DIRECTOR'S CUT - ANALYSIS OF VR FILM CUTS FOR INTERACTIVE STORYTELLING

Colm O Fearghail[1], Cagri Ozcinar[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

The usage of film cuts, or transitions, is a powerful technique in interactive storytelling to express the film story by leading the viewers attention. To explore how existing transition techniques are currently being used by professional 360° filmmakers, this paper investigates the impact of transitions and additional graphical elements from a storytelling perspective. We base this on the recently published Director's Cut dataset [1] which contains professional 360° films with the directors' intending viewing direction (*i.e.,* director's cut) and test subjects' scan-paths. Our objective is to examine widely used transition techniques in professional 360° film, and with our finding guide filmmakers in the storytelling and editing process. To the authors knowledge, this is the first study to analyze professionally prepared VR film cuts concerning the distance of viewers' scan-paths and director's cut. We observed that the intended viewing direction is required for the viewer to understand the story best and not miss the director's plot points. The transition is a point where the viewers are presented with a new scene and are required to orientate themselves within it. Thus if there is a considerable distance mismatch between the intended and actual viewing, the viewer is not in the best position to understand the scene properly. Our results show that the use of simple graphics can serve as a reference for viewers as the transition happens and they are presented with a new immersive environment.

Index Terms— 360° film, storytelling, director's cut, virtual reality, transitions

1. INTRODUCTION

In line with recent technical developments in immersive video technologies, 360° virtual reality (VR) video is an increasingly important format for content creators. The omnidirectional representation and the interactive consumptive nature of 360° video for VR are transforming the storytelling experiences. To this end, well-known continuity editing techniques in cinema, wherein the film is cut from one scene to another to tell the story conclusively [2], namely *film cuts* or transitions, may require comprehensive scientific analysis as well as further improvements to its adaptation to enhance its ability to create immersive experiences in VR film.

Visual narrative in cinematic VR, however, is still undergoing a stage of experimentation as filmmakers and their audience explore new aspects of the medium. The continuity editing techniques which were developed at the beginning of the traditional cinema have been used as the basis for narative developments of cinematic VR film. While there are aspects of cinematic visual language that can be applied to VR film, the research into the cognitive aspects of what works in traditional cinema can offer a perspective to cinematic VR and support VR video technological developments.

To understand how viewers consume a given VR film, we recently introduced the Director's Cut dataset [1] by analyzing the visual attention of cinematic VR videos [1] and aspects of interactive storytelling [3]. The dataset has eight professional 360° video films, five of which contain different film transition techniques to tell their story effectively. The dataset has recorded participants' scan-paths for each film in conducted subjective experiments, and the director's cut, which is the intended viewing direction of the filmmakers, collected from professional VR filmmakers.

Given the rapid technological development in film industry, it is crucial that filmmakers in the medium understand how the use of the transition techniques in VR video affects the viewers' ability to follow and immerse themselves in a given narrative. To this end, this paper investigates the visual attention of viewers across transitions within professionally produced 360° films by taking advantage of the Director's Cut dataset [1]. To the authors knowledge, this is the *first study* to analyze professionally prepared VR film cuts with respect to viewers' scan-paths and director's cut. We analyze various cuts to understand effective transitions and graphical elements in VR narratives. Among our findings are that, while continuity editing can effectively transport a viewer across a transition if the viewer is lead to the intended viewing position, it does not offer an ability to create a cut that is 'invisible' or not explicitly noticed, to the same extent as traditional cinema.

2. BACKGROUND

There are a wide number of visual styles of transitions in traditional cinema which allow the story to move from one scene to another. The most popular of these is a *straight cut* or when the video immediately moves to the next scene. Another type

is a *dissolve* or where one scene slowly fades into the next. A third type is when an element from the previous scene is present as the film moves to the next one. Also, a *wipe* describes a transition where one scene is replaced by other by moving across the frame, this movement can be directional or in the form of shapes. Furthermore, *dip-to-black* is a transition where the scene fades to black between the two scenes, and *dip-to-white* is another transition that occurs in the same way but the scene fades to white instead.

In traditional cinema, these cuts can be ordered into three levels as seen in [4]: first order in which small displacements of the camera or small changes in image size happen; second order when there is a reversal of the camera position, resulting in a change to left/right and/or a change in background; third order cuts which create a linear sequence of actions in the narrative story. The authors' findings were that second and third order cuts cause an increase of eye movements from 200 to 400 ms, which is not obtained after first order cuts.

This differs in $360°$ films as a method to create a first order cut has not been developed to date. As seen in the films in the dataset, the current cuts happen to move the viewer between locations, which results in a entirely new background meaning that they are closer to the nature of second order cuts. The increase found in eye movements would suggest a greater level of cognitive activity to process these and that these order of cuts are not as 'invisible' as first order cuts.

