Exploring Visual Guidance in 360-degree Videos

Marco Speicher
DFKI GmbH
Saarland Informatics Campus
marco.speicher@dfki.de

Christoph Rosenberg
Saarland University
Saarland Informatics Campus
s8chrose@stud.uni-saarland.de

Donald Degraen
DFKI GmbH
Saarland Informatics Campus
donald.degraen@dfki.de

Florian Daiber
DFKI GmbH
Saarland Informatics Campus
florian.daiber@dfki.de

Antonio Krüger
DFKI GmbH
Saarland Informatics Campus
krueger@dfki.de

Figure 1: Visual Guidance Methods, respectively *Object to Follow, Person to Follow, Object Manipulation, Environment Manipulation, Small Gestures,* and *Big Gestures.* Both *Forced Rotation* and *No Guidance* are not shown here.

ABSTRACT

360-degree videos offer a novel viewing experience with the ability to explore virtual environments much more freely than before. Technologies and aesthetics behind this approach of film-making are not yet fully developed. The newly gained freedom creates challenges and new methods have to be established to guide users through narratives. This work provides an overview of methods to guide users visually and contributes insights from an experiment exploring visual guidance in 360-degree videos with regard to task performance and user preferences. In addition, smartphone and HMD are used as output devices to examine possible differences. The results show that using viewers preferred HMD over smartphone and visual guidance over its absence. Overall, the Object to Follow method performed best, followed by the Person to Follow method. Based on the results, we defined a set of guidelines for drawing the viewers' attention in 360-degree videos.

TVX '19, June 5–7, 2019, Salford (Manchester), United Kingdom

ACM ISBN 978-1-4503-6017-3/19/06…$15.00
https://doi.org/10.1145/3317697.3323350

CCS CONCEPTS

• **Human-centered computing** → **User studies**; **Virtual reality**; *Usability testing*.

KEYWORDS

Virtual Reality; Visual Guidance; 360-degree Video.

ACM Reference Format:
Marco Speicher, Christoph Rosenberg, Donald Degraen, Florian Daiber, and Antonio Krüger. 2019. Exploring Visual Guidance in 360-degree Videos. In *ACM International Conference on Interactive Experiences for TV and Online Video (TVX '19), June 5–7, 2019, Salford (Manchester), United Kingdom.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3317697.3323350

1 INTRODUCTION

Omni-Directional Video (ODV) or 360° video is a medium in which viewers are offered an immersive 360° experience. Literature has described many different scenarios for using ODV or 360° video, e.g. as a portable dome setup in which users can interact with applications such as a video conferencing system, a multi-user game or an astronomical data visualization system [2]. Recent efforts such as Microsoft's Illumiroom [9] provide interesting possibilities, as they show how a living room environment could be turned into a small CAVE-like theater.

As this medium is becoming commonplace, it should be considered that passively enjoying 360° video is more than just an extension of traditional film. In an ODV environment, users are granted the freedom to freely adjust the orientation of their field of view (FoV). This freedom makes it increasingly difficult for filmmakers to show what is essential to

the story, because what is depicted on screen influences the viewer's understanding of the narrative [28]. In order to maintain immersion, i.e. the "sensation of being there" [18], and avoid distraction and confusion, it is therefore crucial to guide the viewer's attention towards important parts of the narrative while taking motion sickness into account. However, literature investigating visual guidance specifically for 360° video scenarios is limited. This paper aims to extend existing research by investigating and comparing methods for guiding the viewer's attention. Based on existing literature, we define a broad set of visual guidance methods and consider their applicability for 360° video. Through a user experiment, we assessed each method's performance, the viewer's experience, task load, motion sickness and immersion. These results informed a set of guidelines for guiding the viewer's attention; while directed at creators of 360° video, most of these are also applicable to cinematic VR in general. Hence, the main contributions of this paper are:

- A **set of visual guidance methods** specifically aimed at 360° video.
- A **study comparing** the implemented visual guidance methods with regard to task performance and user preferences.
- A **series of guidelines** for viewer attention guidance in 360° video.

Our results inform the design of visual guidance methods and although our contributions are focused on 360° video, our findings can also be applied in the more general domain of Cinematic Virtual Reality.

2 RELATED WORK

We approach the investigation and evaluation of visual guidance in 360° video from different domains, specifically (1) characteristics of omni-directional videos, (2) guidance in traditional film, and (3) guidance in virtual environments.

Characteristics of Omni-Directional Videos

Omni-Directional Video (ODV) or 360° video offers viewers an immersive 360° panoramic recording. This novel medium is becoming more common thanks to the availability of affordable consumer cameras able to record such types of content, and additionally, the support for sharing such recordings through video platforms and social media websites. This format has recently gained much interest in the field of HCI by exploring content delivery [31], presentation [32] and interaction [20, 24, 25] among other things. However, literature investigating visual guidance specifically for ODV scenarios is lacking. In an ODV environment, users are granted the freedom to freely adjust the orientation of their Field of View (FoV). This freedom has been shown to distract viewers and makes it increasingly difficult for filmmakers to show what is essential to the story [26, 28]. In order to maintain immersion, it is therefore crucial to guide the viewer's attention towards important parts of the narrative, while taking into account the possibility of motion sickness. We consider related work on **visual guidance for traditional film** as ODV extends traditional video content with a much broader FoV. Additionally, we consider work done towards **visual guidance in Virtual Reality (VR)** as ODV can be considered a special case of VR where the audience is presented with the content by placing them in the center of a sphere or cylinder onto which images are projected.

Guidance in Traditional Film

In traditional film theory, directors are provided with a varying set of cinematographic techniques to visually guide the viewer's attention [1]. Paramount to these techniques is to effectively and unobtrusively guide the viewer's attention while maintaining the user's sense of immersion and presence. In classical filmmaking, different scenes have to be put together in order to create a fluent visual narrative during the editing process. Scene transitions define the manner in which these scenes are combined, which influences narrative experiences [16]. They are powerful tools for shaping stories and guiding the attention of the viewers.

Our work focuses on visual cues that aim to guide the user's attention. Visual cues used for guidance are typically considered to be either diegetic or non-diegetic, which informs the manner in which these cues are rooted within the narrative. Diegetic cues are part of the story and can be perceived by the actors within the story, rather than for the viewers or listeners only. Examples of these cues are traffic lights, sun rays, music coming from car radio or night clubs, or movements within a scene. Previous research focusing on non-VR environments has explored the use of salient objects, sounds, lights, or moving cues [3, 6, 29]. Contrary to diegetic, non-diegetic cues are external to the narrative in which the viewer is immersed. Coming from the outside, these external cues are played or visualized over the action for the viewers and listener, but not perceptible by the characters being part of the narrative. The visual story as narrated by a filmmaker, is defined by how the image is framed, recorded, edited and presented to the viewer. Camera settings such as focal length or aperture are manipulated to focus on important parts in the scene while blurring out the non-essential.

