ANALYSIS OF INTENDED VIEWING AREA VS ESTIMATED SALIENCY ON NARRATIVE PLOT STRUCTURES IN VR FILM

Colm O Fearghail[1], *Sebastian Knorr*[1,2] *and Aljosa Smolic*[1]

[1]V-SENSE, School of Computer Science and Statistics, Trinity College Dublin, Ireland.
[2] Communication Systems Group, Technical University of Berlin, Germany.

ABSTRACT

In cinematic virtual reality film one of the primary challenges from a storytelling perceptive is that of leading the attention of the viewers to ensure that the narrative is understood as desired. Methods from traditional cinema have been applied to varying levels of success. This paper explores the use of a saliency convolutional neural network model and measures it's results against the intending viewing area as denoted by the creators and the ground truth as to where the viewers actually looked. This information could then be used to further increase the effectiveness of a director's ability to focus attention in cinematic VR.

Index Terms**—** Storytelling, Virtual Reality, Saliency, Visual Attention.

1. INTRODUCTION

Cinematic virtual reality (VR) film is one of the best developed formats for consumers of virtual reality entertainment. It does still have problems in the area of visual storytelling when compared to the ability of traditional cinematic films to engage an audience [8]. With the ability of the viewers to turn their attention to any area of the 360 degree environment, the content creators need to consider the visual elements of what is being presented. This is crucial in order to relay the narrative to the viewer in a manner that is engaging and to create an immersive environment [18].

The director's intended viewing of a scene is the area where the narrative is taking place. This has been explored by the development of datasets that give the intended viewing direction, the so called Director's Cut [17], as denoted by the content creators alongside the ground truth of viewers, and compared how successful the intended viewing was in commanding the viewers attention [17]. While reasons for why certain directing techniques can be inferred [9], it would also be of use for feedback to occur while there might still be time for the creators to adjust the films to ensure that among the most salient areas, are areas that they have intended the viewers to watch.

One way in which the visual nature of an environment can be evaluated is through the use of saliency applications [7]. Saliency is the term given to the area of computer vision that creates computer models that are inspired by psychological theories of the human perception system [24]. Currently the best performing models are neural networks. One way of describing the process of how perception occurs is by there being a top down process and a bottom up process. Top down is a task driven process that is goal driven and happens after the bottom up process, which is the initial process that occurs and operates on the physical properties of the visual scene [16].

Saliency in 360 degree environments has also been explored, and one such model is SalNet360 which extended convolutional neural networks for saliency prediction on 2D images to onmidirectional images (ODIs) [19]. By using such a model on imagery from cinematic virtual reality films an idea as to how they will perform for a viewer can be reached. An important part of the cinematic process is that of the test screening [11]. This is when a film is shown to a group of people before its general release. The feedback given by this selected audience from the screening can have a big impact on the final form of the film. The nature of cinematic VR makes these test screenings more difficult to perform and an indication as to where the viewers might look is of paramount importance. Providing the saliency information of the imagery within the film would be one way to gain an insight.

By understanding visually the salient elements of the scene as it would be presented to a viewer, the director could then be in the position to either add visual elements or adjust the visual properties in order to create a more salient area in the direction that he wishes for the viewer to look in [12].

This paper explores the ability that a saliency prediction would give a creator as to how effective his intending viewing of a scene would work. This would alert the creator to areas of the film where the intending viewing direction might not be followed, something that is highly important at times that a plot point crucial to the narrative is occurring.

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under the Grant Number 15/RP/2776.

2. RELATED WORK

Narrative theory has been applied to a wide range of different media formats [14]. Virtual reality is one of these formats and the application as to how narrative can be applied to it has been an area of research [1]. Due to unique properties not found in that of screen based media, it has been argued that virtual reality should be considered a specific narrative medium. The process of creating a visual narrative is also an area that has been explored across the use of a wide range of different formats, where a sequence of images are arranged in order for a story to be told. This can been seen from cave paintings to modern day comic books [5]. The principles of how this visual narration can engage a viewer in virtual reality and the potential the medium has to enhance the experience for the viewer has been looked at in [4]. In traditional film, various methods have been developed over its history to relate the narrative to the audience in a variety of visual techniques, including the style of editing and the technique such as montage and visual association [6]. These techniques and the stylistic development of such, have resulted in particular styles becoming more dominant [2], such as a Hollywood style seen in blockbusters.