3. RELATED WORKS

There have been several research works carried out on traditional cinematic editing and the cognitive processes that allow viewers to create a mental scene to comprehend the visual sequence as they view it on a screen. Hasson *et al.*, for instance, investigated the difference between natural and directed scenes to create suspense [5]. Other research works focused on the ability of continuity editing to create a continuous stream of action from separate clips [6], investigating the difference in visual processing based on the type of cut [4], and analyzing the perceived continuity of film using participants' eye tracking data [7].

To maintain a consistent representation of an ongoing event in the face of dramatic low-level visual discontinuity, editing of film pushes the viewer to engage a set of higher order visually driven processes. For instance, Magliano *et al.* [8] argued that the viewers expect continuity of events in a film, and continuity editing works be substantiating these expectations. Schwan and Ildirar [9] suggested a substantial overlap between cognitive processes involved in interpreting real-world conditions and those involved in interpreting cinematic conditions by measuring how adult viewers of a film for the first time interpret its visual narrative. Valuch *et al.* [10] also investigated memory-guided attention during edited guided scenes. Their findings showed that cuts serve as tool to effectively transport the narrative and that conceptual cognitive

processing is critical to this process.

The presence of a familiar line of action is also essential for the interpretation processes of inexperienced viewers. Different ranges of actions help them overcome perceptual discontinuities and this existence of a familiar line of events is what establishes semantic coherence across the shots. For example, the research row in [11] investigated the first-time viewers comprehension of transitions. One reason for the difference in film and viewing a natural scene was the suggestion that viewers switched from resource consuming spatial updating that is typical for natural perception, to a simplified heuristic, which is based on the specific visual characteristics implied by the 180-degree rule. In terms of understanding of a given film, highly experienced viewers resolved the inconsistency by constructing a narrative, and resolved inconsistency by "looking ahead" in an ongoing story by constructing predictive inferences.

However, none of the existing works investigate the impact of VR film cuts with respect to users' viewport trajectories and director's cut, which is the main contribution of our work.

4. ANALYSIS OF VR FILM CUTS

4.1. Dataset

To analyze the effect of various cuts in VR film, we utilized our recently published dataset [1] [1], which contains various levels of film cuts. The dataset also contains an additional information in the form of plot points and directional cues used for user guidance. The dataset has a wide range of content types including documentary, advertisement, tourism, and education. Each video in this dataset has different film cuts. Table 1 describes the characteristics of the film cuts available in the $360°$ narrative videos in the dataset. Evaluations are carried out on five $360°$ narrative videos, as illustrated in Fig. 1: *Luther*, *DB*, *War*, *360Partnership*, and *Vaude*.

Each VR film is available along with director's cut and participants' scan-paths. In total, 20 participants (16 males and four females) participated in conducted subjective experiments. The test-bed in [12] was utilized to record each participant's scan-path using the Oculus Rift Consumer version. The Director's cut was collected from professional filmmakers with The Foundry's commercial compositing software *Nuke* [2] using the *Tracker* node. More details about the process can be found in [1].

4.2. Procedure

To objectively analyze each VR film cut with respect to director's cut and 20 scan-paths, we calculated the Euclidean

---

| (a) *Luther* | (b) *DB* | (c) *War* | (d) *360Partnership* | (e) *Vaude* |

Fig. 1: Sample frames from the five 360° videos used for the experiment.

	Video				
Cut type	*Luther*	*DB*	*War*	*360partnership*	*Vaude*
Dip-to-white	$\{c_4, c_{10}, c_{26}\}$	$\{c_8\}$	–	–	$\{c_2, c_6 \ldots c_{10}\}$
Dip-to-black	$\{c_6\}$	–	–	$\{c_5, c_8, \ldots c_{11}, c_{13}\}$	–
Dissolve	$\{c_3, c_5, c_8, c_9, c_{12} \ldots c_{25}, c_{27}, c_{28}, c_{29}\}$	–	$\{c_2\}$	$\{c_1\}$	$\{c_3, c_4\}$
Wipe	–	–	–	$\{c_2, c_3, c_4, c_6, c_7, c_{12}, c_{14} \ldots c_{16}\}$	–
Straight	–	$\{c_2 \ldots c_7\}$	$\{c_1, c_3\}$	–	–
Others	$\{c_1, c_2, c_7, c_{11}\}$	$\{c_1\}$	$\{c_4\}$	$\{c_1, c_{17}\}$	$\{c_1, c_5\}$

Table 1: Description of film cuts c_k in the five videos from the Directors' Cut dataset [1].

distance (L^2) between director's cut and the median location of the 20 scan-paths for each frame f as follows:

$$Dist_f = ||D_f - T_f||, \qquad (1)$$

where D_f represents the Cartesian coordinates of the director's cut at the f-th frame, and T_f denotes the median value of the scan-paths in terms of Cartesian coordinate at the f-th frame.

