



# Stereo Inverse Brightness Modulation for Guidance in Dynamic Panorama Videos in Virtual Reality

Steve Grogorick,  Jan-Philipp Tauscher,  Nikkel Heesen, Susana Castillo  and Marcus Magnor 

TU Braunschweig, Germany  
{grogorick, tauscher, heesen, castillo, magnor}@cg.cs.tu-bs.de

## Abstract

*The peak of virtual reality offers new exciting possibilities for the creation of media content but also poses new challenges. Some areas of interest might be overlooked because the visual content fills up a large portion of viewers' visual field. Moreover, this content is available in 360° around the viewer, yielding locations completely out of sight, making, for example, recall or storytelling in cinematic Virtual Reality (VR) quite difficult.*

*In this paper, we present an evaluation of Stereo Inverse Brightness Modulation for effective and subtle guidance of participants' attention while navigating dynamic virtual environments. The used technique exploits the binocular rivalry effect from human stereo vision and was previously shown to be effective in static environments. Moreover, we propose an extension of the method for successful guidance towards target locations outside the initial visual field.*

*We conduct three perceptual studies, using 13 distinct panorama videos and two VR systems (a VR head mounted display and a fully immersive dome projection system), to investigate (1) general applicability to dynamic environments, (2) stimulus parameter and VR system influence, and (3) effectiveness of the proposed extension for out-of-sight targets. Our results prove the applicability of the method to dynamic environments while maintaining its unobtrusive appearance.*

**Keywords:** assistive interfaces, interaction, perceptually based rendering, rendering, immersive VR, virtual environments

**ACM CCS:** • Human-centred computing → Virtual reality; Empirical studies in HCI; Empirical studies in interaction design; • Computing methodologies → Perception; Virtual reality; Graphics input devices

## 1. Introduction

Since its renaissance in 2013, driven by the release of the Oculus Rift DK1, VR has reached a broad audience. Various devices have been developed and released by many manufacturers, ranging from small start-ups to large-scale well-established tech companies. Besides expensive high-performance products that offer best visual experience, various low-budget versions have been released to introduce VR to a broader audience, yielding a massive amount of people that have, at least, tried some VR experience by now. Moreover, an ever increasing number of video games are released with support for immersive VR devices offering a large Field Of View (FOV), and even the movie industry has already presented first entire movies that are shot in 360°. So it seems like VR, this time, finally made its way out of the science-experimental-stage into the consumer market. This brings up new possibilities in numerous fields, like scientific visualization, architectural presentations, entertainment or even telepresence as in today's challenging times.

However, due to the absence of the frame that typically restricts viewers' attention to small pre-determined regions of the observed world, presentation of content in virtual environments is a non-trivial task. This can be further distinguished into two challenges onto which we will focus on in this work. First, the wide FOV—much wider than we are typically used to from desktop monitors or TV screens. Second, the possibility to literally turn around and look around. Of course this is not entirely new, as also in 3D video games one could always turn and look around, but video game developers have the option to force a camera reorientation when required. Unfortunately, redirecting the view point of the user is not as easy in VR environments, as externally enforcing virtual self-motion without matching the user's real body motion is known to increase simulator sickness [KLB\*89].

Therefore, appropriate means need to be developed to help viewers in immersive environments to find the right information at the right time. This yields the field of attention guidance, and more



**Figure 1:** Left and right eye images (try cross/parallel viewing) with SIBM stimulus. Intensity is enhanced for illustration purposes.

specifically in case of this paper, subtle visual gaze guidance, that is, guiding viewers' gaze using unobtrusive visual stimuli to not interfere with the actual scene content. Concrete applications that might profit from such mechanisms include highlighting details in cluttered three-dimensional visualizations of scientific data sets; barely visible elements of architectural concepts during virtual walkthroughs; or hard-to-track details in sports broadcasts, such as flying golf balls—without obscuring any adjacent parts or interfering with design intentions.

Stereo Inverse Brightness Modulation (SIBM) [GTA\*19], was specifically designed to overcome this issue in static scenarios, that is, SIBM is a method to visually guide viewers within their visual field when watching panoramic still images in VR. This method builds on an effect in human vision called binocular rivalry. It describes the situation when the perceived images of both eyes do not perfectly match, for example, if something is visible for the left eye but not for the right eye. For stereo vision, the human brain often profits from the occurrence of this phenomenon as an additional depth-cue along object boundaries. There, only one eye can see a focused object while the line-of-sight of the other eye is blocked by another closer object. SIBM elicits this effect by modulating the brightness of a specified circular target region, as shown in Figure 1. The circular region is brightened for one eye while it is darkened for the other. This unnatural discrepancy in visual input attracts our attention. To prevent strong interference with the actual scene content, the strength of the brightness modulation is chosen to only barely exceed the threshold of the effect being visible at all.

In this paper, we contribute a thorough investigation of the applicability of SIBM in *dynamic virtual environments*, via multiple perceptual studies. Therefore, in contrast to the original work [GTA\*19], the method is evaluated within panoramic videos instead of static panorama images. In detail, we investigate:

**Applicability:** We show general applicability of the method in dynamic virtual environments (in comparison to static ones).

**Parameter space exploration:** We explore the different parameter configurations and their effects towards achieving optimal balance between guiding performance and perceptual subtleness.

**Platform impact:** We analyse its usability in two different VR platforms, that is, a head mounted device, and a fully immersive dome projection system.

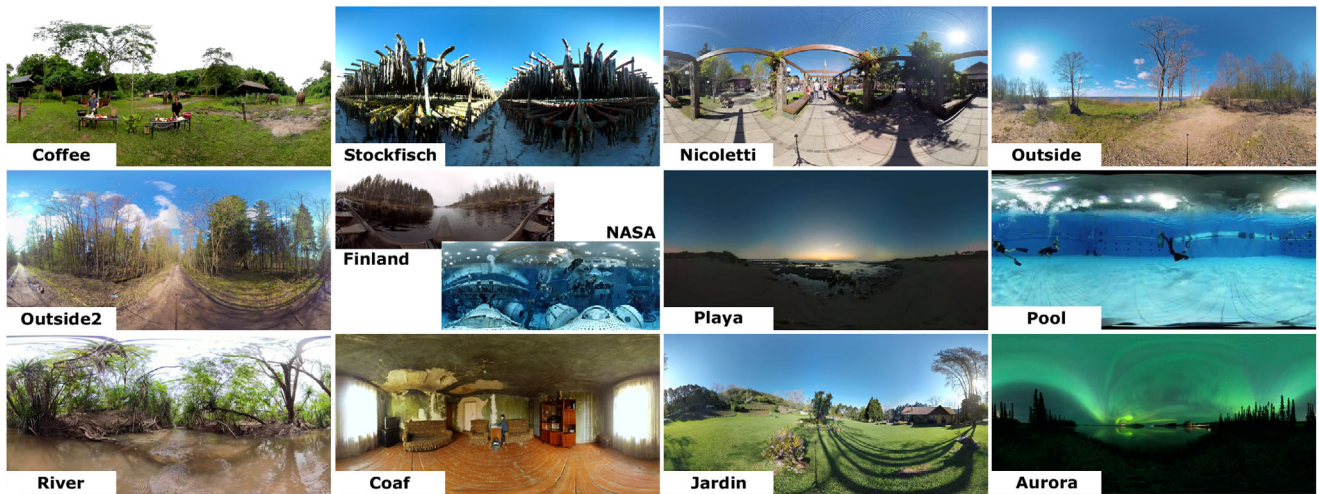
As a second contribution, we also propose an extension of the method to not only consider intra-FOV, but also *extra-FOV attention guidance*, that is, we address the problem of having the areas of interest located completely out of sight.

