

Comparing film-editing

Quentin Galvane¹, Rémi Ronfard¹ and Marc Christie²

¹Inria, Univ. Grenoble Alpes & CNRS (LJK), Grenoble, France

²Irisa & University of Rennes 1, France



Figure 1: Same shot from *Back To The Future* from different sources

Abstract

Through a precise 3D animated reconstruction of a key scene in the movie "Back to the Future" directed by Robert Zemeckis, we are able to make a detailed comparison of two very different versions of editing. The first version closely follows film editor Arthur Schmidt original sequence of shots cut in the movie. The second version is automatically generated using our recent algorithm [GRLC15] using the same choice of cameras. A shot-by-shot and cut-by-cut comparison demonstrates that our algorithm provides a remarkably pleasant and valid solution, even in such a rich narrative context, which differs significantly from the original version more than 60% of the time. Our explanation is that our version avoids stylistic effects while the original version favors such effects and uses them effectively. As a result, we suggest that our algorithm can be thought of as a baseline ("film-editing zero degree") for future work on film-editing style.

1. Introduction

Research on automatic film-making has been conducted for many years. More specifically, the problem of automatic film-editing has been addressed several times with different approaches [HCS96, ER07, GRLC15]. This paper presents a comparative analysis of human-made editing and automatically computed editing. The analysis aims to provide a better understanding of the decision process in film-editing and discusses areas of improvement for automatic editing.

As discussed in [LRGG14], the evaluation of editing systems is a difficult task that remains open to discussions. In previous work, evaluation mostly consisted of user studies to assess the quality of the results. Here we adopt a qualitative approach to analyse both computed and human made editing.

After presenting the essential concepts of film-editing and introducing different editing systems, we focus on the com-

parison between two edits of the same sequence. We compare them based on a qualitative evaluation of the shots, cuts and rhythm used by the automatic system. We analyse the differences to better understand the motivations of an editor and consider possible improvements for the editing system.

2. Background

In this section we first present an overview of basic editing principles later used in the paper for the purpose of the analysis. We then review different automatic editing approaches and detail the one used in this paper. Finally we present the sequence from *Back To The Future* used in this paper to perform the comparison.

2.1. film-editing

film-editing is the task of selecting shots to combine them into sequences that finally create a finished motion picture.

There exist many different styles of video editing, but for this study we focus on continuity editing, the predominant style of editing in Hollywood. The goals of continuity editing are to *minimize the awareness of cuts, create the perception of "continuity" across a cut and ensure that "continuity" is not violated as a consequence of a cut* [Smi05]. To accomplish these goals, continuity editing relies on rules: left-to-right continuity (often referred to as the 180° rule), spatial continuity (position, movement and gaze) and jumpcuts.

But obviously, editing is not limited to making cuts. The core of editing is to select the proper shot that best conveys the story. This idea is expressed by the Hitchcock principle [TS67, Haw05, DeL09, GRLC15] which claims that *the size of a character on the screen should be proportional to its narrative importance in the story*. Selecting a shot then consists in finding the one that best balances the importance of the characters with their perceived size on the screen to reach an "Hitchcock equilibrium".

For live-action movies the problems of cinematography and editing are strongly correlated. Directors often work from storyboards. They usually take decisions beforehand and only use a limited number of cameras (reducing the work of editors). Most of the editing rules mentioned above are already considered during the shooting phase. Figure 2 from [ZKSC85] shows how precisely the shots are defined. Before M.J. Fox another actor had been chosen to interpret Marty McFly in *Back to the Future*. After five weeks of shooting, the director decided to change the main actor and had to start over. Figures 2a and 2b show how similar the shots are: the director already knew exactly what he wanted for the end result and shot the sequence only from very specific viewpoints.



(a) E. Stoltz as Marty McFly (b) M.J. Fox as Marty McFly

Figure 2: Shots of Eric Stoltz and Michael J. Fox as Marty McFly in *Back To The Future*

Automating such subjective decision processes is obviously a complex task.

