Visualization and Analysis of Head Movement and Gaze Data for Immersive Video in Head-mounted Displays

Thomas Löwe *Student Member, IEEE*, Michael Stengel *Student Member, IEEE*,
Emmy-Charlotte Förster *Student Member, IEEE*, Steve Grogorick, and Marcus Magnor *Senior Member, IEEE*

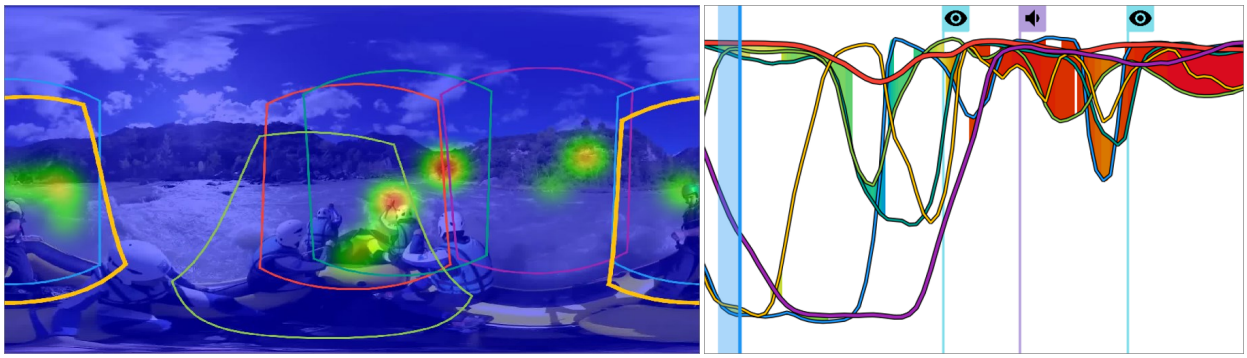

Fig. 1. On the left: Unwrapped frame from an immersive rafting video. The fields of view of individual participants are marked by color coded frames. An attention map allows insight into gaze behavior. On the right: Our specialized View Similarity visualization shows fields of view as lines, branching and joining over time, with the thick vertical blue line indicating the current frame. We observe that in a few seconds all views will converge towards a common region in the video, triggered by two rafters falling overboard.

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 workflow for the combined analysis of head movement and gaze data. A 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 video. 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.

Index Terms—Visual analytics, eye tracking, immersive video

1 Introduction

With the emergence of affordable 360° consumer video cameras, immersive video is becoming increasingly popular [14, 24]. Specialized immersive 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° video, further heightening public interest in the technology [26].

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 [25]. 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 video content.

However, established visualization techniques for gaze data for video assume that all participants observe the exact same stimulus. This is not the case with immersive video, as participants free to choose their individual field of view. 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 be considered. Throughout the rest of this paper we will therefore differentiate between the head orientation (*viewing direction*), and the eye focus of a participant (*gaze direction*).

Of particular interest are moments of attentional synchrony [19], i.e. when the attention of many users is drawn to the same region in the video. In immersive video, we are particularly interested in *joins* and *branches* in viewing experiences. Joins occur when the attention of multiple users is drawn towards a common direction, causing their fields of view to overlap, whereas branches occur when their fields of view diverge. In order to study these moments of attentional synchrony, we propose a View Similarity visualization that illustrates fields of view branching and joining over time, see Fig. 1. Our proposed visual analytics workflow includes three additional visualizations: A limited view from the participant'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.