## 2. Related Work

In the field of visual attention guidance, many different methods have been proposed in recent years for application in images as well as videos; in non-immersive settings [DMGB04, BMSG09, VMFS11, HKS16], in VR settings using Head-Mounted Display (HMD) [LCH\*17, GSEM17, GAM18, RAK18, GTA\*19], and even in an immersive room-scale projection system [GATM18, GTA\*19]. Attention guidance as a support means may be helpful in various application scenarios, such as virtual training or remote teaching [dKJ17, FMS\*19, YKB19b, YKB19a], guided exploration [LH19], multi-monitor surveillance tasks [SKB19] or immersive story telling [SP19, SRD\*19, LSGB20]. For our survey on related work, we will distinguish between passive and active methods, that is, they do or do not actively incorporate real-time gaze tracking data.

Most passive approaches, which do not require the actual gaze direction, use global transformations to shift users' gaze towards target areas in an input image. Kosara *et al.* [KMH02] proposed to simulate depth-of-field from traditional photography, bringing non-target regions of the scene out of focus. Desaturating (and blurring) uninteresting parts of a scene was suggested by Cole *et al.* [CDF\*06]. Smith *et al.* [ST13] presented further investigations of the effectiveness of blur to redirect users' attention. A similar approach was proposed by Hata *et al.* [HKS16], that gradually blurs uninteresting regions while keeping target regions unblurred. Another set of transformations for attention guidance is based on computational visual saliency models. A saliency model estimates how well which image regions will attract viewers' attention [IKN98, SSP\*18]. Although Latif *et al.* [LGCM14] have shown local textural contrast enhancement in paintings to attract attention to the respective regions, the method proposed by Su *et al.* [SDA05] reduces spatial variations in textures of image regions that are meant to distract attention. Hagiwara *et al.* [HSK11] used such models to directly increase saliency of desired image regions. Similarly, Veas *et al.* [VMFS11] utilized per-frame saliency maps to modify visual saliency of video sequences. Using a local modulation on the target region instead of its surrounding, Waldin *et al.* [WWV17] proposed a temporal modulation stimulus that, due to its specific frequency, is visible only in the peripheral vision. Their method exploits the fact that the human eye can resolve higher frequencies in the periphery than in its foveal vision, but this requires high-frequency displays with refresh rates of around 144 Hz. Just recently, Lange *et al.* [LSGB20] suggested utilizing animated swarms, for example, of bees, as diegetic guidance means that appears as part of the original scene.

In the field of active methods, some approaches again focus on differences between peripheral and foveal vision, as minor changes in the peripheral visual field may remain nearly unperceived by the viewer due to the comparatively poor spatial resolution compared to foveal vision [BMSG09]. Such approaches usually capture real-time eye tracking information of the user to stop guidance as



**Figure 2:** Representative frames of all the 360° panorama video sequences used in our experiments.

soon as the target region enters the viewer's foveal vision. Some stimuli are shown for a specific maximum duration as long as the gaze is not successfully attracted to the desired region. For desktop environments, Dorr *et al.* [DMGB04] proposed two methods: a tiny red square and a magnification stimulus; both presented for 120 ms on static images. Barth *et al.* [BDB\*06] presented an analogous method for video sequences. They similarly used red squares in viewers' peripheral visual field that disappear on saccades towards the target region. They exploited an effect in the human eye, called saccadic masking [Dod00], to keep the stimulus from being consciously perceived with foveal vision. Akin to some of the passive methods [KMH02, CDF\*06, ST13], a method that initially blurs non-target regions of a picture was presented by Lintu and Carbonell [LC09]. In addition, they suggested to then gradually deblur the picture as soon as a fixation on the area of interest was detected. Another example that exploits saccadic masking to ensure the stimuli being perceived only in the periphery is the work from Bailey *et al.* [BMSG09, MBG09]. They proposed bright-dark (luminance) and warm-cold (colour) modulation that alternates temporally. Their method was additionally adapted for a controlled real-world environment by Booth *et al.* [BSM\*13], and was evaluated as an assistance for deaf or hard-of-hearing individuals when watching videos [LLAk\*20].

When applying attention guidance in VR, an additional challenge comes into play: the limited FOV. Although content is available 360° around the viewer, one is limited to their FOV and, thus, there might be areas of interest located completely out of sight. For successful storytelling, Pausch *et al.* [PST\*96] reported the necessity of navigating viewers through virtual scenes. They suggested to either employ actors to point in pre-defined directions or to build scenes such that the composition of scene elements automatically draws viewers' attention to specific spots. Lin *et al.* [LCH\*17] proposed two obtrusive methods to resolve this: a green arrow acting as an indicator pointing towards the target area, and an autopilot to dynamically rotate the virtual world around the viewer. Grogorick *et al.* [GSEM17] suggested to repeatedly showing a moving stimulus near the edge of the viewers' FOV that leads towards the target region.

Rothe *et al.* [RAK18] further investigated whether this moving stimulus can be used to increase recall in cinematic VR. Additionally, recent advances of next-generation HMDs manufacturers like Pimax, Samsung [Sam19] or VRgineers (XTAL) increase the available FOV to almost reach the dimensions of the actual human visual field. To complement this brief overview of related work in the field of visual guidance methods, we like to refer to the extensive review by Rothe *et al.* [RBH19].

The SIBM method [GTA\*19], which we investigate in this work, builds on binocular rivalry to attract users' attention. This phenomenon was already thoroughly studied by, for example, Wheatstone, Breese, Levelt and Wolfe [Whe38, Bre09, Lev65, Wol83]. From these works, we know that two rivaling information will not simply blend into each other. Instead they will be perceived alternately, with the more intense of the two versions being perceived over longer periods of time than the other. Even our perception of three-dimensional objects is strongly driven by binocular rivalry, caused by different perspective projections of (nearby) objects onto the two retinæ. According to Arnold *et al.* [AGW07] it is also a means of our visual system to support visibility in cluttered scenes. Ooi and Zijiang [OH99] found an influence of voluntary attention on which of the conflicting images is perceived dominantly. Zhang *et al.* [ZJE\*11] reported attention being a requirement for binocular rivalry, contradicting the findings of Platonov and Goossens [PG14] who showed that binocular rivalry occurs even with complete absence of visual awareness of such a conflicting stimulus.

In comparison to other visual attention guiding methods, SIBM presented a novel approach with respect to maintaining its subtleness, as its stimulus is difficult to observe and inspected directly by viewers, as it is almost completely invisible within the image of one eye alone. Techniques using stimuli like the small but solid red squares [DMGB04, BDB\*06] introduce comparatively clearly visible elements that might distract viewers from the actual scene content. Even approaches like temporal brightness modulation [BMSG09, MBG09, BSM\*13, LLAk\*20], that try to smoothly blend into the scene, introduce temporal variation in return to remain

perceivable. This might quickly act against their subtleness, as the human visual system is known to be highly sensitive to temporal variation in the peripheral visual field [ERH\*18, HW19].

### 3. General Methods

As previously mentioned, the goal of this paper is four-folded:

- G1 Assessing the applicability of SIBM to dynamic virtual environments.
- G2 Examine the influence of the stimulus' parameters.
- G3 Study the impact of the platform on the effectiveness of the method.
- G4 Explore the extension of the method to guidance towards target regions outside the initial FOV.

To achieve such objectives, we conducted three distinct perceptual studies for which, in the following, we describe their general psychophysical methodology.