2.2. Automatic film-editing

In the past decades, several approaches have offered interesting solutions to automatic film-editing. Starting in 1996 the Declarative Camera Control Language (DCCL) by [CAH*96] first introduced idiom-based approaches (also developed in [HCS96]). An idiom is a stereotypical way of conveying a specific action in a scene through a sequence

of shots. Solutions based on film idioms are close to live-cinematography as they try to imitate and simplify the process by combining director and editor: decisions can be made on the fly. Nevertheless, such approaches fail to be extensible due to the burden of creating new idioms for each style, action and context. Moreover, they cannot be considered fully automatic as they still require expert knowledge for the creation of idioms.

Other works addressed the issue of automatic film-editing [TBN00, JY11, KM02] but mostly focused on specific aspects of the editing without seriously considering the cinematography (quality of the shots) or editing rules. Another approach consists of considering film-editing as an optimization problem. The Cambot system, presented in [ER07], optimizes editing using heuristics for shot selection and cuts. Though being novel and efficient, this work does not account for the pacing and does not provide any details on the heuristics used for the optimization. Finally, in a previous paper [GRLC15], we introduced a new solution based on a semi-markov model. It also uses a dynamic programming approach, but precisely describes the evaluation functions used for the optimization. Moreover the semi-markov model used allows control over the global pacing of the edit.

In this last paper, a user study was conducted to prove the necessity of each of the criteria used to optimize the edit. This "subjective analysis" proved the validity of the methods without searching for improvement. Our goal here is to complete this study with an extensive qualitative comparison over the detailed criteria.

2.3. Back To The Future

For the purpose of this comparative analysis, we needed to compare two different edits of the same sequence. To produce these two edits we need to have access to the unedited footage of the sequence. Raw footage is not easy to come by. To overcome this difficulty we used a dataset we made public recently [GRLC15]. It contains a 3D animation that recreates a complete sequence of the movie *Back To The Future* directed by R. Zemeckis and edited by H. Keramidas and A. Schmidt. It also contains the camera used in the movies along with some extra cameras. The recreated sequence contains interesting interactions between the characters (dialogs, motions, etc.) that could be filmed from many angles. The set of 25 cameras placed in the scene offers a large range of possibilities to the system.

3. Comparison

In this section, we analyze the differences between an automatically generated edit and the original sequence of the movie. The generated sequence was computed using the semi-markov model presented in 2.2. Figure 17 in the appendix illustrates the two edits and will be used as refer-

ence for this analysis. We observe that 35% of the shots are shared by the two edits. Thus, 65% of the time, the director and/or editor took a different decision. To better understand these differences we now present a detailed comparison of the three aspects of editing: shot selection, cut and rhythm of the sequence. The first aspect is the core of the shot selection process, and so we detail it more extensively by looking at each of its criteria. We then analyze the main cutting decisions and mistakes made by R. Zemeckis that are detected by our system. Finally, we compare the shot durations of the two sequences.

For this analysis, we computed the cost of each criterion for the two edits. To highlight the differences we display these costs using the colormap in Figure 3 (blue for a low cost and red for a high cost).

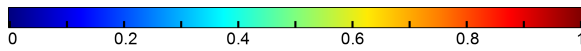


Figure 3: Colormap used to display the cost values. Lowest cost (0) are displayed in blue and highest cost (1) in red.

3.1. Action Visibility

The first analyzed criterion is the action visibility. The cost function devised in [GRLC15] penalizes shots that do not properly capture unfolding actions. It looks at the *bodyparts* of a character taking part in an action and compute their visibility on the screen. The cost is computed as the sum of the occluded area of these *bodyparts*, relative to their total area and weighted by their narrative importance.

Figure 4 highlights several significant differences in action visibility between the original and generated sequences.

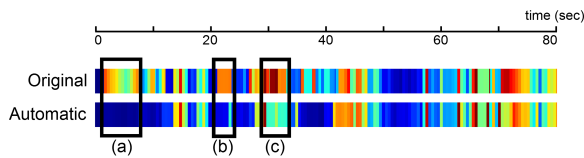


Figure 4: Action visibility costs computed throughout the whole sequence for the original movie and the automatically generated sequence. Main differences are highlighted in (a),(b) and (c) where the visibility of the characters in Zemeckis' version is bad.

