GazeRecall: Using Gaze Direction to Increase Recall of Details in Cinematic Virtual Reality

Sylvia Rothe
LMU Munich
Munich, Germany
sylvia.rothe@ifi.lmu.de

Felix Althammer
LMU Munich
Munich, Germany
felix.althammer@campus.lmu.de

Mohamed Khamis
LMU Munich, University of Glasgow
Glasgow, UK
Mohamed.Khamis@glasgow.ac.uk

ABSTRACT
In this work, we show how the use of flickering in Cinematic Virtual Reality (CVR) content can not only guide the user's attention, but also significantly improve recall of details. The findings are particularly useful for practitioners who generate educational and corporate training content, as well as for directors of CVR videos who want to draw the user's attention to certain details in movies or virtual tours. We report on a between-subjects user study, in which we experimented with flickering methods to increase user's recall of certain details in CVR videos. Participants had to report details of objects on which the method was applied. In our experiments, the intensive flicker improved the recall of details significantly. We discuss how the studied methods can improve recall of details in different use cases.

Author Keywords
Cinematic Virtual Reality; 360° movie; omnidirectional guiding attention; Subtle Gaze Direction

ACM Classification Keywords
H.5.1 Multimedia Information Systems: Artificial, augmented, and virtual realities

INTRODUCTION
360° videos are attracting widespread interest and have many possible applications. For example, they can be used to tell stories about exciting locations in the world, give virtual tours of museums or ancient places, and can be used for educational and corporate training purposes.

In Cinematic Virtual Reality (CVR) the viewer watches a 360° movie using a Head-Mounted Display (HMD). Thus, the viewer is inside the scene and can freely choose the direction of view. Accordingly, the viewer determines the visible section of the movie – the field of view (FoV). Therefore, it is not always possible to show the viewer what is important for the story. Several conventional filmmaking methods for guiding the viewer's line - such as close ups or zooms - are not practical in CVR. For other methods, it needs a closer analysis of whether they are suitable to direct the attention of the viewer to important details in a CVR environment. In this paper, we focus on subtle guidance of user's gaze and explore how it can impact recall of details.

When wearing a HMD, users often look around and could potentially miss important details [16]. Acknowledging this problem, researchers have developed ways to guide the user's attention to ensure they gaze at desired targets [6,11,12]. Subtle methods for gaze direction were investigated for static images on flat displays [1,7]. Such methods can improve the success in search task [7] and reduce the error rate in remembering to regions and their locations [1]. However, little attention was given to studying how well users can recall the details to which they were directed in CVR. Without an understanding of how well users recall the details, it is not clear if gaze direction results in attending to the details, or simply looking towards them, which are two distinct things that cannot be distinguished through eye tracking alone [20].

We report on a user study (N=42), in which we studied the effect of a method for gaze direction on the recall of details that are directed to. We experimented with several flickering intensities and found that higher flicker intensity can increase the recall rate for details in the guided area. Since the ideal method should not negatively influence the VR experience, we investigated its impact on presence and enjoyment. Based on the results, we discuss the tradeoff between ensuring high enjoyment and increasing recall of details.

The contribution of this work is two-fold: 1) we report on a user study in which we evaluated if subtle gaze direction can be adapted to guide users using flickering in CVR videos. 2) we investigate the effect of a state-of-the-art gaze direction method on the recall of details in CVR videos.

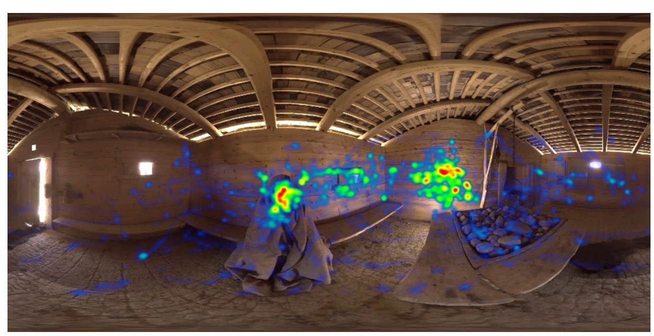

Figure 1: For investigating how the guiding methods influence the gaze of the viewer we recorded head and eye movements and generated heatmaps.

