COMPARING UNOBTRUSIVE GAZE GUIDING STIMULI IN HEAD-MOUNTED DISPLAYS

Steve Grogorick *Georgia Albuquerque* *Marcus Magnor*

TU Braunschweig

ABSTRACT

In this paper we investigate the efficiency of five different image-space post-processing gaze guidance techniques adapted for immersive environments, probing our peripheral vision's sensitivity to different stimuli embedded in complex, real-world panorama still images. We conducted an extensive perceptual study for the commercially available HTC Vive Head-Mounted Display that is equipped with a high-quality eye tracking system to monitor the subjects' gaze direction and saccades in real-time. Despite evaluation reveals no outstanding winner, the local magnification effect by Dorr et al. performed the best with up to 30 % target-directed saccades within the first second.

Index Terms— Virtual Reality, Post-Processing, Perception, Eye Tracking

1. INTRODUCTION

Virtual Reality (VR) has been predicted as one of the most promising recent technologies. Real-world 360° image and video footage can be easily generated with consumer-grade equipment like smart phones and are available in popular online video and social media applications. 360° images allow for a great immersive experience of arbitrary. However, consuming these media formats includes the risk of missing interesting parts of a captured scene, due to the unfamiliar wide range of the content — all around the viewer. The image frame that typically constrains the attention to the desired content is missing and has to be replaced by some other means. Different post-processing approaches have been proposed in the last years, trying to *unobtrusively guide* user attention for traditional desktop settings [1, 2, 3] , and recently for head-mounted displays [4, 5].

Many approaches exploit strengths of the human peripheral vision for tasks like visual search, object recognition, and scene perception. Despite this, subtle changes in the periphery may be nearly unperceived by viewers [2] due to the relatively poor visual acuity compared to foveal vision. For desktop applications, Dorr et al. [1] proposed a small red square and a magnification stimulus on still images. Similarly, Barth et al. [6] presented a red square as peripheral stimulus for video sequences that disappears when a saccade towards the stimulus is detected. Exploiting the saccadic masking phenomenon [7], gaze-contingent deactivation keeps stimuli invisible for foveal vision. Bailey et al. [2, 8] proposed bright–dark (luminance) and warm–cold (color) modulations, both in the periphery. This method was also adapted for controlled real-world [9] and in virtual reality [5] environments.

Other approaches apply global transformations to draw the users gaze towards target locations. These methods are mostly independent from the users' current gaze and can be adapted for multi-user environments. An approach by Kosara et al. [10] is to simulate depth-of-field from traditional photography, bringing desired regions of the image in or out of focus. Cole [11] proposed desaturating and blurring uninteresting regions of the scene. Another field of transformations use computational *visual saliency* models, which estimate how well image regions will catch viewers' attention [12]. Hagiwara et al. [13] use such models to specifically increase saliency of desired image regions. Similarly, Veas et al. [14] make use of salience maps to modify frame-wise visual saliency of video sequences. Lintu and Carbonell [15] proposed a method that initially blurred pictures and deblurred them gradually as soon as a fixation on the area of interest was detected. Similarly, Hata et al. [3] presented an approach to gradually blur uninteresting regions while keeping target regions unblurred. We use blur modulation from Hata et al. [3], as an example of global transformation stimuli.

One of the main challenges in gaze guidance in 360° Head-Mounted Displays (HMDs) is the limited Field Of View (FOV), because areas of interest may be located completely outside the FOV. Lin et al. [4] proposed a green arrow pointing towards the target direction, and an autopilot that automatically rotates the virtual view towards the target. Grogorick et al. [5] suggest to repeatedly move a stimulus towards the target region. Although these works could prove effectiveness for gaze guidance in VR, they concentrate on obtrusive methods or a unique approach.

In this paper we go one step further and adapt multiple existing gaze guidance stimuli to wide-FOV immersive environments to provide an current state-of-the-art overview. Our

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"). We thank Linda Eckardt (wi2, TU Braunschweig) for recruiting participants.