Guidance in Virtual Environments

The field of Cinematic Virtual Reality (CVR) has investigated the applicability of attention guidance techniques in Immersive Virtual Environments (IVE) and its special case scenarios such as 360°or Omni-Directional Video (ODV) [14, 23].

Focusing on diegetic visual effects, Rothe et al. [23] explored three such cinematic methods to guide the viewer's

attention: lights, sounds, and movements. They found that some viewers were induced by new sounds to search for the source of the sound, while a moving light cone changed the viewing direction considerably. Danieau et al. [4] designed four non-diegetic visual effects and compared two of them to guide the viewer's attention. Concluding that it remains difficult to force the user to move her head unconsciously, they propose that future methods have to be refined or combined with other modal cues. Furthermore, Lin et al. [12] also compared a forced rotation method with a method indicating the direction of the target by a visual cue. While forced rotation was preferred over the no guidance condition almost in all aspects when watching a sport movie, it was less advantageous in watching a video tour. In contrast to visual guidance, *Project Orpheus* [30] was used to investigate aural cues and methods to guide the viewer's attention in an immersive VR experience inspired by traditional TV media. In order to keep 3D sound as unobtrusive as possible, it needs to match the image, instead of being used as an announcement of action. Kjaer et al. [10] extend the usage of editing towards CVR and conclude that editing need not pose a problem in relation to CVR, as long as the participants' attention is appropriately guided at the point of the cut. Moghadam and Ragan [16] designed three scene transitions to change the viewer's location and direct their attention. Their preliminary results indicate a correlation between the viewer's experience with 3D gaming environments and the speed of scene transitions without a notable influence on motion sickness. Rothe and Hußmann [22] presented methods for collecting and analyzing head tracking data in CVR. They explored how spatial and non-spatial sound affect the viewing behavior by tracking the head movements.

Inspired by existing literature and established film theory, Nielsen et al. [17] positioned guidance using three different axes, namely explicit vs. implicit, diegetic vs. non-diegetic and limiting vs. non-limiting interaction. They used their taxonomy for guiding users' attention to define and compare a diegetic cue (firefly), a non-diegetic cue (forced rotation) and no guidance. Results concluded that the diegetic cue was more helpful than the non-diegetic one, while the non-diegetic cue may decrease the feeling of presence.

3 VISUAL GUIDANCE METHODS

Most existing literature on visually guiding the user's attention has either focused on distinctly comparing the usage of either diegetic or non-diegetic cues, or weighed one diegetic cue with one non-diegetic cue. To the best of our knowledge, literature aimed at comparing a multitude of diegetic and non-diegetic cues is lacking.

Our work determines a diverse set of visual guidance techniques by utilizing the taxonomy as defined by Nielsen et al. [17]. The three orthogonal, dichotomous dimensions consider explicitness, narrative integration and interaction constraints. Cues can either be directly communicated to the user (*explicit*) or indirectly aimed to guide the viewer's attention (*implicit*). Based on integration into the narrative, cues are either inherently part of the story (*diegetic*) or fall outside the scope of the mise-en-scène (*non-diegetic*). Lastly, cues can be differentiated based on whether or not they limit the user's interaction capabilities by preventing the user from performing certain actions or by completely taking over control of the viewing environment (*allowing* vs. *limiting interaction*).

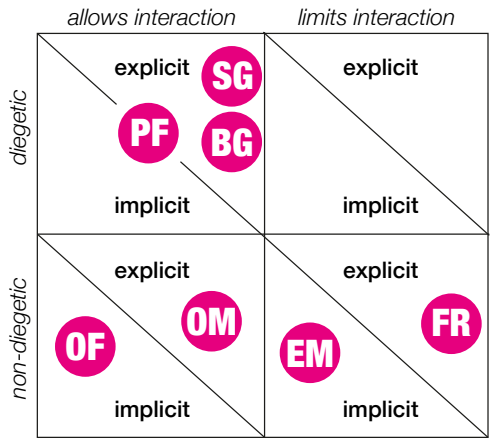

Figure 2: Classification of our guidance methods within the taxonomy by [17]. Note that the *No Guidance* scenario is not present as it falls outside the scope of the taxonomy.

As the taxonomy is aimed at visual guidance within Cinematic Virtual Reality (CVR), it does not explicitly take into account the characteristics of 360° video. Therefore, we additionally consider the applicability for ODV as the pivotal factor to include a potential method. The resulting set of cues, partly known from related studies [4, 12, 17] and visualized within the taxonomy in Figure 2, are defined as *Forced Rotation (FR)*, *Object to Follow (OF)*, *Person to Follow (PF)*, *Object Manipulation (OM)*, *Environment Manipulation (EM)*, *Small Gestures (SG)*, and *Big Gestures (BG)*. As a baseline indicator, we preserve the *No Guidance* scenario.

No Guidance

Technically seen as outside of the scope of visual guidance methods, this technique states there is no explicit guidance of the user's attention. No restrictions in terms of context, isolated use, naturalness of manipulations or the potential size of the area of interest are in effect. The main advantage of this approach remains that no conceptual model for attention guidance is needed. Seeing that no additional editing

of the video, planning or implementation is required, the workload of the director is not influenced. However, in the case of 360° video, the freedom of adjusting the FoV may cause confusion and frustration as the understanding of the narrative can be negatively influenced.

Forced Rotation

In *Forced Rotation*, the area of interest remains in focus as the user's FoV is explicitly moved through rotation of the virtual environment. While the user is free to look around otherwise, the system takes control during the rotation. More precisely, it mimics grabbing the user's head and forcing it to look in a specific direction, i.e. the user cannot control her head rotation during the animation. This method is an example of an *explicit non-diegetic* cue that *limits interaction*, common in traditional film, as the camera is not controlled by the user [4, 17]. Using *Forced Rotation* to guide the viewer's has been explored in several related studies [4, 12] and was implemented by Facebook in 2016[1]. This method does not require input from the user, but may present a cause for confusion. Users misunderstanding the rotation of the camera might try to counter the effect, potentially inciting sickness [4].