#### 3.1. Dynamic virtual environments

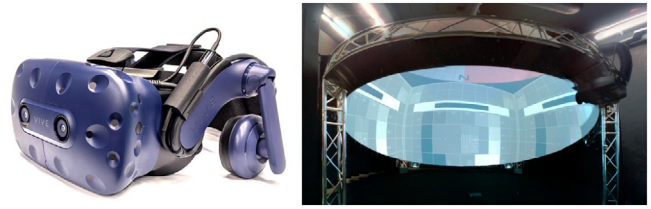
Throughout the three experiments, we used 13 different real-world 360° panorama video recordings (see Figure 2) representing a broad range of real environments, from indoor rooms to outdoor areas (e.g., *Coaf* and *Outside*), narrow spaces to wide open fields (e.g., *Stockfish* and *River*), during bright day or dark nights (e.g. *Coffee* and *Aurora*), empty surroundings to cluttered places (e.g. *Playa* and *Nicoletti*) and even underwater scenes (e.g. *Pool*). From each video, a sequence of 20 s length was selected. All these sequences included dynamic content (e.g. people swimming and diving around in the *Pool* scene).

The target region, that is, the viewing direction in which the stimulus will appear during the trials, was manually pre-selected per video and study, and was identical for all participants within an experiment. It was ensured that the target regions were not occluded by other moving in-video objects during the whole sequence.

The distribution of the target regions covered three different ranges, one per study. These ranges described the angular distance of the regions to the initial viewing direction (fixation cross). They were set to be inside (Studies 1 and 2) or outside (Study 3) the participants' initial FOV for both VR systems (see Section 3.2).

When deciding on the target locations, we considered the fact of visual dispersion being known to be influenced by stimulus-dependent features [LMBR11]. We aimed for our target regions to cover a broad spectrum of these features. Although the SIBM stimulus can be used on arbitrary (high or low saliency) objects, we ensured to not compromise the objective of having challenging targets, that is, regions that typically would not attract much attention. At the same time, the target regions should contain at least some information, so that the guidance does not seem to lead anywhere. Thus, they were required to meet either of the following criteria:

**Low-saliency but still providing some structure:** Regions that show some content, but do not stand out from their surrounding, for example, one of the pool skimmers in the *Pool* scene.



**Figure 3:** The HTC Vive Pro HMD (left) and the DOME (right).

**Typical regions of interest:** Humans or animals [RMFT03, JEDT09] with reduced saliency due to their position, for example, the person sitting by the building in the background far away in the *Coffee* scene.

#### 3.2. Apparatus

The experiments were conducted using two different types of real-time, eye-tracking-enabled, immersive VR systems: a current state-of-the-art HMD and a Dome Projection System (DOME), serving as a prototype for future-generation HMDs offering higher resolution and a nearly full FOV.

**The Head-Mounted Display.** that we have used, is an HTC Vive Pro head-mounted display (Figure 3, left). It contains two OLED displays (one per eye) with a resolution of  $1440 \times 1600$  px each and supports refresh rates up to 90 Hz. External base stations offer sub-millimetre precision for tracking participants' head motions. To gather real-time binocular eye tracking data (up to 120 Hz), the device was extended with the PupilLabs HTC Vive Binocular Add on.

**The Dome Projection System.** is our second VR system, a tilted full-dome real-time video projection system with a diameter of 5 m (Figure 3, right). It is powered by six projectors, each showing  $2560 \times 1600$  px, yielding an overall surrounding horizontal resolution of more than 8K (8855 px), at up to 120 Graphics Processing Unit (GPU) [GÜT\*19]. Active shutter glasses are used for stereoscopic rendering. For eye tracking while using shutter glasses, we assembled a combination of motion capture and eye tracking systems. We used Infrared Light-Reflective Marker (IR-marker) attached to the frame of the glasses, to be able to capture the users' head motion. A manually attached PupilLabs Hololens Addon provided real-time head motion-contingent eye tracking.

#### 3.3. Participants

Participants were recruited in a university environment, including but not limited to students. They took part either voluntarily (Studies 1 and 2) or were compensated with 10 Euro or a participation hour credited to their course degree (Study 3). It was ensured that all participants took part in only one of the three studies, to prevent any participant to be biased by prior knowledge.

### 3.4. Psychological Methodology and Procedure

Experiments were conducted participant-by-participant, one at a time. The participants signed a consent form before taking part in the experiment. For the scenario where the HMD was used, participants were seated in a dimly lit room between the external base stations at a distance of about 1.5 m to both of them. They were provided with the head set and a Vive controller, along with a short introduction to controller keys before putting on the headset. In the DOME scenario, participants were seated on a fixed position approximately in the middle of the DOME. They were provided with stereo shutter glasses and an XBox controller, along with a short introduction to the controller keys.

Then, the participants were presented with a virtual screen describing the detailed instructions for the calibration and the experimental task. During each trial, their task consisted solely of freely exploring the presented virtual environment. They were not informed about the presence of the SIBM stimulus. Before the actual experiment started, they were given another opportunity to ask questions and the controller was taken away to avoid any distractions. All experiments were controlled by a custom-built OpenGL rendering software.

Each experimental session started with a 9-point calibration routine of the eye tracking system. Afterwards, participants got introduced to the task (free viewing/exploration) in an additional example scene. After clarifying remaining questions about the device or procedure, the sequence of actual trials started. All trials began with a grey screen, showing only a fixation cross straight ahead. After 2 s, the video sequence started to smoothly fade in and the fixation cross vanished. After an additional second, the actual trial time (20 s) started. To prevent the SIBM stimulus from drawing attention due to temporal variation when appearing, it was always active right from the beginning and was faded in together with the scene. We recorded the participants' gaze in real-time and continuously tracked the angular distance between gaze and target. When participants fixated the desired target region, that is as soon as the angular distance fell below 10° [BMSG09], the stimulus was smoothly faded out. The video continued to play after the stimulus vanished.

For experiments that were conducted in both VR systems, that is, Studies 2 and 3, the environment (HMD and DOME) was a within-subjects factor and its order was balanced between subjects. Once they finished the first session on one system, and before they moved to the second system in a neighbouring room, participants were given a 5–10 min recreational break.

### 3.5. Evaluation

To evaluate guiding performance of the tested SIBM method, that is, whether attention of our participants could be influenced, we analysed the recorded gaze data of our participants for differences between with- and without-stimulus trials. More specifically, we compared fixations on the pre-selected target regions (see Subsection 3.1).

Also, for analysis of larger scale attentional shifts, we additionally analysed our participants' head motion to keep track of their visual field, especially for Study 3. Within an HMD, the available visual field is given by what is rendered onto the screen(s), while inside a DOME the visual field is constrained by the frame of the stereo shutter glasses.

To evaluate the subjective perception of stimulus presentations with different parameter configurations in the following studies, participants had to answer questions (see Table 1) at different moments during the experiments.

## 4. Study 1: Assessing SIBM in Dynamic Environments

The goal of this first study is to obtain initial evidence whether or not SIBM—which was previously shown to be effective in static environments [GTA\*19]—can be successfully employed to visually guide viewers within dynamic virtual environments (G1). Hence, a perceptual study has been conducted in which 360° videos served as dynamic surroundings.

### 4.1. Participants

A total of 18 people participated voluntarily in this experiment (age range 21 – 39;  $M = 26.5$ ,  $STD = 5.26$ ; 4 females). Each participant reported normal or corrected-to-normal vision. They reported a mean VR experience level of  $\sim 2.39$  ( $STD = 1.06$ ) on a 5-point Likert scale (1: never tried before; 5: regular use).