MUM '18, November 25–28, 2018, Cairo, Egypt
© 2018 Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-6594-9/18/11…$15.00
https://doi.org/10.1145/3282894.3282903

RELATED WORK

Various methods for guiding the users gaze in images or movies are explored for non-VR environments. Some of these methods modify image properties as colour, intensity and acuity [5,10,14,17] Smith et al. [14] investigated nonblurred regions for guiding the viewer's attention. They showed that the viewer tends to regard regions with little or no spatial blur, if the rest of the image is more blurred. This approach is very similar to a method used in traditional filmmaking. Hata et al. [5] extended this method for visual guidance with unnoticed blur effects for images. A threshold was found at which the viewer noticed the blur and could be guided below it. We performed a test of this method with a CVR movie. However, at the present time the resolution of the HMD-displays is not high enough to produce a higher interest in the non-modulated region.

Mendez et al. [10] described a Salience Modulation Technique (SMT) for directing a viewer's attention to a target object. For analyzing the original images, saliency maps are used and the material was modulated dependent on the results of the analysis. So, it was possible to apply minimal changes. This method is working on videos in real time. However, it can only be used, if the target is in the viewer's field of view. Additionally, the method is less strong in environments with moving and blinking objects. Veas et. al [17] investigated SMT regarding modulation awareness, attention and memory. They showed that SMT can shift the attention to selected targets without the viewer noticing the modulation. Additionally, SMT can increase the recall rate.

In the case, the user is not aware of a method, the method is called subtle. In contrast to this, overt techniques will be noticed by the user [15]. Bailey et al. [2] described a method for subtle gaze direction (SGD). They combined subtle image modulation with eye-tracking to direct the viewer's gaze. The method exploits that the peripheral vision has less acuity than the foveal vision (vision in the center of the field of vision). Two methods were investigated: luminance modulation and warm-cool modulation (switching between warm and cool colors), both with a rate of 10Hz in a circle region of approximately 1cm diameter. The modulated region was placed in the peripheral area of the FoV. The modulation stopped when the viewer changes the view in the direction of the modulated region. It was shown that this method is working for digital images.

Ben-Joseph and Greenstein investigated SGD without using eye tracking [3]. They took advantage of the fact, that eyes tend to lead the head [13] and used the head direction for deactivating the flicker if the point of interest (PoI) is near the viewing direction.

In the experiments of Namara et al. [8] modulations of luminance were more effective than warm-cool modulations. They showed that SGD improves the performance for search tasks in images without the participants noticing the modulation. The same method was investigated for guiding in narrative art, with static images [9].

This method was expanded to virtual environments [4]. A luminance modulation was used and the circle shape was dynamically adapted to ellipses for the wide FoV in VR. Additionally, the stimulus was dynamically positioned, so the method can be used also for targets, which are not in the FoV in the beginning of the stimulation. The experiments showed that research tasks results can be improved for hidden objects. In our study, we want to investigate if it is possible to guide the viewer in a CVR video, without any task. Additionally, the environment in CVR is more dynamic.

Waldin et al. [18] took advantage of the fact that the peripheral vision is more sensitive for high-frequent flickering than the foveal vision. So, certain flickers can only be detected in the peripheral region. If the viewer is changing the view in the direction of the flicker, the flicker fuses to a stable image. In this way, no eye-tracking is necessary for stopping the modulation (as it is in SGD). The experiments used flickers of 60Hz and 72Hz, so a display of 120Hz and 142Hz was needed. In a first experiment images with cycles were used, in a second experiment a high complex image. The method worked effectively in both cases. It seems to be necessary to execute a personal calibration routine to find the size and luminance of the flicker modulation.