Object to Follow

This method exploits the movement of an object to guide the user through the scene. While the user is still able to freely look around in the environment, a firefly flying within the scene aims to "offer clues as to where the user should focus" [17]. This approach is classified as an *implicit non-diegetic* cue that *allows interaction*. Similarly, Peck et al. [19] proposed a *non-diegetic* version which distracts the user from manipulations of the virtual environment by using a sphere. Improved versions of this method replaced the sphere with a butterfly or hummingbird, as these are more natural in a virtual environment and could be integrated more easily into the narrative, effectively making them *diegetic* cues. Our *Object to Follow* approach uses a colored sphere that could be integrated during post-production. Since no specific planning is required during the early production stages, this method can be added on easily afterward.

Person to Follow

In the *Person to Follow* method, a person (i.e. an actor) guides the user within the scene. This "guide" can walk next to the object of interest to draw attention to it. This method is an example of an *explicit diegetic* cue that *allows interaction*. Depending on the scene and narrative, this method can be very immersive, as the visual guidance is inherent to the plot. As it is integrated in the narrative, it could also be interpreted as implicit to the viewer, because the character's movement does not directly act as a visual guidance method. *Person to Follow* is not compatible with scenes not containing actors or scenes where actors have limited movement space. Post-production is not required for this method and there are no visual discrepancies as the guide is already present in the scene. However, the implementation may be very complicated, as a certain amount of planning is required to seamlessly integrate the guide's movements in the narrative.

Object Manipulation

Object Manipulation exploits visual salience to guide the viewer's attention by manipulating perceptual properties of an object of interest, e.g. color, luminance or size [15, 29]. This method is classified as an *explicit non-diegetic* cue that *allows interaction*. Areas containing unpredictable contours or unusual details are known to attract the user's attention [13]. To seamlessly integrate the method in the scene, there has to be context for the object of the manipulation, e.g. a television capable of displaying images on its screen or light bulbs that can flicker. Planning and implementation of the method require additional effort during the production process and potentially also the post-production.

Environment Manipulation

In contrast to *Object Manipulation*, the manipulation of the environment considers anything excluding the area of interest in the scene [1]. Based on techniques used in traditional film, examples are fading to black, desaturating or blurring non-important parts of the scene [4, 8]. All these effects can vary in intensity and can be animated to increase the precision of the guidance. This method is an example of an *implicit non-diegetic* cue that *limits interaction*. Since most of the scene is manipulated, the area of interest has to be identifiable by the user and should be not too small. In order to identify the area of interest, it is crucial that the user is able to perceive and understand the changes. If the manipulation is too subtle, fade to black can be understood as turning off the lights[4]. Similar to confusion in traditional film, the viewer may start to feel annoyed or uncomfortable [1].

Small Gestures

The *Small Gestures* method uses facial expression and head pointing. These small, subtle gestures are performed by an actor who indicates an object of interest by broadly "pointing" or turning the head in a certain direction. This method is an example of an *explicit diegetic* cue that *allows interaction*. Similar to *Person to Follow*, *Gestures* are easy to comprehend as no conceptual model needs to be introduced [21]. Seen intuitively, at least one person has to be part of the scene to perform the gesture. *Gestures* need to be recorded during the production, which requires careful planning.

[1] https://media.fb.com/2016/04/12/facebook-360-updates/

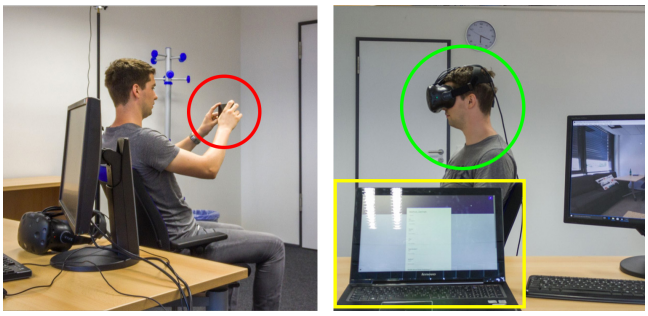

Figure 3: Experimental setup showing all devices used: smartphone (red), HMD (green), and laptop (yellow) to fill out the questionnaires.

Big Gestures

In the *Big Gestures* method the indication of direction is realized by big, ponderous gestures (e.g. waving or pointing with the arm) that are performed by an actor within the scene. This method is an example of an *explicit diegetic* cue that *allows interaction*. *Big Gestures* also do not require the introduction of a conceptual model [21]. Again, at least one person has to be part of the scene to perform the gestures, which must be recorded during the production phase.

4 EXPERIMENT

The purpose of this experiment was to explore a broad range of visual guidance techniques. The following hypotheses were defined based on Nielsen et al. [17]:

- H_1 Visual guidance improves performance and preference.
- H_2 *Forced Rotation* performs worse than all other methods.
- H_3 *Forced Rotation* causes disorientation in the user.
- H_4 *Object to Follow* outperforms all other methods.

Design

The experiment consisted of a within-subjects design with two independent variables, i.e. output device (mobile device, HMD) and visual guidance method (*No Guidance*, *Forced Rotation*, *Object to Follow*, *Person to Follow*, *Object Manipulation*, *Environment Manipulation*, *Small* and *Big Gestures*). We measured five dependent variables related to task performance (error rate) and user preference (task workload, user experience, motion sickness, immersion). Each participant watched 16 videos, 8 on each device, from a total collection of 79 videos including multiple different versions of each narrative. Each video was randomly assigned per participant, while no video was watched twice by the same participant, which was possible as two takes of each narrative (party game) were recorded. The conditions were counterbalanced using a Latin square design, which resulted in 512 trials (32 participants $\times$ 2 devices $\times$ 8 methods).

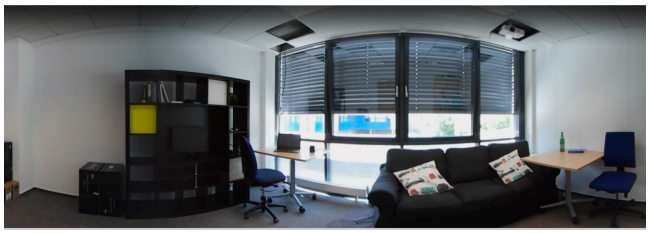

Figure 4: The room featured in the recorded stories. In this scene, the left cabinet door of the TV rack is the object of interest (yellow).

Participants

The experiment included 32 participants (14 female, 18 male) aged between 22 and 29 years ($M = 25.19, SD = 1.94$). All participants were students and received a financial compensation. 37.5% of the participants had no or almost no experience with 360° videos or VR, whereas 46.9% had tried it, but not on a regular basis and 15.6% were very experienced, e.g. developers or heavy users. The frequency of watching 360° videos was rated on a 5-point scale (1: Never; 5: Very often), whereas 93.8% had never or only rarely watched 360° videos before ($M = 1.59, SD = 0.71$). No participant suffered from color-blindness, but 59.4% needed to wear glasses or contact lenses during the experiment.