Continuity editing is one these techniques which has a defined set of rules, the cognitive foundation that this rules operate on has been explored in [26]. Furthermore, research has been performed on the behaviour of visual attention in 360 degree video in order to gain a better understanding as to how viewers interact with imagery in such a space [22]. Also, the use of guidance in a 360 degree environment and the differences between this and traditional film has been an area of research, in particular the manner in which visual cues could form the basis of a new grammar for storytelling within the medium of virtual reality [23].

In [20], new forms of visual cues to act as this guidance could be applied to a narrative in a 360 degree environment have been explored alongside different methods of its implementation and their effectiveness. These included diegetic and non diegetic cues that were further defined by being explicit or implicit. Implicit cues where being more contingent on the bottom up nature of saliency in the imagery. This application of visual guidance in 360 degree video has also been applied to a number of different devices, e.g. in [27], where it was found that the 'object to follow' method performed the best.

The amount of research in the area has also been collected in light of the challenges involved in [25], which offers a comprehensive breakdown of the methods and devices used and the environment that they were tested in. This area of research has roots in the study of a viewer in relation to traditional film, where the manner in which the narrative affects the gaze and the comprehension of a viewer in traditional film has also been measured by the collection of gaze data [15].

A database for the understanding of viewer attention in relation to the intended viewing by the creators of 360 degree content has been collected alongside the plot points that consist in the narrative [17]. Additionally, the authors have explored the visual narrative techniques in relation to storytelling within this database in [9]. Finally, in [10], they have also researched the use of various styles of transitions in the 360 degree videos and their correlation with viewer attention in this database.

Saliency has been an area of active research in the computer vision community. A number of different approaches have been applied with varying levels of success [3]. More recently with the development of omnidirectional images, the area has been expanded in light of this new challenge [7]. Finally, machine learning has had a large impact in the area and has been part of the most successful models presently. In [19], the expansion of these models into omnidirectional images, which takes into account the spherical coordinates of the pixel in the image and the manner that the central prior becomes a horizontal one has been developed and implemented in an end to end manner.

3. METHODOLOGY

3.1. Director's cut

The dataset published in [17] [1] was used, which contains several films across a number of different content types. The intending viewing as indicated by the director of the film was included alongside the actual viewing direction of 20 participants (16 males and four females).

In order to record the viewing direction of each participants the test-bed in [21] was utilized. The intended viewing direction from each of the films directors was collected, using the *Tracker* node, within The Foundry's commercial compositing software *Nuke*[2]. Further details on the process can be found in [17].

Additional information about the films was also included as part of the dataset in the form of plot points, where the position within the film was given alongside the frame range in which the plot point took place. These were times within the films that the directors wished to have the viewers look in a particular place in order that a crucial event in the plot of the film would be understood.

3.2. SalNet360

SalNet360 [19] presented an architectural extension to convolutional neural networks (CNN) for traditional 2D images to more accurately predict the saliency of ODIs by addressing the biases that are inherent to the format.

This included subdividing the ODI into undistorted patches and providing spherical coordinates to the CNN for each pixel

[1] `https://v-sense.scss.tcd.ie/?p=2477`
[2] https://www.foundry.com/products/nuke

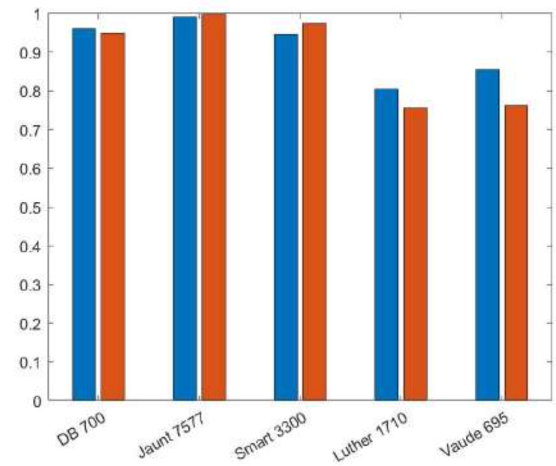


Fig. 1: AUC for each of the *best* performing plot points, where the viewers are *in blue* and the Director's Cut is in *red.*