### 4.2. Design and Procedure

The experiment was conducted only within the HMD, using six video sequences (*Coffee*, *Finland*, *Nicoletti*, *Outside*, *Outside2* and *Stockfish*, see first six images in Figure 2. *Nasa* was used in the explanations to the participants.). The target regions were distributed between 29.99° and 45.98° ( $M = 39.20^\circ$ ,  $STD = 4.83^\circ$ ) off the initial viewing direction (fixation cross). This range was selected to be inside the participants' initial FOV within the HMD, which was

**Table 1:** Different questions used in the self-report parts of our experiments.

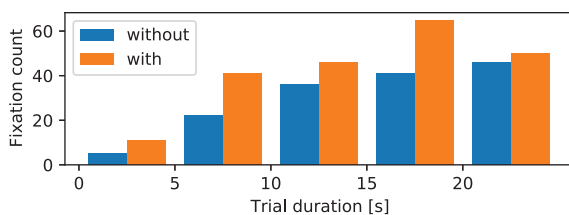
Identifier	Question	Nature	Asked after
Q1	Did you notice any degradation in image quality or distracting artefacts?	Implicit	Trial
Q2	Did you notice the stimulus? ( <i>The participant was given an explanation on its appearance</i> )	Explicit	Experiment
Q3	How distracting was the stimulus to explore the scene?	Explicit	Experiment
Q4	How distracting was the stimulus in comparison with stitching artefacts or chromatic aberrations?	Explicit	Experiment
Q5	How often did you notice the stimulus?	Explicit	Experiment

reported to be up to  $110^\circ$  (i.e.  $55^\circ$  centre outwards), but can decrease with varying Interpupillary Distance (IPD). Following the suggestion in the original work, an individual parameter set (intensity =  $0.15 - 0.51$ ;  $M = 0.33$ ,  $STD = 0.11$  and size =  $2 - 3$ ) was selected for each target region based on the respective image complexity [GTA\*19]. Additionally the intensity was selected higher for scenes with high complexity in combination with a lot of motion. We split the set of six videos into two subsets with three videos each. These two subsets were the same for all participants. During the experiment, scenes of one subset were shown with a guiding stimulus, whereas the others were presented unmodified. Which subset (first or second) to present with stimulus was alternated between participants, such that every second participant got the same sets of altered and unaltered videos. This way we got a balanced distribution of trials with and without stimulus for all tested scenes. The actual presentation order was a per-participant randomized mix of all six videos from both subsets. Per participant this yields a total of  $1_{\text{System}} \times (3_{\text{Scenes}} \times 1_{\text{With}} + 3_{\text{Scenes}} \times 1_{\text{Without}}) = 6$  trials, which corresponds to an experimental duration of  $\sim 3$  min (including scene transitions; excluding introduction).

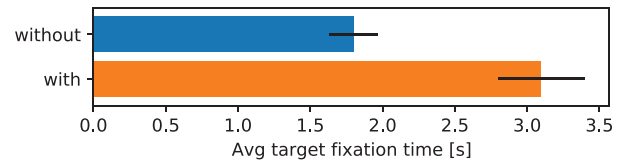
Extending the general procedure, we assessed subtleness by collecting participants' responses to several questions between sequential trials and at the end of the experiment (see the first three questions in Table 1). After each trial, we intentionally imprecisely asked them to report any degradation in image quality or distracting artefacts they might have noticed (Q1) to not bias participants towards the stimulus. At the end of the experiment, we specifically asked whether the stimulus has been noticed (Q2) in any trial and, if so, how distracting it felt while exploring the scene (Q3).

### 4.3. Results and Discussion

As Figure 4 shows, the amount of target fixations during free-viewing increases noticeably ( $n = 225$  with;  $n = 169$  without stimulus) in presence of the guiding stimulus. This indicates an existing positive correlation between stimulus presence and number of target fixations. Moreover, target fixations during trials comprising a guiding stimulus ( $M = 13.79$  s,  $SE = 0.29$  s) occurred significantly earlier on average ( $t(394) = -2.39$ ,  $p < 0.02$ ), as compared to target fixations during trials with no stimulus presentation ( $M = 15.19$  s,  $SE = 0.29$  s). Additionally, the recorded data show an increase in overall target fixation duration, by  $+72$ ,  $22\%$ , as depicted in Figure 5. On average, participants' gaze spent significantly more time ( $t(90) = 2.66$ ,  $p < 0.01$ ) at the target region when a stimulus was present ( $M = 3.10$  s,  $SE = 0.30$  s) compared to no stimulus presentation ( $M = 1.80$  s,  $SE = 0.17$  s).



**Figure 4:** Distribution of recorded target-directed fixations over time, with/without guidance accumulated over all trials. [6:R2]



**Figure 5:** Average target fixation time (standard error (SE) indicated by error bars) per trial, with and without stimulus guidance.

Previous research suggested female participants being less sensitive to the presented method (e.g. [CTR02]). Thus, we analysed the data regarding any potential gender difference. There are no significant differences to report with respect to gender of the participants. We have evaluated the recorded data of this study for males versus females, aggregated over all conditions with stimulus, as well as without stimulus. None of these tests revealed a significant difference ( $p >> 0.05$ ) in average fixation time.

Furthermore, regarding the first fixation onto the target region per trial, participants required around 11 s without the stimulus, while the presence of the stimulus lead to faster reaction around 7 s. Before the participants started to pay attention to the target region for the first time without being guided, they already noticed it around 4 s earlier with the stimulus being present. There was a significant association between the presence of stimulus and whether participants would look at the target region  $\chi^2(1) = 4.18$ ,  $p < 0.05$ . This seems to represent the fact that, based on the odds ratio, the chances of participants fixating the target were 2.12 times higher when it was presented with stimulus than if it was not. Thus, we claim that *the SIBM method is able to effectively attract users' gaze to specific target regions in dynamic VR environments (videos)*. We also asked implicitly about the perception of the stimulus after each trial and explicitly in a post-experiment questionnaire. Regarding subtleness, for the explicit question Q2, participants reported that they noticed the actual stimulus in  $\sim 0.94$  out of the six scenes on average ( $STD = 1.17$ ). Note that 10 out of the 18 participants even reported that they did not consciously take note of the stimulus in any trial. Regarding how distracting the stimulus was (explicit question Q3) participants reported a mean of  $\sim 1.53$  ( $STD = 0.69$ ) on a 5-point Likert scale (1: Not at all; 5: Extremely). The results seem to confirm the *subtleness of the SIBM method for dynamic virtual environments*.

## 5. Study 2: Perceptual Thresholds for SIBM in Dynamic Environments

The goal of the second study is to examine the influence of SIBM's parameters (G2) and the used VR system (G3) on the guiding performance. We conducted a perceptual study to evaluate *efficiency* and *subtleness* of the SIBM method in dynamic virtual environments, that is,  $360^\circ$  panorama video sequences.

### 5.1. Participants

A new set of 25 participants took part in this experiment (age range 21 – 46;  $M = 28.72$ ,  $STD = 5.89$ ; 9 females). Each participant reported normal or corrected-to-normal vision. They reported a mean

VR experience level of  $\sim 2.52$  ( $STD = 0.85$ ) on a 5-point Likert scale (1: never tried before; 5: regular use).

