Probabilistic Editing with Virtual Cinematography

Luiz Velho, Leonardo Carvalho, Djalma Lucio

Technical Report TR-18-04 Relatório Técnico

May 2018 Maio

The contents of this report are the sole responsibility of the authors.
O conteúdo do presente relatório é de única responsabilidade dos autores.
Probabilistic Editing with Virtual Cinematography

LUIZ VELHO, IMPA
LEONARDO CARVALHO, IMPA
DJALMA LUCIO, IMPA

This technical note introduces Probabilistic Editing with Virtual Cinematography. It is part of the VR Kino+Theater platform and provides high level authoring tools for cinematic presentations. The director acts as a VJ controlling in real-time an A/V switcher interface that selects the camera views of a theatrical performance in a virtual reality experience.

CCS Concepts: • Computing methodologies → Graphics systems and interfaces; • Applied computing → Media arts;

Additional Key Words and Phrases: virtual reality, real-time cinema, immersive theatre, storytelling, montage

ACM Reference Format:

1 INTRODUCTION

Entertainment is presently going through radical transformations as a consequence of technological developments, advances in the state of the art, and social changes. This will shape the way cultural audio-visual products will be delivered and consumed.

A direct consequence of the above scenario is that new creative possibilities become available for content producers. Here we explore one of these avenues, related to editing and virtual cinematography.

2 RELATED WORK

A seminal work in the area of interest is "Live Cinema", an initiative of the film director Francis Ford Coppola [2]. His proposal is an attempt to combine theater, film and television as an experimental form of storytelling.

In this setting, performances are acted live and viewed by an audience in real time on a movie screen. The goal is to achieve a more cinematic look and feel than what is typically employed for dramatic broadcasts. It employs professional television technology borrowed from TV sports. (See Fig. 1.)

In order to demonstrate the concept, Coppola promoted a workshop with American Zoetrope at the UCLA School of Theater, Film and Television in 2016. During an one-month period, 75 UCLA students and faculty produced a 27 minute piece called "Distant Vision" for a live broadcast to a limited audience. The production involved operating over 40 cameras, acting and working on sound, set design and construction, costume, props, editing, and stage management.

The main issue of the project is the coordination of all these practical aspects restricted by the traditional video technology. In that respect, we can arguably say that Live Cinema was a vision 'ahead of its time', impaired in many ways by the physicality of the medium.

In the case of non-live video a recent related work is [4].
3 VR KINO+THEATER

VR Kino+Theater [6] is a new platform for storytelling that shares various aspects of Coppola’s vision of Live Cinema. The main difference is that it is based on 3D Computer Graphics and Digital Network Communications.

The platform we propose integrates traditional forms of entertainment, such as Theater and Cinema, with advanced technology, more specifically Virtual Reality and Gaming.

The main components of VR Kino+Theater exploits the concepts of Situated Participatory Virtual Reality and also Live 3D Digital Cinema. We believe this points to the directions for the future of media.

3.1 Situated Participatory VR

Situated Participatory Virtual Reality [7] is a modality of VR that allows the creation of Shared Multi-User Virtual Environments. For this purpose, it combines real and virtual objects in tangible spaces, where the participants, represented by digital avatars, are completely immersed in a simulated world. They use VR headsets and markers for full body motion capture.

The above setting implements the Theater component of the platform. As such, the actors perform in a VR stage that is mapped into a virtual set.

Fig. 2 shows the real actors performing in the VR stage and the corresponding action of their avatars in the CG virtual set.
3.2 Live 3D Cinema


The virtual cinematography framework includes Pre-Programmed Cameras and Interactive Editing for generating the cinematic content.

The above setting implements the Kino component of the platform. In this context, the director selects in real-time the views that are shown on the live movie projection screen.

Fig. 3 shows the director operating a multi-camera switcher during a live presentation.

Fig. 3. Director operating the camera switcher and detail of the interface.

Fig. 4 shows the image selected by the director at a moment of the presentation (the lower right camera in the interface).