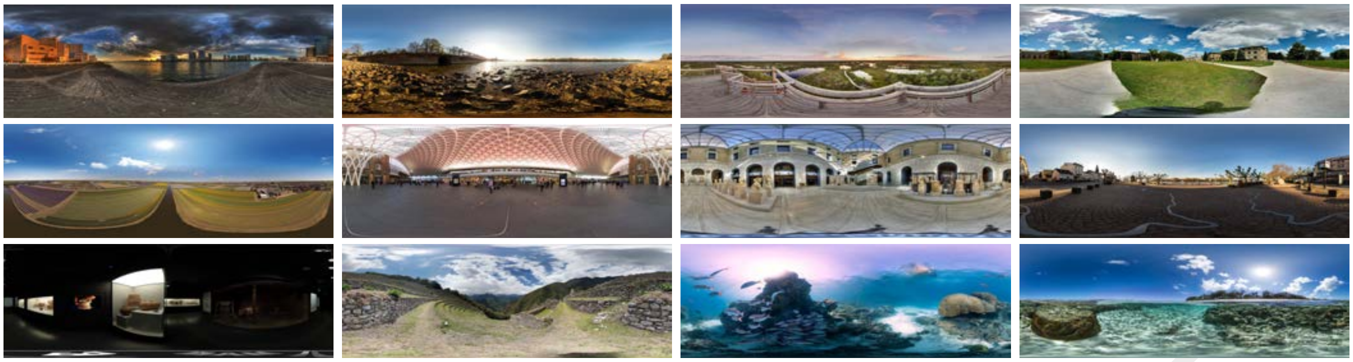

Fig. 1: Selection of 360° panoramas covers a broad range of real-world environments, e.g. in-/outdoor, mid air and under water.

adaptation includes handling of perspective distortions due to the wider FOV and target areas completely outside users' visual field. We use eye tracking data to analyze effectiveness in a perceptual study with 102 participants and 12 distinct 360° images in an HMD. In summary, we contribute the following:

- adaptation of five stimuli to immersive displays
- extensive user studies in a wide-FOV immersive environment using real-time eye tracking
- comparative evaluation of gaze guidance efficiency

2. EVALUATED GAZE GUIDANCE TECHNIQUES

In the following we give an overview of the image-space post-processing stimuli that have been adapted for immersive environments. To date, there is no completely subtle gaze guidance technique but we would like to avoid obvious approaches like green arrows [4] that would distract viewers from the actual task. Nevertheless, we selected stimuli based on the criteria of subtleness. This resulted in a selection of 5 stimuli which will be referred to as *ColorDot*, *SGD*, *ZoomRect*, *ZoomCircle* and *SpatialBlur* for the remainder of this work. The shape of all stimuli is adjusted to prevent perception of perspective distortions which occur in peripheral regions for large FOVs [5].

ColorDot refers to the small red squares proposed by Dorr et al. [1]. We use the proposed duration of 120 ms to exploit saccade masking for subtleness. The stimulus is shown repeatedly with a break of 2 seconds in between to allow free exploration of a virtual environment but still eventually achieve guidance, when the target enters the viewers' FOV. In accordance with previous work we add *gaze-contingent deactivation* to increase subtleness, i.e. the stimulus is switched off permanently as soon as a saccade towards the stimulus is detected [2]. To counteract acuity degradation of the human eye towards the periphery we apply *eccentricity-based scaling* [5]. Scaling is adjusted to retain a size of 1° visual angle at 12° eccentricity as in the original implementation. The final stimulus can be seen in Figure 2 (ColorDot).

Subtle Gaze Direction (SGD) refers to the luminance modulation proposed by Bailey et al. [2]. Except the introduction of eccentricity-based scaling, this stimulus is kept as in the original implementation, as depicted in Figure 2 (SGD).