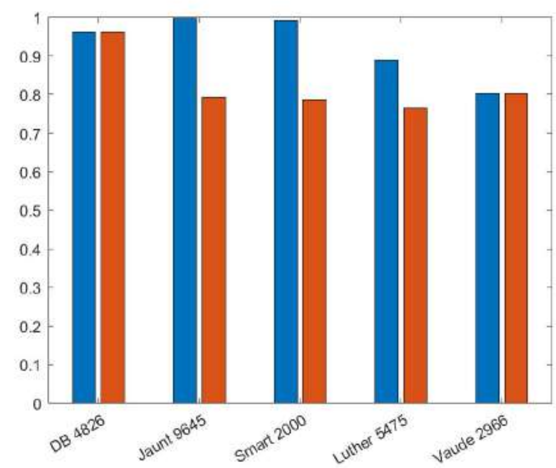


Fig. 2: AUC for each of the *worst* performing plot points, where the viewers are in *blue* and the Director's Cut is in *red.*

in those patches.

This allowed the heavy distortions to the image that result from its mapping to a sphere to be catered for, and also allowed for the position of objects within the ODI to be considered by using the spherical coordinates of the pixels in the patches created.

The patches were created from the ODI by rendered frustums with a field of view of 90 degrees. This allowed each patch to be of an equal size and also corresponds approximately to the field of view of an Oculus Rift head mounted display.

SalNet360 was presented at the Salient360! challenge that was organised as part of the ICME 2017 conference.

3.3. Saliency maps on plot points

From the films containing plot points in the director's cut dataset, frames were collected from the best and the worst performing plot points in relation to the viewers attention and the intended viewing as denoted by the creators of the videos. Alongside the frame range of the plot points the creators also explained the device at that time that was used in order to attract the attention of the viewer. The frame used in order to gain a probability map using SalNet 360 [19], was a frame that was within the frame range of the plot point when the attracting device the director was using was present within it. Plot points that fell within title screens where removed for the purpose of analysing the saliency of visual storytelling techniques within the films. This meant that plot point 1 within the film *Luther* was dropped and that plot point 5, which was the second worst performing plot point was used instead.

The AUC (area under the receiver operating characteristic curve) was gathered from the best performing plot points and also the worst performing plot points when measured against the saliency result. Using information provided by the Director and the points collected from where the viewers actually looked, the AUC (area under the receiver operating characteristic curve) was used to compare how well the viewers and the director agreed with the probability map computed by SalNet360 which are shown in Fig. 1 and Fig. 2. The results for the AUC were collected by using the code published alongside [13]. As can be seen in Fig. 1 the directors made good use of the salient points within the films in other to steer the attention of the viewers, in two cases getting a better result than the viewers. In the plot points that did not do so well in Fig. 2, the viewers were noticeably closer to the saliency results in three of the plot points.

4. ANALYSIS AND DISCUSSION

For five of the films within the director's cut database, the creator included the intended viewing direction and several plot points, which were needed to be viewed for the narrative to be understood. In this section, the ability of the director to command the attention of the viewers is compared to the results from the convolutional neural network during the plot points where the director was most successful and least successful in drawing the attention of the viewers to the intended viewing direction.

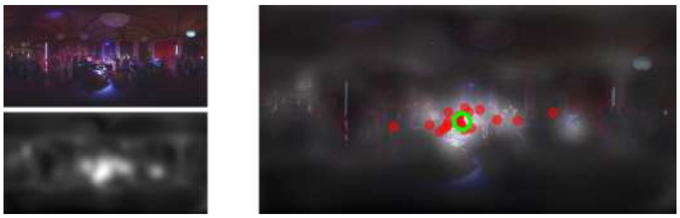
4.1. Saliency at best performing plot points

The following are the scenes in which the director was most successful in commanding the attention of the viewers.