Compared to previous work, the novelty of our work is that we investigate if SGD can be adapted for CVR, and how it affects the recall of details. To this end, two extensions to previous work are necessary 1): adjusting SGD for 360° videos, and 2) adapting SGD for targets that might not necessarily be in the user's FoV.

USERSTUDY: SGD WITH EYE TRACKING

We extended the work of Bailey et al. [2] and Grogorick et al. [4] to investigate subtle gaze direction in cinematic virtual reality. The point of interest was added with a 10Hz flicker, which disappears if the viewer is looking at it. For CVR this method consists of two parts: on the one hand the component for guiding the viewers to the right field of view, on the other hand, the component for guiding them to the exact PoI inside the field of view. In case the PoI is not in the field of view, a flicker appeared at the right or the left edge of the display depending on the position of the PoI relative to the user's field of view. A tolerance range was added to prevent the flicker from jumping between sides when the PoI is behind the viewer.

Apparatus

We used an HTC Vive with an integrated Pupil Labs eye tracker. The participants watched a video (resolution 4096x2048, 25fps), which consists of three scenes. The first one shows a Celtic village where a man talks about the Celtic life. In the second scene the man is sitting in a Celtic house and explains some details about the daily life. Afterwards, the room is filled with Celtic people who are performing different crafts (Figure 2). The first two scenes were used for accustoming the user to the system. In the third scene three objects were chosen as target for the guiding: a woman with an apron at a loom, a man with beard and a spear behind another man. For adding the flickers and recording head and eye tracking data Unity 3D was used.

Figure 2: Scene for guiding. The areas where the viewer was guided are marked red, left: spear, middle: man with beard, right: woman with apron.

Parameters

The frequency (10Hz) and the size (0.76°) of the circular flicker was adapted from the original method of Bailey [2]. In an informal pilot-study we tested several color intensities to find out which values are effective and subtle at the same time. The color intensity is a value in the interval [0,1], which determines the intensity of the modulations. For the pilot study we used the same procedure as Bailey et al. in their original paper [2]. Five participants watched a short video and were instructed to look at a yellow circle in the middle of the FoV. For a flicker in their periphery the intensity was increased by 0.1 steps until the participant noticed it. For every participant the same eight places for the flicker were chosen, with different backgrounds from dark to light. Participants saw the flicker with an average intensity of 0.37, which is higher than in the pilot study of Bailey (i=0.1), where still images were used. For moving picture, the noticed frequency seems to be higher.

For our main study we applied two intensities, one subtle with the intensity determined by the pilot study and one less subtle with the intensity of 1, which was recognized by all participants of the pilot study.

Measurements

The questionnaire consisted of several parts: general part (demographics), enjoyment, presence and recall questions. For the question "How much did you enjoy the experience?" a Likert scale from 1="not at all" to 7="very much" was used. To investigate the presence, we added parts of the presence questionnaire (PQ) of Witmer and Singer [19]. Since the PQ was developed for general Virtual Environments with interactivity and movement, we chose some of the questions which are relevant for Cinematic VR: (Q1) How involved were you in the virtual environment experience? (Q2) How much did the visual aspects of the environment involve you? (Q3) How much did the auditory aspects of the environment involve you? (Q4) How quickly did you adjust to the virtual environment experience? (Q5) How much did your experiences in the virtual environment seem consistent with your real-world experiences? The questions were answered on a seven-level Likert scale from 1="not at all" to 7="completely"(Q1-3) / "less than one minute"(Q4) / "very consistent"(Q5).

The participants were guided to three different PoI and asked about visual details in these areas (recall part): the colour of the apron, the form of the beard and the presence of the spear. Additionally, we recorded head and eye movements, for analyzing where the viewers are looking (Figure 1).

