Attention Guidance for Immersive Video Content in Head-Mounted Displays

Fabien Danieau* Technicolor, France Antoine Guillo ENSAM, France Renaud Doré Technicolor, France

ABSTRACT

Immersive videos allow users to freely explore 4π steradian scenes within head-mounted displays (HMD), leading to a strong feeling of immersion. However users may miss important elements of the narrative if not facing them. Hence, we propose four visual effects to guide the user's attention. After an informal pilot study, two of the most efficient effects were evaluated through a user study. Results show that our approach has potential but it remains challenging to implicitly drive the user's attention outside of the field of view.

Keywords: HMD, immersive movies, visual attention

Index Terms: H.5.1 [HCI]: Multimedia Information Systems— Artificial, augmented, and virtual realities; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual Reality;

1 INTRODUCTION

Immersive systems allowing the exploration of 4π steradian scenes provide novel ways to experience multimedia content. However, this yields new constraints on the side of the content creation. As the user gazes all around, the point of interest (PoI) defined by the content creator may be not seen. Hence important elements of the narrative could be missed. In such conditions, the content creator may want to attract the user's attention to the PoI of the scene when something crucial happens (potentially behind the user).

Filmmakers already use techniques to attract the audience's attention to specific elements of a frame. The control of the depth of field is one extensively used [3]. The focal length of the camera is adjusted in a way to make clear the important elements while the remaining of the scene is blurred. Similar techniques have been implemented in virtual reality systems [4], but these solutions intend to clarify the location the user is looking at rather than attracting the gaze toward a specific direction. Another well-known cinematographic technique consists in moving the camera in order to center the PoI in frame such as a "Pan" or a "Tilt" [3]. But in the context of an immersive system covering the whole field of view, this could lead to motion sickness effects. Nevertheless, Bolte and Lappe proposed to rotate or translate the camera during saccades (quick eye movement between fixation) [2]. To a certain extent $(\pm 5^{\circ})$, a camera rotation is not perceived. This technique requires an eye tracking system though. The user's attention is also attracted by certain features of an image (color, motion, etc.). Based on this principle, Bailey et al. used luminance or color modulation to subtly direct the user's gaze [1]. Finally, elements of the mise-en-scene may be used to drive the user's attention [3]. Pausch et al. relied on characters pointing at or moving toward the PoI within their immersive system [5]. This approach appears to be the less invasive for the user but requires a control of the mise-en-scene which may not be possible in certain situation (live or documentary for instance).

2 ATTENTION GUIDANCE IN VR

In this paper we propose to smoothly drive the user's gaze toward a PoI in a 4π steradian scene using a visual effect. To illustrate our approach we have designed four effects (see Figure 1). A neutral 3D scene has been created and integrated into the Unity engine. Points of interest have been programmed to appear outside of the users field of view. These techniques were prototyped with ShaderForge and tested with the HMD Oculus Rift CV1.



Figure 1: The four effects. The intensity of each effect is gradually increased to drive the user's attention toward the Pol (the explosion here). The black rounded square represents the user's field of view.

Fade to black (Figure 1, top-left). Based on the principle of lightning in cinematography [3], we hypothesize that the user looks at the highlighted elements within an image rather than the ones in the shadow. Here this effect changes the color of the pixels. The farther they are from the PoI the darker they become. At the end, only the PoI seems illuminated. The shadowing of the image is linearly applied.

Desaturation (Figure 1, bottom-left). This effect is designed to increase the saliency of the zone of interest. The pixels outside of this area are progressively desaturated. The hypothesis is that the user will look at the colored area, more salient than the other part of the image.

Blur (Figure 1, top-right). This effect is inspired from the change of the depth of field used in cinematography. It is often used to guide the user's attention toward a PoI (for instance from one character in foreground to another in background). Here pixels far from the PoI are gradually blurred while the others stay clear.

Deformation (Figure 1, bottom-right). The user's peripheral vision is exploited in this effect. We propose here to display a wavelike effect on one side of the user's field of view to incite him to turn his head. Contrary to the others, this effect is local (i.e. not applied on the whole rendered image).

An informal pilot study was conducted to get a first user feedback and to perform a fine tuning of each effect. Objective (head's orientation) and subjective (questionnaire) measures were collected. We observed that Blur and Deformation were not successful at directing one's gaze. The Blur was perceived as progressively blurring the image, without any clear direction. The Deformation effect was still visible and disturbing although at the edge of the field of view. It seems that the field of view of current HMDs (110°) is too narrow. Hence only the two effects Fade to black and Desaturation were selected for a user study.

^{*}e-mail: fabien.danieau@technicolor.com

3 USER STUDY

Two immersive videos were used, a movie $(Help^1)$ and a sport event $(GoPro^2)$, with two simultaneous PoI present on the left and right. Extracts of 40s were selected with the PoI occurring at 16s for Help, and 20s for GoPro. Two effects were then tested: "Fade to black" (referred as *EB*) and "Desaturation" (*ED*). Besides, two controls effects were added: *EN* where no effect was applied on the camera, and *EF* where the camera was automatically oriented toward the PoI as it is traditionally done in movies. With EN we expected the user to miss the PoI while with EF we expected the quality of experience to be decreased. Each effect was then tuned to guide the user toward the left or right PoI. The duration for each effect was 1s to be in line with the narrative speed. In this within-subjects study, a total 2 videos \times 2 PoI \times 4 effects =16 conditions had to be experienced, for a total duration of 35 minutes.