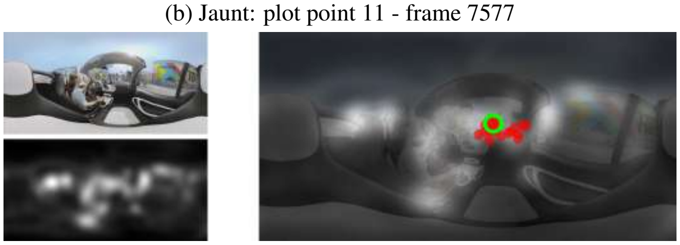
4.1.1. DB

From the frame at the best performing plot point for *DB* (see Fig 3a), we can see the one of the most salient features is centered around the area of the intended viewing position. While there are a number of other features that also create salient points, the attracting device used at this point is that of a top

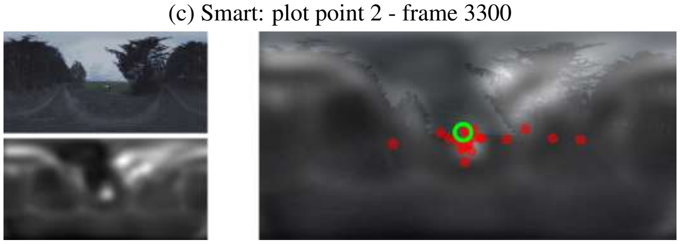
(a) DB: plot point 1 - frame 700



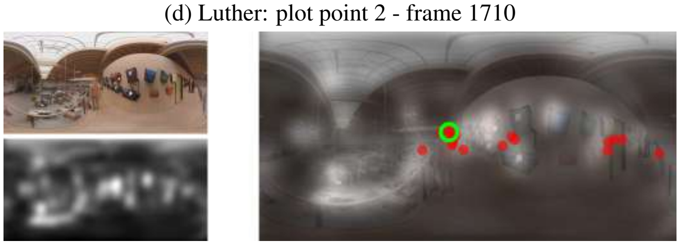
(b) Jaunt: plot point 11 - frame 7577



(c) Smart: plot point 2 - frame 3300



(d) Luther: plot point 2 - frame 1710



(e) Vaude: plot point 1 - frame 695

Fig. 3: *Best* performing plot points. Left: RGB image and saliency map, right: RGB image with overlaid saliency map, director's cut (green circle) and user tracks (red circle).

down feature which takes the appearance of the principle actor talking directly to the camera position. The position of a strong light to the rear of the actor also helps this area become one of the most salient areas. From the AUC score in Fig. 1, the director had a high score close to that of the viewers meaning that the intended viewing direction made good use of the salient regions within the frame.

4.1.2. *Jaunt*

The best performing plot point in the *Jaunt* film (see Fig 3b) occurs towards the latter half of the film. There were a number of additional cues used at this point, however, it is also clear from the saliency map that the most salient areas of the scenes are gathered together in the area that is the region of intended viewing for this point in the film. Lighting was used in a manner to attract the viewers by being more brightly lit in this area than in others. From Fig. 1, the director was closer to the salient regions than the viewers in this case also, which gives a good indication as to the effective use of salient regions within the frame.

4.1.3. *Smart*

Multiple points of saliency are seen in the plot point that performed most highly in the *Smart* film (see Fig 3c). The position of the driver is one of these alongside the area that is outside the windscreen of the car, and the movement of the car in the scene is towards the front. This area in front of the car was that of the intended viewing position for the scene. The multiple members of a band in the scene gathered in this area were the device used in order to attract the attention of the viewer. Top down features at this point included the driver's reaction to the crowd that had gathered. However, as motion cues are not included in the saliency map, the influence of the salient area for attracting the viewers attention compared to the motion cue might not be as strong as the saliency map indicates. The results in Fig. 1 show that the director predicted which regions would be the more salient ones and made use of them in forming the intended viewing direction.