## 5.2. Design and Procedure

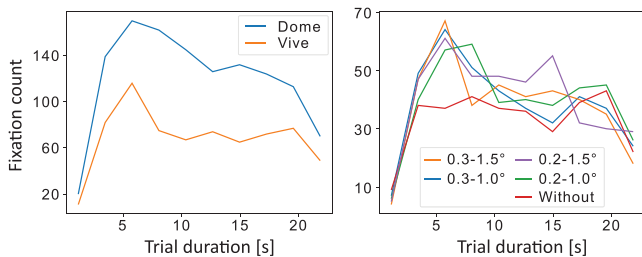
The experiment was conducted within two systems: the HMD, and the DOME. The system was a within-subjects factor and its order was balanced between subjects. For each system, we used the same 12 videos (see Figure 2, with the exception of *Nasa* which was used for the training of the participants). The target regions were distributed between  $20.05^\circ$  and  $45.98^\circ$  ( $M = 34.56^\circ$ ,  $STD = 7.73^\circ$ ) off the initial viewing direction (fixation cross). This range was set to be inside the participants' initial FOV for HMD and DOME. For each video sequence, four combinations of the stimulus parameters intensity (0.2, 0.3) and size ( $1.0^\circ$ ,  $1.5^\circ$ ) were tested. An additional trial per scene was shown without the stimulus. The order of the trials, for all scenes and all parameter permutations (incl. no stimulus trial), was randomized per participant. In total, this yields  $2_{\text{Systems}} \times 12_{\text{Scenes}} \times (2_{\text{Intensities}} \times 2_{\text{Sizes}} + 1_{\text{Without}}) = 120$  trials per participant, which corresponds to an experimental duration of  $\sim 50$  min (including scene transitions; excluding introduction and a 10 min recreational break while switching the system).

Similarly to Study 1, we assessed subtleness by collecting, at the end of the experiment, participants' responses to the last four questions in Table 1. We specifically asked whether the stimulus had been noticed at all (Q2). In the case of getting a positive answer, we also asked how distracting the stimulus felt while exploring the scene (Q3)—compared to how distracting the participant perceived regular panorama video artefacts, like stitching artifacts or chromatic aberration (Q4)—and how often the participant noticed the stimulus (Q5).

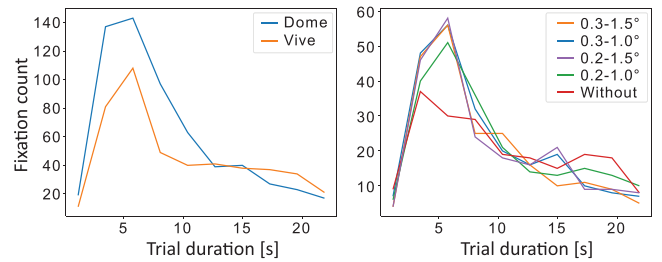
## 5.3. Results and Discussion

To be able to evaluate *efficiency* of SIBM in dynamic virtual environments, we compared users' gaze behaviour with and without the stimulus being present in the same scene.

In average, over the tested parameter combinations, the stimulus was able to increase the total number of target-directed fixations about 18%—from 331 to 389.5 (ranging between 378 and 401). Figure 6 shows the time-dependent distribution of fixations within the target region for different data slices. First (Figure 6, left), we can



**Figure 6:** Distribution of all fixations (count) on the target region over time (s), by VR system (left) and stimulus parameter combination (right).



**Figure 7:** Distribution of first fixations (count) on the target region over time (s), by VR system (left) and stimulus parameter combination (right).

observe that generally more fixations are recorded in the DOME environment as compared to the HMD, which is in line with the results of previous work on static environments [GTA\*19]. On average, accumulated target fixation duration per trial was significantly higher ( $t(1889) = 4.02$ ,  $p < 0.001$ ) when the experiment was conducted in the DOME ( $M = 2.73$  s,  $SE = 0.05$  s) compared to presentation using the HMD ( $M = 2.32$  s,  $SE = 0.04$  s). Second (Figure 6, right), we can report that even with low intensities and small sizes, compared to the tested values of Study 1, there is evidence that the stimulus has a measurable influence within dynamic environments, especially within the first 10 s.

As illustrated by Figure 7, considering only the first fixations of the participants on the target, there was a significant association between the presence of the  $0.3\text{--}1.0^\circ$  stimulus and whether participants would look at the target region  $\chi^2(1) = 9.95$ ,  $p < .001$ . This seems to represent the fact that, based on the odds ratio, the odds of participants fixating the target were 1.45 times higher when they were presented with stimulus than if they were not.

Moreover, Kruskal–Wallis tests revealed that targets received generally more attention in the presence of a stimulus than without ( $H = 7.7074$ ,  $p < 0.0055$ ). Regarding the impact of the intensity of the stimulus, both tested conditions (0.2 and 0.3) showed statistically significant differences with the absence ( $H = 3.9235$ ,  $p < 0.048$  and  $H = 9.6676$ ,  $p < 0.0019$ , respectively) but not between themselves ( $H = 2.1950$ ,  $p > 0.1$ ). Mann–Whitney U tests were used to follow up this finding. A Bonferroni correction was applied and so all effects were tested against a 0.0167 level of significance. It appeared that the 0.3 intensity differs most to no stimulus presentation ( $U = 36862.5$ ,  $p < 0.00094$ ). Comparing the 0.2 intensity to the absence of a stimulus does not reach significance after Bonferroni correction. However, it still shows a tendency towards a measurable difference ( $U = 39375.5$ ,  $p < 0.024$ ), which indicates the average perceptual threshold between intensities of 0.2 and 0.3. Nevertheless, there was no significant effect between both intensities themselves ( $U = 87671.5$ ,  $p > 0.069$ ).

Finally, the size of the stimulus has also shown to have an influence when compared with the absence of the stimulus ( $H = 7.8464$ ,  $p < 0.0051$  for  $1.5^\circ$ , against  $H = 5.2816$ ,  $p < 0.022$  for  $1.0^\circ$ ). Differences among both sizes do not show a significant effect ( $H = 0.4729$ ,  $p > 0.49$ ). Bonferroni-corrected *post hoc* tests for stimulus size follow the same result as for the intensity. The stimulus size of  $1.5^\circ$  exhibits the most significant effect against no

stimulus presentation ( $U = 36630.5$ ,  $p < 0.0026$ ). When comparing a stimulus size of  $1.0^\circ$  to stimulus absence, a weaker but still significant effect was observed ( $U = 39607.5$ ,  $p < 0.011$ ), indicating the average perceptual threshold to be slightly below a stimulus size of  $1.0^\circ$ . Again, both sizes compared to themselves did not reach the level of significance ( $U = 90524.0$ ,  $p < 0.246$ ).

There are no significant differences to report with respect to gender of the participants. We have evaluated the recorded average fixation time of this study for males versus females, aggregated over all conditions with stimulus, as well as without stimulus. None of these tests revealed a significant difference ( $p > 0.05$ ).

Regarding *subtleness*, for question Q5 participants reported that they noticed the stimulus in  $\sim 39.38\%$  of all trials in average ( $STD = 27.09$ ). Note that 16% of the participants even reported that they did not consciously take note of the stimulus at all. Regarding how distracting the stimulus was (question Q3), participants reported a mean of  $\sim 2.13$  ( $STD = 0.97$ ) on a 5-point Likert scale (1: Not at all; 5: Extremely). For distractions that resulted from regular panorama artifacts (question Q4), our participants reported a mean of  $\sim 2.46$  ( $STD = 1.11$ ) on the same 5-point Likert scale. This means that, in comparison, the stimulus was about 8.33% less distracting than regular artefacts that appear in current-state panorama videos. The results seem to confirm the *subtleness of the SIBM method for dynamic virtual environments*.

