



HAL
open science

Live Probabilistic Editing for Virtual Cinematography

Luiz Velho, Leonardo Carvalho, Djama Lucio

► **To cite this version:**

Luiz Velho, Leonardo Carvalho, Djama Lucio. Live Probabilistic Editing for Virtual Cinematography. 17th International Conference on Entertainment Computing (ICEC), Sep 2018, Poznan, Poland. pp.40-51, 10.1007/978-3-319-99426-0_4 . hal-02128590

HAL Id: hal-02128590

<https://inria.hal.science/hal-02128590v1>

Submitted on 14 May 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Live Probabilistic Editing for Virtual Cinematography

Luiz Velho¹, Leonardo Carvalho², and Djama Lucio³

IMPA

Abstract. This paper introduces Probabilistic Editing for 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 DJ controlling in real-time an A/V switcher interface that selects the camera views of a theatrical performance in a virtual reality experience.

Keywords: virtual reality · real-time cinema · immersive theatre · storytelling · montage

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 live 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.

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.

Another related work is [4], where probabilistic editing is used for video in a post-production phase, consequently it cannot be applied to a live show.

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 initiative 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. 1 shows the real actors performing in the VR stage and the corresponding action of their avatars in the CG virtual set.



Fig. 1. VR Theater - VR Stage and CG Virtual Set.

3.2 Live 3D Cinema

Live 3D Digital Cinema is the technology behind the non-immersive Audio-Visual presentation format of VR Kino+Theater. It consists of the Computer Graphics infra-structure for Animation, Real-Time Simulation and Rendering of the experience.

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

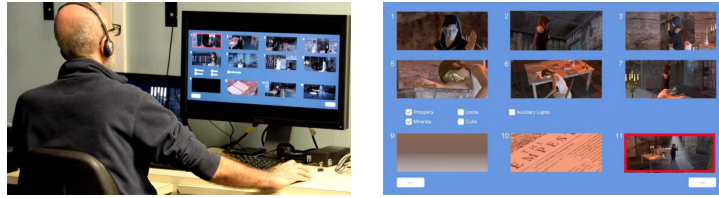


Fig. 2. Director operating the camera switcher and detail of the interface.

Fig. 3 shows the image selected by the director at a moment of the presentation (the lower right camera in the interface).



Fig. 3. Image of the selected camera exhibited on the movie screen.

4 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.

4.1 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.

4.2 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 knows about the entities in a Unity scene and controls the camera specification based on visual composition rules.

The two main control mechanisms are the Composer and the Transposer, they allow to specify the camera in screen and scene space respectively.

4.3 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.

4.4 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. 4 shows the Kino+Theater Camera Switcher interface.

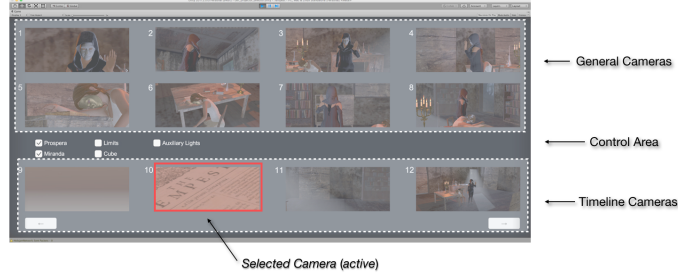


Fig. 4. Multi-camera switcher interface.

5 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 motivates 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 DJ, 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.

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.

5.1 Stylistic Control

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

The Shot Classes are related to visual characteristics of the movie image. 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.

5.2 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. 4 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. 5 shows the complete camera abstraction hierarchy.

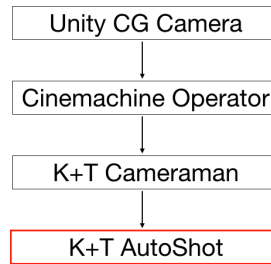


Fig. 5. Camera Abstraction Hierarchy.

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.

6 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.

6.1 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. Furthermore, they are defined per scene, i.e., they depend on the narrative content and the staging of specific scenes.

For example, Fig. 6 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. 6. Camera Group for CloseUp Shots of the Cell Scene in The Tempest.

6.2 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.

Close Miranda	Close Prospera
0.2	0.8

Table 1. Node R1

As a whole, the cut graph models a probability 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. 7 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.

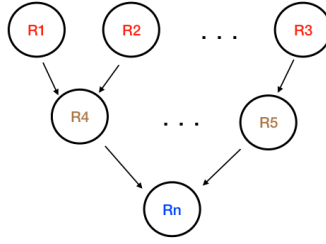


Fig. 7. Example of a Cut Graph.

6.3 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 $\mathcal{P} = \{P_\theta \in \mathcal{P} : \theta \in \Theta\}$, where Θ 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).

