Dialectical Polyptych: an interactive movie

Rui António, Bruno Silva and João M. F. Rodrigues

Abstract—Most of the known video games developed by big software companies usually establish an approach to the cinematic language in an attempt to create a perfect combination of narrative, visual technique and interaction. Unlike most video games, interactive film narratives usually involve an interruption in time whenever the spectator has to make choices. “Dialectical Polyptych” is an interactive movie included in a project called “Characters looking for a spectator”, which aims to give the spectator on-the-fly control over film editing, thus exploiting the role of the spectator as an active subject in the presented narrative. This paper presents a system based on a 3D sensor for tracking the spectator’s movements and positions, which allows seamless real-time interactivity with the movie. Different positions of the body prompt a change in the angle or shot within each narrative, and hand swipes allow the spectator to alternate between the two parallel narratives, both producing a complementary narrative.

Keywords—Audiovisual, Interactive film, Multinarrative, 3D sensors

I. Introduction

Video games have been developing an approach to the cinematic language to create a perfect combination of narrative, visual technique and interaction. Unlike video games, interactive film narratives usually involve an interruption in the time of the narrative, whenever the spectators want/have to make choices. This paper presents an interactive experience of film viewing without this interruption.

According to Deleuze [1], the power of cinema is in the editing; see also [2][3]. So why not give the spectator control over this? In an interactive film the spectator becomes co-author of the work by deciding which part of the narrative is being consumed at any moment, giving it a new status in relation author/work/reception. Weissberg [4], see also Penafria [5] and Bréandon & Renucci [6], introduced the term “Spect-acteur” (spect-actor) wherein “actor” refers to act in the direction of action. The “see” (spect) is complemented by the gesture of one who acts on the work. The spectator in the role of spect-acteur leads, performs, completes, alters the structure, immersed in the work environment, and engages in acts of transformation and creation [7].

Rui António
UAlg/Uab, CIAC, University of the Algarve, Faro, Portugal

Bruno Silva
ESEC, CIAC, University of the Algarve, Faro, Portugal

João M. F. Rodrigues
ISE, LARSyS (ISR – Lisbon), CIAC University of the Algarve, Faro, Portugal

For Primo [8], systems that anticipate the spectator response and limit their action cannot be classified as interactive, but as reactive. As for Fragoso [9], interactivity should not be seen with an overvaluation of symmetry in two-way flow of communication; see also [10].

Thinking about interactivity in cinema, should lead to new forms of expression and not necessarily the new forms of two-way communication. In this paper, it is presented “Dialectical Polyptych” consisting of an interactive movie prototype with two parallel narratives. Each narrative has five different shooting angles/shot. The spectator can choose the narrative and the angle/shot at every moment, through interaction, with no interruption in the narrative flow. The interaction is made with the body and there is no contact with physical devices. The body tracking is performed by a three-dimensional (3D) sensor that picks up the skeleton of the spectator, and according to his/her position and movements the narrative flow to be viewed is determined. Thus, it is the spectator that plays the role of the movie editing in real time.

The main contribution of the paper is the prototype itself, i.e., the real time movie editing by the spectator with no interruption in the narrative flow. To the best of your knowledge, there is no similar work in the literature.

Section 2 presents the contextualization of the work as well as the state of the art. Section 3 detailed all the steps for the development of the “Dialectical Polyptych” prototype, finally, in Section 4, conclusions and future work are presented.

II. Contextualization

The editing is one of two basic components in audiovisual production [11], the other component in the audiovisual language is the framework or selection unit. According to Deleuze [12], the evolution of cinema will make up through the editing, the mobile camera and the emancipation of taking views Griffith also had contributed to the editing by ranging of shot to give emotional impact through the long shot, medium shot, close-up, subjective camera (point of view of the character) and travelling (moving camera) [13]. Thus, Griffith wanted to involve the spectator emotionally through scale changes in the shot, giving the public a progressive emotion. Porter expanded the idea of linear narrative with the use of parallel-editing to depict two simultaneous events or points of view [14][15]. This technique alternates two or more scenes that often happen simultaneously but in different locations. Like Porter’s technique, “Dialectical Polyptych” prototype also invokes parallel editing but from the result of spectator interaction.