## 6. Study 3: Extra-FOV SIBM Guidance

A third study is conducted to evaluate an extension of the original SIBM, that might greatly improve applicability in terms of reasonable use cases. To allow guidance towards target locations outside the user's initial FOV (G4), we examine a dynamic stimulus placement mechanism. We propose to dynamically generate *intermediate stimulus positions*, that indicate the direction towards the current actual target location far outside the viewer's FOV. Specifically, we suggest to place the stimulus at an intermediate position on the direct axis between the current viewing direction and the actual target. The distance from the current gaze point is set to be just outside the visual field, that is, in case of an HMD it is placed such that it is blocked by the frames of the converging lenses. Firstly, this placement prevents gathering unintended overt attention due to temporal variance, that is, the stimulus appearing within a viewer's FOV. Secondly, it ensures that already slight head motion towards the intended direction will move the stimulus inside the FOV. Upon recognition of further head rotations towards the intermediate position, the stimulus is relocated to a newly generated intermediate position, closer to the actual target location. This is repeated until the actual target location is reached. We suggest to also update the intermediate position as soon as it enters the central region (up to  $30^\circ$  off centre) of a viewer's visual field. This was reported to be the threshold above which head rotations will almost always be involved towards a fixation [Sch11].

In the following, we show our experimental results for the aforementioned technique to redirect the person's attention towards extra-FOV target regions.

### 6.1. Participants

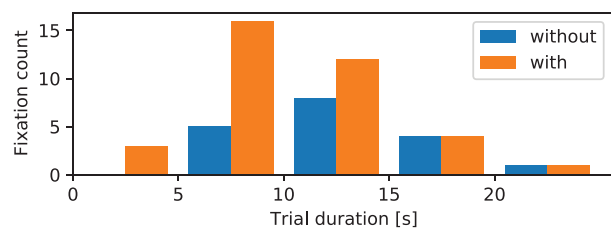
A new group of participants, for a total of 20 people participated voluntarily in our experiment (age range 19 – 27;  $M = 22.3$ ,  $STD = 2.53$ ; 11 females). Each participant reported normal or corrected-to-normal vision. They reported a mean VR experience level of  $\sim 1.6$  ( $STD = 0.73$ ) on a 5-point Likert scale (1: never tried before; 5: regular use).

### 6.2. Design and Procedure

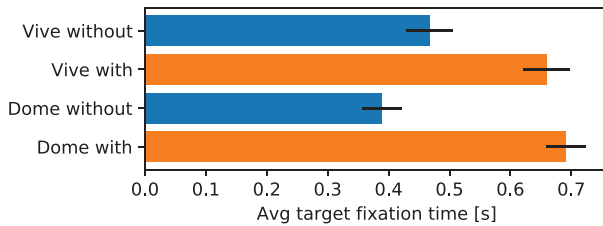
Following the experimental design of Study 2, the experiment was conducted for the HMD, and the DOME, with system as within-subjects factor, order balanced between participants. In both systems, participants saw the same set of 12 panorama sequences (see Figure 2, with the exception of *Finland* which was used for the exemplary trial in the explanations to the participants). The target regions were distributed between  $41.54^\circ$  and  $83.12^\circ$  ( $M = 72.65^\circ$ ,  $STD = 10.67^\circ$ ) off the initial viewing direction (fixation cross). This range was selected to be outside the participants' initial FOV within the HMD and DOME, which was constrained by the size of the display/lens and the frame of the shutter glasses, respectively. Similar to Study 1, the stimulus parameters (intensity = 0.2 – 0.6;  $M = 0.37$ ,  $STD = 0.10$  and size =  $1.5^\circ$ ) were selected individually per scene, based on each scene's visual complexity, as suggested in the originating work [GTA\*19]. We selected the more conspicuous size from the previous experiment to match the degraded acuity of the human vision in the more eccentric target regions. We selected slightly more conspicuous parameter values than in the previous experiments to match the even stronger degradation of visual acuity in the more eccentric parts of the visual field. Similar to previous studies, trials with and without a stimulus were shown in pseudo-randomized order. In total, this yields  $2_{\text{Systems}} \times 12_{\text{Scenes}} \times (1_{\text{With}} + 1_{\text{Without}}) = 48$  trials per participant, which corresponds to an experimental duration of  $\sim 20$  min (including scene transitions; excluding introduction and a 5 min recreational break while switching the system).

### 6.3. Results and Discussion

In order to determine whether SIBM is suitable for guidance to target regions outside a user's current FOV, we investigate the distribution of fixations onto the target region ( $10^\circ$ ). As can be seen in Figure 8, the probability of a first fixation increases considerably within the first half of a trial, if the guiding stimulus is present. In absence of the stimulus, first fixations are shown to occur much more



**Figure 8:** Distribution of first target fixations over time, with and without guiding stimulus, accumulated over all trials. [6:R2]



**Figure 9:** Average target fixation time (standard error (SE) indicated by error bars) per trial, with and without stimulus guidance, separated for both systems.

evenly distributed over time. On average, participants reached the target significantly earlier ( $t(54) = -2.52, p < 0.02$ ) when a stimulus was present ( $M = 10.23$  s,  $SE = 0.62$  s) as compared to trials with no stimulus presentation ( $M = 13.47$  s,  $SE = 0.55$  s).

Moreover, analysis of the duration of target fixations per trial reveals a noticeable (+52, 27%) increase for trials with a guidance stimulus compared to trials without. On average, participants focused significantly longer ( $t(524) = 3.25, p < 0.002$ ) at the target region when a guiding stimulus was present ( $M_{\text{with}} = 0.67$  s,  $SE_{\text{with}} = 0.04$  s) as compared to trials without stimulus ( $M_{\text{without}} = 0.44$  s,  $SE_{\text{without}} = 0.02$  s).

With respect to the tested systems, this effect was more pronounced for the DOME (+76, 92%;  $M_{\text{with}} = 0.69$  s,  $SE_{\text{with}} = 0.07$  s;  $M_{\text{without}} = 0.39$  s,  $SE_{\text{without}} = 0.04$  s;  $t(201) = 2.74, p < 0.01$ ) than it was for the HMD (+40, 43%;  $M_{\text{with}} = 0.66$  s,  $SE_{\text{with}} = 0.06$  s;  $M_{\text{without}} = 0.47$  s,  $SE_{\text{without}} = 0.03$  s;  $t(323) = 2.02, p < 0.05$ ), as depicted in Figure 9. A possible reason for this might be the larger FOV that is available in the DOME as compared to the FOV within the HMD.

Once more, there are no significant differences to report with respect to gender of the participants. We have evaluated the recorded data of this study for males versus females, aggregated over all conditions with stimulus, as well as without stimulus. None of these tests revealed a significant difference ( $p > 0.05$ ) in average fixation time.

Overall, we therefore assume the stimulus to induce better target identification, showing the potential of our proposed mechanism, enabling successful application of SIBM for visual guidance to out-of-FOV target regions.

## 7. General Conclusions

In this paper, we have investigated SIBM, a recent technique specifically designed for visual guidance in stereoscopic virtual environments. In contrast to the originating work, our investigation evaluated the applicability of the method within dynamic (video) instead of static (photo) 360° panorama recordings.

We, therefore, conducted a series of three perceptual studies using 13 distinct 360° panorama videos and two VR systems, to evaluate its efficiency and subtleness. The results of Study 1 and 2 show that the technique is suitable for gaze guidance in VR systems offering a wide FOV and can be successfully adjusted for a wide

variety of dynamic environments while remaining subtle. In accordance with the originating work of the SIBM method [GTA\*19], our results confirm that also in dynamic (video) surroundings parameter values need to be selected on a per-scenario basis. This is to compensate for different levels of scene complexity (e.g. colour intensity, contrast, speed and amount of motion, or ratio of fine and coarse structures) and specifications of used display systems (e.g. contrast or brightness).

