Gaze Visualization for Immersive Video

Thomas Löwe, Michael Stengel, Emmy-Charlotte Förster, Steve Grogorick, and Marcus Magnor

Abstract In contrast to traditional video, immersive video allows viewers to interactively control their field of view in a 360° panoramic scene. However, established methods for the comparative evaluation of gaze data for video require that all participants observe the same viewing area. We therefore propose new specialized visualizations and a novel visual analytics framework for the combined analysis of head movement and gaze data. A novel View Similarity visualization highlights viewing areas branching and joining over time, while three additional visualizations provide global and spatial context. These new visualizations, along with established gaze evaluation techniques, allow analysts to investigate the storytelling of immersive videos. We demonstrate the usefulness of our approach using head movement and gaze data recorded for both amateur panoramic videos, as well as professionally composited immersive videos.

1 Introduction

With the emergence of affordable 360° consumer video cameras, immersive video is becoming increasingly popular [10, 20]. Specialized 360° video players allow users to interactively rotate the viewing direction during playback. Alternatively, head-mounted displays (HMDs) can be used to provide deeper immersion and a more natural control scheme, in which the viewing direction is controlled by the rotation of the head. Recently, YouTube launched support for 360° panoramic video, further heightening public interest in the technology [30].

While immersive video has since been used in sports, marketing and also creative filmmaking, efforts to generate knowledge about storytelling in immersive video have only recently emerged [19]. No specialized methods to evaluate the perception and viewing behavior of the viewer have yet been developed. One of the most common approaches to analyze user attention in traditional video is eye tracking. By recording and aggregating gaze data from multiple participants, experts can gain insight into the viewing behavior of users, e.g. how the eye is guided by content.

T. Löwe (✉) • M. Stengel • E.-C. Förster • S. Grogorick • M. Magnor
Computer Graphics Lab, TU Braunschweig, Braunschweig, Germany
e-mail: loewe@cg.cs.tu-bs.de

However, established visualization techniques for gaze data for video assume that all participants receive the exact same stimulus. This is not the case with immersive video. While all participants are watching the exact same video, each participant is free to choose their individual field of view. Thus, content that occurs outside of this field of view is missed. In order to gain insight into the viewing behaviour for immersive video, both the eye gaze and the head orientation must therefore be considered. Throughout the rest of this chapter we will differentiate between head orientation (*viewing direction*), and eye focus (*gaze direction*).

Storytellers working with immersive video often layer each frame with multiple subplots. They are particularly interested in knowing how many viewers will follow each subplot and understanding which elements of their video may induce a switch between subplots. We therefore focus on *joins* and *branches* between participants' fields of view. Joins occur when the attention of multiple viewers is drawn towards a common direction, causing their fields of view to overlap, whereas branches occur when their fields of view diverge.

We propose a novel View Similarity visualization that illustrates fields of view branching and joining over time. Our proposed visual analytics workflow includes three additional visualizations: A limited view from the viewer's perspective, a 3D sphere-mapped version of the video to provide spatial context, and an unwrapped view of the entire frame to provide global context. All of these views can be additionally overlaid with established gaze visualizations, such as attention maps [16] or scan paths [17], in order to equip experts with a familiar set of analysis tools.

This chapter is organized as follows: Sect. 2 introduces related work. Section 3 describes our proposed visualizations and details the visual analytics workflow. In Sect. 4 we demonstrate the usefulness of the proposed framework using head movement and gaze data we gathered in a user study (Fig. 1). Section 5 concludes this chapter and outlines future work.

Fig. 1 *Left*: A participant watching a 360° video using a head-mounted display. *Right:* The video is mapped to both eyes, putting the observer at the center of the scene

2 Related Work

Eye tracking is an established tool in many fields of research, and has been used to analyze visual attention in several real-world scenarios, including video compression [4, 18], medicine [23], visual inspection training [7], and commercial sea, rail and air vehicle control [11, 12, 31].

Recently, Blascheck et al. presented a comprehensive State-of-the-Art survey of visualization for eye-tracking data [2], citing numerous methods for the visualization of gaze data for traditional video. Among the most common representations for eye-tracking data in video are attention maps [8, 16] and scan paths [17]. However, these methods can not be directly applied to immersive video, where each viewer controls an individual viewing area.

There have been gaze data visualizations that allow participants to individually inspect static 3D scenes in an interactive virtual environment [21, 22, 27, 29]. Here synchronization between participants is achieved by mapping scan paths or attention maps onto the static geometry. However, in immersive video there is no actual 3D geometry, but rather a recorded 360° video that is mapped onto a sphere around the observer. Additionally, the observer's position is fixed to the center of the sphere, since a free view point is not appropriate for immersive video recorded from a fixed camera position. Thus, immersive video falls into the mid-range between traditional video and 3D scenes, and neither approach directly applies. Our framework therefore combines methods from both scenarios using multiple views.

We further reduce the problem of synchronizing participants to finding moments when the attention, i.e. viewing direction, of many users is drawn to a certain region in the video. These moments are also commonly referred to as moments of attentional synchrony [25].

In traditional video, attentional synchrony is also analyzed by monitoring gaze transitions between manually annotated *areas of interest* (AOI) [3, 13, 14]. However, annotating these AOIs is often time-consuming and exhausting. This is particularly true for immersive video, where the unintuitive distortion of popular texture formats (e.g. equirectangular projection) makes selection more difficult, e.g. AOIs moving around the observer will have to wrap around from the right to the left edge of the video frame. Additionally, multiple stories often occur simultaneously in different parts of the video, further increasing the workload for the annotation.