ZoomRect refers to the magnification stimulus proposed by Dorr et al. [1]. Similar to the ColorDot stimulus, out implementation is extended with repeated presentation, gaze-contingent deactivation and eccentricity-based scaling. Original square shape and presentation time of 120 ms are kept and eccentricity-based scaling is adjusted to retain the suggested size of 2° visual angle at 12° eccentricity. The final stimulus can be seen in Figure 2 (ZoomRect).

ZoomCircle is a novel stimulus inspired by the ZoomRect stimulus. Similarly, it uses a magnification effect and is shown repeatedly. However, the stimulus is presented as a cir-

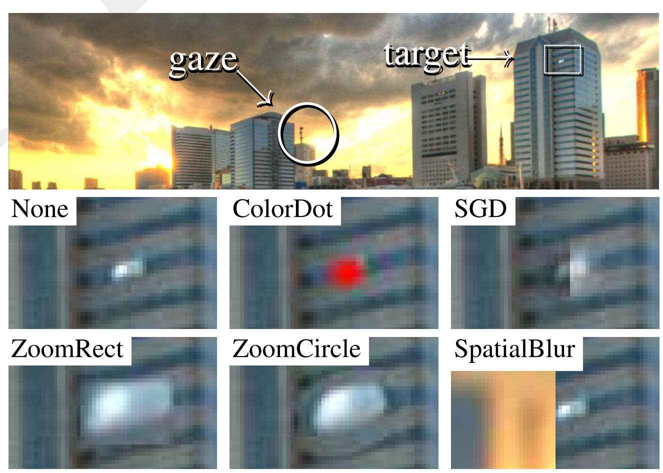

Fig. 2: View straight ahead within a 360° picture, with current gaze point (◯) and target area (□) highlighted (top). Target area closeups below show the different stimuli. For illustration, SGD stimulus intensity was increased and shows both states (dark and bright) side by side. SpatialBlur closeup shows blurred and unblurred regions side by side.

cular shape instead of a rectangular and the presentation time is increased to 500 ms, with a break of 2 s in between. Slowing down the zoom effect to a more perceivable level shall strengthen the perception of optical expansion of an object moving towards the observer. Finally, to prevent introduction of suddenly appearing sharp edges around the stimulus, the zoom factor to enlarge the content follows a radial gaussian fall-off from the center to the border.

Spatial Blur refers to the blur filter from Hata et al. [3] using spatial filtering to reduce details in non-target image regions. Our implementation follows the original work.

3. SYSTEM SETUP

We use an *HTC Vive* (Figure 3) comprising two OLED displays with a resolution of 1080 × 1200 px each and a refresh rate of 90 Hz. Two external base stations offer sub-millimeter precision for tracking head movements. The device is upgraded with an integrated infrared eye tracking solution from *SensoMotoric Instruments (SMI)*[1] offering binocular eye tracking at 250 Hz with an accuracy of 0.5°.

Fig. 3: HTC Vive with eye tracking.

4. EXPERIMENT

Participants are instructed to watch a sequence of 6 different scenes. A gaze guidance stimulus is presented in five these scenes while one scene serves as control. The experiment procedure consists of a pre-questionnaire, a main part and a post-questionnaire.

A pre-experiment questionnaire was answered by the participants before the main experiment collecting demographic information, i.e. gender, age and if a visual aid is required. Participants' current health condition was captured with the Simulator Sickness Questionnaire (SSQ) suggested by Kennedy et al. [16].

[1] https://www.smivision.com/wp-content/uploads/2016/11/smi_prod_eyetracking_hmd_HTC_Vive.pdf

The experiment started with an introduction to the HMD. Then participants were instructed to explore 360° photographs of six different places for 20 s each. Stimuli and scenes are paired pseudo-randomly such that all combinations of stimulus and scene appear approximately equally often. Fixed stimuli locations are manually chosen per scene, but equally distributed around the viewing position throughout all scenes. Before presentation of each picture, a ring with decreasing size serving as a fixation point for a common initial gaze location is shown. After calibration of the eye tracking system, the experiment started with the fixation ring, followed by the first scene and so on, until all stimuli were presented.