Figure 5 illustrates the difference in visibility highlighted in Figure 4(a). At this frame, the only occurring action is Marty, staring at George. The lack of visibility on Marty's face was obviously orchestrated by the director in order to slowly reveal Marty's reaction. While our system safely chose a shot with the proper visibility on the two characters for the whole duration of the action (see Figure 5b), R.

Zemeckis uses this lack of visibility to drag the audience's interest toward Marty's appearing face and emphasize his reaction.



(a) Shot from the movie (b) Computed shot

Figure 5: Shot taken at $t = 3s$ when Marty stares at George. Zemeckis uses the lack of visibility to drag the audience's attention (a). The generated sequence uses a shot with perfect visibility over the two characters (b).

When trying to automate a process as complex as video editing, one is bound to make simplifying assumptions. Even though assuming that poor visibility is synonymous with poor shot quality might sound reasonable, in some circumstances, it might not be the case. With this optimization based approach, the goal is only to avoid making mistakes. Handling such a motivated and complex shot would require a lot more reasoning on the actions and computation of importance.

3.2. Hitchcock Principle

The next important criterion used in shot selection is the action ordering. It is based on the Hitchcock principle mentioned in Chapter 2.1. This term penalizes the shots where the on-screen importance of a character does not match its narrative importance.

Figure 6 highlights some strong differences regarding hitchcock principle's cost.

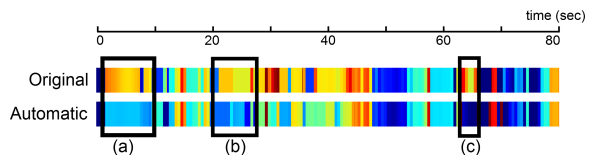


Figure 6: Hitchcock costs computed through the whole sequence for the original movie and the automatically generated sequence. Main differences are highlighted in (a),(b) and (c) where the generated version has a better *Hitchcock equilibrium*.

The shots in Figures 5a and 5b also illustrate the difference (a) of Figure 6. At this specific moment in the story, Marty is the most important character, followed by George. In Figure 5a, it is obvious that the narrative importance of the characters does not match their screen sizes, as Marty barely appears in the screen. During this shot, the cost slowly decreases with the appearance of Marty in the frame which

slowly reaches a "Hitchcock equilibrium". This example illustrates one of the current limitations of the system. It does not allow this form of intensification. It would require a variable importance within the action itself.

Figure 7a offers another illustration of Hitchcock principle violation (see Figure 6(b)). It is taken during a dialog between George and Goldie and yet most of the screen space is occupied by Marty in the foreground. Figure 7b shows the automatically selected shot. This one satisfies the Hitchcock principle as it focuses on the two characters involved.

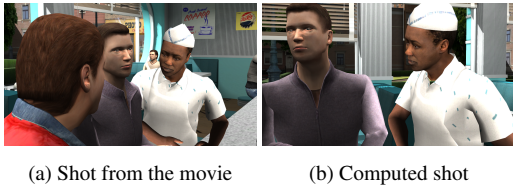


Figure 7: Shot taken at $t = 23s$ when Goldie talks to George. *Zemeckis* included the main protagonist in the frame (a). Only characters involved in occurring actions appear in the generated version (b).

Here Marty is not involved in this specific dialog, but his presence on the screen is important since he is the main protagonist of the movie. It shows that he is listening to the conversation and, thus, gives information to the audience on his understanding of the situation. This limitation does not come from the Hitchcock principle but rather from the computation of importance itself. It does not consider any higher level of importance or involvement in the situation (such as a three person dialog) or the global story.

3.3. Action Proximity

Finally, the last term used for the evaluation of a shot is the action proximity. It aims at maximizing the amount of actions visible in the screen by penalizing shots with unused screen-space (i.e. that does not contain characters or objects). This term has less importance than the other two and is essentially used to decide between cameras with similar visibility and action ordering.

Figure 8 shows the computed cost of action proximity for both the original movie and the computed sequence. Unlike the other two terms, the results are similar.

The few differences are due to different shot selection based on other criteria. For example, Figure 9 shows that the difference in "occupied screen-space" is due to a different shot selection. The automatic approach chooses to focus on George and Goldie, whereas *R. Zemeckis* selected the shot with Marty's reaction.