- *Thomas Löwe is with the Computer Graphics Lab, TU Braunschweig, Germany. E-mail: loewe@cg.cs.tu-bs.de.*
- *Michael Stengel is with the Computer Graphics Lab, TU Braunschweig, Germany. E-mail: stengel@cg.cs.tu-bs.de.*
- *Emmy-Charlotte Förster is with the Computer Graphics Lab, TU Braunschweig, Germany. E-mail: foerster@cg.cs.tu-bs.de.*
- *Steve Grogorick is with the Computer Graphics Lab, TU Braunschweig, Germany. E-mail: grogorick@cg.cs.tu-bs.de.*
- *Marcus Magnor is with the Computer Graphics Lab, TU Braunschweig, Germany. E-mail: magnor@cg.cs.tu-bs.de.*

Manuscript received 31 Mar. 2015; accepted 1 Aug. 2015; date of publication xx Aug. 2015; date of current version 25 Oct. 2015.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

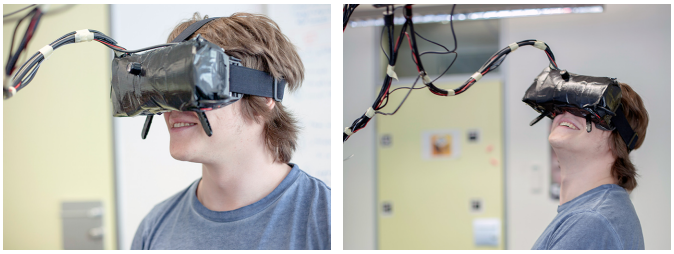

Fig. 2. Our experimental setup: A participant is watching an immersive video using our custom-built head-mounted display with integrated eye tracking. The participant is seated on a rotatable chair, in order to allow safe and free 360° body and head movement.

All of these views can be overlaid with the established attention map [11] and scan path [12] visualizations, in order to equip analysts with a familiar set of evaluation tools.

This paper is organized as follows: Section 2 introduces related work. Section 3 describes our proposed visualizations and details the visual analytics workflow. In Section 4 we demonstrate the usefulness of the proposed workflow using head movement and gaze data we gathered in a user study, using a custom-built HMD with integrated eye tracking [21] (Fig. 2). Section 5 concludes this paper and outlines future work.

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 [3, 13], medicine [17], visual inspection training [5], and commercial sea, rail and air vehicle control [7, 8, 28].

Recently, Blascheck et al. presented a comprehensive State-of-the-Art survey of visualization for eye tracking data [1], 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 [6, 11] and scan paths [12]. However, when comparing gaze data from multiple participants, these visualization techniques require that all viewers receive the exact same stimulus, which is not the case for immersive video, as each participants controls an individual viewing area.

There have been gaze data visualizations that allow users to inspect static 3D scenes in an interactive virtual environment [15, 16, 20, 22]. Here synchronization between participants is achieved by mapping scan paths or attention maps onto the static geometry. Since these methods assume the 3D stimulus to be static—which is not the case for video—they are not applicable in our case. Moreover, these approaches use a free 3D viewpoint, whereas we only need to consider the orientation of the participant's head, since a free camera position is not appropriate for immersive video. We can thus 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 [19].

In traditional video, attentional synchrony is also analyzed by monitoring gaze transitions between manually annotated *areas of interest* (AOI) [2, 9, 10]. 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 contrast to traditional video, immersive video allows participants to freely control their field of view. It can thus no longer be assumed that all participants share the same viewing direction, making traditional comparative eye tracking analysis difficult. We therefore propose a new visual analytics workflow that considers both the recorded gaze direction, as well as the recorded head orientation of participants watching immersive videos.

Figure 3 provides 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 frames, i.e. those frames in which many participants are focusing on similar regions of the scene. A specialized View Similarity visualization allows discriminating between individual participants, and adds spatial context. The viewing direction of each participant is represented by a line, with the proximity of lines representing the view similarities over time. The closer the lines, the more similar the viewing directions.

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 analysts 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 such as animated attention maps or scan paths, with gaze data aggregated over a user-controlled temporal window.

In the following we give a detailed description of each visualization and discuss it's usage and technical details.

(1) View Similarity The most prominent visualization in our interface is the View Similarity visualization (Fig. 1 right, Fig. 3 bottom). It shows the angular proximity of all participants' viewing directions over time. This allows analysts to quickly identify moments of attentional synchrony between individual participants.

