



Tracing the Visual Path: Gaze Direction in the 360 Video Experience

Valentina Rozo-Bernal^(✉)  and Pablo Figueroa 

Universidad de los Andes, Bogota, Colombia
v.rozob@uniandes.edu.co

Abstract. In traditional storytelling, a kind of predefined structure typically guides the narrative. However, when we move beyond the constraints of a rectangular screen, new possibilities emerge. 360-degree videos offer a unique opportunity to narrate with every element visible to the user in each frame. Nevertheless, some questions arise: What exactly are the end users seeing? Where is their attention directed? This paper deepens into an analysis of the data captured from the position and rotation of the headset worn by viewers while watching 360-degree videos. To accomplish this, two videos created by university students were examined, where gaze information was captured by testing it with a group of participants. Alongside watching the videos, participants provide their feedback on what elements they focused during the video. The findings revealed that participants tend to focus on elements intended to draw their attention. Furthermore, when the camera was stationary, participants found it easier to explore their surroundings, highlighting the value of utilizing the 360-degree format.

Keywords: 360 Videos · Narratives · Gaze Detection · Visual Attention · Focus · Context

1 Introduction

Currently, there are multiple ways of doing videos thanks to the various formats available. A new alternative is the 360 video format, which allows users to explore their surroundings. This experience is called cinematic Virtual Reality (VR) or live-action VR [8]. The 360-degrees camera model used for this study, is known as spherical or equirectangular and is based in the Structured for Motion (SfM) photogrammetry workflow [2]. This means that the workflow of 360-degree videos is similar to the composition of 3D images through the collection of 2D images. In the next figure it is shown the 360-degree video pipeline.

In the Fig. 1 it is displayed the 360-degree video pipeline, and Table 1 provides detailed explanations for each term depicted in the diagram to enhance comprehension.

To provide a clearer understanding of the 360-degree video pipeline in Table 1, are the explanations for each term described in the diagram:

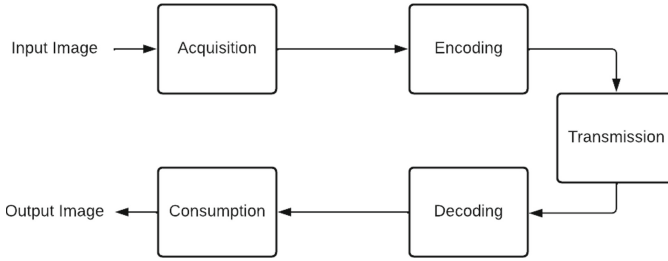


Fig. 1. 360-degree video pipeline

Table 1. Overview of 360-Degree Video Steps

Name	Description
Acquisition	To capture the image through the optical system. This step involves stitching the image. However this step may introduce some distortions within the visual content [1]
Encoding	Reduce redundancy in the signal. Use encoding tools as H.264, H.265, VP9, AVI, depending on the camera [1]
Transmission	It is the process of delivering the video content from the source to the end-user's device [1]
Decoding	Display the geometry and unpack the data into the end-user device [1]
Consumption	Is the final step where the end-user can watch the 360-degree video in a VR environment [1]

In addition to the technical specifications, to create videos in this format there is still room for improvement. For instance, when designing the story, from the script to the storyboard there should exist tools to pre-visualize it in VR, making it so that creators can define the context and focus on the story they are trying to tell.

However, there is not a clear guide to follow when creating content for this format. This raises questions about how to effectively highlight focus and context within the content. So to find some light in this topic, and provide insights into creating content for the 360-degree format, an analysis of gaze direction is conducted.

To achieve this, an app captures gaze direction data while users watched two 360-degree videos. Then the data is extracted and processed, to assess viewers head rotations. The aim of this analysis is to determine if people are utilizing the full 360-degree field of view. This paper is structured as follows. First, there is a brief literature review on how the narratives are applied to 360-degree videos and how this works for traditional storytelling. Then, the methodology to conduct the investigation is explained. Finally, the results and conclusions of the study are presented.

2 Literature Review

There are some examples of studies based on 360 videos in which attention is evaluated, such as, an application designed to minimize distraction among viewers, this investigation contributed to the optimization of adaptive video streaming [13]. Other similar study, is one, which aimed to capture viewing information data to identify popular areas in a 360-degree video, as the researchers believed that users should be guided to achieve a good level of immersion and to reduce the risk of viewers getting lost within the storyline. The objective was to provide insights into guiding viewers and enhancing immersion in 360-degree videos. [9]. Other example, is an investigation in which the researchers extracted a dataset of gaze trajectories from 360-degree videos, this study aims to provide guidance for working with this format and develop a model for predicting eye direction in such videos. One of the findings revealed that people usually have the same view patters, key element to make possible the prediction of gaze direction in 360-degree videos [3].