Apparatus

All videos used in this experiment were filmed with a Samsung Gear 360 (3840 $\times$ 1920; 29.97 fps) and the final videos used 4096 $\times$ 2048 pixels with the same frame rate and no sound in the equi-rectangular format. As a mobile device, an LG Nexus 5X was used (5.2 inch; 1080 $\times$ 1920 pixels; 60Hz) with Android 6.0. Users were able to look around in the virtual environment by moving the device (see Figure 3). For the HMD, we used the HTC Vive with a resolution of 2160$\times$1200 pixels at 90 Hz. The FoV was set to 100° for all devices to guarantee equal conditions.

Task

A static camera was positioned in the center of the room (see Figure 4). In the videos, two men played eight different party games (e.g. hit-the-pot, hide and seek, or ping-pong) and a visual guidance method was used in each video to guide the viewer's attention to a certain object of interest. We provided two variations of each game to guarantee that the participants didn't watch the same video on both devices. The games served as a narrative to provide a simple and easy to understand story in the short videos, which lasted between 35 and 60 seconds. Each video contained an object that was manipulated by changing its color and blinking 3 times for 3 seconds in the video (see Figure 1). Every game

Figure 5: Example scene depicting the Hit-the-Pot party game. Here, the tile in the upper left corner of the ceiling is the object of interest (magenta).

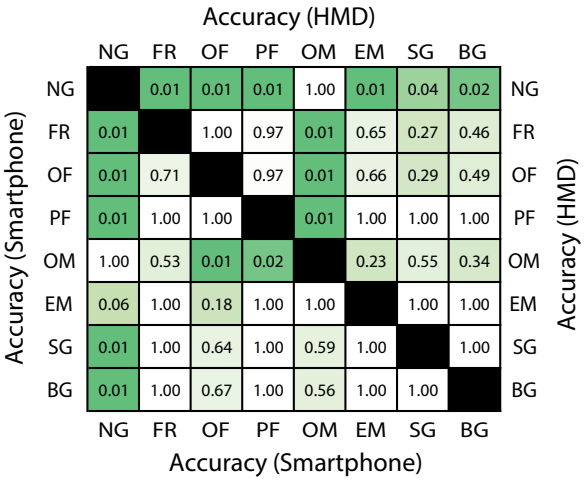

Figure 6: Bonferroni-Holm corrected pairwise comparisons with Accuracy for smartphone condition (left triangle) and HMD (right triangle) as dependent variables.

featured a unique combination of a manipulated object and color, e.g. a tile in the ceiling, a cabinet door, a waste bin, the screen of a television or a window (see Figure 5). The objects of interest were equally distributed in the room. The task goal for each video was to identify (1) *if* something was manipulated, (2) *what* was manipulated and (3) *how* it was indicated. After each video was watched completely, participants were asked if something had been indicated to them. If yes, they were asked to identify the indicated object and name the guidance method used in the video.

Procedure

First, the participants were briefed on the experiment and had the opportunity to get familiar with 360° videos by watching an introductory video, where nothing was manipulated, or no method was used. After that, the first eight videos were shown on the mobile device or via the HMD, depending on the Latin square order. The participants were seated on a swivel chair during the whole experiment to make it easier for them to look around. No specific task and no conceptual model was given, however, after the completion of each video, participants were asked if something had been indicated to them, what and how. Then, specific post-task questionnaires were filled out for measuring user experience, motion sickness, immersion and task workload. When the participants finished watching the first set of videos and filled out all questionnaires, the device was changed and the same procedure was repeated on the second device. After all 16 videos (8 per device) had been watched, a final demographic post-study questionnaire was filled out. The whole experiment per participant took about 60 minutes.

5 RESULTS

This section presents the results of the experiment using the following abbreviations for the methods: *No Guidance (NG)*, *Forced Rotation (FR)*, *Object to Follow (OF)*, *Person to Follow (PF)*, *Object Manipulation (OM)*, *Environment Manipulation (EM)*, *Small Gesture (SG)*, and *Big Gesture (BG)*. Although participants had a variety of experiences with 360° videos and VR, a multivariate ANOVA regarding experience with 360° videos and VR showed no significant differences between methods or output devices ($p = 1.0$). Overall, among all visual guidance methods, the FR method got the lowest user experience rating and caused significantly more disorientation, while OF was best for all measurements.

Performance

Concerning task performance, we focused on accuracy measurement. Right after watching a video, the participant was asked **if** something had been indicated to them. If the answer was positive, they were asked what and how it was indicated. The overall accuracy is represented by a score from 0 to 5 according to the three performance measurement questions (0: negative; 1: positive indication; 3: positive indication and method; 4: positive indication and object; 5: positive indication, method and object).

There was a statistically significant difference in overall accuracy depending on which method was used for visual guidance, using a smartphone ($\chi^2(7) = 54.49, p < 0.01$) or wearing a HMD ($\chi^2(7) = 68.96, p < 0.01$). Post-hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni-Holm correction applied for every pair of methods using the smartphone and wearing the HMD (see Figure 6). Pairs using the smartphone showed significant differences between NG and all methods except for OM and EM, as well as OM-OF and OM-PF. There were also significant differences between all pairs of NG except for NG-OM,

Table 1: Objective measurements and subjective feedback ratings with significant differences between the methods. The best method per scale is shown in dark green, the second in green, and a ranking by Roman numerals.