Fig. 4. Image of the selected camera exhibited on the movie screen.
4 BASIC CONCEPTS
In this section introduce some basic concepts that are necessary for the understanding of our framework. They are related to Storytelling in Theater and Cinema.

In terms of storytelling, we will focus on the structure of narrative, while in theater we concentrate on the staging process and in cinema we discuss cinematography and editing.

4.1 Storytelling
Storytelling is a means for sharing experiences. It is a cultural tradition that predates writing. However, here we are particularly interested in the more structured form of storytelling that is used in theater and cinema.

The Classical Narrative, usually traced back to Aristotle’s Poetics, has a three-part structure for the events in a story. That chain of events correspond to the: Setup; Development; and Resolution. This is the basic formula for classical theater plays and movies.

The Theatrical Structure divides the play into Acts which is further subdivided into Scenes.

The Cinema Structure divides a movie into Sequences, Scenes and Shots. A shot consists of a single continuous action recorded by the camera. A scene is composed of several shots, and a sequence is a made of scenes.

Fig. 5 illustrates the above structures.

![Fig. 5. Narrative film structure.](image-url)
4.2 Theater

In a Theatrical production *Staging* is the process of designing the presentation of the narrative. It involves determining the space (Set) and timing (Action) for the actor’s performance.

In the above context, *Blocking* consists of the precise planning of the action in the set.

To help the staging process, the creative people involved in a production usually employ a Storyboard. (See Fig. 6.)

For the VR Kino+Theater we have developed also a *Planner App* to facilitate the process. (See Fig. 7.)
4.3 Cinema

In Cinema, besides the actions that compose the flow of events in a narrative, two other aspects are relevant: Cinematography and Editing.

Cinematography concerns the visual aspects of the content captured by the camera to be shown on the movie screen.

The essential element of cinematography is the Shot. The visual characteristics of the shot form the Grammar of the Shot. These characteristics correspond to: Framing (Full / Medium / Closeup); Composition (Symmetry, etc); Position / Angle (Low / High / Birdseye); and Camera Movement (Track / Pan-Tilt / Zoom).

Fig. 8 shows an example of shots with different framing and same symmetry.

![Fig. 8. Full, Medium and Close Shots with Left Assymetry.](image)

Editing is the task of assembling individual shots into scenes and sequences. Basically it amounts to a series of decisions of What and When. There is, which shots to place at each moment of the visual narrative. In that context, the Cut refers to the transition (and type) from one shot to the next.

Fig. 9 shows the interface for an editing program (Apple’s Final Cut).

![Fig. 9. Final Cut Interface.](image)
5 VR KINO+THEATER CINEMATOGRAPHY

In this section we give an overview of the VR Kino+Theater Cinematography. It is composed of a camera specification infrastructure and an interface for the camera selection by the director in real-time.

The camera specification infrastructure is implemented through a layered architecture with three levels: Unity CG Cameras; Cinemachine Camera Operators; and K+T Virtual Cameras. The director interface consists of a live camera switcher.

5.1 Unity Camera

The Unity Camera Layer corresponds to the low level of the camera specification infrastructure. It consists of a standard Computer Graphics Camera of the Unity Game Engine [5]. The camera is defined by the usual parameters, such as position, orientation, field of view, etc. Fig. 10 shows the camera object in the Unity Editor.

![Camera in Unity Editor](image)

Fig. 10. Camera in Unity Editor.

5.2 Cinemachine Operators

The Cinemachine camera operators constitute the intermediate level of the camera specification infrastructure. These camera operators embody a framework for smart, programmable cameras. In that respect the operator know about the entities in a Unity scene and control the camera specification based on visual composition rules.

The two main controls are the Composer and the Transposer, they allow to specify the camera in screen and scene space respectively. Fig. 11 shows examples of controlling the composer and transposer.

![Controlling the Composer (left) and Transposer (right)](image)

Fig. 11. Controlling the Composer (left) and Transposer (right).
5.3 Kino+Theater Cameras

The VR Kino+Theater Camera layer forms the higher level of the camera specification infrastructure. It consists of an object based abstraction for creating the entity of a Virtual Cameraman and allowing to instantiate these objects for specific purposes.