A post-experiment questionnaire again assessed the participants current health condition with respect to simulator sickness, using the same SSQ. Additionally, they were asked to identify "possible visual artifacts" that may have occurred during presentation.

5. RESULTS

A total of 102 participants (31 female, 71 male) aged between 19–41 (average: 25.7, σ = 3.79) took part in the experiment. Normal vision was reported by 55 participants and 47 reported corrected-to-normal vision. We excluded one participant that was severely affected by simulator sickness, identified by a *Total Severity* (TS_{SSQ}) [16] score difference before and after the experiment of more than 50, indicating a strong increase of motion sickness:

$$\mathrm{TS}_{SSQ}^{after}(p) - \mathrm{TS}_{SSQ}^{before}(p) > 50, p \in Participants$$

Figure 4 shows an exemplary 360° image, overlaid with stimulus-wise averaged heatmaps of captured gaze data. Without gaze guidance, the target region barely received any attention (subfigure *None*), whereas with active guidance, all techniques significantly increased the attention for this spot.

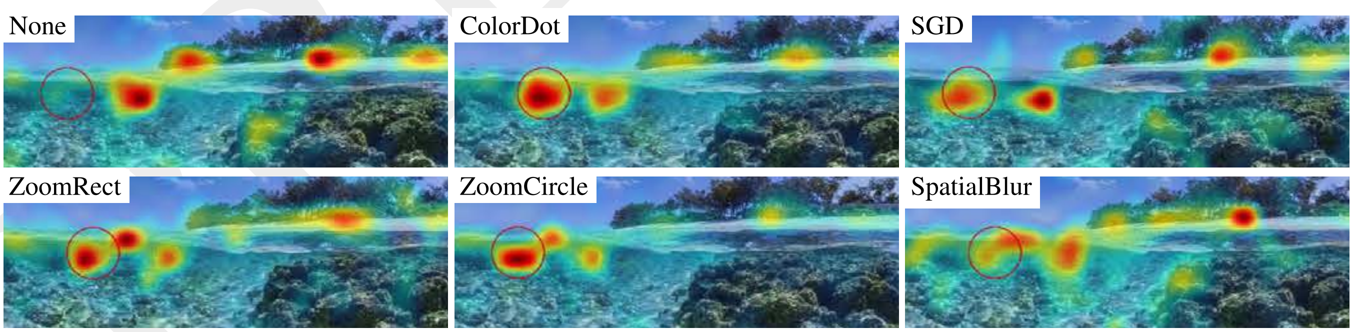

Fig. 4: Exemplary 360° picture with heatmap overlay of accumulated gaze data for all participants when presenting the different stimuli (or *None*). The thin red circle marks the position where the stimulus was presented.

