGV '05)*. Association for Computing Machinery, New York, NY, USA, 164. https://doi.org/10.1145/1080402.1080445

[54] Jonathan Sutton, Tobias Langlotz, and Yuta Itoh. 2019. Computational Glasses: Vision augmentations using computational near-eye optics and displays. In *2019 IEEE ISMAR-Adjunct*. IEEE, IEEE, New York, U.S., 438–442.

[55] Natsumi Suzuki and Yohei Nakada. 2018. Effects selection technique for improving visual attraction via visual saliency map. *2017 IEEE Symposium Series on Computational Intelligence, SSCI 2017 - Proceedings* 2018-January (2018), 1–8. https://doi.org/10.1109/SSCI.2017.8280808

[56] Hironori Takimoto, Syuhei Hitomi, Hitoshi Yamauchi, Mitsuyoshi Kishihara, and Kensuke Okubo. 2017. Image modification based on spatial frequency components for visual attention retargeting. *IEICE Transactions on Information and Systems* E100D, 6 (2017), 1339–1349. https://doi.org/10.1587/transinf.2016EDP7413

[57] Hironori Takimoto, Tatsuhiko Kokui, Hitoshi Yamauchi, Mitsuyoshi Kishihara, and Kensuke Okubo. 2015. Image modification based on a visual saliency map for guiding visual attention. *IEICE Transactions on Information and Systems* E98D, 11 (2015), 1967–1975. https://doi.org/10.1587/transinf.2015EDP7087

[58] Hironori Takimoto, Katsumi Yamamoto, Akihiro Kanagawa, Mitsuyoshi Kishihara, and Kensuke Okubo. 2019. Attention retargeting using saliency map and projector–camera system in real space. *IEEJ Transactions on Electrical and Electronic Engineering* 14, 6 (2019), 853–861. https://doi.org/10.1002/tee.22874

[59] Antonio Torralba, Aude Oliva, Monica Castelhano, and John Henderson. 2006. Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Pyschol Rev* 113, 4 (2006), 766–786.

[60] Anne M. Treisman and Garry Gelade. 1980. A feature-integration theory of attention. *Cognitive Psychology* 12, 1 (1980), 97 – 136. https://doi.org/10.1016/0010-0285(80)90005-5

[61] Mihran Tuceryan, Yakup Genc, and Nassir Navab. 2002. Single-point active alignment method (spaam) for optical see-through hmd calibration for augmented reality. *Presence: Teleoperators & Virtual Environments* 11, 3 (2002), 259–276.

[62] T. Ueda, D. Iwai, T. Hiraki, and K. Sato. 2020. Illuminated Focus: Vision Augmentation using Spatial Defocusing via Focal Sweep Eyeglasses and High-Speed Projector. *IEEE Transactions on Visualization and Computer Graphics* 26, 5 (May 2020), 2051–2061. https://doi.org/10.1109/TVCG.2020.2973496

[63] Eduardo E. Veas, Erick Mendez, Steven K. Feiner, and Dieter Schmalstieg. 2011. Directing Attention and Influencing Memory with Visual Saliency Modulation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 1471–1480. https://doi.org/10.1145/1978942.1979158

[64] N. Waldin, M. Waldner, and I. Viola. 2017. Flicker Observer Effect: Guiding Attention Through High Frequency Flicker in Images. *Computer Graphics Forum* 36, 2 (2017), 467–476. https://doi.org/10.1111/cgf.13141

[65] Manuela Waldner, Mathieu Le Muzic, Matthias Bernhard, Werner Purgathofer, and Ivan Viola. 2014. Attractive flicker-Guiding attention in dynamic narrative visualizations. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 2456–2465. https://doi.org/10.1109/TVCG.2014.2346352

[66] W. Wang, J. Shen, and L. Shao. 2018. Video Salient Object Detection via Fully Convolutional Networks. *IEEE Transactions on Image Processing* 27, 1 (2018), 38–49.

[67] Jeremy M. Wolfe and Todd S. Horowitz. 2004. What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience* 5, 6 (June 2004), 495–501.