Participants and Procedure

Our study was designed as a between-subjects experiment. We divided 42 participants randomly into two groups: Group A (5 female, 16 male, average age 24.9), Group B (9 female/12 male, average age 25.4). All participants watched the same video, but each group experienced a different flicker intensity. Group A experienced a lower flicker intensity. Group B watched the same video with a more obvious flicker that had a higher intensity. Our movie consisted of three scenes. The first two scenes were used for acclimatization and 15s after the third scene started, the first flicker appears. The flicker guided the viewer to the point of interest for 20 seconds. After that, the next point of interest was aim of the guiding process. The order of the targets was counterbalanced using a Latin square.

The participants watched the videos in a standing posture turning their head and body. No task was assigned, the participants should look around and follow the objects they are interested in. After watching the video, a part of the questionnaire was filled out.

Results

Enjoyment

To find out how the flicker intensity influence the enjoyment, we performed a t-test. The score for the question "How much did you enjoy the experience?" for Group A (mean=5.19, SD=1.36) was significantly higher than for group B (mean=4.43, SD=1.4), p=0.04 (Figure 3). This suggests that the higher the flickering, the less the users enjoy the videos.

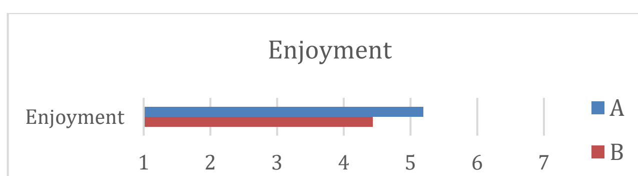

Figure 3: Means for the enjoyment items. The enjoyment is higher when using the flicker that has a lower intensity.

Presence

For comparing the presence across the two groups, we performed t-test for each item, which showed no significant differences (p> 0.1) (Figure 4). This means we did not find any evidence that the flickering influences presence in CVR.

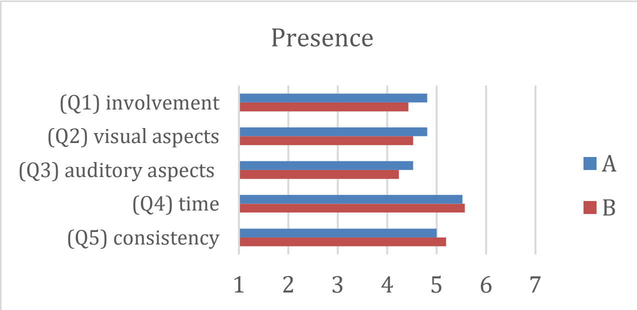

Figure 4: Means for the presence items from the presence questionnaire (PQ) of Witmer and Singer [19]**. No significant effect of the flickering intensity was found on the presence.**

Recall Rate

The recall rate was better for group B (Figure 5). More people could reproduce the form of the beard and the colour of the apron when the flicker with the higher intensity was used. The spear was not perceived, possibly because of its position (higher than the other two) or the bright area nearby (the open door), which makes it harder to notice it. For all recall questions the participants had 5 possibilities for answering. So, a rate of 20% is also probably by guessing. The exact fisher test showed a significant difference between the recall rates of group A and B: $\chi^2 = 6.61$ and p=0. 01.

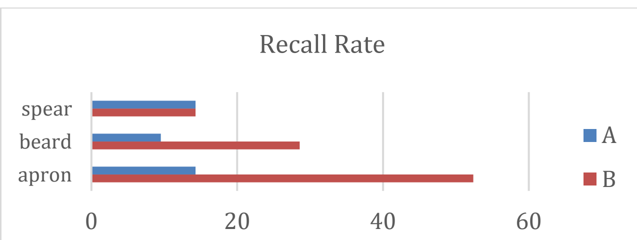

Figure 5: Recall Rate for the three objects in percent. The spear was hard to see. For the other two objects the higher intensity increased the recall rate significantly.

Heatmaps

Analyzing the heatmaps for head and eye tracking data, we found that the movements while watching the video were more in-line with the flickering when it had a higher intensity (Group B) than it was subtle (Group A).