In that respect, the Kino+Theater Cameras are designed for Cinematography Storytelling and allow the director to compose shots for each scene of the narrative.

There are two classes of camera objects: General; and Timeline. General cameras refer to shot types that can be used freely during a scene, while Timeline cameras are meant to be used at certain times along the scene and are choreographed for specific events of the action. Fig. 12 shows an example of a general camera object for close-up shots of Miranda in the Cell scene.

![Fig. 12. Close-up Shot of Miranda.](image)

5.4 Kino+Theater Camera Switcher

The Director’s Interface allows the control a live image on the movie screen by selecting the active view using an special-purpose multi-camera switcher.

This interface contains the views of 12 pre-programmed cameras showing in real-time the CG simulation. A view is activated by a simple click. The director interface also contains additional controls for triggering simulation events. The cameras are divided in two blocks: one block with 8 general multi-purpose cameras, such as close ups, medium shots and characters points of view; and another block with a sequential list of timeline cameras which are custom designed for specific parts of the action. Fig. 13 shows the Kino+Theater Camera Switcher interface.

![Fig. 13. Multi-camera switcher interface.](image)
6 REACHING TO A HIGHER LEVEL

The VR Kino+Theater Cinematography infrastructure implements a powerful mechanism for live editing of cinematic experiences. Nonetheless, the image director operates the camera switcher interface using a rather explicit control. He/she has to exercise every single cut of the visual piece (at precise moments in real-time).

The scenario described above motivate us to reach to a higher level of control that could provide more expressive power. In that sense, the goal is to allow the director to act as a VJ, using an interface designed for stylistic control and live improvisation.

Such an interface has to expose the “right parameters” in a concise and intuitive way. It should be pre-configured based on the scene content and the desired cinematic style variations. The key for creating this device is to exploit the concept of generative interfaces, as we shall see further in this section.

Another important point is that the proposed functionality should be built on top of the VR Kino+Theater Application Framework and rely on the layered architecture of its Cinematography infrastructure.

6.1 Stylistic Control

The Stylistic Control is based on Shot Classes and Timeline Events.

The Shot Classes are related to visual characteristics as discussed earlier in Sec. 4.3. For example: Close-Ups; Middle; and Wide shots. The Timeline Events are related to specific moments when an action occurs.

These two style elements are combined using Cinematic Rules that are part of the Cinematographic language. Together they deal with aspects such as pacing of cuts, etc.

An example of cinematic rule is Cut-on-Action.

Fig. 14 shows a cut-on-action rule, where from a Mid Shot it transitions to a Wide Shot, triggered by the action of the character that turns his head looking to the right and focusing its attention on a discussion in the parking lot.

6.2 Generative Interfaces

Generative Interfaces provide parametric controls for the generation of different types of objects, such as audio/visual media in our context.
There is a new research area devoted to this type of interface. One example is the, so called, Artificial Intelligence Augmentation (AIA) that builds the parameterization from data using machine learning. Fig. 15 shows a generative interface for typographic font design. The controls allow to specify the following style aspects of a typeface: weight (i.e., bold, light), slant (i.e., italic, normal) and aspect (i.e, condensed, etc).

6.3 Architecture
The Cinematography Infrastructure of VR Kino+Theater is implemented using a layered architecture. It consists of several layers for the camera entities.

As presented in Sec. 5 the first three layers correspond respectively to the levels of abstraction for: the Unity Camera; Cinemachine Operator; and K+T Cameraman.

In order to incorporate style control, we extend this hierarchy to include a higher order level: the K+T AutoShot. Fig. 16 shows the complete camera abstraction hierarchy.

The K+T AutoShot embodies a high level cinematic style control that is the basis of probabilistic editing, as will be discussed in the next section.
7 PROBABILISTIC EDITING

Probabilistic Editing is a framework for the design of cinematic style in live audio/visual performance presentations. The framework is based on Film Grammar, Stylistic Edit Patterns and Flow of Action in order to create a Live High-Level Control mechanism.

In this framework the Film Grammar maps to the concept of Camera Groups, the Stylistic Patterns are represented by a Cut Graph and the Flow of Action follows mark-up sequences in the Timeline.