Method	Accuracy (Smartphone) Median [0, 5]	Accuracy (HMD) Median [0, 5]	User Experience Mean (SD) [−3, 3]	Task Workload Mean (SD) [0, 100]	Motion Sickness (%) Mean (SD) [0, 100]	Immersion Mean (SD) [1, 7]
No Guidance (NG)	V: 0.0	V: 0.0	VII: $0.34_{\pm 0.92}$	$33.79_{\pm 22.85}$	$19.20_{\pm 13.60}$	$4.21_{\pm 1.91}$
Forced Rotation (FR)	II: 2.0	III: 2.0	VIII: $0.23_{\pm 1.12}$	$34.35_{\pm 21.24}$	$22.71_{\pm 14.71}$	$4.22_{\pm 2.07}$
Object to Follow (OF)	II: 2.0	I: 4.0	I: $0.83_{\pm 0.95}$	$27.65_{\pm 17.39}$	$18.63_{\pm 12.90}$	$4.34_{\pm 2.01}$
Person to Follow (PF)	I: 3.0	II: 3.0	II: $0.78_{\pm 0.87}$	$29.91_{\pm 20.28}$	$18.10_{\pm 12.28}$	$4.53_{\pm 1.97}$
Object Manipulation (OM)	V: 0.0	V: 0.0	VI: $0.51_{\pm 0.90}$	$31.74_{\pm 21.59}$	$20.06_{\pm 15.81}$	$4.41_{\pm 2.03}$
Environment Manipulation (EM)	II: 2.0	III: 2.0	V: $0.59_{\pm 0.86}$	$35.29_{\pm 24.04}$	$19.57_{\pm 13.63}$	$4.24_{\pm 2.08}$
Small Gestures (SG)	IV: 1.0	III: 2.0	III: $0.71_{\pm 0.98}$	$29.94_{\pm 18.69}$	$18.06_{\pm 11.84}$	$4.43_{\pm 2.08}$
Big Gestures (BG)	III: 1.5	IV: 1.5	IV: $0.60_{\pm 0.74}$	$31.06_{\pm 19.29}$	$17.93_{\pm 11.64}$	$4.43_{\pm 1.97}$
$(\chi^2(7) = .. \mid F_{(7,528)} = ..)$	54.49	68.96	3.47	1.06	0.96	0.27
$p < .. (\eta^2 = ..)$	0.01	0.01	0.01 (0.02)	0.39 (0.02)	0.46 (0.02)	0.99 (0.01)

as well as significances for OM-FR, OM-OF, and OM-PF. An overview of the median results of overall accuracy using the smartphone and wearing the HMD can be found in Table 1. In summary, the fewest indications were recognized in the NG and OM conditions ($Median = 0$) for both output devices. Overall, using FR , in 73.4% trials it was noted that something was indicated and 68.8% identified the method correctly, but in only 39.1% of the videos could the manipulated object be named. The median of the overall accuracy for OF ($Median = 2.0, 2.0$ to 4.0), FR ($Median = 2.0, 0.0$ to 3.25), and EM ($Median = 2.0, 0.0$ to 2.0) were the second best results using the smartphone, while OF performed best ($Median = 4.0, 2.0$ to 4.0) and PF achieved the second best results wearing the HMD ($Median = 3.0, 0.0$ to 3.0). When using OF and disregarding the output device, participants stated in 87.5% of the videos that something had been indicated, and in a total of 84.4% the method could also be named correctly, whereas the manipulated object could only be identified in almost every second trial (48.4%). In contrast, the object could be identified more often (71.9%) using PF , whereas the method itself could only be identified in 21.9% of the trials.

Overall, NG had the worst overall accuracy compared to all other methods when considering the output devices individually (smartphone: $Median = 0.0, 0.0$ to 0.0; HMD: $Median = 0.0, 0.0 to 0.0$), followed by OM (smartphone: $Median = 0.0, 0.0$ to 0.25; HMD: $Median = 0.0, 0.0$ to 0.25). No significant differences could be found between the devices ($p > 0.05$). Nevertheless, 68.8% of the participants did not name the correct object that was indicated in the videos using the mobile device, and 61.7% when using the HMD. The method of guidance was not correctly recognized by 59.8% when watching the videos on the smartphone and 54.3% on the HMD.

User Experience

Our variation of the UEQ [11] consisted of six questions that represented each factor of the original questionnaire (on a 7-point scale from −3 to 3). A univariate ANOVA regarding the overall user experience rating showed significant differences between the methods ($F_{7,528} = 3.47, p < 0.01, \eta^2 = 0.05$), as well as the output devices ($F_{7,528} = 7.45, p < 0.01, \eta^2 = 0.01$). NG ($M = 0.34, SD = 0.92$) and FR ($M = 0.23, SD = 1.12$) achieved the lowest ratings, while OF ($M = 0.83, SD = 0.95$) was rated highest, followed by PF ($M = 0.78, SD = 0.87$) (see Table 1). Bonferroni-Holm corrected pairwise comparisons showed significant differences between FR-OF ($p < 0.01$), as well as significant effects between NG-OF ($p < 0.05$) and FR-PF ($p < 0.02$). Concerning the output devices, HMD achieved a higher score ($M = 0.68, SD = 0.96$) on average than using the smartphone ($M = 0.47, SD = 0.90$), but the difference was not significant. Overall, the best rating was achieved by

OF with the HMD ($M = 0.98, SD = 0.81$) while FR using the smartphone was rated worst ($M = 0.22, SD = 0.96$).

A multivariate ANOVA showed significant differences between the methods regarding the following UEQ subscales: unpleasant/pleasant ($F_{7,528} = 16.17, p < 0.01, \eta^2 = 0.03$), inefficient/efficient ($F_{7,528} = 3.75, p < 0.01, \eta^2 = 0.05$), and confusing/clear ($F_{7,528} = 4.78, p < 0.01, \eta^2 = 0.06$), as well as significant effects concerning inferior/valuable ($F_{7,528} = 2.53, p < 0.02, \eta^2 = 0.03$) and conventional/inventive ($F_{7,528} = 2.25, p < 0.03, \eta^2 = 0.03$). With regard to attractiveness (unpleasant/pleasant), FR was rated worst ($M = 0.20, SD = 1.58$) and NG ($M = 0.83, SD = 1.45$) and EM ($M = 0.88, SD = 1.25$) were the only other methods rated below 1. Moreover, NG ($M = 0.02, SD = 1.25$), FR ($M = 0.28, SD = 1.50$) and OM ($M = 0.19, SD = 1.31$) were rated as most inefficient. Furthermore, FR got the only negative values in perspicuity (confusing/clear; $M = -0.14, SD = 1.73$) and dependability (obstructive/supportive; $M = -0.13, SD = 1.48$). Concerning the novelty question (conventional/inventive), BG was rated worst at 0.09 ($SD = 1.48$) and OF as the most novel method ($M = 0.75, SD = 1.40$).

Task Workload

The overall task workload using NASA TLX [7] was rated between 27 (OF: $M = 27.65, SD = 17.39$) and 36 on average (EM: $M = 35.29, SD = 24.04$), but without significant differences ($p < 0.40$). With regard to the output condition, the overall task load was rated as very similar on both devices (mobile: $M = 30.23, SD = 18.75$; HMD: $M = 29.29, SD = 19.66$). No significant differences could be found between the output conditions ($p = 0.54$). A multivariate ANOVA showed a significant effect between the output devices in the NASA TLX sub-scale "Performance" ($F_{7,528} = 4.42, p < 0.04, \eta^2 = 0.01$). Here, the participants rated their own subjective performance higher for the mobile device ($M = 31.52, SD = 22.36$) than for the HMD ($M = 27.15, SD = 23.40$).