The film editing continues to present new forms and technologies and to propose new questions regarding the association and organization of cinematic images. Regardless of the technique used, the editing will always have a dominant role in the structure and meaning of a film. The new digital technology has facilitated the deconstruction of the narrative to an interactive format. The
digital image becomes easily manipulated allowing the change of order in which the different segments of the film are presented.

For interactive cinema technological advances is not enough. It is also necessary to point out new solutions to the narratives. In the beginning of the XX century, Pudovkin said: “the editing builds scenes from the separate pieces” [16]. The combination of these pieces forms different meanings depending on the chosen sequence. Thus, if given the user the possibility, there will be a similar editing to what happens on the Internet with hypertext, where the user also chooses the “pieces” of text to be read. The Russian cinematographic school proposed from its inception the structural fragmentation of audio-visual work through editing procedures. When an editor builds a film (non-interactive), what he/she does is to choose a sequence from fragments, choosing a path of many possible. This is the film shown to the public. If the other sequence was being chosen it would be a different film, its meaning would be different, even though the raw material is the same. Giving the spectator control of the editing is to allow a different individual experiment, to each spectator in each view.

Currently, in movie or in an audio-visual product, there is a corresponding author for the final result, although there are other stakeholders as the screenwriter, cinematographer, and editor, among others. For example, the editor according to André Bazin’s concept consists in renewed powers in the message [17]. Even the simplest editing gives a unit of meaning to certain events and could be considered somehow a work of author. To Sergei Eisenstein [18][19] the editor becomes co-author of the work, it is necessary to “guide the viewer in the desired direction”. Currently there is a need for participation by the spectator. With the new concepts of authorship and co-authorship, from the advent of hypertextuality, the spectator is not satisfied with the passivity [20]. In the case of interactive cinema or an interactive audio-visual product, the question arises of who is the author of this product: the director or the spectator? [13]. The director who developed the product enabling the spectator involvement in the construction and reconstruction of the work, or the spectator that “conduct” the new work? With the advent of interactive and digital capabilities, we can consider the spectator co-author of the work, as it builds and rebuilds the work at the time of fruition through the choices of the various paths and achieving new experiences.

In fact, when a participant interacts with the machine, he is not only interacting with the machine but also with himself. In this interaction, the participant is involved in the discussion with himself internally. The participant is provoked by stimuli that cause him to question about the resulting stimulation of interaction and machine answers to his action [21][22]. The first interactive experience produced for film, was “Kinoautomat” in 1967 [23], it was an example where the spectators chose, through buttons placed on chairs, the action to be taken by the characters at certain times of the narrative. Many other examples appeared over the years, one of those was in 2010 the “Last Call” [24] realized by Jung Von Matt. It was an interactive cinema project/game, created as an advertising piece for NBCA channel. The film presents the story of a character stalked by a serial killer. When trying to escape she finds a mobile phone that she uses to make contact with the viewers/spectators of the film, and the spectators can help in the decisions that she should take over the narrative. The spectator, upon entering the room, gets an invitation to register his mobile phone number in a particular digital platform, so he/she is ready to receive protagonist calls at any time. At certain times of the film, the main character calls randomly someone from the audience and asks a question about the path or decision to make. A voice recognition program captures the decision of the spectator.

Based on this protagonist will respond to the direction suggested by the spectator so a specific sequence history is displayed. Thus “Last Call” features a number of possible paths and endings based on the interaction of spectators that establish changes in the narrative paths giving them the impression of history control.