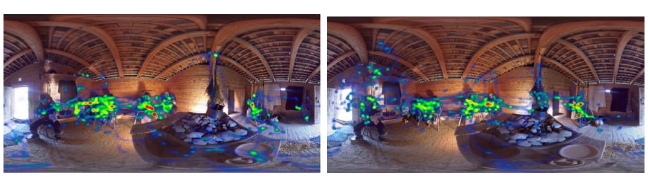

Figure 6: Heatmaps for group A (left) and group B (right) in the time interval, when the flicker was on the women on the right side. The intensive flicker caused a hotspot at the target.

As already mentioned earlier, participants seldom focused on the spear, which can also be seen in the heat maps (Figure 7). This explains why the recall rate of the spear's details has not been improved.

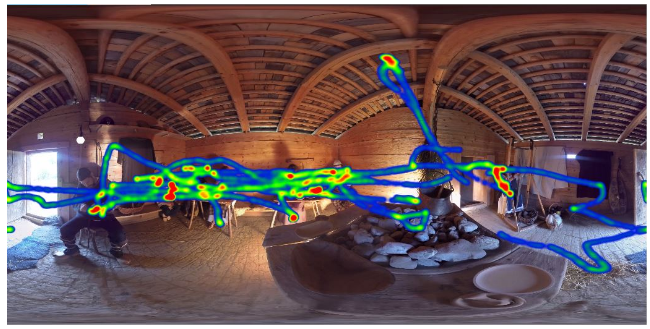

Figure 7: The flicker is on the spear the left side of the image (white point over the man next to the door). Data forms heatmaps around moving things, not near the flicker.

DISCUSSION AND FUTURE WORK

In our experiments the subtle flicker was not stronger than movements or interesting things in the movie. Although subtle gaze guiding is suitable for guiding in still images [2], in moving pictures they compete with cues in the movie which attract the attention of the viewer. However, the flicker with the higher intensity, which was not subtle, could guide the user in the most cases. While the higher flicker intensity reduces enjoyment, it increases the recall of details.

The effectiveness of flickering to guide gaze depends on the video material and hardware. The brightness of the area of interests, where the viewer should be guided, has a big influence on the effectiveness of the flicker. Even though we used the same video material in the pilot and main study, and the parameter were determined for this material, the same parameters were applied for all backgrounds. For future work, we plan to investigate how flickering can be adapted to the background to improve its effectiveness.

Methods such as SGD with high frequent flickers [18] require a display frequency higher than 90Hz and a wider field of view. The resolution and frequency of today's displays restricts the potential of subtle flicker methods. However, advances in hardware promise higher frequency displays in the future, which would result in better control of the subtlety of flickering for gaze direction and recall improvement.

CONCLUSION

We experimented with flickering in CVR. A user study was conducted where guiding gaze with a subtle flicker was compared with a more obvious flicker. We did not find significant impacts on presence. Subtle flickers seem to be less effective in CVR than in still pictures, since movements draw the attention of the viewer. While more obvious flickering decreases enjoyment, it increases the recall rate. So, methods with high intensity flickering could be used for serious use cases, where remembering details is more important than enjoying the video.

REFERENCES

1. Reynold Bailey, Ann McNamara, Aaron Costello, Srinivas Sridharan, and Cindy Grimm. 2012. Impact of subtle gaze direction on short-term spatial information recall. *Proceedings of the Symposium on Eye Tracking Research and Applications*: 67–74.

2. Reynold Bailey, Ann McNamara, Nisha Sudarsanam, and Cindy Grimm. 2009. Subtle gaze direction. *ACM Transactions on Graphics* 28, 4: 1–14.

3. E Ben-Joseph and E Greenstein. 2016. Gaze Direction in Virtual Reality Using Illumination Modulation and Sound. Retrieved July 30, 2018 from https://web.stanford.edu/class/ee267/Spring2016/report_benjoseph_greenstein.pdf.