The mismatch between the director's cut and the median of the scan-paths is also evaluated according to *before* and *after* film cuts. For this purpose, we define two windows, w_1 and w_2, and apply the distance metric in Eq. (1) for each window. Here, w_1 and w_2 are represented as before and after cut periods, respectively. Distance values for each defined period (window) are estimated as follows:

$$w_1 = Dist(i_s - (t_1 + t_2) : i_s - t_1) \qquad (2)$$

and

$$w_2 = Dist(i_f + t_1 : i_f + (t_1 + t_2)) \qquad (3)$$

where i_s and i_f are the starting and finishing frame numbers for a given cut, respectively. Threshold values for windows are denoted as t_1 and t_2, which were selected empirically as constant values (t_1=50 and t_2=500 frames) for analysis purposes. This 'window' can be set to the scene between the cuts or do other instances that are wished to measure. As seen in Fig. 2 'c' is used to denote a cut, while 'g' is used to show when a graphical element was used. To compare the distribution of the estimated distance data, we utilize the empirical cumulative distribution function (ECDF) as statistical visualization method for each window $w = \{w_1, w_2\}$. The ECDF shows the fraction of a given w less than or equal to a particular value of the estimated distance, $Dist$. ECDF's x- and y-axes are calculated as follows:

$$x = w_l \quad \text{and} \quad y = l/N, \quad \forall l : w_{l-1} \leq w_l \leq w_{l+1} \qquad (4)$$

where N is the total number of data points in a given w.

4.3. Individual evaluation of films

To analyze each film, we estimated the distance between the director's cut and the median of scan-paths using Eq. (1), and plot the result for each film in Fig. 2.

4.3.1. Luther

Looking at the Fig. 2 (a), we see that *Luther* video clip has the most amount of cuts out of the films. The cuts took the form of dissolves in-between the scenes, with three of them using dip-to-white at $\{c_4, c_{10}, c_{26}\}$, and a dip-to-black at c_6. A number of these dissolves also had a graphical element across them; this occurred at $\{c_5, c_7, c_{11}, c_{16}\}$, where the Playmobil character of the Luther was present, at c_{20} and c_{21} the graphical element across the cuts is that of two walkers.

c_1 begins the video with a fade up into the intro scene where the viewers are introduced to the character of Luther that will be one of the methods used to direct their attention. At g_1, the title appears in place of Luther, and at c_2, the first scene fades up behind these before the title disappears at g_2. Looking at Fig. 2 (a), we notice that the director was successfully able to guide the viewers' attention during this sequence.

There is a dissolve at c_3 which takes the video from Luther on a tower overlooking the landscape to being on a paddle boat. The position of Luther matches that of a sailing boat in the scene that the video dissolves into. At c_4, there is a dip-to-white, the style of which is like a strong flare of sunlight, which takes the viewer to a room interior where the animated character of Luther is present. The first transition can be observed at c_5, wherein the animated character is present across the dissolve. This dissolve is forewarned by imposing an element of the new scene, the altar of the church interior, first and then later a wipe to fill the screen. From the Fig. 2 (a),
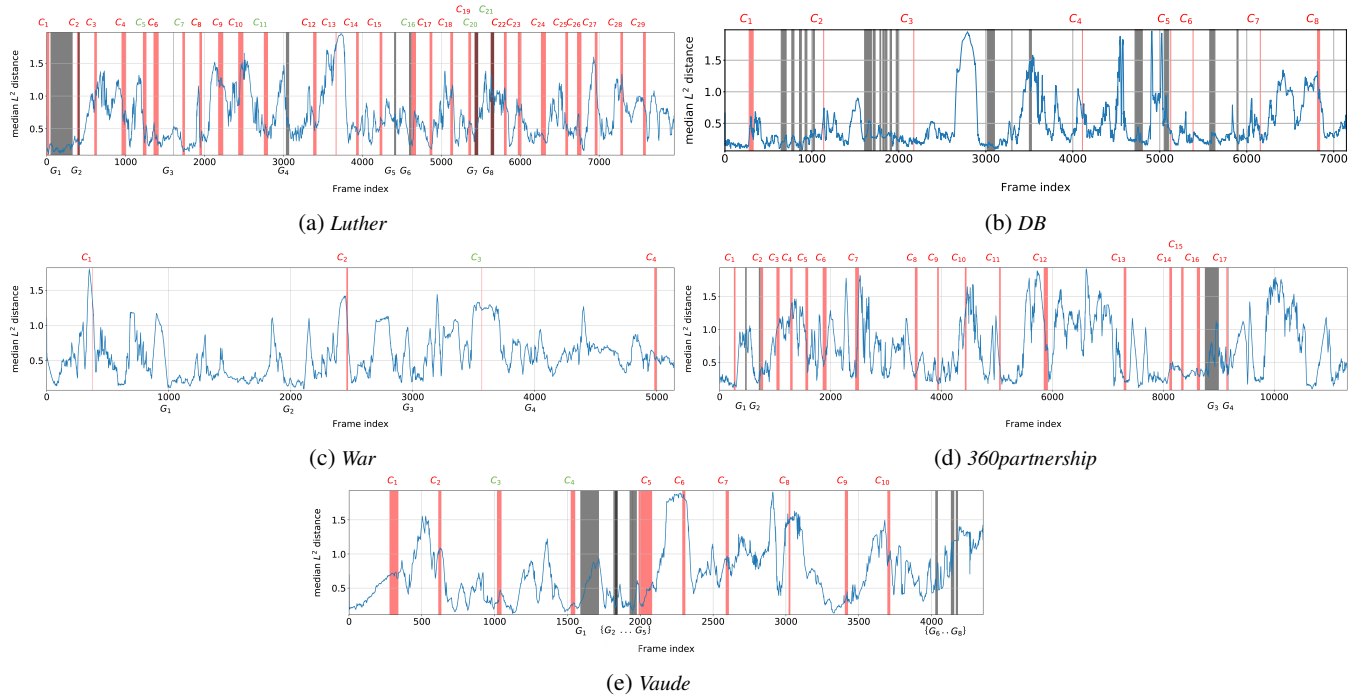
Fig. 2: Euclidean distance between director's cut and median of users' scan-paths 'C' are cuts, 'G' are graphical elements, cuts with graphical elements are colored green.