While we believe that AOIs can be beneficial for the evaluation of immersive video, specialized annotation tools would be required to make working with AOIs feasible. Therefore, our approach avoids dependency on manually annotated AOIs and instead gauges attentional synchrony based on the similarity of the individual viewing directions.

3 Workflow and Visualizations

In an immersive video, each frame is an equirectangular 360° panorama. During playback, the video is mapped onto a sphere, with the observer at its center. In contrast to traditional video, the observer can interactively control their viewing direction. With regular 360° panoramic video players, this is usually done by clicking and dragging the video using a trackball metaphor. In a head-mounted display on the other hand, the viewing direction is directly controlled through head rotation, which can create a deeper sense of immersion. While we focus on users watching immersive video using a head mounted display, our proposed workflow also holds for regular 360° panoramic video players.

In order to be able to evaluate viewing behavior for such immersive video it is not enough to only consider the recorded gaze direction, but also the recorded viewing direction and field of view. An easy way to simultaneously visualize these aspects is to unwrap the video and map both the gaze position, as well as the field of view onto it. However, the warped equirectangular frame is often difficult to interpret (Fig. 2). The otherwise rectangular field of view becomes distorted in this visualization; particularly near the top and bottom (poles) of the frame. Additionally, the frame wraps horizontally, occasionally splitting the field of view. While the frame could be warped in such a way as to rectify and to center a single users field of view, this would further complicate any comparative analysis. We therefore conclude that

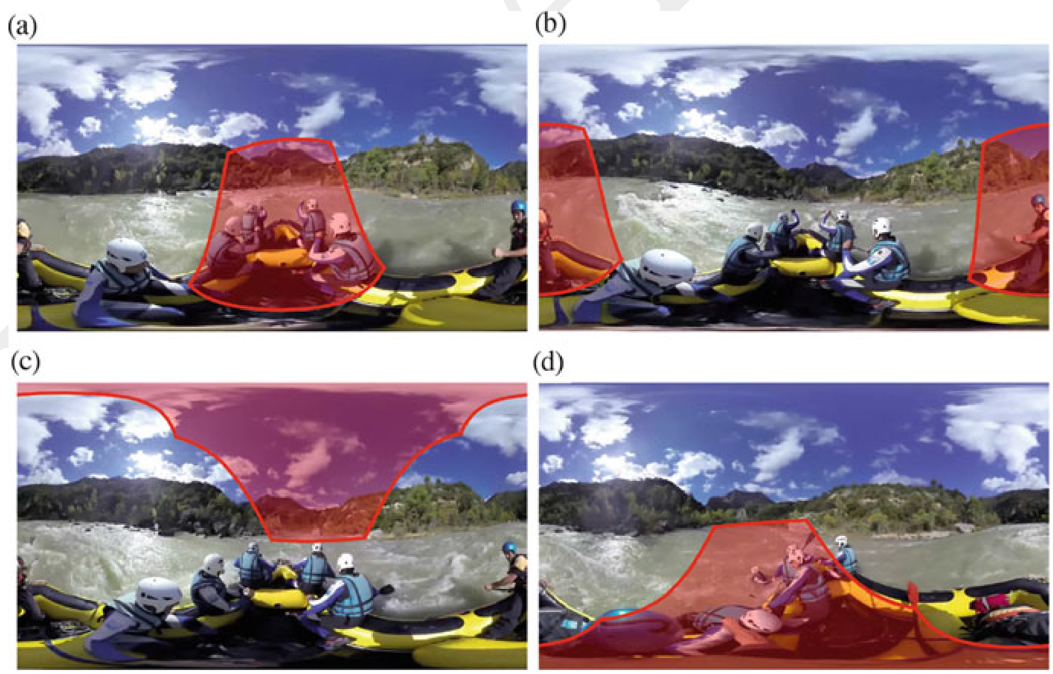

Fig. 2 Distortions introduced by unwrapping a spherical video frame can complicate the interpretation of a participant's otherwise rectangular field of view. The *red area* represents a participant's field of view with (**a**) minor distortions, (**b**) horizontal wrapping, (**c, d**) polar distortions