4.1.4. *Luther*

The plot point at this stage in the film had the animated character of Luther in a dark field (see Fig 3d). One of the more salient areas in the scene is directly located upon this character, this also forms the intended viewing direction of the scene. The edge of the trees against the brighter background is also shown as a salient region, which helps to frame the area in which the character is present, the character also being a top down feature. This use of bottom up saliency areas to re-enforce that of a top down feature would appear to help orientate the viewer to look in the intended area. In this case the viewers were closer than the director to the salient regions as can be seen in Fig. 1.

4.1.5. *Vaude*

By looking at the saliency maps generated from the best performing plot point in the film *Vaude* in Figure 3e, it is clear that there are multiple points of saliency across the frame, similar to *DB*. The intended direction as noted by the director is also shown as a salient point. From the frame itself there are properties that contribute to this saliency result. The lighting in the frame is even and the setting of a factory interior gives plenty of features outside of the intended viewing that add salient points to the scene. The fact that this was one of the best performing plot points in the film shows the ability of a top down feature, namely an actor talking directly to the camera to override the bottom up features, which consist of salient points across the frame. This top down feature becomes the most effective in terms of attracting the viewer's attention by being an essential part of the stories narrative. From Fig. 1 we can see that the viewers were closer to the salient regions of the frame that the intended viewing direction as desired by the director.

4.1.6. *Overall features present at best performing plot points*

From the saliency maps generated from the best performing plot points described above there are a number of salient features that seem to help contribute to their performance. One of these features is lighting, the lighting in most of the frames was conductive to drawing the eye to the area of intended viewing. The other feature is the use of top down features being present in areas that are also salient in a bottom up manner. While the most effective plot points in each film were selected, the plot points in *Jaunt* and *DB* were the most effective overall in attracting the viewer's attention. The saliency map generated for the frames in these plot points shows that the salient areas within these frames are in the intending viewing areas. Also these salient areas had less competition from other areas in the scene. The AUC scores across the plot points in the films, Fig. 1, do show that the directors in this case made good use of the salient regions within the frame while plot points crucial to the narrative were occurring. In two cases the directors scored more highly that the viewers as can be as can be seen in the scores *Jaunt* and *Smart*.

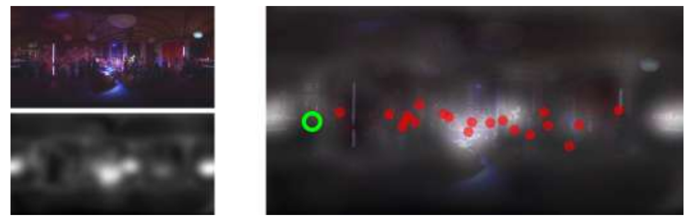
4.2. Saliency at worst performing plot points

The following are the scenes in which the director was least successful in commanding the attention of the viewers.

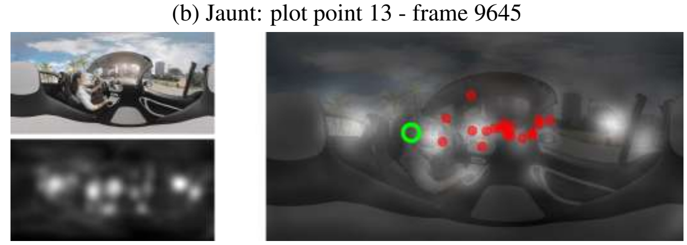
4.2.1. *DB*

The saliency map generated from one of the frames of the worst performing plot points in the *DB* film shows multiple salient points (see Fig 4a). One of the stronger areas of these salient features was at the intended viewing of the scene, which at this point of the film was the infographic which is present to the left of the frame. The other features that appear salient within the frame are more diffused and are across the horizontal axis of the scene. These features combine to give a less clear area of salient points than that was present in the best performing plot point. From Fig. 2 we can see that the director and the viewers were very close in the AUC score for the frame.