Finally, talking about capturing attention, there is an investigation that addresses the challenges of transmitting 360-degree videos, and how the attention is influenced by video quality. The study highlights the importance of not only narrative elements but also technical quality in creating a positive viewer experience. Additionally, surprise elements during the transmission were found to be effective in capturing attention, in this way narrative and technical aspects need to be considered when creating a 360-degree video [7].

2.1 Tools in 360-degree Videos

As there are studies to identify and reduce distraction in 360 videos, there are also tools that help to create and adequate narrative to show what is worth watching in this format. For example, there is a Storyboard app to plan and collaborate with other creatives in the process of making videos. This storyboard is presented in a paper in which the researchers propose a workflow and introduce a multi-device storyboard tool to work in virtual reality as well as in traditional formats during the pre-production phase of film-making [4]. Other tool in 360-degree videos was built for evaluating different head predictors in the context of streaming. The primary focus of the study was to compare the performance of various head predictors for 360-degree videos, so it would be easier to pin focus on a given video in this kind of format [11].

As there are tools built for 360-degree videos and investigations which evaluate attention and distraction, there are also, some studies in which the tools are the key to evaluate the areas of interest within such videos. One of this tools is 360RAT a software tool designed to identify areas of interest in this video format. The researchers conducted a experiment in which they observed that individuals tend to focus and highlight regions with minimal object amount in the scene [10].

2.2 Traditional Storytelling

In traditional format videos, there is a structured process that creators can usually follow, there are also different shoots and camera angles, there is also a narrative structure to tell what is off-screen and on-screen. However, in 360-degrees videos following these “instructions” it is quite difficult, first because it is impossible to take some things off screen, there is always something happening, something that the viewer might see when they turn around. Nevertheless, there is something that can be followed and is the narrative structure, always ensuring that the user does not lose track of the storyline. For this purpose there are some guides like the book *The Writer’s Journey: Mythic Structure for Writers* in which it is presented a way to structure a story for the characters as well as for the lineal structure once posed by Aristotle [12].

3 Methodology

To investigate the direction of viewers attention, A specific set of steps were followed. First, two 360-degree videos created by a group of students were carefully selected for analysis. The criteria for video selection prioritized those with a clear main plot and rich contextual elements that allowed users to observe the plot from various angles in every frame.

Subsequently, gaze information was collected from a group of students while they watched the selected videos. Additionally, short surveys were conducted with the test viewers to gather further insights. The main goal of this analysis was to determine how frequently the viewers changed their rotation and position during the video, thus revealing patterns of their attention and engagement.

3.1 360 Video Player: A Tool to Capture Gaze Direction

To gather the information about gaze direction from the participants, a 360-degree video player for Oculus Quest 2 was built, it is noted that this headset is the one used for this experiment. To capture the gaze direction, it is used the rotation of the headset in the X, Y and Z axis. The video player was developed in Unity, with a user interface (UI) where eleven 360-degree videos where displayed, as it is shown in Fig. 2, however participants where only allowed to watch two videos during the experiment: one in which there was a stationary camera and other in which the camera was moving in a particular direction.

During the experiment, participants select a video. Once the video start, they have limited control and couldn’t fast forward or rewind it. When they finish, they return to the main screen, however as a mechanism to exit the video if participants present motion sickness they may press any button in the controllers to safely return to the initial interface. Data capture begin as soon as the participant start watching the video, the position and rotation are recorded each second to finally be saved in a CSV file. These files are stored with the key name of the video and a random number in order to prevent overwriting of the file. Finally, once the data is captured, the files are stored in the main application files for future analysis.