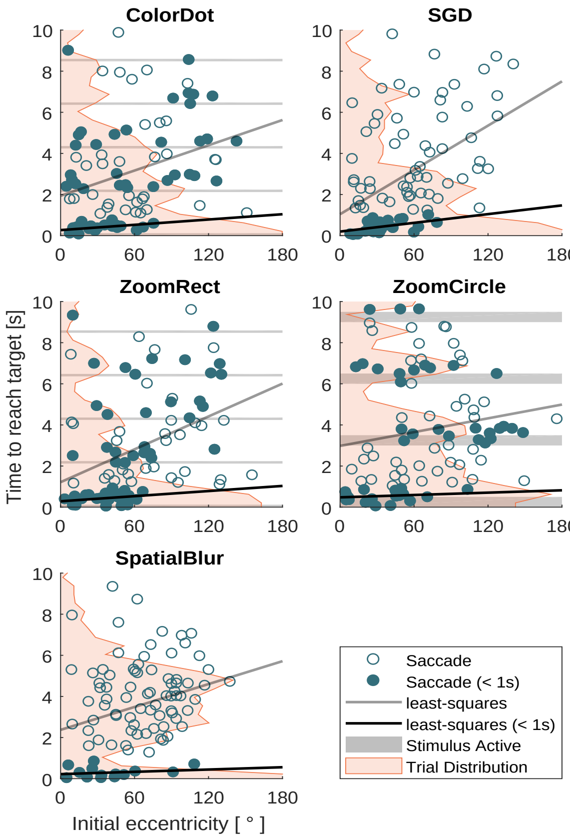

Fig. 5: Time to Reach the Target (TRT) plotted against initial target eccentricity (outlined circles). For repetitive stimuli, the presentation periods are highlighted in light gray. Trials with target-directed saccades within one second after a stimulus onset are highlighted as solid circles. Target-directed saccades become less frequent over time as the stimulus does not reappear after its first fixation.

Because a sudden onset of a stimulus in the viewer's peripheral visual field should trigger a saccade almost instantly, later saccades are most probably not related to the stimulus. Therefore, we inspect only the first second of gaze data after stimulus onsets. By design ColorDot, ZoomRect and ZoomCircle are shown repeatedly for fixed short durations (gray bars in Figure 5) with long phases of invisibility in between. Repetition is canceled on first target-directed saccade detection. For those, the 1 s timespan after the corresponding stimulus onset is analyzed. Without this repetition, ZoomRect outperforms the other stimuli with up to 30 % successfully induced target-directed saccades within 1 s — twice as much as SpatialBlur, depicted in Figure 6 (solid lines). A reason may be that SpatialBlur does not precisely highlight target locations and thus requires more extensive search for users to spot the target. With repeated presentation, ZoomRect and ColorDot reach almost 100 % success rate (within the full 20 s period), as depicted in Figure 6 (dotted lines), rendering our stimulus repetition a valuable extension.

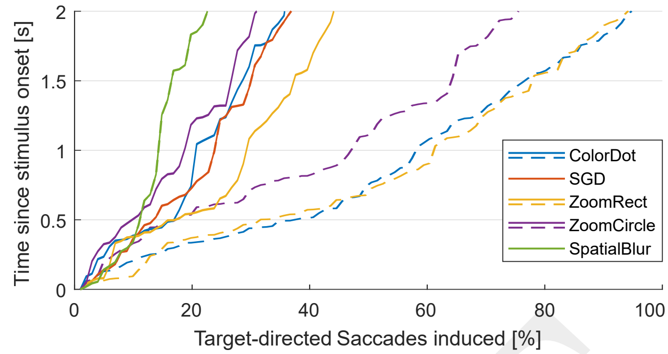

Fig. 6: TRT after stimulus onsets with (dotted lines) and without (solid lines) accumulation for repeated stimuli.

Besides guidance performance, we are also interested in the level of subtleness of the tested methods, i.e. how good they are at *not* being consciously perceived by viewers. Evaluation of the post-experiment questionnaire with regard to subtleness shows that 26 participants reported *no image artifacts* overall. Irritating *blur on the whole image* was reported by 59 participants while a suddenly appearing *red dot* was identified 12 times and a *zoom effect* was correctly recognized by only 3 participants. Additionally, 26 participants provided vague descriptions of something flickering, which may hint at SGD or ZoomRect but could also describe ZoomCircle or ColorDot. We intentionally did not ask for specific visual artifacts to not bias the participants.

Based on our data, we infer that to date there is no stimulus that remains completely imperceivable. From our results we recommend SGD, ZoomRect or ZoomCircle for gaze guidance in VR.

6. CONCLUSION

We compare the effectiveness of several adapted image-space post-processing subtle gaze guidance stimuli in an HMD. In particular, we analyze eye tracking data recorded in a psychophysics study with 102 participants and 12 distinct 360° images. Our results demonstrate that the adapted ZoomRect stimulus slightly outperforms the other stimuli in such wide–FOV immersive environments, inducing target–directed saccades in up to 30 % of the trials within the first second. Furthermore, the presented stimulus repetition method yields noticeable performance improvements for suitable stimuli.