It is also worth to mention that, even when previous research indicates that the human visual system shows gender differences regarding, in between others, FOV or peripheral vision and reaction time [VS17, CTR02, Chr13, VVB95], the results of our studies indicate that the presented technique seems to be also robust against gender bias.

Although the original approach has a good performance in well-posed situations, it still has its limitations. As people are able to freely explore scenes in VR, by using solely this method it is not guaranteed that a predetermined target region will always be present within a viewer's FOV. This connects to our second contribution, that is, on how to direct users in virtual environments beyond their own FOV.

Thus, in this paper we also proposed a modification to the original method addressing this kind of situations, SIBM for extra-FOV guidance. For cases when the target is out of the FOV of the user, we introduced a real-time reallocation of the stimulus (based on real-time head-tracking data) to still forward the participant to the target. Our results of Study 3 indicate the effectiveness of this solution, which could address the challenge of having a dynamic viewpoint. Moreover, we plan to extend the method to address other exciting scenarios like targeting moving in-video content. Assuming the availability of object tracking for the scene, the presented method could be modified to bind the stimulus to non-static targets. This renders the modified SIBM a promising method for upcoming VR applications in fields like storytelling or data visualization.

## Acknowledgements

The authors gratefully acknowledge funding by the German Science Foundation (DFG MA2555/15-1 'Immersive Digital Reality' and DFG INST 188/409-1 FUGG 'ICG Dome'). Open access funding enabled and organized by Projekt DEAL.

## References

- [AGW07] ARNOLD D. H., GROVE P. M., WALLIS T. S.: Staying focused: A functional account of perceptual suppression during binocular rivalry. *Journal of Vision* 7, 7 (2007), 7–7.
- [BDB\*06] BARTH E., DORR M., BÖHME M., GEGENFURTNER K., MARTINETZ T.: Guiding the mind's eye: Improving communication and vision by external control of the scanpath. *Human Vision and Electronic Imaging* 6057, (2006), 60570D–60570D.
- [BMSG09] BAILEY R., MCNAMARA A., SUDARSANAM N., GRIMM C.: Subtle gaze direction. *ACM Transactions on Graphics* 28, 4 (2009), 100.

- [Bre09] BREESE B.: Binocular rivalry. *Psychological Review* 16, 6 (1909), 410.
- [BSM\*13] BOOTH T., SRIDHARAN S., MCNAMARA A., GRIMM C., BAILEY R.: Guiding attention in controlled real-world environments. In *Proceedings of SAP* (New York, NY, 2013), ACM Press, pp. 75–82.
- [CDF\*06] COLE F., DECARLO D., FINKELSTEIN A., KIN K., MORLEY R. K. & SANTELLA A.: Directing gaze in 3d models with stylized focus. *Eurographics Symposium on Rendering 17* (2006), 377–387.
- [Chr13] CHRAIF M.: Gender influences in peripheral and central visual perception for the young. *Procedia - Social and Behavioral Sciences* 84 (2013), 1100–1104. (The 3rd World Conference on Psychology, Counseling and Guidance, WCPCG-2012. <https://doi.org/10.1016/j.sbspro.2013.06.707>.)
- [CTR02] CZERWINSKI M., TAN D. S., ROBERTSON G. G.: Women take a wider view. In *CHI'02: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, 2002), ACM Press, pp. 195–202.
- [dKJ17] DE KONING B. B., JARODZKA H.: Attention guidance strategies for supporting learning from dynamic visualizations. In *Learning from Dynamic Visualization*. R. Lowe and R. Ploetzner (Eds.). Springer, Berlin, Germany, 2017, pp. 255–278.
- [DMGB04] DORR M., MARTINETZ T., GEGENFURTNER K., BARTH E.: Guidance of eye movements on a gaze-contingent display. In *Dynamic Perception Workshop GI Section "Computer Vision"*. U. J. Ilg, H. H. Bülthoff and H. A. Mallot (Eds.). IOS Press, Amsterdam (2004), pp. 89–94.
- [Dod00] DODGE R.: Visual perception during eye movement. *Psychological Review* 7, 5 (1900), 454.
- [ERH\*18] ESCOBAR M.-J., REYES C., HERZOG R., ARAYA J., OTERO M., IBACETA C., PALACIOS A. G.: Characterization of retinal functionality at different eccentricities in a diurnal rodent. *Frontiers in Cellular Neuroscience* 12 (2018), 444.
- [FMS\*19] FENG Y., MCGOWAN H., SEMSAR A., ZAHIRI H. R., GEORGE I. M., PARK A., KLEINSMITH A., MENTIS H.: Virtual pointer for gaze guidance in laparoscopic surgery. *Surgical Endoscopy* (2019), 1–7.
- [GAM18] GROGORICK S., ALBUQUERQUE G., MAGNOR M.: Comparing unobtrusive gaze guiding stimuli in head-mounted displays. In *Proceedings of the International Conference on Image Processing* (Piscataway, NJ, October 2018), IEEE, pp. 2805–2809.
- [GATM18] GROGORICK S., ALBUQUERQUE G., TAUSCHER J.-P., MAGNOR M.: Comparison of unobtrusive visual guidance methods in an immersive dome environment. *Transactions on Applied Perception* 15, 4 (October 2018), 27:1–27:11.
- [GSEM17] GROGORICK S., STENGEL M., EISEMANN E., MAGNOR M.: Subtle gaze guidance for immersive environments. In *SAP '17: Proceedings of the ACM Symposium on Applied Perception* (New York, NY, 2017), ACM Press, pp. 4:1–4:7.
- [GTA\*19] GROGORICK S., TAUSCHER J.-P., ALBUQUERQUE G., KASSUBECK M., MAGNOR M.: Towards vr attention guidance: Environment-dependent perceptual threshold for stereo inverse brightness modulation. In *SAP '19: ACM Symposium on Applied Perception* (New York, NY, 2019), ACM Press, pp. 22:1–22:5.
- [GÜT\*19] GROGORICK S., ÜBERHEIDE M., TAUSCHER J.-P., BITTNER P. M., MAGNOR M.: Gaze and motion-aware real-time dome projection system. In *IEEE Virtual Reality (VR) Workshop PerGraVAR* (2019).
- [HKS16] HATA H., KOIKE H., SATO Y.: Visual guidance with unnoticed blur effect. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (New York, NY, 2016), ACM Press, pp. 28–35.
- [HSK11] HAGIWARA A., SUGIMOTO A., KAWAMOTO K.: Saliency-based image editing for guiding visual attention. In *PETMEI '11: Proceedings of the 1st International Workshop on Pervasive Eye Tracking & Mobile Eye-Based Interaction* (New York, NY, 2011), ACM Press, pp. 43–48.
- [HW19] HIMMELBERG M. M., WADE A. R.: Eccentricity-dependent temporal contrast tuning in human visual cortex measured with fmri. *NeuroImage* 184 (2019), 462–474.
- [IKN98] ITTI L., KOCH C., NIEBUR E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 11 (1998), 1254–1259.
- [JEDT09] JUDD T., EHINGER K., DURAND F., TORRALBA A.: Learning to predict where humans look. In *ICCV'09: 2009 IEEE 12th International Conference on Computer Vision* (Piscataway, NJ, 2009), IEEE, pp. 2106–2113.
- [KLB\*89] KENNEDY R. S., LILIENTHAL M. G., BERBAUM K. S., BALTZLEY D., MCCAULEY M.: Simulator sickness in us navy flight simulators. *Aviation, Space, and Environmental Medicine* 60, 1 (1989), 10–16.
- [KMH02] KOSARA R., MIKSCH S., HAUSER H.: Focus+ context taken literally. *IEEE Computer Graphics and Applications* 22, 1 (2002), 22–29.
- [LC09] LINTU A., CARBONELL N.: Gaze guidance through peripheral stimuli. Institut national de recherche en informatique et en automatique (INRIA). Preprint, 2009. Hyper articles en ligne (HAL). <https://hal.inria.fr/inria-00421151>
- [LCH\*17] LIN Y.-C., CHANG Y.-J., HU H.-N., CHENG H.-T., HUANG C.-W., SUN M.: Tell me where to look: Investigating ways for assisting focus in 360° video. In *CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (New York, NY, 2017), ACM Press, pp. 2535–2545.
- [Lev65] LEVELT W. J.: *On Binocular Rivalry*. PhD thesis, Van Gorcum Assen, 1965.