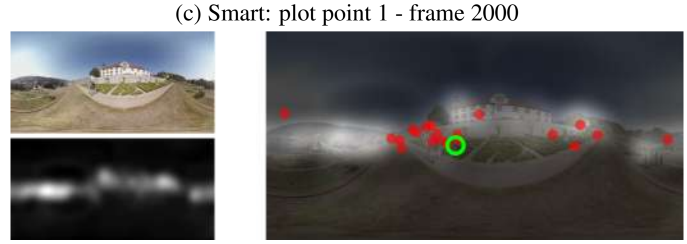
(a) DB: plot point 5 - frame 4826



(b) Jaunt: plot point 13 - frame 9645



(c) Smart: plot point 1 - frame 2000



(d) Luther: plot point 5 - 5475

(e) Vaude: plot point 3 - frame 2966

Fig. 4: *Worst* performing plot points. Left: RGB image and saliency map, right: RGB image with overlaid saliency map, director's cut (green circle) and user tracks (red circle).

4.2.2. *Jaunt*

The worst performing plot point in the the *Jaunt* film (see Fig 4b) is constructed much in the same manner as the best performing plot point was, however with the addition of extra graphical elements that appear to the far left and right of the frame. This was a device to ensure that the information was related to the viewer. While this could be an effective way to ensure that the viewer sees it, it also means that the director has less control as to which area actually does attract the viewers attention, which could form part of the reason for the low performing score. The viewers were far closer to the salient regions within this frame than the area that the director wanted to direct their attention to as can be seen in Fig. 2. The manner in which the graphics were displayed could have been part of the reason for this result, as each graphic formed a separate salient region.

4.2.3. Smart

The worst performing plot point in the *Smart* video (see Fig 4c) takes place at the point in which there are multiple people outside the car using signs that point towards a 'portal', which is a device used to transport the car to a new area. The saliency map generated at this point shows how the salient points within the frame are spread out in a larger area than was seen in the best performing plot point. These multiple salient points are spread out across the horizontal axis. Here, the strong motion cue of the car also seems to have a larger influence than that of the top down feature, i.e. the face of the actor, which was the intended viewing area. From Fig. 2 the viewers here were again closer to the salient regions of the frame that the intended viewing that the director had wanted them to observe.

4.2.4. Luther

The plot point that performed most poorly within the *Luther* film is that of two walkers as they traverse across the frame (see Fig 4d). One of the more salient areas is just to the left of these two walkers and the salient points within the frame then extend right across the horizontal axis of the frame in a more diffused manner. In the frame the viewers were closer to the regions that were salient than the area that the director wanted them to look at as can be seen from Fig. 2.

4.2.5. Vaude

From the worst performing plot point in *Vaude*, we can see from the saliency map in Fig 4e that there are two competing areas of strong saliency that are combined with other areas of more diffused salient points to the left of the frame. It should also be noted that at this point in the film, the camera was positioned on a moving bicycle and the movement also contributed to the poor result. The intended viewing position for this plot point was that of a person wearing a panda suit, and the presence of such was also further emphasized by the one of the cyclists waving in that direction. One area of salient points in the frame was this cyclist and the position of other salient features was spread out across the horizontal axis. From Fig. 2, we can see that the director and the viewers were close in the AUC score for this frame, when measured against the saliency map.

4.2.6. Overall features present at worst performing plot points

From the saliency maps generated at the worst performing plot points in the films there are a number of similarities between them. The salient points occur across a wider area that can be found in the best performing plot points. These areas are also found across the horizontal region of the frame. Except for the *Smart* video, there is also not a strong or dominant top down feature within them to attract and command the viewers attention. From the worst performing plot points the two that had the worst overall performance in comparison to the others, were the plot point in *Jaunt* and the plot point in *Vaude*. From Fig. 2 we can see that the viewers had a closer score to the salient regions of the scene as opposed to the area that the director wished them to view, as can be seen in the results for *Jaunt*, *Smart* and *Luther*.

4.3. Saliency as an indication of storytelling intent

From the salient points located within the maps generated, it is clear that the various devices used by the directors of the films were considered to be areas that are salient. The area of these salient points are also grouped closer in the better performing plot points than in those that performed less well. This goes some way to show that the storytelling intent within the frames of the plot point areas can be considered salient, which is the result of the director's planning and the film making techniques used in the production of the films. This intent from the films creators is also the reasoning behind the idea of a director's cut within cinematic virtual reality. From the AUC results in Fig. 2 and Fig. 1 it can be seen that in the better performing plot points the director had a closer score to the areas considered salient than the viewers in two of the films but that in the worst performing plot points the viewers had a higher score than that directors intended viewing direction. This does give a good indication that the viewers are influenced by regions of considered salient by the network in the films. It also suggests that these areas can offer a good ability to guide the viewer if this areas are also the areas that the director intended for the viewer to observe.