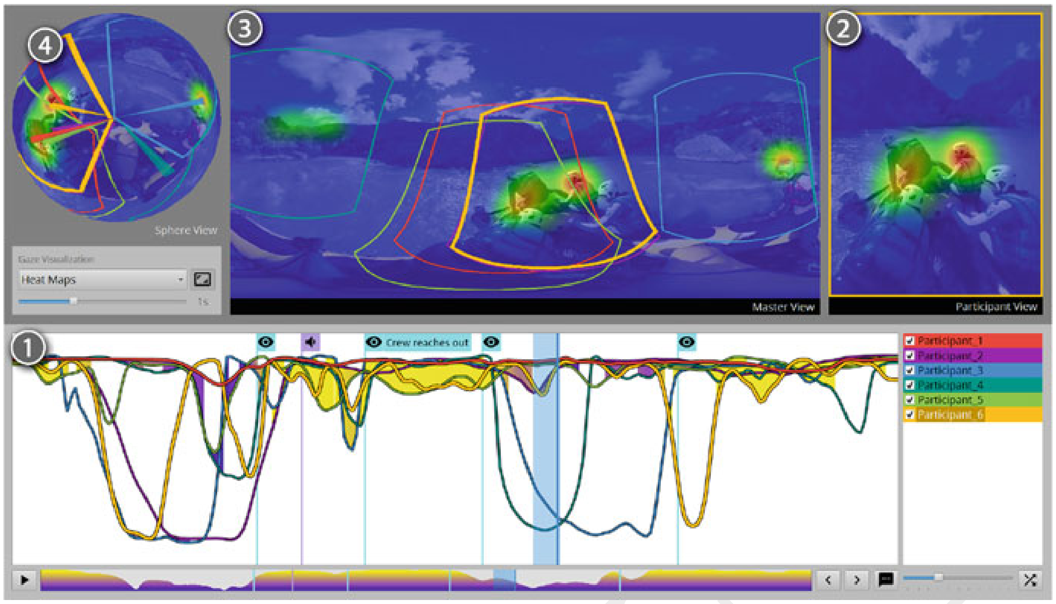

Fig. 3 Overview of our user interface for the immersive video clip "VIDEO 360: RAFTING" [1]. (*1*) View Similarity visualization for the entire clip. (*2*) Participant View for *Participant 6*. (*3*) Master View and (*4*) Sphere View. The color-coded frames indicate the individual field of view for each participant. Additionally, an attention map allows determining the overall gaze distribution

simply unwrapping the frame and applying traditional gaze visualization techniques for regular video is insufficient for analyzing immersive 360° video.

Figure 3 shows an overview of our proposed user interface.

The bottom half of our interface is dedicated to providing a temporal overview of the head-tracking data. A seek slider can be used to select a frame in the video. This slider is additionally overlaid with a quality metric that guides analysts towards potentially relevant sections, i.e. those sections in which many participants are focusing on similar regions of the scene. A specialized View Similarity visualization allows discriminating between individual participants, and combines temporal with spatial context. The viewing direction of each participant is represented by a line, with the proximity of the lines representing the view similarities over time. The closer the lines, the more similar the viewing direction in that frame.

The upper half of our interface is dedicated to analyzing gaze data, and to providing a spatial overview. On the right, a limited user view shows the scene from the currently selected participant's perspective. In the middle, an unwrapped view of the entire scene provides global context. On the left, an interactively rotatable 3D sphere-mapped version of the video allows the analyst to view the frame in a more natural projection. This allows for a better understanding of rotational context that is commonly lost in the unnaturally distorted unwrapped view.

Each of theses views can additionally be overlaid with established gaze visualizations commonly used for the analysis of regular video, such as animated attention maps or animated scan paths. For these traditional gaze visualizations, gaze data is aggregated over a user-controlled temporal window.

In the following we give a detailed description of each visualization and discuss its intended usage and technical details.

(1) View Similarity The most prominent visualization in our interface is the View Similarity visualization. It shows the proximity of all participants' viewing directions over time. This allows analysts to quickly identify moments of attentional synchrony between individual participants. In order to be able to visualize the relationship of multiple 3D viewing directions over time, we use a dimensionality reduction technique. First, we create a distance matrix of all recorded viewing directions, regardless of participant and timestamp. The distance Δ between two viewing directions $v_{1,2} \in \mathbb{R}^3$ is determined using their cosine dissimilarity:

$$\Delta(v_1, v_2) = cos^{-1}\langle v_1, v_2\rangle,$$

where $\langle\cdot,\cdot\rangle$ denotes the scalar product.

We then use nonmetric multidimensional scaling [6] to create a 1D embedding of all viewing directions, as this method preserves relative distances as well as possible. Depending on the temporal resolution of the head-tracker and the number of participants, this matrix may become arbitrarily large. The approach therefore strongly depends on the stability and performance of the underlying MDS algorithm. On the other hand, head-tracking data is not subject to as fine and rapid changes as eye-tracking data. Therefore, reducing the temporal resolution for the purpose of calculating the View Similiarity, will still yield adequate results.

Finally, we reintroduce time as a second dimension. By connecting all records for a participant over time, we obtain a line graph. In this resulting graph the proximity of lines at each frame is an approximate representation of the similarity between viewing directions in that frame.

To further highlight attentional synchrony, similarities between viewing directions that are above a user-defined threshold are additionally marked, by visually connecting the lines into clusters. Readability is further enhanced by additionally coloring these clusters using a simple quality metric and color gradient.

This quality metric q uses the sum of normalized distances from each participant's viewing direction $p \in P$ to it's k-nearest-neighbor $k_p \in P$:

$$q = 1 - \frac{1}{|P|}\sum_{p \in P}\left(\frac{\Delta(p, k_p)}{\pi}\right)^2, \quad q \in [0, 1]$$

The smaller the distances, the more clustered the viewing directions, and the higher the quality. We found that $k = 2$ worked well for our smaller test data sets. We also empirically found that a default threshold value of one third of the field of view of the display worked well, which for our setup was approximately 30°.

A simple seek slider at the bottom allows selecting a frame in the video, as well as zooming and panning the View Similarity visualization.

Analysts can also place annotations at key frames, in order to mark important audio or visual cues in the video, thus adding semantic context to the visualization. Each annotation is represented by a colored flag on the timeline.

(2) Participant View The Participant View shows the scene as it was experienced by a single selected participant. The limited perspective is intended to prevent analysts from erroneously assuming information that is provided to them by the more global views, but that would not have been visible to the participant during the trial. This view allows analysts to study the attention of an individual participant, and to investigate which elements in the scene might have influenced that participant to move their field of view.

Unfortunately, this limited view does not allow the analyst to differentiate between movement of the participant's head and movement of the camera in the video. While the camera is often fixed in immersive video—as simulated self motion has a tendency to cause discomfort for some participants [15, 26]—there are a large number of fast moving amateur sports and drone videos, as well as an increasing number of artistic videos that make use of slow, deliberate camera movements. Therefore, our framework supplies two additional spatial visualizations that provide global spatial context in relation to the video content.