Motion Sickness

For the motion sickness assessment we used a modified version of the MSAQ [5] with one question (1: lowest, 9: highest) for each of the four original categories. While all other methods were rated between 18% and 23%, FR was rated highest with 22.71% ($SD = 14.71$) overall motion sickness. However, significant differences were found only with regard to the subscales "disorientation" or "dizziness" ($F_{7,528} = 2.74, p < 0.01, \eta^2 = 0.04$). Concerning the **disorientation** factor, most methods were rated between 20.89% and 23.44%, while EM was rated at 24.44% ($SD = 1.74$) and FR as the highest at 32.78% ($SD = 2.17$). A univariate ANOVA showed significant differences in the overall motion sickness rating regarding the output devices ($F_{7,528} = 9.02, p < 0.01, \eta^2 = 0.02$), with the HMD ($M = 21.53\%, SD = 15.74$) being more prone to induce motion sickness than the mobile device ($M = 17.97\%, SD = 10.81$).

Immersion

The immersion questionnaire [27] was reduced to a single question with a comment section. Users rated their sense of being in the room on a 7-point scale (from 1:low immersion to 7: high immersion). A univariate ANOVA showed no significance between the methods with regard to the immersion rating ($p = 0.99$). NG was rated 4.21 ($SD = 1.91$) on average, FR 4.22 ($SD = 2.07$) and EM 4.24 ($SD = 2.08$) on average compared to the other methods (between 4.41 and 4.53). There was a significance between the devices ($F_{7,528} = 82.58, p < 0.01, \eta^2 = 0.14$) with the smartphone at 3.59 ($SD = 1.90$), lower than the HMD at 5.15 ($SD = 1.60$).

6 DISCUSSION

The study was designed to gather insights on visual guidance methods, including diegetic, non-diegetic and no guidance, using two conventional types of output devices, i.e. mobile and HMD. The chosen techniques cover a broad range of methods aimed at watching 360° videos.

In the following, we (1) discuss the experimental results with regard to task performance and user preference, (2) formulate guidelines for guiding the viewer's attention, and (3) address production characteristics of visual guidance methods for 360° videos.

Performance & Preference

We could confirm findings by prior work [12], because all tested visual guidance methods were more efficient in drawing attention to objects of interest and were more preferred by the participants than using no guidance at all (H_1). Contrary to our assumption of H_2 and findings by prior work [4, 17], participants did not perform significantly worse in identifying and naming the objects of interest when *Forced Rotation* was used, compared to all other visual guidance methods, not considering *No Guidance*. In fact, *Forced Rotation* had third best overall accuracy in the HMD case. Thus, H_2 needs to be rejected. Considering all visual guidance methods and the *No Guidance* scenario, *Forced Rotation* got the lowest average user experience rating. Results indicated that *Forced Rotation* caused significantly more disorientation than all other tested methods, including *No Guidance* (H_3). *Object to Follow* (best for HMD as output device, second for smartphone) and *Person to Follow* (best for smartphone as output device, second for HMD) performed best for guiding the participant's attention to areas of interest, reflected by the significantly higher accuracy compared to all other tested methods (H_4). In summary, we were able to accept all hypotheses, except for H_2 and H_4 partly. Based on these results, we formulate the following guidelines for guiding the viewer's attention.

Guidelines for Viewer Attention Guidance

Concerning the output devices, we found no significant differences for performance and task load while watching 360° videos on mobile or HMD. However, the performance results (see Table 1) indicate that more participants were able to identify the indicated object and the method of guidance when using the HMD as output device. While the overall task workload was on the same level for both devices, the subjective performance rating was significantly better for HMD, which indicates that the participants felt that they were better off using the HMD. Although they expected high motion sickness ratings for the HMD as output device due to its technical limitations, the distance between the ratings was smaller than expected, and here, disorientation was the crucial factor. However, it is worth mentioning that a static camera was used in our experiment, so using a non-static camera might increase the effects. But as technical developments for HMDs continue to proceed quickly, it is merely a matter of time until this gap is filled.

Head-mounted Displays (HMD) should be preferred over mobile devices when watching 360° videos, in particular with regard to user experience and immersion.

All methods tested in the study performed better than the videos with no guidance. Only half of the participants were able to indicate the manipulated object by chance when they were looking around in the environment, whereas the other half misinterpreted gestures in the video.

At least visual guidance is necessary if something is supposed to be seen by the viewer and not just by chance.

When *Forced Rotation* was used, some participants interpreted the rotation in the video as an error, which consequently led to lower user experience and accuracy. A common behavior was that they tried to compensate for the forced video rotation so that their perceived area of interest (e.g. the actors) remained in their viewport instead of the manipulated object to be identified. This behavior was also observed in prior work using forced rotation in 360° videos [4, 12]. Furthermore, *Forced Rotation* induced the highest motion sickness on average, but not significantly. In addition, it was the only method, which led users to mention that they felt unpleasant. Furthermore, the lower immersion ratings can be explained by the restriction of the user's freedom of camera control.

Interfering with camera control can help to guide the user's attention to a certain extent, but might have a negative impact on the user's preference and should only be used in exceptional cases.

It turned out that using an object to follow as visual guidance in a video (e.g. a firefly [17, 19]) is the most efficient way to guide the viewer's attention to a certain area of interest due to the significantly higher recognition rate compared to the other methods. However, more than a third of the participants did not follow the object, which could have occurred for various reasons. After analyzing the participants' behavior using recorded videos and the post-study questionnaires, we came up with two possible explanations. First, some participants identified the colored spherical object as visual guidance, but decided not to follow it, because the narrative (here: actors playing a party game) had greater importance to them. Second, have tried to follow it, but were unable to do so, because of the object's velocity or movement. Overall, when using *Object to Follow* to guide the viewer's attention, the task load was rated lowest and user experience highest, but without impact on motion sickness and immersion.

Using an object to follow turned out to be the best method to choose regarding performance and preference. This method also has the potential to guide users in an unobtrusive way, but the implementation is the crucial point.

Despite the issues that come up when using an object to follow (e.g. a firefly), only one method was more successful than *Object to Follow* in guiding the participants to the area of interest, namely the *Person to Follow* method. Here, half of the participants could name the manipulated object, but couldn't identify the person in the video as part of the guidance method. Nevertheless, the highest immersion rating on average among all methods indicates that this guidance method could be the most subtle and natural way for guiding the viewer's attention. Furthermore, the highest user experience, low motion sickness and task load on average show a clear user preference for this method. Although the method performed very well in the study, the implementation within a scene could be very complex, as a person acting as the guide is required. And ideally, that person should be part of the narrative, which therefore makes it difficult to edit in the post-production.