- [LGCM14] LATIF N., GEHMACHER A., CASTELHANO M. S., MUNHALL K. G.: The art of gaze guidance. *Journal of experimental psychology: human perception and performance* 40, 1 (2014), 33.
- [LH19] LANKES M., HASLINGER A.: Lost & found: Gaze-based player guidance feedback in exploration games. In *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts* (2019), pp. 483–492.
- [LLAK\*20] LI J., LUETTGEN M., AL-KHAZRAJI S., HUENERFAUTH M., BAILEY R., ALM C. O.: Gaze guidance for captioned videos for dhh users. *Journal on Technology & Persons with Disabilities* 8, (2020). <https://doi.org/10.2111.3/215980>.
- [LMBR11] LE MEUR O., BACCINO T., ROUMY A.: Prediction of the Inter-Observer Visual Congruency (IOVC) and Application to Image Ranking. In *MM'11: Proceedings of the 19th ACM International Conference on Multimedia* (2011), pp. 373–382.
- [LSGB20] LANGE D., STRATMANN T. C., GRUENEFELD U., BOLL S.: Hivefive: Immersion preserving attention guidance in virtual reality. In *Proceedings of the Conference on Human Factors in Computing Systems* (2020), pp. 1–13.
- [MBG09] MCNAMARA A., BAILEY R., GRIMM C.: Search task performance using subtle gaze direction with the presence of distractions. *ACM Transactions on Applied Perception* 6, 3 (2009), 17.
- [OH99] OOI T. L., HE Z. J.: Binocular rivalry and visual awareness: The role of attention. *Perception* 28, 5 (1999), 551–574.
- [PG14] PLATONOV A., GOOSSENS J.: Eye dominance alternations in binocular rivalry do not require visual awareness. *Journal of vision* 14, 11 (2014), 2–2.
- [PST\*96] PAUSCH R., SNODDY J., TAYLOR R., WATSON S., HASELTINE E.: Disney's aladdin: first steps toward storytelling in virtual reality. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, 1996), ACM Press, pp. 193–203.
- [RAK18] ROTHE S., ALTHAMMER F., KHAMIS M.: Gazerecall: Using gaze direction to increase recall of details in cinematic virtual reality. In *MUM 2018: Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia* (New York, NY, 2018), ACM Press, pp. 115–119.
- [RBH19] ROTHE S., BUSCHEK D., HUSSMANN H.: Guidance in cinematic virtual reality-taxonomy, research status and challenges. *Multimodal Technologies and Interaction* 3, 1 (2019), 19.
- [RMFT03] ROUSSELET G. A., MACÉ, M. J.-M., FABRE-THORPE, M.: Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *Journal of Vision* 3, 6 (07 2003), 5–5.
- [Sam19] SAMSUNG: Head-mounted electronic device. U.S. Patent Application 2019/0026871. <http://pdfaiw.uspto.gov/aiw?docid=20190026871> (2019). (Online; accessed 10-May-2019).
- [Sch11] SCHOR C. M.: Neural control of eye movements. In *Adler's Physiology of the Eye*. L. A. Levin, S. F. E. Nilsson, J. Ver Hoeve, S. M. Wu, P. L. Kaufman and A. Alm (Eds.). Saunders Elsevier, Amsterdam, 2011, pp. 220–242.
- [SDA05] SU S. L., DURAND F., AGRAWALA M.: *De-Emphasis of Distracting Image Regions using Texture Power Maps*. Tech. Rep. 025, MIT CSAIL, 2005.
- [SKB19] STRATMANN T. C., KEMPA F., BOLL S.: Lame: light-controlled attention guidance for multi-monitor environments. In *Proceedings of the 8th ACM International Symposium on Pervasive Displays* (New York, NY, 2019), ACM Press, pp. 1–5.
- [SP19] SALSELAS I., PENHA R.: The role of sound in inducing storytelling in immersive environments. In *Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound* (2019), pp. 191–198.
- [SRD\*19] SPEICHER M., ROSENBERG C., DEGRAEN D., DAIBER F., KRÜGER A.: Exploring visual guidance in 360-degree videos. In *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video* (2019), pp. 1–12.
- [SSP\*18] SITZMANN V., SERRANO A., PAVEL A., AGRAWALA M., GUTIERREZ D., MASIA B., WETZSTEIN G.: Saliency in vr: How do people explore virtual environments? *IEEE Transactions on Visualization and Computer Graphics* 24, 4 (2018), 1633–1642.
- [ST13] SMITH W. S., TADMOR Y.: Nonblurred regions show priority for gaze direction over spatial blur. *The Quarterly Journal of Experimental Psychology* 66, 5 (2013), 927–945.
- [VMFS11] VEAS E. E., MENDEZ E., FEINER S. K., SCHMALSTIEG D.: Directing attention and influencing memory with visual saliency modulation. In *CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, 2011), ACM Press, pp. 1471–1480.
- [VS17] VANSTON J. E., STROTHER L.: Sex differences in the human visual system. *Journal of Neuroscience Research* 95, 1–2 (2017), 617–625.
- [VVB95] VOYER D., VOYER S. D., BRYDEN M. P.: Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin* 117 2 (1995), 250–270.
- [Whe38] WHEATSTONE C.: XVIII. Contributions to the physiology of vision.—Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions* 128, (1838), 371–394.
- [Wol83] WOLFE J. M.: Influence of spatial frequency, luminance, and duration on binocular rivalry and abnormal fusion of briefly presented dichoptic stimuli. *Perception* 12, 4 (1983), 447–456.

- [WWV17] WALDIN N., WALDNER M., VIOLA I.: Flicker observer effect: Guiding attention through high frequency flicker in images. *Computer Graphics Forum* 36, (2017), 467–476.
- [YKB19a] YOSHIMURA A., KHOKHAR A., BORST C. W.: Eye-gaze-triggered visual cues to restore attention in educational vr. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Piscataway, NJ, 2019), IEEE, pp. 1255–1256.
- [YKB19b] YOSHIMURA A., KHOKHAR A., BORST C. W.: Visual cues to restore student attention based on eye gaze drift, and application to an offshore training system. In *Symposium on Spatial User Interaction* (New York, NY, 2019), ACM Press, pp. 1–2.
- [ZJE\*11] ZHANG P., JAMISON K., ENGEL S., HE B., HE S.: Binocular rivalry requires visual attention. *Neuron* 71, 2 (2011), 362–369.