(3) Master View The Master View shows the entire scene as an unwrapped video frame. In this equirectangular mapping—also known as latitude-longitude mapping—the azimuth is mapped to the horizontal coordinate, while the elevation is mapped to the vertical coordinate of the image. This mapping format is commonly used, since the sphere is flattened into a rectangular area, and thus traditional compression methods for rectangular images and videos can be applied. In this unwrapped view, the center of the view is the relative *front* of the scene (Fig. 2a), and the left and right edges of the view are the relative *back* of the scene (Fig. 2b). The top and bottom of each frame exhibit the most distortion, as these map to the *top* and *bottom* poles of the sphere (Fig. 2c, d). Additionally, the individual fields of view of each participant are marked by color-coded frames. This view is intended to give analysts global context, since all events that are occurring in a frame of the video can be observed at once.

As previously discussed, the distorted perspective and the fact that the image wraps around can make interpretation difficult. Therefore, our framework provides an additional more natural mapping.

(4) Sphere View The Sphere View maps the immersive video to the inside of a sphere, which can be rotated using the trackball metaphor [24]. As with the Master View, the fields of view of each participant are marked by color-coded frames. An arrow from the center of the sphere to the eye focus position of each selected participant additionally marks the gaze directions in 3D. This grants the analyst an intuitive spatial understanding of which direction each participant is facing in the immersive scene. For fulldome videos, this sphere view can also be used to obtain the original dome master mapping (Figs. 8 and 9).

4 Results

Our attention analysis framework relies on head movement and gaze data being recorded while the participant is immersed in the video. While HMDs with integrated eye tracking have been costly and therefore suited for professional use only, consumer-grade devices have been announced [9] and will allow a larger community to perform eye-tracking studies in virtual reality. In order to be able to develop suitable visualizations for such future studies, we have developed and built our own HMD with integrated eye tracking [28].

Our HMD provides binocular infrared-based eye tracking with low-latency and a sampling frequency of 60 Hz. In a calibrated state the gaze direction error of the eye tracker ranges from 0.5° to 3.5°, increasing at the edges of the screen. The head tracker provides 100 Hz for updating the head orientation with a viewing direction error of 0.5° to 1.0°. The display has a native resolution of 1280 × 800 pixels and a refresh rate of 60 Hz. The field of view is 110° vertically and 86° horizontally per eye.

We recorded data from 6 participants (5 males, 1 female) of which 4 had normal vision and 2 had corrected-to-normal vision. Our perceptual study was conducted as follows: First, we explained the HMD and the concept of 360° video to the participant. The participant was then seated on a rotatable chair and the HMD was mounted on their head, while still allowing for free head and body movement (Fig. 4). After calibrating the eye tracker, different monocular 360° panoramic videos were shown to the participant, while their head orientation and gaze data was being captured.

We showed a total of four immersive videos. Three videos consisted of equirectangular frames with a resolution of 1280 × 640, the fourth video consisted of dome master frames with a resolution of 2048 × 2048.

Fig. 4 Our experimental setup: A participant is watching an immersive video using our custom-built head-mounted display with integrated eye tracking. The participant is seated on a rotatable chair, in order to allow safe and free 360° body and head movement

In the following we take an in depth look at our results for two of these videos: The Creative Commons YouTube video “VIDEO 360: RAFTING”, and a clip from the artistic fulldome short film “UM MENINO”.

4.1 Video: RAFTING

“VIDEO 360: RAFTING” [1] is a short immersive video available under the Creative Commons CC BY license (Fig. 5). The clip is 42 s long and shows a rafting scene with a static camera centered in the raft. At 11 s into the video, the raft and the camera tilt, and two of the rafters fall into the water. The remainder of the video shows the crew reaching out and pulling the two back into the raft safely.

Figure 5 shows the View Similarity visualization with three frame annotations. The first annotation marks the moment the raft and the camera begin to tilt, the second marks an audible scream, and the third marks the moment the crew reaches out and starts helping their crewmates.

We observe that initially, all participants are individually exploring the immersive scene. The Master View in Fig. 3 shows that during this time, most attention is focused towards the travelling direction of the raft.