it would appear that this method of transition was effective at gaining the attention of the viewer and increasing their viewing position to the intended one of the director.

A dip-to-black transition is available at c_6 before revealing a dark field with the Luther character present in it. The placement of the Luther character in this new scene is just right of where he was present in the previous one. After this transition, there is a lightning strike at g_3, which would appear to have brought the attention of the viewer back towards the intended view of the animated character. In c_7, the use of the animated character is present across the transition. The first scene is the character in a dark field the transition then happens in the form of a dissolve and in the second scene, that of a street exterior, the character is partially obscured by a door which soon opens. The position of the character does not change across the dissolve. This process of directing the attention of the viewer first towards the animated character and then using this as a reference across the cut, as observed in Fig. 2 (a), gave one of the best performing scores for the distance between the directors' intended viewing and the viewers' median value.

At c_8, this scene dissolves into the exterior of a church. The position of Luther in the previous scene matches that of a tourist taking a photo. There is another dissolve at c_9 to a scene of people relaxing by a river. This time there are not any elements that match positionally, which may explain the increase in the distance from the intended viewing. A dip-to-white room can be seen at c_{10} that contains the animated character, which serves as a reference to bring the viewers back to

the intended viewing direction. Looking at the Fig. 2 (a), this scene at c_{11} has a lower distance value as the cut happened as opposed to before or after the cut happened. This can be explained as the animated character was present in the scene, and the new scene (the one that the viewer will transition to) appears first as an insert, in the form of a screen that contained an element of the forthcoming scene, which later expands by use of a wipe to cover the screen as the animated figure walks towards it. This allowed for the viewers to use the animated character as a visual reference as the scene changed and also forewarned the viewer that a transition was about to happen. The church in the scene is empty at the beginning, but, a choir along with their conductor and audience is faded in at g_4. This appears to have refocused the viewers back closer to the intention of the director.

Another dissolve can be observed at c_{12}. This dissolve is appeared into a street at night where the camera moves along with a rickshaw that travels along it. This movement may have contributed to the increase of mismatch with the intended viewing as the scene continued, along with the dark nature of the environment. At c_{13}, this scene is dissolved into another street where a firework show is taking place. While after the transition mismatch increased further, the use of fireworks appeared to be salient enough to attract the viewers back closer again to the intended viewing as the scene continued. Fig. 3 shows how the scene was able to bring the viewers' attention back to the intended position before the scene transitioned again. By limiting the amount of areas of visual interest in the scene c_{13} dissolved into, the distance to the in-
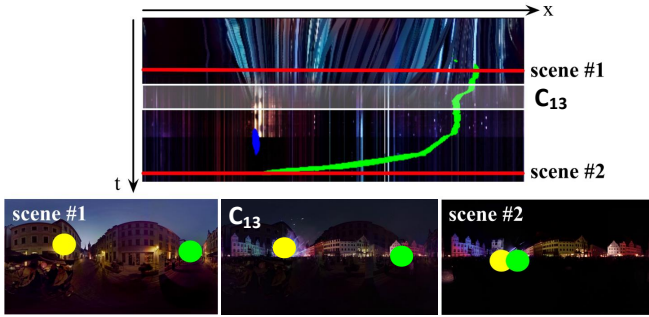
Fig. 3: An impact of transitions: Top; an equiangular image blended from the frames illustrated underneath. Director's cut (yellow) and the median of users' scan-paths (green) for *Luther*.

tended viewing position decreased, allowing for an effective transition to happen then at c_{14}.