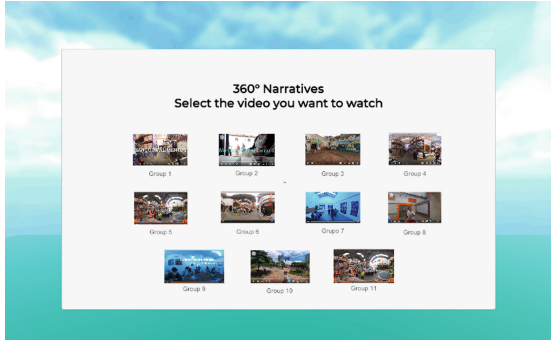


Fig. 2. 360 Video Player

3.2 Cameras for Capturing 360-degrees Videos

To capture the images, the creators used the Insta360 One X2. This camera is designed to simplify the process of recording videos in the 360-degree format. It has a maximum 5.7k recording resolution and is equipped with movement stabilizer, along with a dual lens. It is important to note that each lens of this camera has a 200-degree field of view. This feature is intended to create a seamless effect where the tripod or selfie stick becomes invisible in the captured images. This effect is achieved by capturing more than 360-degrees of the scene, causing the images to overlap and effectively erase the stick [5].

The Insta360 One X2 offers users the flexibility to record in various formats. However, the chosen mode for this project was the 360-degree mode, with this mode, users can capture video or images and then use the Insta360 App to select their preferred angles and easily edit the video to create final product. Users can also use other editing software, such as Adobe Premiere Pro. In this software, there is a plugin called GoPro GX ReFrame, designed to assist in a process known as reframing. Reframing let users to select the desired viewing angle of the video, transforming it into a flat plane. When editing, users have the possibility to edit based on a specific angle or in the equirectangular view [6]. For this experiment, it is essential to mention that the videos will be exclusively analyzed within a software designed for a virtual reality environment.

3.3 Information from the Creators

The two videos described by it creators follow these main plots:

- Video 1: The video of Fig. 3 show the reality of many people and how they access to benefits from an institution that works to end hunger.
- Video 2: The video of Fig. 4 makes the viewer feel the experience of volunteering at the institution and what this involves.

Taking this into account, the first video emphasizes in the people that access to food, while viewers listen to some interviews about how the beneficiaries of



Fig. 3. Four moments of the first video



Fig. 4. Four moments of the second video

the program access to the resources. This video is narrated from a static camera while people start appearing in the shoot. Thus, the viewer can turn their head 360-degrees to see what is happening all the time. As for the second video, it portrays a day in the life of a volunteer. The narrative is presented in first person to show the viewer as the volunteer. The 360-video makes a tour through the facilities to learn every step they must follow to participate in this initiative. As a result, this video intend to make the viewer turn in every direction as if they were in their first day as volunteer in real life.

3.4 Information from the Viewers

To gather the information from the viewers, there was a test group of 32 students, ages form 19 to 23. The method of the experiment it is as follows; the group was divided in two, the first group, watched the video number one and then the video number two, while the second group watched them in opposite order. Then, they were asked to fill a short survey in which the focus and context

of the video were asked. Due motion sickness the data of the 32 participants wasn't collected. As a result, a total of 26 CSV files were extracted for each video, but not all files contained the complete information, the average watched time was 220s. Despite incomplete data, analysis was conducted on at least the first minute of each video, as this was the minimum duration observed by the participants before experiencing motion sickness.

In order to collect the data an application in Unity for the Oculus Quest 2 was developed, in which the participants could select the video. While they were in this process, every second the position and rotation of the head was captured. It should be noted that, out of the eleven available videos, only two were selected, due the correct use of the 360-degree format, factors such as the position of the camera and the use of additional resources such as voice-overs and on-camera interviews. Finally, the data was extracted from the headsets to continue with the study.

4 Results

As for the results, once the data was analyzed, it was obtained that people tend to look up and down in different positions while watching the first video. In the next plots, for the first video, it is shown that users usually rotate their heads in every direction. The plots from Fig. 5, 6, 7 and Fig. 8 correspond to the data gathered in the CSV files for rotation in x and y axes for video number one and video number two.

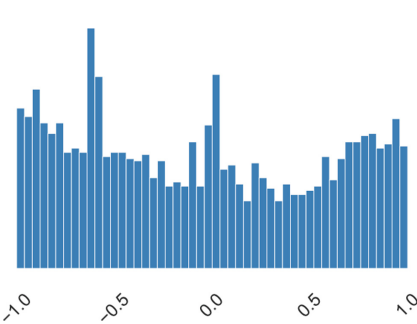


Fig. 5. Rotation in the y-axis for the first video: frequency vs rotation

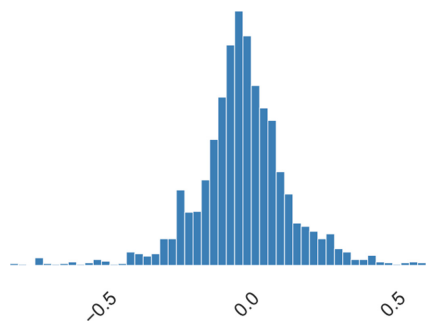


Fig. 6. Rotation in the x-axis for the first video: frequency vs rotation

On the other hand, the analysis of the same variables in the second video reveals that fewer users tend to look right and left, the most of them remained in the same position in the y axis. In the Fig. 8 there is a smaller range of rotation in x. However, the difference in this axis between the two videos is not significant, as indicated by the similar standard deviation values. For video 1 is 0.1574 and