If Object to Follow *cannot be used,* Person to Follow *should be chosen to guide the viewer's attention with regard to accuracy and user preference, if the narrative allows it.*

Apart from *No Guidance*, *Object Manipulation* performed poorly throughout all metrics in the study. A possible cause for the bad performance and preference measurements could be the limited area in which the manipulation is visible. In our implementation, the size of the area of interest was increased by a large blinking colored halo around the manipulated object to be identified. However, even the enlarged halo turned out to be ineffective when the user is looking in a different direction, for example if the narrative is happening somewhere else. However, this method could be an easy way to increase the size of the area of interest in the post-production, albeit only to a certain degree.

Object Manipulation is recommended in the post-production to highlight the area of interest, but should only be considered if there is no other choice or the area of interest is close to the narrative.

Another method which has a good potential to be effectively used in the post-production process is environment manipulation. Despite altering the appearance of almost the whole scene by darkening it, less than two thirds of the participants recognized it as the method of guidance. However, some participants kept watching without any kind of reorientation, whereas some interpreted the reduced brightness as a notice from the device. The overall below average preference and performance indicate that the method caused confusion. Although, there are other implementations for changing the environment, large-scale possible manipulations of the scene could make the videos more prone to motion sickness and immersion breaks, when immersion was already rated below average among all methods. Here, our results are in line with the findings of prior work [4].

Environment Manipulation can be seen as the inverted version of Object Manipulation and is more likely recommended in post-production to highlight the area of interest, if the area of interest is not in the user's field of view.

The results showed that there is no significant difference between using an arm (*Big Gestures*) or only the head (*Small Gestures*) for pointing to the area of interest. When using the pointing gestures for visual guidance in our experiment, less than half of the participants were able to indicate that there had been guidance and name one of the gestures as the method. Moreover, only a third could identify the manipulated object, which shows the gesture methods to be ineffective. Furthermore, these methods also led to confusion as some participants stated that they did not understand that they were meant to follow the pointing direction and interpreted it as part of the narrative. As expected, there was no impact on immersion or motion sickness, but pointing with an arm (*Big Gestures*) has been rated as most conventional. In contrast, facial expressions and pointing with the head (*Small Gestures*) were perceived as a very subtle approach. However, *Small Gestures* seemed to be good enough for guidance only if merely a rough direction is needed, whereas arm pointing (*Big Gesture*) is required for more specific guidance.

Small or Big Gestures (here: facial expressions, head and arm pointing) as part of the narrative can be an appropriate alternative to the Person to Follow method, even though they significantly degrade user experience.

Production Characteristics

The implementation of the methods in the 360° videos production workflow highly depends on the method. While some effects can be added during post-production to arbitrary videos[4], other methods require specific planning before shooting (e.g. *Person to Follow*). This can include the movement of persons in the scene, which usually cannot be changed easily in the post-production process. Adding effects in post-production requires less planning, because changes to the video can be undone or altered. Here, it is worth to mention that post-production does not imply that characters did not perceive something in the narrative, as their actors could have pretended to notice it.

However, the amount of work for implementing a method subsequently also varies. To keep the methods as subtle as possible, the amount of visual differences between the original video and edited version should be minimal, but still perceivable. The results of our study indicate that subsequent changes had a negative impact on immersion. Moreover, the results show that the less subtle the changes are, the greater the impact. Some methods can be used during production, as well as in the post-production (e.g. *Forced Rotation* or *Manipulation*). However, this might lead to a number of drawbacks. For example *Forced Rotation*, as implemented in Facebook's player, cannot be transferred or exported, as it is not part of the video. Moreover, adding the rotations in post-production, as it was also done in our study, can lead to inaccurate rotations as the user's exact field of view is not known.

7 CONCLUSION

This work investigated the guidance of viewers' attention in Omni-Directional or 360° Video. Based on prior literature, we defined a set of three diegetic cues, i.e. *Person to Follow* and *Small & Big Gestures*, and four non-diegetic cues, i.e. *Object to Follow*, *Object Manipulation*, *Forced Rotation*, and *Environment Manipulation*. As a baseline indicator, we maintained a *No Guidance* scenario. For each type of cue, we recorded a scenario implementing the specific visual guidance technique, which resulted in 79 videos, including introduction videos. We compared the performance and user preference of the proposed techniques in a within-subjects experiment.

Using visual guidance was more effective to draw attention to objects of interest than no guidance. Viewers preferred visual guidance over its absence. Of all visual guidance methods, the *Forced Rotation* method got the lowest user experience rating and caused significantly more disorientation. The *Object to Follow* method performed best. Based on the results, we defined a set of design guidelines for guiding the viewers' attention in ODV. While our results extend existing literature, both in explored methods and in significant findings, many properties and methods for visual guidance remain unexplored. A taxonomy specifically focused on 360° video might be able to identify potential method combinations while taking production costs into account.

REFERENCES

[1] Steven Ascher and Edward Pincus. Ascher2012. *The filmmaker's handbook: A comprehensive guide for the digital age.* Penguin.

[2] Hrvoje Benko and Andrew D. Wilson. 2010. Multi-point Interactions with Immersive Omnidirectional Visualizations in a Dome. In *ACM International Conference on Interactive Tabletops and Surfaces (ITS '10)*. ACM, New York, NY, USA, 19–28. https://doi.org/10.1145/1936652.1936657

[3] Stanley Coren. 2003. *Sensation and perception.* Wiley Online Library.

[4] Fabien Danieau, Antoine Guillo, and Renaud Doré. 2017. Attention guidance for immersive video content in head-mounted displays. In *2017 IEEE Virtual Reality (VR)*. 205–206. https://doi.org/10.1109/VR.2017.7892248

[5] Peter J Gianaros, Eric R Muth, Toby J Mordkoff, Max E Levine, and Robert M Stern. 2001. A questionnaire for the assessment of the multiple dimensions of motion sickness. *Aviation, Space, and Environmental Medicine* 72, 2 (2001), 115.