Also, in 2010 was released an interactive 3D film entitled “Scenario” [25][26] developed at the iCinema Center for Interactive Cinema Research at the University of New South Wales. The movie is projected on a 360° panoramic screen with motion sensors that track the audience. Interaction is given between the human participants and the humanoid creatures on the screen controlled by the film’s artificial intelligence engine. The narrative unfolds depending on how the audience interacts with the film. In 2014 the Filmstrip production company started the project “Biosuite” [27] in collaboration with Queens University’s Sonic Arts Research Center (SARC). This project explores the way audience’s emotional reactions control the narrative of the film using ECG (Electrocardiogram) signals and GSR (Galvanic Skin Response) which measures the change in conductance of a person skin. These signals are interpreted through computational software and determine the changes in the film narrative as well as the generation of the music score.

As it is possible to perceive above, there are many ways to develop the interaction, one of the paradigms for human-computer interaction are the 3D sensors, such as the Microsoft Kinect [28], the Asus Xtion [29], the Leap Motion [30], or the Structure Sensor [31]. Those sensors can be used to interpret specific human gestures, enabling a completely hands-free control of electronic devices, the manipulation of objects in a virtual world or the interaction with augmented reality applications. Many of these tracking and gesture recognition sensors have a huge importance in the videogames industries. Hence, with the appropriate software, they have also the capability to detect the user skeleton and/or tracking a single or several users, while replicating with accuracy e.g., the hands and the user movements in a 3D mesh.

As expected, in the literature there are several interactive installations that integrate art, technology, image, film and volumetric sensors [32][33][34] with also analytic frameworks to evaluate such interactions in public installations [35]. In terms of gesture recognition and tracking using 3D sensors, we can refer for instance, pose and hand control of interactive art installations [36], air painting application [37]. For the present prototype, it was chosen for the interaction Microsoft Kinect sensor due to the human skeleton tracking capabilities. Kinect is a device consisting of a RGB camera, depth sensors and microphones [28]. This sensor can be connected to any personal computer via USB, it has an RGB camera with the resolution of 640 x 480 at 30 Hertz, and an infrared sensor with the same resolution which allows to extract the depth information to...
track the movement of objects and persons in 3D space in front of the sensor (for the present work it was used Kinect sensor, not Kinect 2 sensor).

### III. Dialectical Polyptych

The implementation of the “Dialectical Polyptych” prototype, has two major phases (see Fig. 1): (a) Pre-production, which includes a brief description of the script preparation, filmic storyboard, and technical aspects that were taken into account in the shooting of the prototype. (b) Prototyping and installation, it consists in the implementation of the prototype, as well as some considerations about the installation.

![Figure 1: Project development stages.](image)

**A. Pre-production**

For the production of the film the first step was the script writing and respective storyboard. The script consists of two parallel narratives that address the issue of loss of a loved person. Both have the same characters although the action takes place at different time and space. There is a narrative parallelism between them. In the filming script, the different angles and shots were covered for each scene, and for each of the two narratives. Since these are two parallel narratives, there was a special care in the synchronization. After writing the script, the creation of the storyboard was the next step in order to predict the filming shots and avoid possible match cut editing errors. A detailed explanation of the storyboard is out of the focus of this paper. Figure 2 shows the relative positions and angles where the cameras where situated, the shooting sequence was done with six cameras simultaneously.

**B. Filming and editing**

For filming were used DSLR (Digital Single Lens Reflex) cameras, with FullHD 1920x1080 resolution, equipped with 28mm, 50mm, 135mm and 210mm objectives depending on the distance from the camera to the subject (see Fig. 2). The filming was carried out, where possible, with multiple cameras simultaneously (see Section II-A), in order to facilitate synchronization between the images. Each of these cameras captured different camera angles. The cameras were placed on tripods, sliders, shoulder rig, Steadicam stabilizer and on the ground, depending on the requirement of the shooting and camera movement. In other situations we proceeded to separate shots that served as a complement for different flows of the narrative, for example, details or flashback.

For editing, all narrative flow was overlapped in different layers, in order to achieve synchronization of the film action. Each narrative consisting of five flows, a single sound track was maintained. After synchronization and editing of different scenes for the two narratives, all obtained flows were exported.