The transition at c_{14} changes the scene to that of Luther outside a building at night, c_{15} sees this scene change again to the same location during the daytime. Graphical elements are then faded into the scene at g_5, one to Luther immediate left and the others further to the right, the statues to the right are directly faced by the character of Luther. These are then faded out at g_6.

A dissolve transition and the character of Luther is present across it at c_{16}. The scene changes to a location closer to one of the buildings in the background of the previous one. This building is matched in its position across the two scenes. The use of two points of reference across this transition could be a cause of the increasing of mismatch alongside additional graphical elements that were imposed on the scene, resulting in a greater competition between these aspects. The result of this transition can be seen in Fig. 5. Looking at the figure, for example, we see that 75% of scan-path data has 1.2 and 0.4 distance values before and after c_{17}, respectively. After this, the next transition, c_{17}, takes the video to a scenic viewpoint with a building in front. This then at c_{18} dissolves into an interior scene where the animated character of Luther sits in front of a desk.

At c_{19} there is a dissolve where two walkers made their way from left to right across the video as two dissolves transported them across locations. It can be seen from the graph that the first transition at c_{20} was more successful than the latter at c_{21}, the result for this could be that once the viewer had noticed the walkers and their trajectory that they became less important than exploring the scene.

At c_{22} the scene dissolved into that of an exterior of a building with two women in period dress, talking together. This scene then dissolves at c_{23} into the interior of the building where a winding staircase is present in the center of the scene. The same two women then descend around a corner of the staircase. This scene then dissolves into an interior room with Luther present in it at c_{24}. The position of Luther and

the two women match across the dissolve.

The dissolve at c_{25} takes a location where two hikers are making their way up steps alongside a waterfall. At c_{26} then a dip to white occurs and in the next scene, people are enjoying the views of the countryside from a hot air balloon. The position of these elements matches on either side of the cut. The mismatch increases during this scene as it is explored but the dissolve at c_{27} where the position of the hot air balloon matches with a band performing on stage draws the viewer's attention back to the intended position. This concert is then dissolved at c_{28} to a location overlooking a fun fair, the position of the stage from the previous scene matches the location of the fun fairs placement. At c_{29} there is a dissolve into the end logo.

4.3.2. DB

The transitions in the *DB* clip were straight cuts aside from the initial fade up and at the very end a fade-to-white. Graphical elements were used in order to direct attention and also to position a viewer in the right direction before a transition happened. After the fade up into the scene at c_1, the first of these graphics appears at g_1 and goes through a number of animations at $\{g_2, g_3, g_4\}$ before fading at g_5. From the graph, we can see that the use of graphics in this manner does appear to occupy the visual attention of the viewers and keeps them close to the intended viewing of the scene.

At c_2 the actor in the scene was addressing the camera directly, and then after the cut the next scene had a group of football players celebrating after a match in the same position. At the time of the next cut, the direction from the previous scene again got the viewer into position to best understand the next scene that was presented to the viewer. At g_6 a graphic again appears in the scene, this time there are animations within it at $\{g_7, g_8, g_9, g_{10}, g_{11}\}$ before fading at g_{12}. Again from looking at the graph, this imposition of a graphic and a sequence of animations on it does seem to command the attention of viewers. This scene then ends at c_3 with the actors in the scene bunched together in the middle of the screen, the next scene is that of an empty apartment, two actors enter it from the same position that the actors in the previous scene occupied.

Another graphic appears in the scene at g_{13} which takes the from of a video call. At g_{14} the actor walks around the room, which would require the viewer to make a complete turn, which is then populated by furniture by fading in graphics. This would appear not to have been effective in commanding the viewer's attention from the graph. At g_{15} The graphic of the video call fades out of the scene.