4.4. Use for saliency in post-production

One area of application would be in the post-production environment where the frame could be adjusted in regards to the salient properties within it. This could include reducing the competition of multiple salient points by looking at the features that create them, namely intensity and colour. Additional elements could also be added to the scene such as graphics to further increase the saliency of the area that the director wishes to be viewed. Fig 5 shows a mockup of a possible application of saliency in a post-production environment like The Foundry's Nuke.

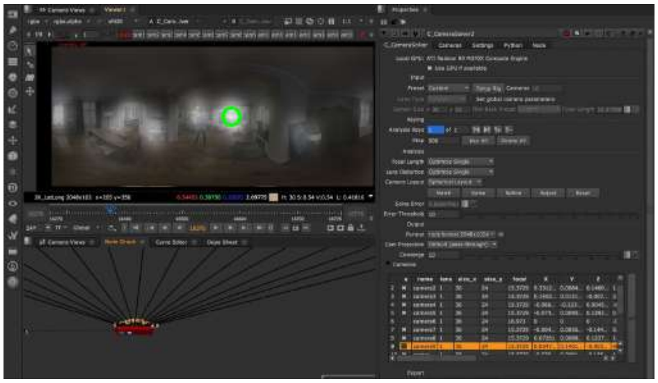


Fig. 5: Possible application of saliency in a post-production environment, namely The Foundrys commercial com-positing software Nuke, the RGB image with overlaid saliency map, director's cut (green circle)

5. CONCLUSION

Directing the attention of a viewer in cinematic VR is one of the most crucial tasks in order for the director to convey the story that they wished to tell. From the results shown, a saliency model can detect regions that will draw the viewers eye. If the intended area that the director wanted to guide the attention of the viewer to is outside these regions, there is less chance that they will successfully do so. There generally is a high agreement between saliency and viewers proving the claim that saliency can be used to estimate the behaviour of audiences. The director's cut or where the viewers where intended by the director to look, did not agree as highly as the audience at certain points, in particular during the plot points in the worst performing scenes, which suggests that a greater awareness of the saliency could have been utilitized at these points for the director to increase the probability of the viewers to look in the intended viewing area.

This means that understanding the salient elements of the frame can help to improve upon the ability to guide the viewer's attention in a manner that is implicit and non diegetic, meaning that it would be not tied to elements within the scene. This would give a creators in the medium the ability for feedback and a greater degree of confidence in their ability to ensure that the viewers follow the story as they intended them to.

Further work will concentrate on improving the saliency results by incorporating sound and movement elements within the plot point areas.

6. REFERENCES

[1] Ruth Aylett and Sandy Louchart. Towards a narrative theory of virtual reality. *Virtual Reality*, 7(1):2–9, 2003.

[2] David Bordwell. *The way Hollywood tells it: Story and style in modern movies*. Univ of California Press, 2006.

[3] Ali Borji and Laurent Itti. State-of-the-art in visual attention modeling. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):185–207, 2012.

[4] Fiona Carroll, Michael Smyth, and Linda Dryden. Visual-narrative and virtual reality. In *Book of Selected Readings: The International Association Of Visual Literacy (IVLA 2004), Jahoannesburg, South Africa*. Citeseer, 2004.

[5] Neil Cohn. Visual narrative structure. *Cognitive science*, 37(3):413–452, 2013.

[6] Ken Dancyger. *The technique of film and video editing: history, theory, and practice*. Focal Press, 2014.

[7] Ana De Abreu, Cagri Ozcinar, and Aljosa Smolic. Look around you: Saliency maps for omnidirectional images in VR applications. In *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE, 2017.

[8] Kath Dooley. Storytelling with virtual reality in 360-degrees: a new screen grammar. *Studies in Australasian Cinema*, 11(3):161–171, 2017.