**C. Prototyping and Installation**

As mentioned in Section II, the device used for the interaction was the Kinect sensor, this device allows the detection and tracking of the spectator skeleton and then interpret body movement and gestures. Figure 3 top shows an example of the skeleton (top-right) and the overlap (bottom) of the skeleton with a frame from a film (top-left). For the development of the prototype it was used programming environment Processing [38] and the library OpenNI [39] for an integration of Kinect device. The different narrative flows are presented through the interaction of the spectator. For this, Kinect was used for tracking the spectator's position and gestures of the hands.

The Kinect device allows the reading of \((x, y, z)\) coordinates of the spectator. Through these data spectator's position including the distance to the device can be computed. Figure 3 Bottom-left illustrates the relative position of the spectator to the Kinect at the beginning of the film, and on the right, the four movements that the spectator is tracked. The different four spectator body positions in front of the sensor allow change the angle or shot within each narrative (see Fig. 4). The spectator can move laterally and longitudinally. Based on the values of \(x\) (lateral position) and \(z\) (distance), it is computed which the flow to be presented in each narrative, at each moment. The lateral position of the spectator \((x)\) defines the flow to be presented in Fig. 4: right, middle and left. The spectator distance defines the flows forward, mid and back.

![Figure 2: Shot sequence with six cameras simultaneously.](image)

![Figure 3: Top, skeleton on the film image, captured by the sensor during the interaction of the spectator. Bottom-left and right, Kinect camera tracking, and the spectator motion.](image)

![Figure 4: Change of image flow by the spectator motion interaction.](image)
Changes in the position of the spectator result in different shooting angles, perspectives and views (Fig. 4). The horizontal scanning motion of the hand (horizontal swipe) allows the spectator to alternate between the two parallel narratives. Figure 5 shows all the interaction combined, i.e., for each narrative 1 and narrative 2, and within the narrative the different points of view of the spectator in function of his/her position to the Kinect. It is very important to mention that all the interactions and respective narrative flows (streams) occurs on-the-fly without any stop in the film action. For a better understanding, imagine that 10 movies started at the same time and running in parallel (at the same time) in 10 cinema rooms each one telling a different point of the exactly same story at the exactly same time in the story, and the spectator by moving his/her body or hand can travel from one room to the other, exactly to the same time in the narrative only viewing a different point or time of the narrative.

In terms of implementation, having all the streams synchronized, the high level functions available in the OpenNI library were used, since the Kinect range can vary in x between 0 to N = 640 px (pixels) and in z the range distance from \(d_l = 500 \text{ mm} \) to \(d_f = 4000 \text{ mm} \) (millimetres). Considered \(x_s\) and \(z_s\) coordinates of head joint (in px and mm respectively), \(x_{h,l/r}\) and \(z_{h,l/r}\) the coordinates for the left_hand or the right_hand joints [28], again in px and mm respectively. The hand swipe interaction follows the next algorithm:

a) Track the \(x_{h,l/r}\) hand position of the spectator, using left_hand or the right_hand joint, as well as the \(x_s\) head joint;
b) If the head joint doesn’t change the x position significantly between 2 consecutive frames, i.e., \(\Delta x_s = \lfloor x_{s, t} - x_{s, t-1} \rfloor\), and \(\Delta x_s < d_h\), detect if there was a hand swipe (with \(t\) the present time frame).

For a hand swipe occurrence, \(\Delta x_{h,l/r,t} \geq d_h\), again in 2 consecutive frames, with \(\Delta x_{h,l/r,t} = \lfloor x_{h,l/r,t} - x_{h,l/r,t-1}\rfloor\):
d) If a hand swipe occurs, select the narrative;
   a. If the narrative was the “1” pass to “2”;
   b. otherwise pass from “2” to “1”.