[6] E.B. Goldstein. 2009. *Sensation and Perception.* Cengage Learning. https://books.google.de/books?id=2tW91BWeNq4C

[7] Sarah G Hart and Lowell E Stavenland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Human Mental Workload*, P. A. Hancock and N. Meshkati (Eds.). Elsevier, Chapter 7, 139–183. http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20000004342_1999205624.pdf

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

[9] Brett R. Jones, Hrvoje Benko, Eyal Ofek, and Andrew D. Wilson. 2013. IllumiRoom: Peripheral Projected Illusions for Interactive Experiences. In *ACM SIGGRAPH 2013 Emerging Technologies (SIGGRAPH '13)*. ACM, New York, NY, USA, Article 7, 1 pages. https://doi.org/10.1145/2503368.2503375

[10] Tina Kjær, Christoffer B. Lillelund, Mie Moth-Poulsen, Niels C. Nilsson, Rolf Nordahl, and Stefania Serafin. 2017. Can You Cut It?: An Exploration of the Effects of Editing in Cinematic Virtual Reality. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology (VRST '17)*. ACM, New York, NY, USA, Article 4, 4 pages. https://doi.org/10.1145/3139131.3139166

[11] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and Usability Engineering Group*. Springer, 63–76.

[12] Yen-Chen Lin, Yung-Ju Chang, Hou-Ning Hu, Hsien-Tzu Cheng, Chi-Wen Huang, and Min Sun. 2017. Tell Me Where to Look: Investigating Ways for Assisting Focus in 360-degree Video. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 2535–2545. https://doi.org/10.1145/3025453.3025757

[13] Norman H Mackworth and Anthony J Morandi. 1967. The gaze selects informative details within pictures. *Attention, Perception, & Psychophysics* 2, 11 (1967), 547–552.

[14] Andrew MacQuarrie and Anthony Steed. 2017. Cinematic virtual reality: Evaluating the effect of display type on the viewing experience for panoramic video. In *2017 IEEE Virtual Reality (VR)*. 45–54. https://doi.org/10.1109/VR.2017.7892230

[15] Victor A. Mateescu and Ivan V. Bajić. 2014. Attention Retargeting by Color Manipulation in Images. In *Proceedings of the 1st International Workshop on Perception Inspired Video Processing (PIVP '14)*. ACM, New York, NY, USA, 15–20. https://doi.org/10.1145/2662996.2663009

[16] Kasra Rahimi Moghadam and Eric D Ragan. 2017. Towards understanding scene transition techniques in immersive 360 movies and cinematic experiences. In *2017 IEEE Virtual Reality (VR)*. 375–376. https://doi.org/10.1109/VR.2017.7892333

[17] Lasse T. Nielsen, Matias B. Møller, Sune D. Hartmeyer, Troels C. M. Ljung, Niels C. Nilsson, Rolf Nordahl, and Stefania Serafin. 2016. Missing the Point: An Exploration of How to Guide Users' Attention During Cinematic Virtual Reality. In *Proceedings of the 22Nd ACM Conference on Virtual Reality Software and Technology (VRST '16)*. ACM, New York, NY, USA, 229–232. https://doi.org/10.1145/2993369.2993405

[18] Niels Christian Nilsson, Rolf Nordahl, and Stefania Serafin. 2016. Immersion revisited: A review of existing definitions of immersion and their relation to different theories of presence. *Human Technology* 12 (2016).

[19] Tabitha C Peck, Henry Fuchs, and Mary C Whitton. 2009. Evaluation of Reorientation Techniques and Distractors for Walking in Large Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics* 15, 3 (May 2009), 383–394. https://doi.org/10.1109/TVCG.2008.191

[20] Benjamin Petry and Jochen Huber. 2015. Towards Effective Interaction with Omnidirectional Videos Using Immersive Virtual Reality Headsets. In *Proceedings of the 6th Augmented Human International Conference (AH '15)*. ACM, New York, NY, USA, 217–218. https://doi.org/10.1145/2735711.2735785

[21] Syed Shaukat Raza Abidi, MaryAnn Williams, and Benjamin Johnston. 2013. Human Pointing As a Robot Directive. In *Proceedings of the 8th ACM/IEEE International Conference on Human-robot Interaction (HRI '13)*. IEEE Press, Piscataway, NJ, USA, 67–68. http://dl.acm.org/citation.cfm?id=2447556.2447570

[22] Sylvia Rothe and Heinrich Hußmann. 2018. Spatial Statistics for Analyzing Data in Cinematic Virtual Reality. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces (AVI '18)*. ACM, New York, NY, USA, Article 81, 3 pages. https://doi.org/10.1145/3206505.3206561

[23] Sylvia Rothe, Heinrich Hußmann, and Mathias Allary. 2017. Diegetic Cues for Guiding the Viewer in Cinematic Virtual Reality. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology (VRST '17)*. ACM, New York, NY, USA, Article 54, 2 pages. https://doi.org/10.1145/3139131.3143421

[24] Gustavo Rovelo, Donald Degraen, Davy Vanacken, Kris Luyten, and Karin Coninx. 2015. Gestu-Wan - An Intelligible Mid-Air Gesture Guidance System for Walk-up-and-Use Displays. 9297 (09 2015), 368–386. https://doi.org/10.1007/978-3-319-22668-2_28

[25] Gustavo Alberto Rovelo Ruiz, Davy Vanacken, Kris Luyten, Francisco Abad, and Emilio Camahort. 2014. Multi-viewer Gesture-based Interaction for Omni-directional Video. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 4077–4086. https://doi.org/10.1145/2556288.2557113

[26] A. Sheikh, A. Brown, Z. Watson, and M. Evans. [n. d.]. Directing attention in 360-degree video. *IET Conference Proceedings* ([n. d.]), 29 (9 .)–29 (9 .)(1). http://digital-library.theiet.org/content/conferences/10.1049/ibc.2016.0029

[27] Mel Slater, Martin Usoh, and Anthony Steed. 1994. Depth of presence in virtual environments. *Presence: Teleoperators & Virtual Environments* 3, 2 (1994), 130–144.

[28] 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.* Springer International Publishing, Cham, 197–208. https://doi.org/10.1007/978-3-319-49616-0_19

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

[30] Mirjam Vosmeer and Ben Schouten. 2017. Project Orpheus A Research Study into 360Âř Cinematic VR. In *Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video (TVX '17)*. ACM, New York, NY, USA, 85–90. https://doi.org/10.1145/3077548.3077559

[31] Maarten Wijnants, Peter Quax, Gustavo Rovelo Ruiz, Wim Lamotte, Johan Claes, and Jean-François Macq. 2015. An Optimized Adaptive Streaming Framework for Interactive Immersive Video Experiences. , 6 pages. https://doi.org/10.1109/BMSB.2015.7177259

[32] Maarten Wijnants, Gustavo Rovelo Ruiz, Donald Degraen, Peter Quax, Kris Luyten, and Wim Lamotte. 2015. Web-Powered Virtual Site Exploration Based on Augmented 360 Degree Video via Gesture-Based Interaction *(WSICC '15)*.