4. Steve Grogorick, Michael Stengel, Elmar Eisemann, and Marcus Magnor. 2017. Subtle gaze guidance for immersive environments. *Proceedings of the ACM Symposium on Applied Perception - SAP '17*, ACM Press, 1–7.

5. Hajime Hata, Hideki Koike, and Yoichi Sato. 2016. Visual Guidance with Unnoticed Blur Effect. *Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI '16*, ACM Press, 28–35.

6. Yen-Chen Lin, Yung-Ju Chang, Hou-Ning Hu, Hsien-Tzu Cheng, Chi-Wen Huang, and Min Sun. 2017. Tell Me Where to Look. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, ACM Press, 2535–2545.

7. Ann McNamara, Reynold Bailey, and Cindy Grimm. 2008. Improving search task performance using subtle gaze direction. *Proceedings of the 5th symposium on Applied perception in graphics and visualization - APGV '08*, ACM Press, 51.

8. Ann McNamara, Reynold Bailey, and Cindy Grimm. 2008. Improving search task performance using subtle gaze direction. *Proceedings of the 5th symposium on Applied perception in graphics and visualization - APGV '08*, ACM Press, 51.

9. Ann McNamara, Thomas Booth, Srinivas Sridharan, Stephen Caffey, Cindy Grimm, and Reynold Bailey. 2012. Directing gaze in narrative art. *Proceedings of the ACM Symposium on Applied Perception - SAP '12*, ACM Press, 63.

10. Erick Mendez, Steven Feiner, and Dieter Schmalstieg. 2010. Focus and Context in Mixed Reality by Modulating First Order Salient Features. In Springer, Berlin, Heidelberg, 232–243.

11. Lasse T. Nielsen, Matias B. Møller, Sune D. Hartmeyer, et al. 2016. Missing the point. *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology - VRST '16*, ACM Press, 229–232.

12. Sylvia Rothe and Heinrich Hußmann. 2018. Guiding the Viewer in Cinematic Virtual Reality by Diegetic Cues. *International Conference on Augmented Reality, Virtual Reality and Computer Graphics* , Springer, Cham, 101–117.

13. Malcolm Slaney, Andreas Stolcke, and Dilek Hakkani-Tür. 2014. The Relation of Eye Gaze and Face Pose. *Proceedings of the 16th International Conference on Multimodal Interaction - ICMI '14*, ACM Press, 144–147.

14. Wayne S. Smith and Yoav Tadmor. 2013. Nonblurred regions show priority for gaze direction over spatial blur. *Quarterly Journal of Experimental Psychology* 66, 5: 927–945.

15. Evan A. Suma, Gerd Bruder, Frank Steinicke, David M. Krum, and Mark Bolas. 2012. A taxonomy for deploying redirection techniques in immersive virtual environments. *2012 IEEE Virtual Reality (VR)*, IEEE, 43–46.

16. Hannah Syrett, Licia Calvi, and Marnix van Gisbergen. 2017. The Oculus Rift Film Experience: A Case Study on Understanding Films in a Head Mounted Display. In Springer, Cham, 197–208.

17. Eduardo E. Veas, Erick Mendez, Steven K. Feiner, and Dieter Schmalstieg. 2011. Directing attention and influencing memory with visual saliency modulation. *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*, ACM Press, 1471.

18. N. Waldin, M. Waldner, and I. Viola. 2017. Flicker Observer Effect: Guiding Attention Through High Frequency Flicker in Images. *Computer Graphics Forum* 36, 2: 467–476.

19. Bob G. Witmer and Michael J. Singer. 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments* 7, 3: 225–240.

20. Takemasa Yokoyama, Hiroki Sakai, Yasuki Noguchi, and Shinichi Kita. 2015. Perception of Direct Gaze Does Not Require Focus of Attention. *Scientific Reports* 4, 1: 3858.