The result is a generative edit interface that extends the VR Kino+Theater camera switcher.

7.1 Camera Groups

Camera Groups embody the main conceptual entity that is manipulated in the probabilistic setting of our framework. They form the building blocks of Cut Graphs (see next subsection) that are used to represent a cinematic style design.

Typically, camera groups are created by the image director following classification principles that are based on shot classes (see Sec. 4.3). Furthermore, they are defined per scene, i.e., they depend on the narrative content and the staging of specific scenes.

For example, Fig. 17 illustrates a camera group for the CloseUps in the Cell Scene of the experiment "The Tempest". This group includes the close-up shots of the characters Miranda and Prospera.

![Fig. 17. Camera Group for CloseUp Shots of the Cell Scene in The Tempest.](image)

7.2 Cut Graph

Cut Graphs are the probabilistic representation of a cinematic style in our framework. They are mathematically a Probabilistic Graphical Model [3], in which the nodes are random variables and the links are statistical dependencies among these variables.

The nodes describe in probability terms the particular cinematic style rules for a given sequence that is part of the narrative. For example, Table 1 shows a rule for the Camera Group “CloseUp of Cell Scene” mentioned above. Essentially, it specifies that close-ups of Miranda have a 20% chance of being selected, while close-ups of Prospera have 80% chance.

<table>
<thead>
<tr>
<th>Close Miranda</th>
<th>Close Propsera</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Table 1. Node R1
As a whole, the cut graph models a probability density distribution that characterizes a particular cinematic style. Intuitively, this density distribution is a way to decide which shots to select for a cut in a probabilistic sense.

Fig. 18 shows an example of a Cut Graph. The top nodes (R1 to R3) are associated with parametric input decisions, and the bottom node (Rn) is associated with the final cut selection.

7.3 Modeling the Distributions
In order to model the probability distributions in the Cut Graph we can use either parametric or non-parametric models.

In the parametric setting we have the following characterization of distribution functions \( P = \{ P_\theta : \theta \in \Theta \} \), where \( \Theta \) is the set of parameters. while in the non-parametric setting the distribution is given by a table, for instance, in the form of a histogram (as in the example of Table 1).

7.4 Timeline
One important aspect that remains to be considered is related to the Timeline. That is, to answer the question: “When to perform a Cut”? In order to model this temporal aspect we resort to Track Markers. They are associated with Camera Groups and defined by a list of time-stamped annotations indicating the moments to evaluate the cut graph for a decision of which cut to make.

In other words, the several layers of track markers collectively specify how often to perform a Cut. In that sense, the image director can determine the Granularity of Markers, which could also have a Nesting structure. Fig. 19 illustrate the Timeline and Track Marker layers.
7.5 Style Design

The Style Design is accomplished by combining the nodes in a cut graph to set the conditional probabilities. Each node is associated with a track marker layer that specifies both the camera groups involved and type of probabilistic model, as well as, the timeline events for potential cut evaluation using the graph.

The general format of the track mark entries is as follows:

index 1
start time --> end time
[camera group id] (optional list of camera / weight pairs)
.
.
.
index N
start time --> end time
[camera group id] (optional list of camera / weight pairs)

Fig. 20. Track Marker File.

For non-parametric models the list of cameras of a camera group and their associated weights. This allows a computation of the histogram description. (See below).

(camera 1 name : weight) .... (camera N name : weight)

Fig. 21. Camera/Weight list for non-parametric model.

Parametric models are associated with camera groups and a procedural layer that has controls for their parameters, which can be exposed in the live editing interface as we will see in the next subsection. Fig. 22 shows an example of a style design using both non-parametric and parametric models.

Fig. 22. Cinematic Style Design using parametric and non-parametric models.
7.6 Editing Interface

The editing interface extends the live image switcher to provide high levels controls. It is programmable with the style parameters for each scene and is meant to be used in Live Cinema.