A simple seek slider at the bottom allows selecting a frame in the video, as well as zooming and panning the View Similarity visualization. The potential relevance of each frame is automatically gauged by a quality metric. This metric uses the sum of distances from each viewing direction to it's k-nearest-neighbor. In this paper, we define the *distance* between two viewing directions to be their *angular dissimilarity*. The smaller the distances, the more clustered the viewing directions, and the higher the quality. An area chart in the seek slider indicates the result, with an additional color gradient to accentuate exceptionally high scoring frames.

We use dimensionality reduction, in order to be able to visualize the relationship of multiple 3D viewing directions, i.e. head orientations, over time. First, we create a distance matrix for all viewing directions, regardless of participant and time. We then use nonmetric multidimensional scaling [4] to create a 1D embedding of these viewing directions, since this method preserves relative distances as well as possible. Finally, we reintroduce time as a second dimension, resulting in a line plot, in which the distances between lines are an approximate representation of the similarities between viewing directions. 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. In order to increase readability, we color these clusters using the above mentioned quality metric and color gradient. We empirically found that a default threshold value of one third of the angle of view worked well.

Analysts can also place annotations at specific key frames, in order to mark important audio or visual cues in the video, thus adding semantic context to the visualization.

(2) Participant View The Participant View shows the scene as it was experienced by the selected participant. This allows analysts to study the attention of an individual participant, and which elements in the scene might have influenced that participant to move her field of view.

