

# Computational Model of Film Editing for Interactive Storytelling

Christophe Lino, Mathieu Chollet, Marc Christie, Rémi Ronfard

► **To cite this version:**

Christophe Lino, Mathieu Chollet, Marc Christie, Rémi Ronfard. Computational Model of Film Editing for Interactive Storytelling. Mei Si and David Thue and Elisabeth André and James C. Lester and Joshua Tanenbaum and Veronica Zammitto. ICIDS 2011 - International Conference on Interactive Digital Storytelling, Nov 2011, Vancouver, Canada. Springer, 7069, pp.305-308, 2011, <10.1007/978-3-642-25289-1\_35>. <hal-00646405>

**HAL Id: hal-00646405**

**<https://hal.inria.fr/hal-00646405>**

Submitted on 19 Feb 2013

**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.

# Computational Model of Film Editing for Interactive Storytelling

Christophe Lino<sup>1,2</sup>, Mathieu Chollet<sup>1,3</sup>, Marc Christie<sup>1</sup>, Remi Ronfard<sup>3</sup>

<sup>1</sup> MIMETIC, INRIA Rennes Bretagne - Atlantique, France

<sup>2</sup> School of Computing Science, Newcastle University, UK

<sup>3</sup> IMAGINE, INRIA Grenoble Rhone Alpes, France

**Abstract.** Generating interactive narratives as movies requires knowledge in cinematography (camera placement, framing, lighting) and film editing (cutting between cameras). We present a framework for generating a well-edited movie from interactively generated scene contents and cameras. Our system computes a sequence of *shots* by simultaneously choosing which camera to use, when to cut in and out of the shot, and which camera to cut to.

**Keywords:** Camera planning, Virtual Cinematography

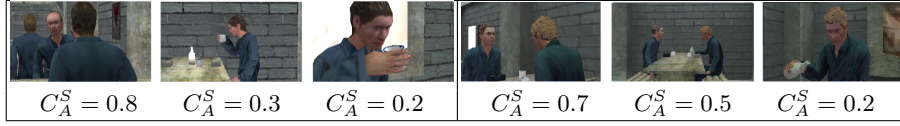
## 1 Introduction

In interactive storytelling, it is useful to present 3D animation in a cinematic style, which means selecting appropriate cameras and appropriate inter-cutting between cameras to properly convey the narrative. We propose an optimization framework for selecting shots and cuts while the narrative unfolds, based on a relatively simple scoring scheme driven by working practices of film and television [3, 6]. We cast the problem of film editing as selecting a path inside an *editing graph* which consists of a collection of evolving film takes (a take is a continuous sequence of images from a given camera) and precisely deciding when to cut in and out of film takes. In contrast to related work, we also account for a precise enforcement of pacing (rhythm at which cuts are performed). We propose an algorithm suitable for online editing which uses an efficient best-first search technique. The algorithm relies on short-term anticipation to improve quality in cuts and produce movies consistent with the rules of cinematography and editing, including shot composition, continuity editing and pacing.

The paper is organized in two parts. The first part describes the score functions used to evaluate shots, transitions and pacing illustrated by a number of examples. The second part explains the search process for exploring the editing graph during the storytelling process with a very minimal lookahead.

## 2 Film Grammar Rules

In our system the score of a movie is built up incrementally as the sum of the scores of its shot fragments and transitions. The cost per fragment (a fragment



**Fig. 1.** Action costs. Left: Three shots of a drinking action. Right: Three shots of a pouring action.

is a part of a take of duration  $\Delta t$ ) is evaluated as a weighted sum of all violations of the rules of frame composition  $w_k \times C_k^S$ . And similarly, the cost of a transition (or cut) is evaluated as a weighted sum of all violations of the rules of editing  $w_l \times C_l^T$  (see equation 1).

We compute a complete sequence  $s$  as a sequence of shots  $s_i$  of durations  $d_i$  and cuts between shots. Each shot  $s_i$  is processed as the concatenation of fragments  $f(t)$  where  $t$  is a time interval of length  $\Delta t$ . We then assume that the cost of  $s$  is the sum of the costs for all of its fragments and cuts, plus a function  $C^P$  of the durations of shots:

$$(1) \quad C(s) = \sum_t \left( \sum_k w_k \times C_k^S(f(t), t) + \sum_l w_l \times C_l^T(f(t), f(t+1)) \right) + \sum_i C^P(d_i)$$

## 2.1 Shot composition

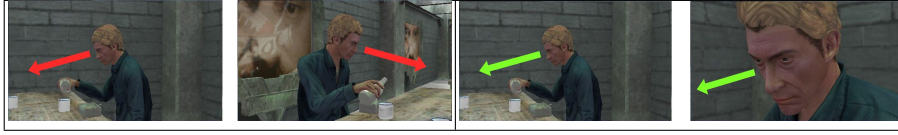
The cost of a shot fragment integrates the violation of three terms: action, visibility, and composition. The action term  $C_A^S(t)$  measures the amount of the scene action which is missed in the given fragment, computed as a sum over all actions  $a$  occurring during the fragment:

$$(2) \quad C_A^S(t) = \sum_a imp(a) \times M_A[type(a), size(a), angle(a)]$$

where  $type(a)$  is the type of action,  $imp(a)$  its importance in the narrative,  $size(a)$  the screen size of its protagonist and  $angle(a)$  the profile angle of its protagonist. The action matrix  $M_A$  contains empirical preferences for shot framings as a function of action types. Figure 1 illustrates the preferences for shot sizes and profiles for the special case of two actions types: pouring and drinking.