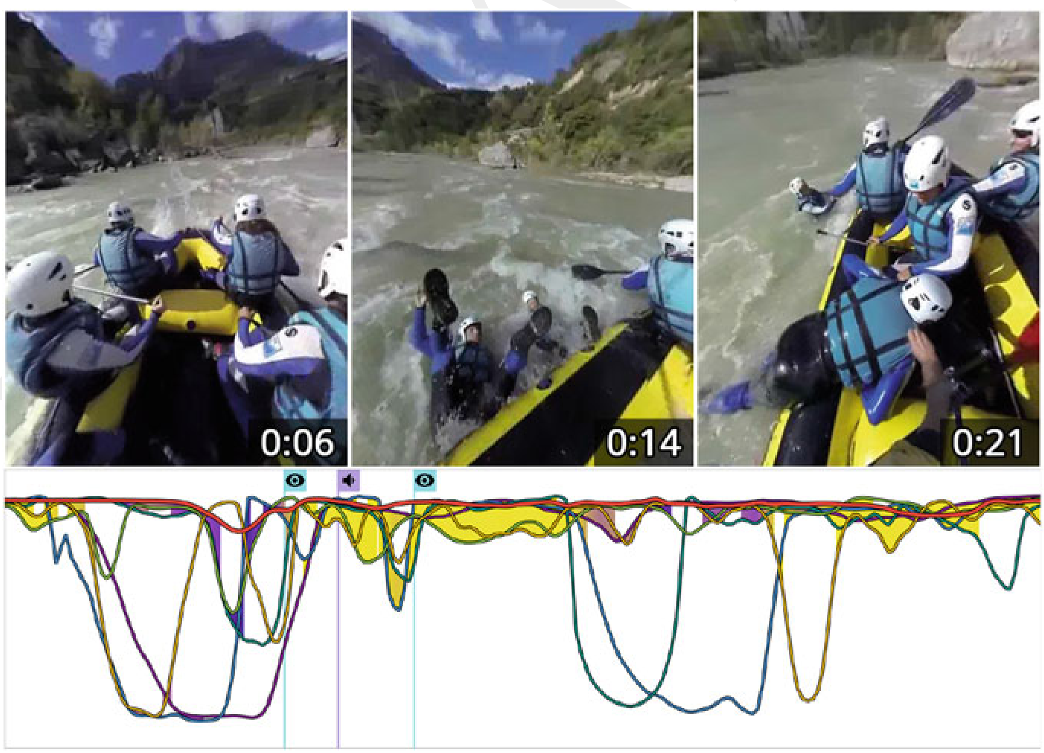

Fig. 5 *Top*: Scenes from the immersive 360° video clip “VIDEO 360: RAFTING” [1]. *Bottom*: View Similarity Visualization for the entire video

From the moment the raft and the camera tilt, all participants begin searching for what happened. We observe, that all participants' views follow the tilt of the raft and thus quickly converge around the two rafters in the water, even before the scream can be heard. Figure 6 shows the Master View during the rescuing efforts, with an overlaid attention map accumulated over all participants.

During the rescuing effort, the field of view of most participants remains centered on the events unfolding on the raft. It is particularly interesting that the gaze of most participants is focused on the helping crewmembers, rather than on the rafters in the water. After the first rafter has been saved, we observe that individual participants briefly turn their heads to check on the crewmember in the back of the raft.

After both rafters are safely back on the raft, we observe that most attention is again directed towards the traveling direction of the raft.

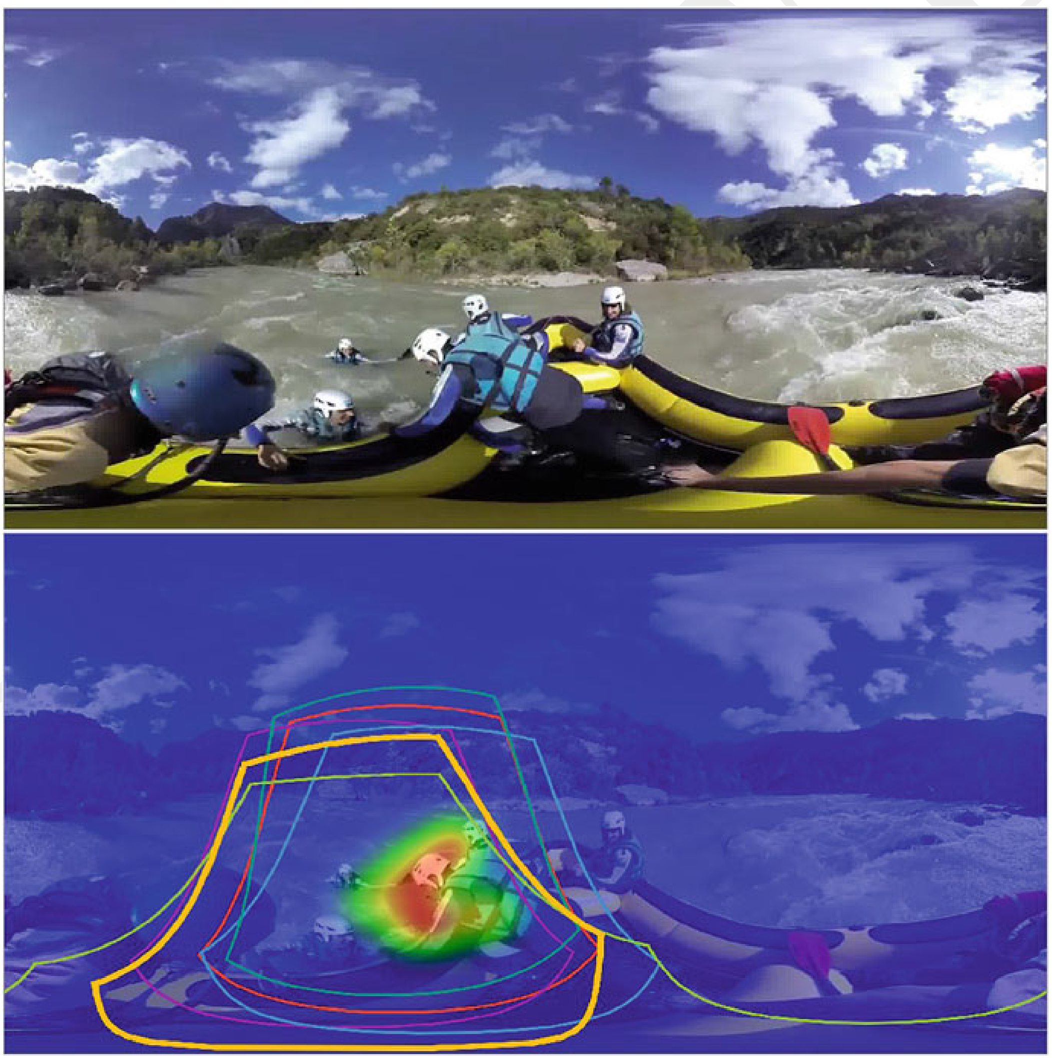

Fig. 6 *Top*: Master View of a frame from "VIDEO 360: RAFTING". Two rafters have fallen into the water, while the crew reaches out to help them. *Bottom*: The same frame overlaid with fields of view and an attention map. All participants are focused on the rescuing effort

In this use case our framework enabled us to quickly identify moments of attentional synchrony and to investigate the gaze behaviour leading up to that moment.