6.4 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. 8 illustrate the Timeline and Track Marker layers.

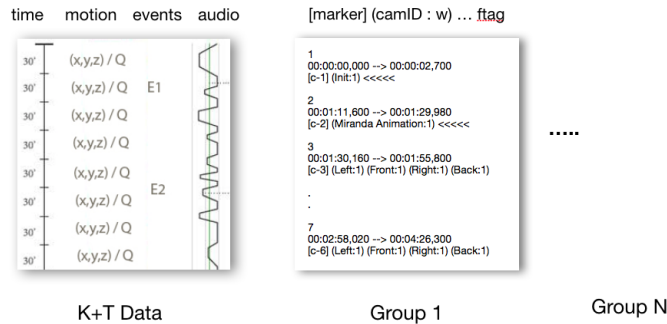


Fig. 8. Timeline and Track Marker Layers.

6.5 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. 9. Track Marker File.

For non-parametric models the specification is 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. 10. 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. 11 shows an example of a style design using both non-parametric and parametric models.

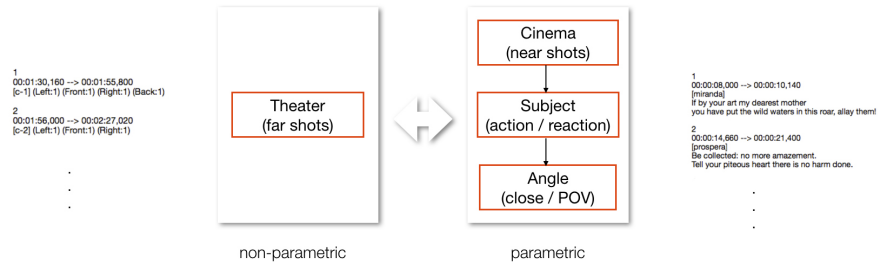


Fig. 11. Cinematic Style Design using parametric and non-parametric models.

6.6 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. 12 shows the Editing interface. Note the highlighted area indicating the style controls.

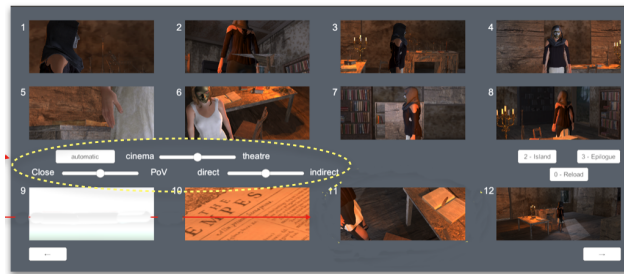


Fig. 12. Probabilistic Editing Interface.

7 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].

7.1 Cameras

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

The cinematic style design is the one depicted in Fig. 11. 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.

7.2 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 100% of Film style and 100% of Theatrical style. In the Film style montage, only near shots have been used and the cuts performed when each characters started a new line 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% / 50% chance).

7.3 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 50% Film and 50% Theatrical montage produced a result very close to a realistic editing setting, with a well balanced combination of wide and near shots and a good cut dynamics.

8 Conclusions and Future Work

In this paper we described a new expressive tool for Audio / Visual Presentations in the context of Live Cinema. It extends the VR Kino+Theater image switcher to provide high level style controls that allows the director to act as a DJ.

Future work goes into two directions: investigate the inverse approach to generate a Cut Graph, and experiment with an editing interface controlled by the viewer.

Our current work proposes 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.

Finally, another way to explore the ideas discussed in this paper is in a live non-linear editing setting where 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.

References

1. Company, E.: Eko studio (2017), <https://studio.helloeko.com/>
2. Coppola, F.: Live Cinema and Its Techniques. Liveright (2017)
3. Jordan, M.I.: Graphical models. *Statist. Sci.* **19**(1), 140–155 (02 2004)
4. Leake, M., Davis, A., Truong, A., Agrawala, M.: Computational video editing for dialogue-driven scenes. *ACM Trans. Graph.* **36**(4), 130:1–130:14 (Jul 2017)
5. Technologies, U.: Web portal (2017), <https://unity3d.com/>
6. Velho, L., Carvalho, L., Lucio, D.: Vr kino+theater: a platform for the future digital media. Technical Report TR-01-2018, VISGRAF Lab - IMPA (2018)
7. Velho, L., Lucio, D., Carvalho, L.: Situated participatory virtual reality. In: Proceedings of XVI Simposio Brasileiro de Jogos e Entretenimento Digital (2017)
8. Velho, L., Prazeres, M., Carvalho, L., Lucio, D., Beato, A., Cronemberger, E., Giannella, J., Rolla, V., Varvaki, H., Abreu, D.S., Yates, R.: Making the tempest. Technical Report TR-02-2018, VISGRAF Lab - IMPA (2018), <https://www.visgrafimpa.br/tempest/>