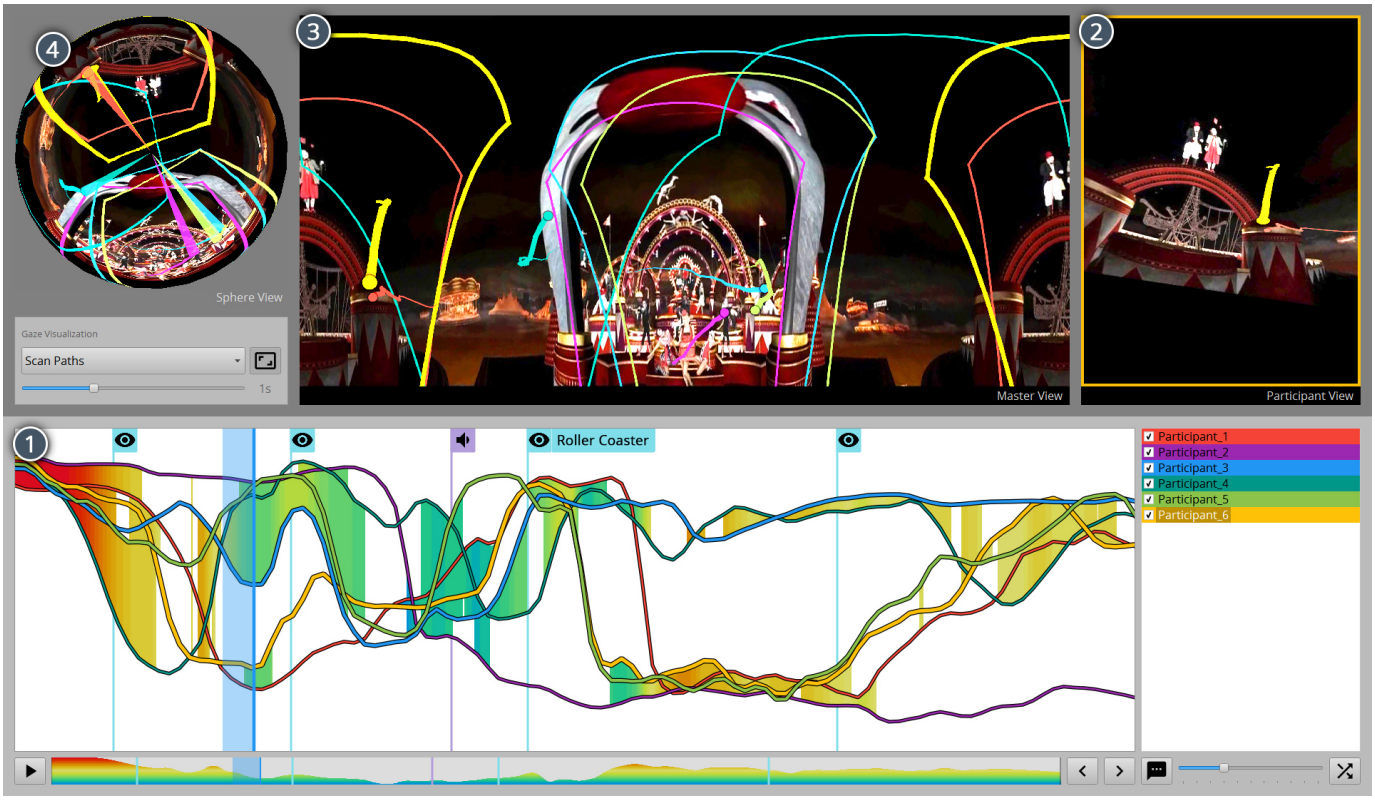

Fig. 3. Overview of our user interface for the fairground sequence from the immersive fulldome short film UM MENINO. (1) View Similarity Visualization for the entire clip. (2) Participant View for *Participant 6*. (3) Master View and the (4) Sphere View. The color-coded frames indicate the individual fields of view. Additionally, color-coded scan paths allow determining the gaze direction of each user.

Additionally, this limited perspective is intended to prevent analysts from erroneously assuming information that is provided to them by the global views, but that would not have been visible to the participant during the trial.

(3) Master View The Master View shows the entire scene as an unwrapped video frame. In this equirectangular mapping, the center of the view is the relative *front* of the scene, and the left and right edges of the view are the relative *back* of the scene. 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. Unfortunately, the warped perspective and the fact that the image wraps around can make interpretation difficult. Therefore, an additional more natural mapping is required.

(4) Sphere View The Sphere View maps the immersive video to the inside of a sphere, which can be rotated using the trackball metaphor [18]. 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 participants is facing in the immersive scene. For fulldome videos, this sphere view can also be used to obtain the domemaster mapping, see Fig. 4.

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 [23] 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 [21].

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 degrees, 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 degrees. The display has a native resolution of 1280x800 pixels and a refresh rate of 60 Hz. The field of view is 110 degrees vertically and 86 degrees 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 user study was conducted as follows: First we explained the HMD and the concept of 360° video to the participant. The HMD was then mounted on the participant's head, while still allowing free head movement. After calibrating the eye tracker, different immersive videos were shown to the participant, while recording head orientation and gaze data.

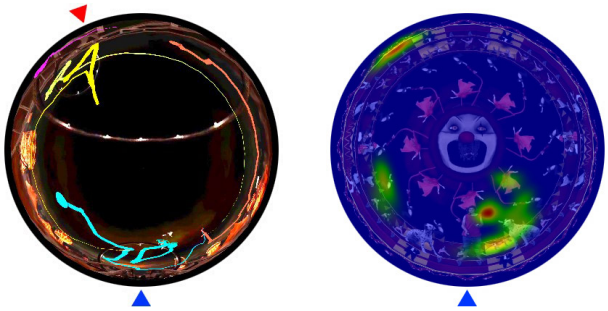

Fig. 4. Two Sphere Views from UM MENINO looking upward into the dome. The artist-indended viewing direction is additionally marked by the blue arrow. On the left: The roller coaster sequence with scan paths. The red arrow marks the travelling direction of the roller coaster. On the right: The kaleidoscopic sequence with a superimposed attention map.

Fig. 5. Scenes from the immersive fulldome short film UM MENINO.

Fig. 7. Scenes from the immersive 360° video clip RAFTING.

4.1 Video: UM MENINO

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

Figure 3 shows a screenshot of our user interface for the initial fairground sequence. In the View Similarity visualization, we observe that during this sequence all participants are individually exploring the immersive scene. Using the Master View and the Participant View we additionally notice that most attention is indeed focused towards the artist-intended viewing direction.