4.2 Video: UM MENINO

"UM MENINO" is an artistic 360° fulldome short film. The complete video is 5:46 min long and shows circus performers composited into a virtual environment. While the video is for the most part designed with a fixed forward direction, it has immersive elements. For our evaluation we selected a 45 s long sequence that starts at 2:23 min. The clip begins with a slow dolly shot moving backwards through a busy fairground. After 15 s the camera accelerates, simulating the viewer speeding away backwards in a roller coaster. After an additional 15 s the ride ends as the viewer emerges from the mouth of a giant clown, leading into an abstract kaleidoscopic sequence.

Figure 7 shows the View Similarity visualization with five frame annotations. The first annotation marks candy being thrown, the second annotation marks additional performers appearing, the third annotation marks a sound of the roller coaster

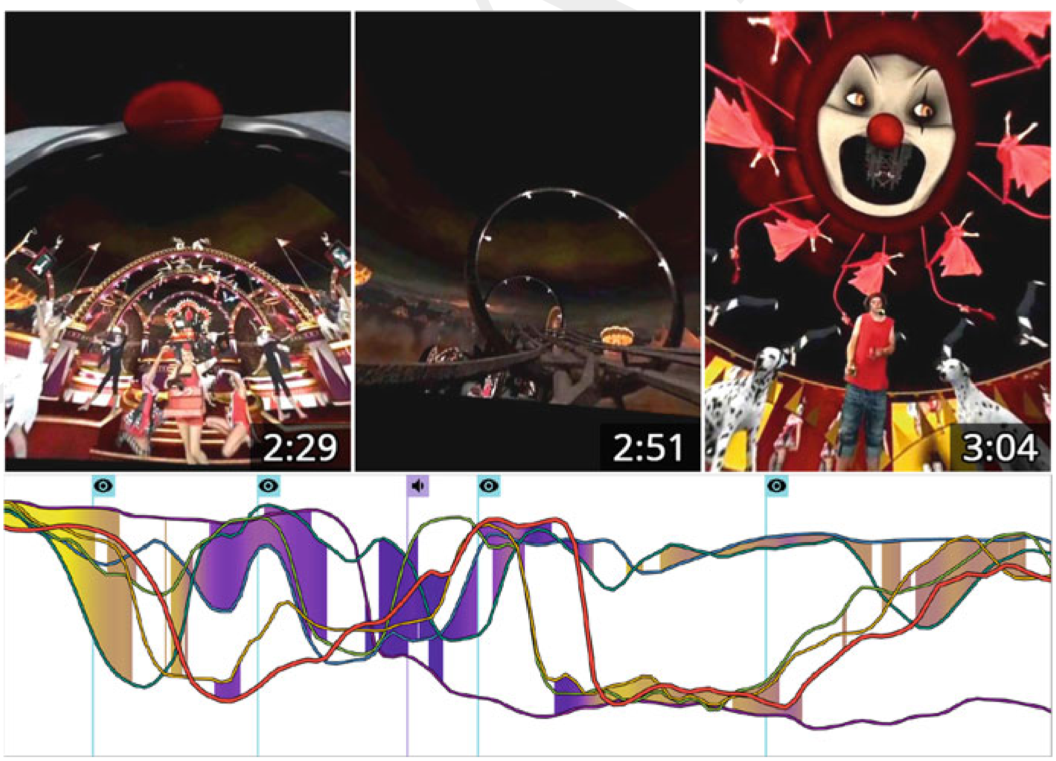

Fig. 7 *Top*: Scenes from the immersive fulldome short film "UM MENINO". *Bottom*: View Similarity Visualization for the entire video

charging up, the fourth annotation marks the moment the roller coaster accelerates, and the fifth annotation marks the beginning of the kaleidoscopic sequence.

The artists also created a regular video version of their short film, in which an observer is simulated [5]. This adaptation was intended to make the otherwise fulldome video available to a broader audience. We use this adaptation as a guide to understand the artist-intended viewing direction.

We observe that initially, each participant is individually exploring the scene. Using the Master View and the Participant View we additionally notice that most attention is indeed focused towards the artist-intended viewing direction.

At approximately 6 s into the video the performer at the center of the fairground reaches down and throws animated candy. This candy flies over the observer to the other side of the dome, and as can be seen in the regular video adaptation, the artists' intention was for the viewer to track it. In Fig. 8 (left) we observe that the participants recognized and initially followed the candy. However, they did not fully turn around, but rather quickly returned their attention to the busy fairground, which can be seen in Fig. 8 (right).

Shortly after the roller coaster sequence begins, the viewing directions form two distinct clusters. In Fig. 9 (left), the scan path visualization shows that most participants turn away from the artist-intended viewing direction, in order to instead face the travelling direction of the roller coaster. While this was not the case during the slow backward movement of the dolly shot of the previous sequence, the sudden acceleration appears to have caused a change in viewer behavior.