The interface operates either in an auto or manual mode. The auto mode selects automatically the cuts based on the probabilistic style graph, without the interference of the director. However, the director can also change the style parameters using the interface controls and these modifications will be reflected in real-time on the cut decisions by the machine. Furthermore, the director can override the probabilistic style machine to select individual cameras at any moment — virtually operating the switcher in manual mode.

Fig. 23 shows the Editing interface. Note the highlighted area indicating the style controls.

8 CASE STUDY: THE TEMPEST

In this section we present a case study of using the high level live editing functionality on an actual A/V experiment.

We used the tool for the Cinema presentation of the Shakespeare’s play, ”The Tempest” [8].

8.1 Cameras

The production of the experiment consisted of three scenes: the Cell; the Clearing and the Epilogue.

Figs. 24, 25, 26, show respectively the live editing interface for the three scenes.

The cinematic style design is the one depicted in Fig. 22. At the top node of the cut graph, the random variables control the decision between a theatrical and film styles – in this case, that means respectively wide shots of longer duration versus near shots with fast paced cuts.

In the Cell scene the camera groups are close-ups and points-of-view for the track of film style and mid shots from fixed angles (front, back, left, right) for the track of theatrical style.

In the Clearing scene the camera groups are also close-ups and points-of-view for the track of film style but for the track of theatrical style we included both wide and mid shots.

The Epilogue scene does not have a probabilistic editing setting, only a deterministic zoom shot.
Fig. 24. Cameras of the Cell.

Fig. 25. Cameras of the Clearing.

Fig. 26. Cameras of the Epilogue.
8.2 Results

In order to verify the effectiveness of the high level cinematic style control of our framework, we performed several laboratory tests and also produced a live presentation with participation of the audience.

The first test was a comparison between Film and Theatrical styles. In that case, we generated two different montages of the experiment “The Tempest”, featuring these two extremes, i.e., respectively 100In the Film style montage, only near shots have been used and the cuts performed when each characters started a new part of the dialogue. In the Theatrical style montage, only mid and wide shots have been used and the cuts selected when the action caused a change in relative position of the characters.

Also, the Film style was controlled by a parametric probability distribution that would decide between action or reaction shots (i.e., showing the character speaking or listening) and between close-up or point-of-view shots of the selected character. In contrast, the Theatrical style was controlled by a non-parametric probability distribution, created by the director to maximize the visibility of the characters in the frame during their action.

The second test was a montage that combined Film and Theatrical styles with equal probability (i.e., a 50

8.3 Evaluation

The evaluation of our tests revealed that the proposed probabilistic cinematic style designed for the experiment “The Tempest” provided a simple and intuitive control of the style variables involved.

The extreme cases: Film and Theatrical montages showed what was expected, in terms of framing coverage and pace of cuts.

The intermediate case: combination of 50produced a result very close to a realistic editing setting, with a well balanced combination of wide and near shots and a good cut dynamics.

9 THE OTHER SIDE OF THE COIN

In this report, we presented a generative interface for high-level probabilistic live editing with virtual cinematography. In that sense, the probability distributions in the cut graph, as well as the parametric controls in the interface, are created by the director. This is the direct approach to solve the problem.

The other side of the coin is the inverse approach to automatically generate the solution. The problem, then, is to generate the editing machinery from examples, using machine learning techniques, such as Deep Learning with Neural Networks.

In this setting, we can have two possible scenarios: the first scenario would be to estimate only the editing style from montages of an author, using supervised learning; the second scenario would entail a full understanding of the editing style structure manifold, including the style, the parametrization and the AIA interface controls.

It is worth noting that this inverse approach to the problem has many potential applications, for example in Live presentations of Sports and Music Shows.

10 PLAYER INTERACTIVE EDITING

Finally, another way to explore the ideas discussed in this technical report is in a non-linear editing setting. In this case, the montage of the story would be controlled by viewer through an interactive interface for selecting the relevant shots and show them on the screen.

This kind of interface has been proposed by the Eko group in the context of interactive storytelling [1]. They recently released the series “War Games” using this approach. Fig. 27 shows the player interactive editing interface for the series.
Fig. 27. War Games interactive editing interface.

REFERENCES


Received January 2018