for video 2 is 0.1597 showing that in this axis participants do not explore the most up and down. The major difference is in the y-axis where the standard deviation for the first video is 0.75 while for the second is 0.61.

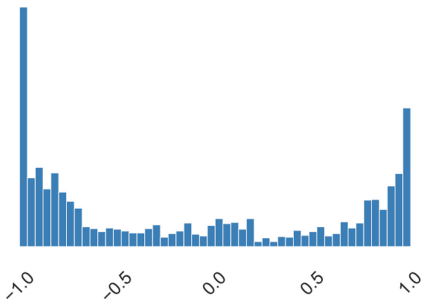


Fig. 7. Rotation in the y-axis for the second video: frequency vs rotation

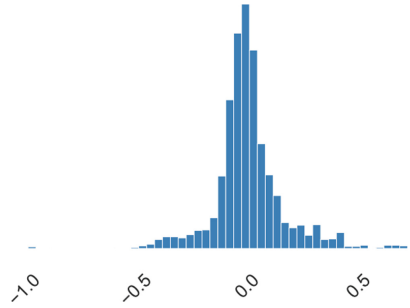


Fig. 8. Rotation in the x-axis for the second video: frequency vs rotation

The findings indicate that complete control over user attention may not let take the full advantages of utilizing a 360-degree video format. However, when there are elements that control the narrative without being constant and allow viewers explore the scene, there is major sense of enjoyment and better comprehension of the narrative.

In order to compare the rotation in both videos, the plots of the rotation in the x and y axes are displayed below. For better comprehension of the data presented, five CSV files were randomly selected out of the 26 available. In Fig. 9 it can be observed that users tend to look around more frequently in the first video compared to the second video. In the video number two participants tend to stay in the same position for longer duration resulting in minimal rotation.

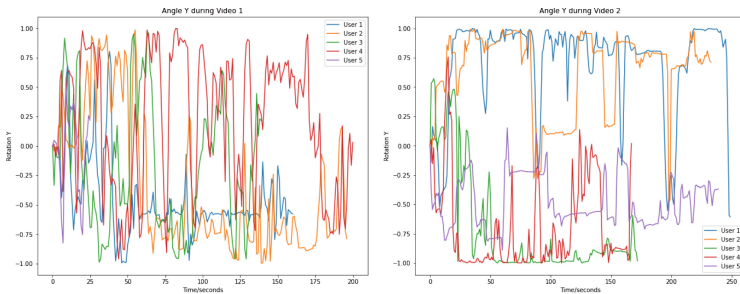


Fig. 9. Rotation in the y-axis for both videos: rotation vs time

In the case of the x-axis in the Fig. 10, it can be observed that participants in the first video exhibit a little higher degree of freedom in their gaze. The

plot shows a bit more rotation in this axis, indicating they looked up and down frequently compared to the second video. So, in the second video the rotation angle is smaller than in the first video, suggesting, participants explore a little less vertically.

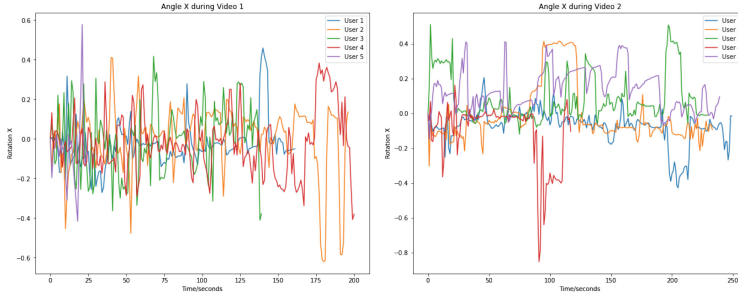


Fig. 10. Rotation in the x-axis for both videos: rotation vs time

5 Conclusions

The observed differences in the previous results can be attributed to the distinct narrative styles employed in the two videos. While the first one is open to explore the scene, the second one is directing the gaze of the viewer to an specific point. Other details that highlight from this is that the only thing trying to catch the attention of the participant in the first video are the audios of interviews, while in the second one has a person in the scene who is talking, so people tend to direct their attention to them. On the other hand, based on the responses from the participants, the surveys indicate that the first video possessed a well-defined storyline, which was crucial for understanding the context. In contrast, feedback from the second video suggests that although context played a significant role, it did not compel the participants to explore their surroundings.