At c_4 the scene cuts from the apartment to an exterior of a garden with a family present in it. The shots were matched in position by the use of the same actor on either side of the cut. The use of this positional matches can be seen on the viewers' behavior in Fig. 4. Here, the distance values of w_1

(before c_4) and w_2 (after c_4) are 0.53 and 0.28 for 75% of scan-path data. At g_{16} another graphic appears close to where the family has gathered in the scene. This stays present in the scene until g_{17} where it fades away. There is a high mismatch in the intending viewing and actual viewing as this graphic is present, at this stage the action in the scene is taking place in a position that would be to the rear of the initial starting position of the video i.e. the viewer would have to look behind them. c_5 cuts from this garden exterior to return to the actor and scene the video had previous cut from at c_2, the use of a graphical element attracted the attention of the viewer, and then by use of movement of the actors present led the viewer into position for the cut. The actor within it again addresses the camera directly.

At c_6 the use of the actor talking directly to the camera was again used to command the attention of the viewer before the cut happened, the position where the actor was being the position that a meeting is taking place in the second scene. At g_{18} a graphic is imposed in the position that the actors are located. It animates until it fades at g_{19}. As the graphic remains in the scene the actors in it conclude their meeting.

The combination of actors and the direction of their movement was again the method used at c_7 to effectively transport the viewer into the next scene while facing the most salient position. The video finally fades to white at c_8.

4.3.3. War

War had a number of transitions with different styles. c_1 was a straight cut, c_2 was a dip to black. c_3 was also a straight cut but this time a bright red flare was used to direct the attention of the viewer upwards, the intention is that the viewer would miss the cut as it happened as the viewer would be visually arrested elsewhere.

At c_1 the first scene shows a landscape depicting World War 1. There are three bright areas in this scene and from the graph, it is clear that the visual attention of viewers differed from the intended view. The scene this is cut to is that of two soldiers having a conversation in a trench. The intended viewing at this point is that of the talking soldier and from the graph we can see that this brought the viewers' attention back to the intent of the director.

At c_2 the cut happened when one soldier was talking. The transition used was a dip to black and in the next scene the director used darkness to block much of the scene leaving the brightest areas around the two soldiers. The effectiveness of this in the graph is clear, where the distance between the intended viewing direction and the direction that the video was viewed dropped from just over 1.4 to just under 0.2.

The effect of the flare at c_3 in the video across the cut is not extremely clear from the graph. Previous to the flare there was a bright blue light and at the time of the flare a whistle was blown to denote the charge from the trenches, all of which could have acted as competitors for the viewers' attention.
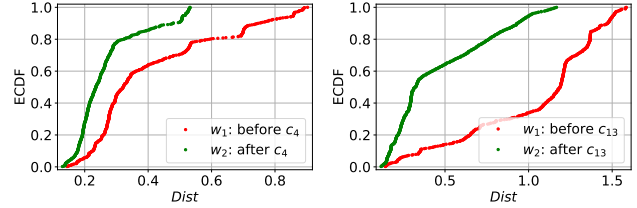

Fig. 4: ECDF graphs for the *DB* clip (left) at c_4 and the *360partnership* clip (right) at c_{13}.
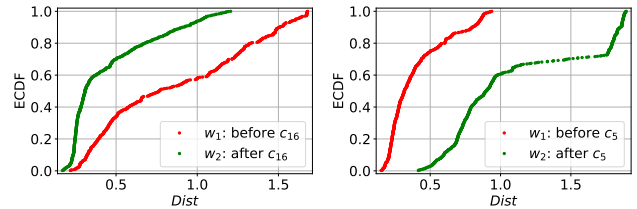

Fig. 5: ECDF graphs for the *Luther* clip (left) at c_{16} and the *Vaude* clip (right) at c_5.

4.3.4. 360partnership

This video contains a number of transitions, directional wipes coming from the left side of the screen to the right were used at $\{c_2, c_3, c_4, c_6, c_7, c_{12}, c_{14}, c_{15}, c_{16}\}$ and also dip to black at $\{c_5, c_8, c_9, c_{10}, c_{11}, c_{13}, c_{17}\}$. Shot in a documentary style this video was the most encouraging for a directional point of view to allow the viewer to freely explore the environment. At g_1 a graphical element, in the form of text, was imposed on the video animated into the video and at g_2 it was animated out. At g_3 text again appeared to deliver more information to the viewer.

The most effective of the cuts in the video in keeping the viewer's attention close to that of the director cuts were those at $\{c_9, c_{11}\}$ and c_{13}. At c_9 the video transitions from a classroom with the teacher demonstrating the method of washing hands to an exterior of the school where the children gather around a sink. At c_{11} the scene presented to the viewer before the cut has one child that is centered in the middle of the video. At c_9 the positions of people were used to match across the transition, c_{11} used the environment that the scene transitions into and at c_{13} people were used again, this time it is possible that linking the semantic meaning of these screen elements i.e. both are children, had an additional effect. The effect of this transition can be seen in Fig. 4, where 0.88 distance reduction is observed after the cut for 60% of the data.