The value \(d_h\) was computed as \(d_h = K \times N \text{ px}\) and \(K=10\%\). The value of \(K\) was empirically selected, and small changes in \(K\) only affects the speed of the movement (faster or slower). For the spectator skeleton tracking (position) it was used the head coordinates \(x_s\) and \(z_s\) that were taken as a reference point for the interpretation of the movement that defines each of the flows to be presented. The tracking algorithm interaction that triggers the change of narrative angles of shot/views, follows the next basic steps:

a) Track the \(x_s\), and \(z_s\) head joint position of the spectator;
b) Divide the x pixels available, \(N\) px, in three regions, each worth \(M = 1/3 \times N\); c) Divide the z distances available \([d_l, d_f]\) mm, in five regions, each worth \(Z = 1/5 \times (d_f - d_l)\). The reason for the five regions is due to the spectator comfort, in a way to “avoid” the Kinect closest and fastest region;

d) Select the view according to mapped values:

a. In case \(x_s < M\) px the spectator sees the left flow, for \(M \leq x_s \leq 2 \times M\) px correspond to middle, and for \(x_s > 2 \times M\) px correspond to right flow;
b. In case \(z_s < (Z + d_z)\) mm the spectator sees the front flow, \((Z + d_z) \leq z_s \leq (2 \times Z + d_z)\) mm corresponds to the middle, \((2 \times Z + d_z)\) mm corresponds to back flow. With, \(d_z = d_l \times K\), and \(K = [1.0, 2.6]\). Again \(K\) is a comfort parameter that can be changed in function of the projection screen, small screens the smaller should be \(K\). In the present installation it was used \(K = 1.5\).

c. In the case that (a) and (b) occurs at the same time it is given preference to (a).

In the passage from one stream to another is given the current position of playhead to ensure a film continuity. The new flow continues from that same exact position the previous stream stopped, as described in the next steps (this is done between frames, in less than 1/30s):

a) Get and save current narrative flow (\(N_{ch}\) playhead position, with \(c\) the current narrative flow, and \(i\) the number existing narratives \(i = \{1,\ldots,10\}\);
b) Pause current narrative flow;
c) Play new narrative flow (\(N_{ch+i}\) from saved playhead position).

In this prototype installation were considered only the values \(x\) and \(z\). However, it is expected in the future to use also the value of \(y\) in order to detect height movements of the spectator, such as jumps or squats, creating new points of view of the spectator in function of those movements.

As usual, for the viewing of the film it is suggested a darkened space. It is also necessary computer, video projector or screen, speakers and a Kinect device. Despite the possibility of interactivity, there is the possibility that the spectator does not interact. If so, the spectator will see the film only in the perspective of one of the characters who will be chosen at random at the start of the movie. If there are two or more viewers, the interaction will be controlled by the first spectator tracked by Kinect device.

### IV. Conclusion

This paper presented “Dialectical Polyptych” an interactive movie prototype with two parallel narratives. Each narrative has five different shooting angles/shot. The spectator can choose the narrative and the angle/shot at every moment, through interaction, with no interruption in the narrative flow. The interaction is made with the spectator body thus, it is the spectator that plays the role of the movie editing in real time.

There weren’t detected errors due to interaction movement, due to all the movement are simple and based on the OpenNI library, also the flow of the movies does not present any “cut” (wait/delay) when a narrative or a shooting angle changes due to an interaction.

It is expected that this work contributes to different forms of film viewing and audio-visual language evolution following the technological evolution and making use of its potential. The use of sensors, due to is transparent interface characteristics, and whose functionality does not need to be understood by the participant although this start to associate
their behaviour with the action or reaction of the installation, enables interactivity without the manipulation of buttons or control devices, freeing the body and for a more natural interaction. To the best of your knowledge, there is no similar work in the literature.

Future work includes, as already mentioned, the implementation of more views in function of different spectator movements/positions. Also in research is how to complement the interaction of more than one spectator (group of spectators).

Acknowledgment

This work was supported by FCT project LARSyS (UID/EEA/50009/2013), CIAC (PEstOE/EAT/UI4019/2013) PRHOLO QREN I&D, nr. 33845, POPH and FEDER. We also thank our project leader SPIC - Creative Solutions [www.spic.pt].

References


Figure 5: Top, different streams for narrative 1 based on the position of the spectator. Bottom, the same but for narrative 2.