In conclusion, depending on the intention of the creators it is worth the use of a 360-degrees format. The findings from the analyzed data indicate that the participants usually tend to concentrate on what the creator intends them to focus on. In other words, the user feel free to take a look around, but if there is something catching their gaze is easier to just follow that point. Nevertheless, if creators let viewers take a look, giving other elements to focus on as the sound, this could enrich the experience and the use of this format.

For future work, it is recommended that guidelines be developed for using focus and context effectively in a given storyline for this particular format. For the researcher conducting this study, it is worth the use of 360-degree videos if the user is not conditioned to always see the same focus point. While it is essential to guide the viewer's attention, it is not obligatory to do it throughout

the entire video. There are some helpers that can prevent the viewer from getting distracted, like sounds or moments of silence and stillness where the viewer is invited to explore the scene.

References

1. De A. Azevedo, R.G., Birkbeck, N., Simone, F.D., Janatra, I., Adsumilli, B., Frossard, P.: Visual distortions in 360-degree videos. *IEEE Trans. Circuits Syst. Video Technol.* (2019). <https://doi.org/10.48550/arXiv.1901.01848>
2. Barazzetti, L., Previtali, M., Roncoroni, F.: 3D modeling with 5k 360° videos. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* (2022). <https://doi.org/10.5194/isprs-archives-XLVI-2-W1-2022-65-2022>
3. Duanmu, F., Mao, Y., Liu, S., Srinivasan, S., Wang, Y.: A subjective study of viewer navigation behaviors when watching 360-degree videos on computers. In: 2018 IEEE International Conference on Multimedia and Expo (ICME) (2018). <https://doi.org/10.1109/ICME.2018.8486537>
4. Henrikson, R., Araujo, B.D., Chevalier, F., Singh, K., Balakrishnan, R.: Multi-device storyboards for cinematic narratives in VR. In: *UIST '16: Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (2016). <https://doi.org/10.1145/2984511.2984539>
5. Insta360: How to use the invisible selfie stick. <https://blog.insta360.com/how-to-use-the-invisible-selfie-stick>
6. Insta360: Reframe plugin tutorial. <https://n9.cl/reframe>
7. van Kasteren, A., Brunnström, K., Hedlund, J., Snijders, C.: Quality of experience of 360 video – subjective and eye-tracking assessment of encoding and freezing distortions. *Multimed. Tools Appl.* **81**(7), 9771–9802 (2022). <https://doi.org/10.1007/s11042-022-12065-1>
8. Knorr, S., Ozcinar, C., Fearghail, C.O., Smolic, A.: Director’s cut - a combined dataset for visual attention analysis in cinematic VR content. In: *CVMP '18: Proceedings of the 15th ACM SIGGRAPH European Conference on Visual Media Production* (2018). <https://doi.org/10.1145/3278471.3278472>
9. Mäkelä, V., et al.: What are others looking at? exploring 360° videos on hmds with visual cues about other viewers. In: *TVX '19: Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video* (2019). <https://doi.org/10.1145/3317697.3323351>
10. Prado, M., et al.: 360RAT: a tool for annotating regions of interest in 360-degree videos. In: *WebMedia '22: Proceedings of the Brazilian Symposium on Multimedia and the Web* (2022). <https://doi.org/10.1145/3539637.3557930>
11. Rondón, M.F.R., Sassatelli, L., Aparicio-Pardo, R., Precioso, F.: A unified evaluation framework for head motion prediction methods in 360° videos. In: *MMSys '20: Proceedings of the 11th ACM Multimedia Systems Conference* (2020). <https://doi.org/10.1145/3339825.3394934>
12. Vogler, C.: *The Writer’s Journey: Mythic Structure for Writers* (1992)
13. Xue, T., Ali, A.E., Viola, I., Ding, G., Cesar, P.: Designing real-time, continuous QoE score acquisition techniques for hmd-based 360° VR video watching. In: *14th International Conference on Quality of Multimedia Experience (QoMEX)* (2022). <https://doi.org/10.1109/QoMEX55416.2022.9900914>