In future work, individual stimuli shall be assessed in more detail with regard to subtleness to derive possible recommendations on how to improve them. We aim to examine systems with even wider FOV to further approach the actual limits of the human visual field. Also repetitive stimuli are to be investigated with respect to priming effects between successive stimulus presentations. Furthermore, we plan to extend our study to investigate subtle gaze guidance in 360° video sequences.

7. REFERENCES

[1] Michael Dorr, Thomas Martinetz, Karl Gegenfurtner, and Erhardt Barth, "Guidance of eye movements on a gaze-contingent display," in *Dynamic Perception Workshop of the GI Section "Computer Vision"*, Uwe J. Ilg, Heinrich H. Bülthoff, and Hanspeter A. Mallot, Eds., 2004, pp. 89–94.

[2] Reynold Bailey, Ann McNamara, Nisha Sudarsanam, and Cindy Grimm, "Subtle gaze direction," *ACM Transactions on Graphics (TOG)*, vol. 28, no. 4, pp. 100, 2009.

[3] Hajime Hata, Hideki Koike, and Yoichi Sato, "Visual guidance with unnoticed blur effect," in *Proceedings of the International Working Conference on Advanced Visual Interfaces*. ACM, 2016, pp. 28–35.

[4] Yen-Chen Lin, Yung-Ju Chang, Hou-Ning Hu, Hsien-Tzu Cheng, Chi-Wen Huang, and Min Sun, "Tell me where to look: Investigating ways for assisting focus in 360 video," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2017, pp. 2535–2545.

[5] Steve Grogorick, Michael Stengel, Elmar Eisemann, and Marcus Magnor, "Subtle gaze guidance for immersive environments," in *Proceedings of the ACM Symposium on Applied Perception*. ACM, 2017, pp. 4:1–4:7.

[6] Erhardt Barth, Michael Dorr, Martin Böhme, Karl Gegenfurtner, and Thomas Martinetz, "Guiding the mind's eye: improving communication and vision by external control of the scanpath," in *Electronic Imaging 2006*. International Society for Optics and Photonics, 2006, pp. 60570D–60570D.

[7] Raymond Dodge, "Visual perception during eye movement.," *Psychological Review*, vol. 7, no. 5, pp. 454, 1900.

[8] Ann McNamara, Reynold Bailey, and Cindy Grimm, "Search task performance using subtle gaze direction with the presence of distractions," *ACM Transactions on Applied Perception (TAP)*, vol. 6, no. 3, pp. 17, 2009.

[9] Thomas Booth, Srinivas Sridharan, Ann McNamara, Cindy Grimm, and Reynold Bailey, "Guiding attention in controlled real-world environments," in *Proceedings of the ACM Symposium on Applied Perception*. ACM, 2013, pp. 75–82.

[10] Robert Kosara, Silvia Miksch, and Helwig Hauser, "Focus+ context taken literally," *IEEE Computer Graphics and Applications*, vol. 22, no. 1, pp. 22–29, 2002.

[11] Forrester Cole, Douglas DeCarlo, Adam Finkelstein, Kenrick Kin, R Keith Morley, and Anthony Santella, "Directing gaze in 3d models with stylized focus.," *Rendering Techniques*, vol. 2006, pp. 17th, 2006.

[12] Laurent Itti, Christof Koch, and Ernst Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[13] Aiko Hagiwara, Akihiro Sugimoto, and Kazuhiko Kawamoto, "Saliency-based image editing for guiding visual attention," in *Proceedings of the 1st international workshop on pervasive eye tracking & mobile eye-based interaction*. ACM, 2011, pp. 43–48.

[14] Eduardo E. Veas, Erick Mendez, Steven K. Feiner, and Dieter Schmalstieg, "Directing attention and influencing memory with visual saliency modulation," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2011, CHI '11, pp. 1471–1480, ACM.

[15] Andrei Lintu and Noëlle Carbonell, "Gaze guidance through peripheral stimuli," working paper or preprint, 2009.

[16] Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lilienthal, "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness," *The International Journal of Aviation Psychology*, vol. 3, no. 3, pp. 203–220, 1993.