4.3.5. Vaude

In the *Vaude* clip, the majority of the cuts came in the from of a dip to white between the scenes. c_1 denotes the end of the intro piece to the video and c_2 happens to transition from an exterior scene of the factory to within it where we meet the main actor of the video. As mentioned before c_3 and c_4 had an actor present across them and the transitions took the form of a dissolve. From the graph at Fig. 2 (e) we can see that this

method again was effective in keeping the viewer close to the intended viewing direction for that part of the video.

At g_1 a graphic of a bag was faded into the scene close to where the actor was addressing the viewer directly by speaking to the camera. At g_2 this graphic faded out. Later at g_3 a separate graphic is introduced to the screen in form of a logo, at g_4 this logo then fades out.

At g_5 a graphical element in the form of a video was imposed into the scene, later a wipe expanded out from this in order to create a transition into it at c_5. This wipe brought the viewer into a new scene however due to the positioning of the camera, there was a difference in the scale of the actors, which could be the reason for worsening mismatch of viewers during the entirety of the next scene. The performance of this transition can be seen in Fig. 2 (e). c_6 is notable for the highest mismatch between the intended view and what people actually watched. At this point, there were environment factors at play that could have attracted the viewers' attention elsewhere, in this case scenic countryside, rather than the main actor addressing the camera. From the figure, we can see that the use of the main actor did work to draw the viewers back towards the intending viewing direction of the scene as it continued. At c_7 the video went from a static scene to one where the actors were cycling, the camera being attached to one of the bicycles. This caused a number of visual discomforts for the viewers, such as the fact that the camera was not stabilized, which could explain the divergence between views as the scene continued. At c_8 the intended view in the first scene was a man in a panda suit, holding a sign. This did not attract attention as effectively as expected for the majority of viewers meaning that they were not in the position as expected when the transition happened. There was a difference in scale of the actors between these scenes too because of the positioning of the camera. From the figure, we can see how this caused exploratory behavior on behalf of the viewers before they found where the story was present in the scene.

At c_9 the cut happened when the most salient object in the previous scene, that of the main actor addressing the camera, matched with the salient area in the next one, the three characters that have been introduced in the video approaching the camera along with a bridge on their bicycles. They continue to the end of this bridge until c_{10}, where a dip to white transitions the viewer to a center position on the bridge where they are addressed again by the main actor. The attention across the cut increases in mismatch, this is as the screen is dipping to white, but the attention of the viewer is lead by movement of the bicycles in the first scene to the position of the actor addressing the camera in the second one.

4.4. Overall findings and discussion

From analyzing the visual attention of the viewers across the videos in the database there does seem to be a number of methods that a director in the medium of $360°$ video can use to effectively transport a viewer across a transition. From the videos in the database and the graphs in Fig. 2 that show the distance between the director's intending viewing and viewer's scan-path, it can be seen that the visual attention of the viewer must be taken into account throughout the video. If the viewer is not orientated across the transition it increases the risk of them not being orientated in the new scene after it and this spatial immersion must be addressed before the viewer can best engage with the scene.

As seen in c_5 in the *vaude* clip, an unsuccessful cut can result in the distance of the viewer and the intended viewing direction as decided by the director increasing in the following scenes, which can then put them in a worsening position as the video continues and more transitions occur. This can result in the viewer missing plot points and other crucial elements from the director's intending viewing of the video, alongside lessening the overall quality of experience.

4.4.1. Leading the viewer to the cut

Some videos in the database also had a number of graphical elements imposed on the video at times in order to direct the attention of the viewer. This was used in a variety of ways in the videos, often they helped to lead and position the viewer so when a transition happened, they would be in the best position to understand the new scene as the director intended. One of the methods used was that of actors movement within the scene. This was used in DB by matching the actors position across the transition with the position of the actors in the new one allowing the viewer to be facing the intended position after the cut happened. In Luther this method was also used at c_{17}, where a building present in the scenes to either side of the transition matched in position across the dissolve.

4.4.2. Use of graphical elements

The *DB* clip contains graphical elements between g_1 and g_5 which do seem to attract the attention of the viewers and animating these graphics can help to retain this attention. Having these graphics in the scene can also allow the viewer to have some degree of confidence that they are viewing the scene in the intended position and that they are not missing something present in a different area.

4.4.3. Use of graphical elements across transitions

Another method used in a number of the videos was to use a graphical element to act as a reference within the video as the transition happened. These elements were used a number of times while a transition occurred in order to guide the attention of the viewer across them. This can be seen in Luther at $\{c_5, c_7, c_{11}, c_{16}, c_{20}, c_{21}\}$. It was also used in War at c_3 and in Vaude at c_3 and c_4.