## 2.2 Shot transitions

A transition between shots causes a visual discontinuity. The art of the editor is to minimize the perception of discontinuity by selecting appropriate shots and moments for cutting [1, 2, 5]. In this work, we compute the cost of a cut as a sum of terms measuring discontinuities in the screen positions, gaze directions and motion directions of actors. Moreover, we weight each term with the screen size  $size(i)$  of each actor  $i$ , so that continuity in the foreground receives a larger reward than in the background.



**Fig. 2.** Gaze continuity. Left: the gaze direction of the main character changes, resulting in a poor cut. Right: keeping the gaze directions consistent results in a better cut.

**Screen continuity** Screen continuity prevents actors in successive shots to appear to jump around the screen. Because the actor’s eyes are the most important center of attention, we favor transitions which maintain the actor’s eyes at the same screen location.

**Gaze continuity** When watching a movie the gaze direction of actors should not change. We thus use a cost function that penalizes camera changes that cause apparent reversals in the actor’s gaze directions.

$$(3) \quad C_{GAZE}^T = \sum_{i \in screen(f_1) \cap screen(f_2)} size(i) \times \delta(sign(x_{i,f_1}^G) - sign(x_{i,f_2}^G))$$

where  $screen(f_k)$  represents the actors that project on the screen during fragment  $f_k$ ,  $x_{i,f_k}^G$  is the horizontal on-screen coordinate of the gaze direction for actor  $i$  in fragment  $f_k$  and  $\delta$  is the Kroneker symbol. Figure 2 shows two cuts with increasing gaze continuity scores.

**Motion continuity** The motion direction of actors in two successive shots should not change also. We thus use a cost function that penalizes camera changes that cause apparent reversals in the actor’s motion, defined as:

$$(4) \quad C_{MOTION}^T = \sum_{i \in screen(f_1) \cap screen(f_2)} size(i) \times \delta(sign(x_{i,f_1}^M) - sign(x_{i,f_2}^M))$$

where  $x_{i,f_k}^M$  is the screen motion of the actor’s eyes in fragment  $f_k$  measured in the horizontal on-screen direction, and  $\delta$  is the Kroneker symbol.

### 2.3 Shot durations

To control the pace of the editing, we introduce a duration cost per shot, measuring the deviation from a log normal law, where  $d$  is the duration of the shot,  $\mu$  and  $\sigma$  are resp. the mean and the standard deviation of the log normal law:

$$(5) \quad C^P(d) = \frac{(\log(d) - \mu)^2}{2\sigma^2}$$

The log-normal distribution is a compact and discriminative representation of shot durations in movies as well as sentence lengths in natural language [4] and its two parameters can be used as a signature of film editing or writing styles.

### 3 Film editing as path finding

The computation of an optimal sequence of shots consists in searching the path of least cost in our editing graph. Exact and efficient algorithms exist for computing a solution offline. For interactive storytelling applications, we instead describe an approximate method that chooses shots and cuts incrementally as the story unfolds and runs at interactive framerates. At a given depth in the search process (i.e. advancement in time over the fragments), a decision needs to be made whether to stay within the current shot or perform a cut to a shot in an other take. We use a short observation window over the next  $W$  fragments to compute the best moment for transition. Given the current shot is  $s_c$ , for a given time  $t$  in the observation window and for each shot  $s_i \neq s_c$ , we compute the cost  $C^{CUT}$  of a possible transition from shot  $s_c$  to shot  $s_i$ , and we compare it to the cost  $C^{NOCUT}$  of staying in the current shot.

If  $C^{NOCUT}(s_c) \leq \min_i C^{CUT}(s_c, s_i)$  (i.e. the cost of staying in  $s_c$  is the minimal cost), we extend the duration of  $s_c$  by  $\Delta t$  and the observation window is shifted a fragment ahead. If there exists a shot  $s_i$  such that  $C^{CUT}(s_c, s_i) < C^{NOCUT}(s_c)$  at time  $t$ , we need to know whether to cut at the current time  $t$  to shot  $s_i$ , or to wait for a better moment. To implement this, we scan successive fragments at  $t + \Delta t, t + 2\Delta t, ..t + W\Delta t$  in the observation window until a cost lower than  $C_{CUT}(s_c, s_i)$  is found. In such case, the best cut occurs later and the observation window can be shifted a fragment ahead. Otherwise,  $t$  represents the best moment for a transition and a cut is performed towards shot  $s_i$ . Results are presented here [http://sites.google.com/site/christophelino/work/film\\_editing](http://sites.google.com/site/christophelino/work/film_editing).

### 4 Conclusion

We have introduced a novel framework for virtual cinematography and editing which adds an evaluation function to previous approaches. Preliminary results demonstrate that our approach is efficient in separating correct from incorrect shot sequences in complex narratives with many actors and actions, and is thus appropriate for future research in film-mediated interactive storytelling.

This work has been funded (in part) by the European Commission under grant agreement IRIS (FP7-ICT-231824).

### References

1. Berliner, T., Cohen, D.J.: The illusion of continuity: Active perception and the classical editing system. *Journal of Film and Video* 63(1), 44–63 (2011)
2. D'Ydewalle, G., Vanderbeeken, M.: *Perceptual and cognitive processing of editing rules in film* (1990)
3. Murch, W.: *In the blink of an eye* (1986)
4. Salt, B.: *Film Style and Technology: History and Analysis*. Starword (2003)
5. Smith, T.J.: *An Attentional Theory of Continuity Editing*. Ph.D. thesis, University of Edinburgh (2005)
6. Thompson, R.: *Grammar of the Edit*. Focal Press (1993)