Shortly after the roller coaster sequence begins, the viewing directions form two distinct clusters. In Figure 4 (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 4 (right) shows that during this sequence most viewers have returned to the artist-intended viewing direction, focusing on the protagonist, with a slight tendency to look up.

In this use case, our framework allowed us to identify variations in the viewing behavior of participants. In particular, 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.

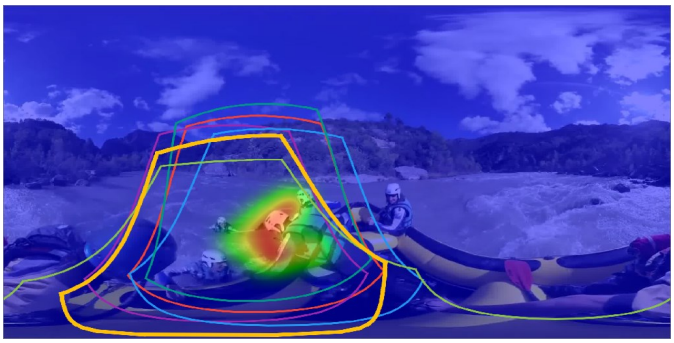

Fig. 6. Master View of a frame from the RAFTING video. All participants are focused on the rescuing effort.

4.2 Video: RAFTING

VIDEO 360: RAFTING [27] is a short immersive video available under the Creative Commons license, see Fig. 7. The clip is 42 seconds long and shows a rafting scene with a static camera centered in the raft. At eleven seconds 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 two getting back into the raft safely.

Figure 1 (right) shows a segment from 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 Figure 1 (left) shows that most attention is focused towards the travelling direction of the raft.

From the moment the raft and the camera tilt, all participants begin searching for what happened, and soon all views converge around the two rafters in the water. Figure 6 shows the Master View during the rescuing efforts, with an overlaid attention map accumulated over all participants. All attention is now focused on the crewmember that is reaching out to help.

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.

5 Conclusion and Future Work

In this paper, we have presented a novel visualization framework for analyzing head movement and gaze data for immersive 360° video. 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 furthermore provided three additional visualizations being appropriate for panoramic video and supporting established gaze evaluation techniques. We evaluated our approach within a small-scale user study including different types of panoramic video, and found that our framework can be used to detect whether the attention guidance of an immersive video works as expected.

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 also like to conduct a larger user study in order to gain further and more statistically significant insight into attentional synchrony for immersive scenes.

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] T. Blascheck, K. Kurzhals, M. Raschke, M. Burch, D. Weiskopf, and T. Ertl. State-of-the-art of visualization for eye tracking data. In *Proceedings of EuroVis*, volume 2014, 2014.

[2] M. Burch, A. Kull, and D. Weiskopf. Aoi rivers for visualizing dynamic eye gaze frequencies. In *Computer Graphics Forum*, volume 32, pages 281–290. Wiley Online Library, 2013.

[3] M. Cheon and J.-S. Lee. Temporal resolution vs. visual saliency in videos: Analysis of gaze patterns and evaluation of saliency models. *Signal Processing: Image Communication*, 2015.

[4] T. Cox and A. Cox. *Multidimensional Scaling, Second Edition*. Taylor & Francis, 2010.

[5] A. T. Duchowski, E. Medlin, N. Cournia, A. Gramopadhye, B. Melloy, and S. Nair. 3d eye movement analysis for vr visual inspection training. In *Proceedings of the 2002 symposium on Eye tracking research & applications*, pages 103–110. ACM, 2002.

[6] A. T. Duchowski, M. M. Price, M. Meyer, and P. Orero. Aggregate gaze visualization with real-time heatmaps. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 13–20. ACM, 2012.

[7] K. Itoh, J. P. Hansen, and F. Nielsen. Cognitive modelling of a ship navigator based on protocol and eye-movement analysis. *Le Travail Humain*, pages 99–127, 1998.