In the final kaleidoscopic sequence the camera slowly moves downwards, away from the giant clown at the top of the dome. This scene is largely symmetric, except for the artist-intended viewing direction, in which the protagonist of the video can be seen juggling. Figure 9 (right) shows that during this sequence most viewers

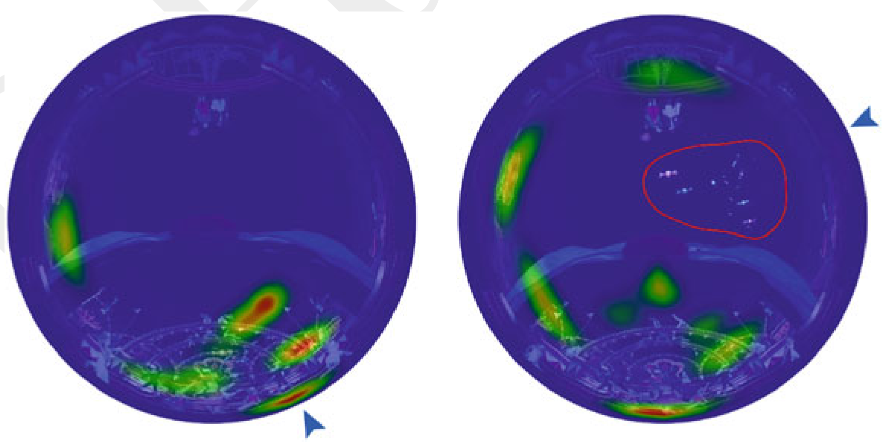

Fig. 8 Two Sphere Views of the candy throwing sequence from "UM MENINO". Both views are looking upward into the dome and are overlaid with attention maps. The *blue arrow* marks the artist-intended viewing direction. On *the left*: The performer at the center of the fairground begins to throw candy. On *the right*: The candy (*circled* in *red*) is floating across the dome along the artist-intended viewing direction, but none of the participants are tracking it

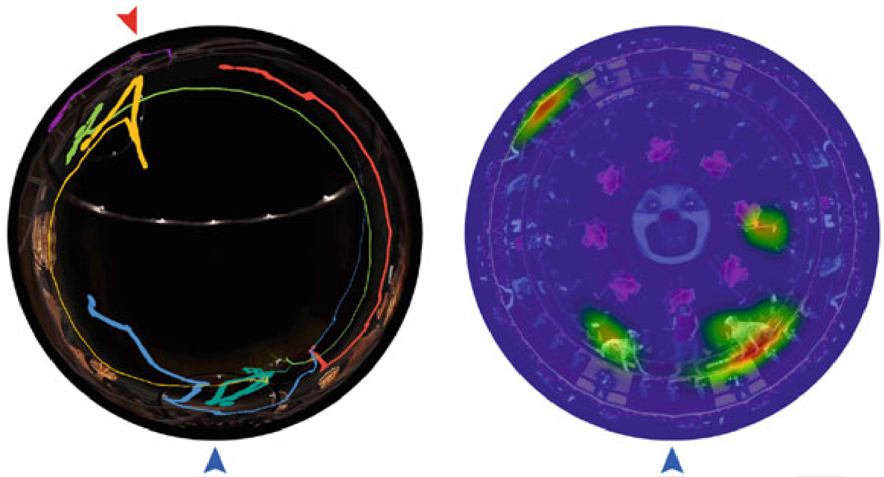

Fig. 9 Two Sphere Views from "UM MENINO" looking upward into the dome. The artist-intended viewing direction is marked by the *blue arrow*. On *the left*: The roller coaster sequence with scan paths. The *red arrow* marks the travelling direction of the roller coaster. On *the right*: The kaleidoscopic sequence with a superimposed attention map

have returned to the artist-intended viewing direction, focusing on the dogs next to the protagonist, with a slight tendency to look up.

In this use case, our framework allowed us to identify moments in which the observed viewing direction differed from the artist-intended viewing direction. Our additional visualizations further allowed us to investigate potential reasons for this difference in behavior.

5 Conclusion and Future Work

In this chapter, we have presented a novel visual analytics framework for jointly analyzing head movement and gaze data for immersive videos. Our design provides a specialized View Similarity visualization which allows analysts to quickly identify moments of spatiotemporal agreement between the viewing directions of individual participants. We also proposed three additional views (participant view, master view, and sphere view) that provide spatial context. These views can be combined with established gaze visualization techniques, in order to investigate viewing behavior in immersive video. We evaluated our approach within a small-scale perceptual study including amateur, choreographed and animated immersive video, and found that our framework can be used to detect whether the attention guidance of an immersive video works as intended.

As future work, we intend to further investigate how our method can be used to review and enhance artistic storytelling in immersive videos of different genres. We would like to extend our approach by supporting annotated areas of interest, in order to obtain an additional semantic context. We would also like to conduct a larger perceptual study in order to gain further and more statistically significant insight into attentional synchrony and, by extension, storytelling in immersive videos.

Acknowledgements The authors thank Laura Saenger, Flávio Bezerra and Eduard Tucholke for permission to use the short film "UM MENINO". The authors gratefully acknowledge funding by the German Science Foundation from project DFG MA2555/6-2 within the strategic research initiative on Scalable Visual Analytics and funding from the European Union's Seventh Framework Programme FP7/2007-2013 under grant agreement no. 256941, Reality CG.