[9] Colm O Fearghail, Cagri Ozcinar, Sebastian Knorr, and Aljosa Smolic. Director's cut - analysis of aspects of interactive storytelling for VR films. In *International Conference on Interactive Digital Storytelling*, pages 308–322. Springer, 2018.

[10] Colm O Fearghail, Cagri Ozcinar, Sebastian Knorr, and Aljosa Smolic. Director's cut - analysis of VR film cuts for interactive storytelling. In *2018 International Conference on 3D Immersion (IC3D)*, pages 1–8. IEEE, 2018.

[11] Catherine Griff. Film audience testing in australia: Capturing the audience before it bites. *Studies in Australasian Cinema*, 6(2):159–174, 2012.

[12] Aiko Hagiwara, Akihiro Sugimoto, and Kazuhiko Kawamoto. Saliency-based image editing for guiding visual attention. In *Proceedings of the 1st international workshop on pervasive eye tracking & mobile eye-based interaction*, pages 43–48. ACM, 2011.

[13] Jonathan Harel, Christof Koch, and Pietro Perona. Graph-based visual saliency. In *Advances in neural information processing systems*, pages 545–552, 2007.

[14] David Herman, JAHN Manfred, and RYAN Marie-Laure. *Routledge encyclopedia of narrative theory*. Routledge, 2010.

[15] John P Hutson, Tim J Smith, Joseph P Magliano, and Lester C Loschky. What is the role of the film viewer?

the effects of narrative comprehension and viewing task on gaze control in film. *Cognitive Research: Principles and Implications*, 2(1):46, 2017.

[16] Laurent Itti. *Models of bottom-up and top-down visual attention*. PhD thesis, California Institute of Technology, 2000.

[17] Sebastian Knorr, Cagri Ozcinar, Colm O Fearghail, and Aljosa Smolic. Director's cut: A combined dataset for visual attention analysis in cinematic VR content. In *Proceedings of the 15th ACM SIGGRAPH European Conference on Visual Media Production*, page 3. ACM, 2018.

[18] John Mateer. Directing for cinematic virtual reality: how the traditional film directors craft applies to immersive environments and notions of presence. *Journal of Media Practice*, 18(1):14–25, 2017.

[19] Rafael Monroy, Sebastian Lutz, Tejo Chalasani, and Aljosa Smolic. SalNet360: Saliency maps for omni-directional images with CNN. *Signal Processing: Image Communication*, 69:26–34, 2018.

[20] Lasse T Nielsen, Matias B Møller, Sune D Hartmeyer, Troels Ljung, Niels C Nilsson, Rolf Nordahl, and Stefania Serafin. 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*, pages 229–232. ACM, 2016.

[21] C. Ozcinar and A. Smolic. Visual attention in omnidirectional video for virtual reality applications. In *10th International Conference on Quality of Multimedia Experience (QoMEX)*, Sardinia, Italy, May 2018.

[22] Cagri Ozcinar and Aljosa Smolic. Visual attention in omnidirectional video for virtual reality applications. In *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE, 2018.

[23] Jayesh S Pillai, Azif Ismail, and Herold P Charles. Grammar of VR storytelling: Visual cues. In *Proceedings of the Virtual Reality International Conference-Laval Virtual 2017*, page 7. ACM, 2017.

[24] Nicolas Riche, Matthieu Duvinage, Matei Mancas, Bernard Gosselin, and Thierry Dutoit. Saliency and human fixations: State-of-the-art and study of comparison metrics. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1153–1160, 2013.

[25] Sylvia Rothe, Daniel Buschek, and Heinrich Hußmann. Guidance in cinematic virtual reality-taxonomy, research status and challenges. *Multimodal Technologies and Interaction*, 3(1):19, 2019.

[26] Tim J Smith. The attentional theory of cinematic continuity. *Projections*, 6(1):1–27, 2012.

[27] Marco Speicher, Christoph Rosenberg, Donald Degraen, Florian Daiber, and Antonio Krúger. Exploring visual guidance in 360-degree videos. In *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video*, pages 1–12. ACM, 2019.