The orientation of the participant's head was captured at the apparition of the PoI (yaw axis). This objective measure allowed us to measure the angle between the head and PoI. A questionnaire was also designed to evaluate the subjective experience. Four assertions were evaluated on a 5-point Likert Scale: the experience was comfortable (Q1), I felt like something was directing my gaze (Q2), I felt dizzy (Q3), the visual rendering was disturbing (Q4).

For each participant the experiment was first introduced, and an informal test of the Oculus was proposed. Then a questionnaire was submitted to collect information on the age, gender and expertise in video games and VR (from "0 - none" to "4 - daily use"). The 16 randomized conditions were then experienced, and for each one the questionnaire was submitted. Finally the participant was incited to freely comment the experience during an open-ended interview.

4 RESULTS & DISCUSSION

10 participants took part in this experiment, aged from 30 to 56 ($\bar{x} = 41.8$, $\sigma_x = 9.61$), including 1 female. They had a medium expertize in video games ($\bar{x} = 2.5$, $\sigma_x = 1.18$), and a low expertize in VR ($\bar{x} = 1.9$, $\sigma_x = 1.20$). Non-parametric tests were used to analyze those data: Friedman Anova and pairwise Wilcoxon tests with Holm-Bonferroni correction.



Figure 2: Average angles between the participants' head and the Pol.

Results of the objective measure were first investigated (see Figure 2). The results suggested that EB performed slightly better than the control condition EN. However even though a F. Anova was significant (Help: $\chi^2 = 13.62, df = 3, p = 0.003$, GoPro: $\chi^2 = 9.78, df = 3, p = 0.021$), the Wilcoxon tests showed differences for EF-EN and EF-ED (p < 0.05 for both videos). Also, no statistical differences were found between the left and right PoI.



Figure 3: Questionnaire results. Means and standard deviations.

The results of the questionnaire were then analyzed for both videos (see Figure 3. No differences were found). The F. Anova was significant for all questions (p < 0.05), but Wilcoxon tests pointed out few statistical differences. For Q1, only EF and EN were different (p = 0.022): the comfort with our effects ED and EB was then not different from the one with EN. For Q2, only EN was different from the others (p < 0.05): our effects were perceived. Regarding Q3, the conditions were not statistically different: the effects were not leading to a feeling of dizziness. Finally, Q4 showed that EN was different from the others (p < 0.05). Besides, EB and EF were different (p = 0.007), suggesting that EB was disturbing.

From the interviews, we observed a trade-off between the efficiency and the visibility of the effects. Implicitly driving the user's gaze is an ambitious challenge and does not seem really possible within a duration of few seconds. Either the effect is disturbing, as with EB, and the user may follow the pointed direction, or it is discreet and might be ineffective as ED was. It has also to be noted that the participants had to watch height times each video. Even if the conditions were randomized, the awareness of the narrative could influence the participants' behavior and will have to be investigated in future studies. Interestingly, the behavior of the participants with EF was not as expected: most of them compensated the forced rotation to stay in front of the PoI they had chosen. This confirms that the orientation of the user's head must be let to him.

5 CONCLUSIONS & PERSPECTIVES

In this paper, new techniques to guide the user's attention during the viewing of immersive videos were presented. Four visual effects were designed and two were evaluated. Results showed that this approach has a potential but also that it is a challenging task to make a user unconsciously move his head. Future work will be dedicated to the improvement of the four visual effects. Besides, others cues (i.e. audio or haptic) will be added to enforce them.

REFERENCES

- R. Bailey, A. McNamara, N. Sudarsanam, and C. Grimm. Subtle gaze direction. ACM TOG, 28(4):100, 2009.
- [2] B. Bolte and M. Lappe. Subliminal Reorientation and Repositioning in Immersive Virtual Environments using Saccadic Suppression. *IEEE TVCG*, 21(4):545–552, 2015.
- [3] D. Bordwell, K. Thompson, and J. Ashton. *Film art: An introduction*, volume 7. McGraw-Hill New York, 1997.
- [4] S. Hillaire, A. Lécuyer, R. Cozot, and G. Casiez. Depth-of-field blur effects for first-person navigation in virtual environments. *IEEE Computer Graphics and Applications*, 28(1):47–55, 2008.
- [5] R. Pausch, J. Snoddy, R. Taylor, S. Watson, and E. Haseltine. Disney's aladdin: first steps toward storytelling in virtual reality. In *Proceedings* of the 23rd annual conference on Computer graphics and interactive techniques, pages 193–203. ACM, 1996.

¹Google Spotlight Story: HELP - https://youtu.be/G-XZhKqQAHU ²GoPro VR: Omni Trailer - https://youtu.be/0wC3x_bnnps