References

1. Ábaco Digital Zaragoza, VIDEO 360: RAFTING, 2015. youtube.com/watch?v=h0xo8QEPrk0, vis. 27 Jul 2015
2. Blascheck, T., Kurzhals, K., Raschke, M., Burch, M., Weiskopf, D., Ertl, T.: State-of-the-art of visualization for eye tracking data. In: Proceedings of EuroVis, vol. 2014 (2014)
3. Burch, M., Kull, A., Weiskopf, D.: Aoi rivers for visualizing dynamic eye gaze frequencies. In: Computer Graphics Forum, vol. 32, pp. 281–290. Wiley Online Library, Chichester (2013)
4. Cheon, M., Lee, J.-S.: Temporal resolution vs. visual saliency in videos: analysis of gaze patterns and evaluation of saliency models. Signal Process. Image Commun. **39**, 405–417 (2015)
5. ChimpanZés de Gaveta, Um Menino Trilha: ChimpanZés de Gaveta, 2015. youtube.com/watch?v=q72AwheNYPk, vis. 18 Feb 2016
6. Cox, T., Cox, A.: Multidimensional Scaling, 2nd edn. Taylor & Francis, Boca Raton (2010)
7. Duchowski, A.T., Medlin, E., Cournia, N., Gramopadhye, A., Melloy, B., Nair, S.: 3d eye movement analysis for vr visual inspection training. In: Proceedings of the 2002 Symposium on Eye Tracking Research & Applications, pp. 103–110. ACM (2002)
8. Duchowski, A.T., Price, M.M., Meyer, M., Orero, P.: Aggregate gaze visualization with real-time heatmaps. In: Proceedings of the Symposium on Eye Tracking Research and Applications, pp. 13–20. ACM (2012)
9. FOVE: The world's first eye tracking virtual reality headset, 2015. getfove.com, vis. 29 Jul 2015
10. Google Jump, 2015. google.com/cardboard/jump, vis. 29 Jul 2015
11. Itoh, K., Hansen, J.P., Nielsen, F.: Cognitive modelling of a ship navigator based on protocol and eye-movement analysis. Le Travail Humain, pp. 99–127. Presses Universitaires de France, Paris (1998)
12. Itoh, K., Tanaka, H., Seki, M.: Eye-movement analysis of track monitoring patterns of night train operators: effects of geographic knowledge and fatigue. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 44, pp. 360–363. SAGE Publications (2000)
13. Kurzhals, K., Weiskopf, D.: Space-time visual analytics of eye-tracking data for dynamic stimuli. IEEE Trans. Vis. Comput. Graph. **19**(12), 2129–2138 (2013)
14. Kurzhals, K., Weiskopf, D.: Aoi transition trees. In: Proceedings of the 41st Graphics Interface Conference, pp. 41–48. Canadian Information Processing Society (2015)
15. LaViola Jr, J.J.: A discussion of cybersickness in virtual environments. ACM SIGCHI Bull. **32**(1), 47–56 (2000)
16. Mackworth, J.F., Mackworth, N.: Eye fixations recorded on changing visual scenes by the television eye-marker. JOSA **48**(7), 439–444 (1958)
17. Noton, D., Stark, L.: Scanpaths in eye movements during pattern perception. Science **171**(3968), 308–311 (1971)

18. Nyström, M., Holmqvist, K.: Effect of compressed offline foveated video on viewing behavior and subjective quality. ACM Trans. Multimed. Comput. Commun. Appl. (TOMM) **6**(1), 4 (2010)
19. Oculus Story Studio, 2015. oculus.com/storystudio, vis. 29 Jul 2015
20. Perazzi, F., Sorkine-Hornung, A., Zimmer, H., Kaufmann, P., Wang, O., Watson, S., Gross, M.: Panoramic video from unstructured camera arrays. In: Computer Graphics Forum, vol. 34, pp. 57–68. Wiley Online Library, Chichester (2015)
21. Pfeiffer, T.: Measuring and visualizing attention in space with 3D attention volumes. In: Proceedings of the Symposium on Eye Tracking Research and Applications, pp. 29–36. ACM (2012)
22. Ramloll, R., Trepagnier, C., Sebrechts, M., Beedasy, J.: Gaze data visualization tools: opportunities and challenges. In: Proceedings of the Eighth International Conference on Information Visualisation, IV 2004, pp. 173–180. IEEE (2004)
23. Schulz, C., Schneider, E., Fritz, L., Vockeroth, J., Hapfelmeier, A., Brandt, T., Kochs, E., Schneider, G.: Visual attention of anaesthetists during simulated critical incidents. Br. J. Anaesth. **106**(6), 807–813 (2011)
24. Shoemake, K.: Arcball: a user interface for specifying three-dimensional orientation using a mouse. In: Graphics Interface, vol. 92, pp. 151–156. Morgan Kaufmann Publishers, San Francisco (1992)
25. Smith, T., Henderson, J.: Attentional synchrony in static and dynamic scenes. J. Vis. **8**(6), 773–773 (2008)
26. Soyka, F., Kokkinara, E., Leyrer, M., Buelthoff, H., Slater, M., Mohler, B.: Turbulent motions cannot shake vr. In: Virtual Reality (VR), pp. 33–40. IEEE (2015)
27. Stellmach, S., Nacke, L., Dachselt, R.: Advanced gaze visualizations for three-dimensional virtual environments. In: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications, pp. 109–112. ACM (2010)
28. Stengel, M., Grogorick, S., Rogge, L., Magnor, M.: A nonobscuring eye tracking solution for wide field-of-view head-mounted displays. In: Eurographics 2014-Posters, pp. 7–8. The Eurographics Association (2014)
29. Tory, M., Atkins, M.S., Kirkpatrick, A.E., Nicolaou, M., Yang, G.-Z.: Eyegaze analysis of displays with combined 2D and 3D views. In: Visualization (VIS'05), pp. 519–526. IEEE (2005)
30. Youtube creator blog, 2015. youtubecreator.blogspot.de/2015/03/a-new-way-to-see-and-share-your-world.html, vis. 27 Jul 2015
31. Weibel, N., Fouse, A., Emmenegger, C., Kimmich, S., Hutchins, E.: Let's look at the cockpit: exploring mobile eye-tracking for observational research on the flight deck. In: Proceedings of the Symposium on Eye Tracking Research and Applications, pp. 107–114. ACM (2012)