In the *Luther* clip, the transitions at $\{c_5, c_7, c_{11}\}$ were all effective in lowering the distance between what was viewed

and the intended viewing. The use of this method at c_{16} did not have the same effect, in fact, the distance increased across the transition which took the form of a dissolve. By looking at the video, it is clear that at this transition there were a number of screen elements that competed for attention. The sequence starting at c_{20} and continuing on until the transition at c_{21} were the two walkers were present across the dissolve also did not command attention in the same manner. A reason for this could be that the animated Luther character was more salient to the viewers as he was introduced and commonly used throughout the video as a device to direct their attention to different areas of the scene.

In the *war* clip, a flare was used across the transition which was a straight cut to direct the attention of the viewer upwards as the scene cut to a different location below. From the graph, this does seem to have affected on viewing behavior denoted by the slight plateau on the graph. It is inconclusive as to whether this can be described as effective use of this method.

In the *vaude* clip, this was used at c_3 and c_4 and these two transitions were two of the most effective cuts in terms of lowering the distance to the intended viewing of the entire video. This time the main actor was the screen element that was used across the cut, like the animated character in the *Luther* video, the main actor had been introduced to the viewers and was used to direct their attention in previous scenes.

5. CONCLUSION

The use of transitions in $360°$ film is still part of ongoing exploration as the process of building a visual narrative in it is still being explored. While continuity editing for traditional cinema can be adapted to the format to create a narrative, by placing scenes together, within these scenes it does not offer the same ability as in traditional cinema.

In traditional cinema transitions are used to transport the viewer within the scene, this information in the shots presented is then used to update the mental model that the viewer has internally created. The use of a graphical element that the viewer recognizes across the transition does seem like a method that can provide the viewer with a reference for their orientation when they find themselves presented with the new scene. It also allows them to know that they are looking in the intended viewing direction if there is not a large number of additional screen elements to compete for their attention.

The challenge for creators in the format is to find a method that can approach the ability a first order cut has to update the mental model. Until there is this ability to create an 'invisible' cut that is intricate to the format, $360°$ film will not have the tools it needs to accurately deliver a visual narrative as effectively as traditional cinematic experiences.

Future work will aim at developing a robust objective metric for evaluation of film transitions with the aim of supporting $360°$ film technology.

7. REFERENCES

[1] S. Knorr, C. Ozcinar, C. O Fearghail, and A. Smolic, "Director's cut - a combined dataset for visual attention analysis in cinematic VR content," in *The 15th ACM SIGGRAPH European Conference on Visual Media Production*, London, UK, Dec 2018.

[2] T. J. Smith, "An attentional theory of continuity editing (PhD thesis)," *University of Edinburgh. College of Science and Engineering. School of Informatics.*, 2006.

[3] C. O Fearghail, C. Ozcinar, S. Knorr, and A. Smolic, "Director's cut - analysis of aspects of interactive storytelling for vr films," in *International Conference for Interactive Digital Storytelling (ICIDS) 2018* , Dublin, Ireland, 2018.

[4] G. d'Ydewalle, G. Desmet, and J. Van Rensbergen, "Film perception: The processing of film cuts," in *Eye guidance in reading and scene perception*, pp. 357–367. Elsevier, 1998.

[5] U. Hasson, O. Landesman, B. Knappmeyer, I. Vallines, N. Rubin, and D. J. Heeger, "Neurocinematics: The neuroscience of film," *Projections*, vol. 2, no. 1, pp. 1–26, 2008.

[6] T. J Smith, D. Levin, and J. E Cutting, "A window on reality: Perceiving edited moving images," *Current Directions in Psychological Science*, vol. 21, no. 2, pp. 107–113, 2012.

[7] T. Swenberg and P. E. Eriksson, "Effects of continuity or discontinuity in actual film editing," *Empirical Studies of the Arts*, vol. 36, no. 2, pp. 222–246, 2018.

[8] J. P. Magliano and J. M. Zacks, "The impact of continuity editing in narrative film on event segmentation," *Cognitive Science*, vol. 35, no. 8, pp. 1489–1517, 2011.

[9] S. Schwan and S. Ildirar, "Watching film for the first time: How adult viewers interpret perceptual discontinuities in film," *Psychological Science*, vol. 21, no. 7, pp. 970–976, 2010.

[10] C. Valuch, P. König, and U. Ansorge, "Memory-guided attention during active viewing of edited dynamic scenes," *Journal of vision*, vol. 17, no. 1, pp. 12–12, 2017.

[11] S. Ildirar and S. Schwan, "First-time viewers' comprehension of films: Bridging shot transitions," *British Journal of Psychology*, vol. 106, no. 1, pp. 133–151, 2015.

[12] 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.