[8] K. Itoh, H. Tanaka, and M. Seki. 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*, volume 44, pages 360–363. SAGE Publications, 2000.

[9] K. Kurzhals and D. Weiskopf. Space-time visual analytics of eye-tracking data for dynamic stimuli. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2129–2138, 2013.

[10] K. Kurzhals and D. Weiskopf. Aoi transition trees. In *Proceedings of the 41st Graphics Interface Conference*, pages 41–48. Canadian Information Processing Society, 2015.

[11] J. F. MACKWORTH and N. Mackworth. Eye fixations recorded on changing visual scenes by the television eye-marker. *JOSA*, 48(7):439–444, 1958.

[12] D. Noton and L. Stark. Scanpaths in eye movements during pattern perception. *Science*, 171(3968):308–311, 1971.

[13] M. Nyström and K. Holmqvist. Effect of compressed offline foveated video on viewing behavior and subjective quality. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 6(1):4, 2010.

[14] F. Perazzi, A. Sorkine-Hornung, H. Zimmer, P. Kaufmann, O. Wang, S. Watson, and M. Gross. Panoramic video from unstructured camera arrays. In *Computer Graphics Forum*, volume 34, pages 57–68. Wiley Online Library, 2015.

[15] T. Pfeiffer. Measuring and visualizing attention in space with 3d attention volumes. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 29–36. ACM, 2012.

[16] R. Ramloll, C. Trepagnier, M. Sebrechts, and J. Beedasy. Gaze data visualization tools: opportunities and challenges. In *Information Visualisation, 2004. IV 2004. Proceedings. Eighth International Conference on*, pages 173–180. IEEE, 2004.

[17] C. Schulz, E. Schneider, L. Fritz, J. Vockeroth, A. Hapfelmeier, T. Brandt, E. Kochs, and G. Schneider. Visual attention of anaesthetists during simulated critical incidents. *British journal of anaesthesia*, 106(6):807–813, 2011.

[18] K. Shoemake. Arcball: a user interface for specifying three-dimensional orientation using a mouse. In *Graphics Interface*, volume 92, pages 151–156, 1992.

[19] T. Smith and J. Henderson. Attentional synchrony in static and dynamic scenes. *Journal of Vision*, 8(6):773–773, 2008.

[20] S. Stellmach, L. Nacke, and R. Dachselt. Advanced gaze visualizations for three-dimensional virtual environments. In *Proceedings of the 2010 symposium on eye-tracking research & Applications*, pages 109–112. ACM, 2010.

[21] M. Stengel, S. Grogorick, L. Rogge, and M. Magnor. A nonobscuring eye tracking solution for wide field-of-view head-mounted displays. In *Eurographics 2014-Posters*, pages 7–8. The Eurographics Association, 2014.

[22] M. Tory, M. S. Atkins, A. E. Kirkpatrick, M. Nicolaou, and G.-Z. Yang. Eyegaze analysis of displays with combined 2d and 3d views. In *Visualization, 2005. VIS 05. IEEE*, pages 519–526. IEEE, 2005.

[23] FOVE: The world's first eye tracking virtual reality headset, 2015. getfove.com, vis. 29 Jul. 2015.

[24] Google Jump, 2015. google.com/cardboard/jump, vis. 29 Jul. 2015.

[25] Oculus Story Studio, 2015. oculus.com/storystudio, vis. 29 Jul. 2015.

[26] Youtube creator blog, 2015. youtubecreator.blogspot.de/2015/03/a-new-way-to-see-and-share-your-world.html, vis. 27 Jul. 2015.

[27] Ábaco Digital Zaragoza, VIDEO 360: RAFTING, 2015. youtube.com/watch?v=h0xo8QEPrk0, vis. 27 Jul. 2015.

[28] N. Weibel, A. Fouse, C. Emmenegger, S. Kimmich, and E. Hutchins. 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*, pages 107–114. ACM, 2012.