This sequence of *Back To The Future* is not appropriate for analyzing this criterion due to the proximity of the characters and the cameras in the confined environment of the bar.

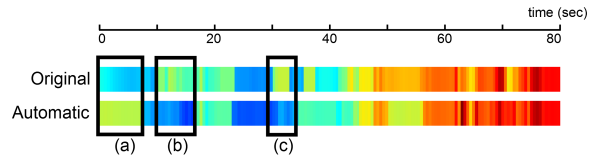


Figure 8: Action proximity costs computed through the whole sequence for the original movie and the automatically generated sequence. Main differences are highlighted where the proximity of the characters is better in *Zemeckis*' version (a) or in the generated version (b) and (c).

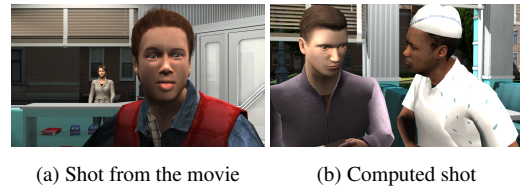


Figure 9: Shot taken at $t = 31s$ when Goldie talks to George and Marty stares at them. Characters occupy more screen space in the generated version (b) than in the original one (a).

3.4. Cuts and continuity editing

In this section, we analyze the quality of the cuts with regards to the continuity editing style. Figure 10 shows the computed costs of each cut (the value displayed for each shot is the cost of the previous cut). This cost is a weighted sum of costs computed from the different continuity rules mentioned in section 2.1 with an emphasis on the left-to-right continuity and jump-cut rules.

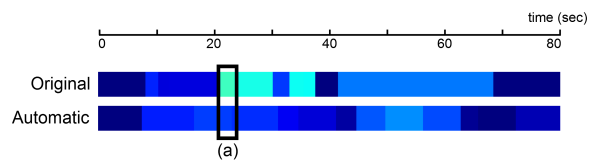


Figure 10: Cut costs computed through the whole sequence for the original movie and the automatically generated sequence. In both version, only minor transgressions can be noticed (a)

For both the original and automatic editing, only minor transgressions can be noticed, as illustrated in Figure 11 with the spatial displacement of Marty in screen space. None of the two sequences have jump-cuts or left-to-right discontinuity.

This result not only confirms that the original sequence from *Back To The Future* satisfies the rules of continuity but also that the optimization based approach gives a proper implementation of the continuity editing style for this dataset.

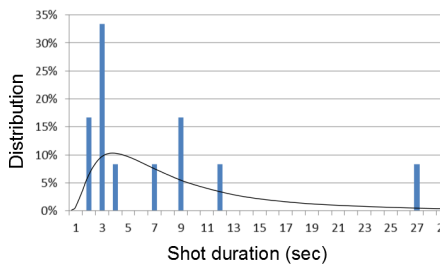


Figure 11: Cutting discontinuity: the position of *Marty* significantly changes from one shot to another, introducing a position discontinuity.

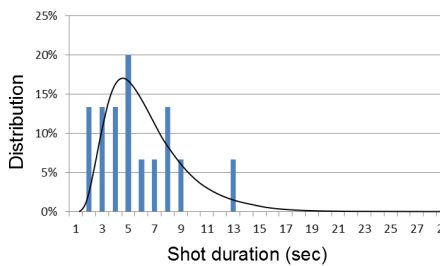
3.5. Pacing

Finally, the last element of film-editing that we analyze and compare with the original sequence of *Back To The Future* is the rhythm. In [GRLC15], the cost function that is used to evaluate the quality of the pacing is based on previous studies ([Sal09, CDN10]) showing that the shot durations in a movie follow a log-normal distribution. For each shot duration, a cost is computed using the density function of the desired log-normal distribution (defined by an ASL and a standard deviation).

For the generated sequence, an average shot length of 5.25s was used as parameter of the cost function. Figure 12 shows the distributions of the shot durations for the two versions.



(a) Back To The Future: ASL = 6.64s; standard deviation = 0.82



(b) Generated sequence: ASL = 5.31s; standard deviation = 0.51

Figure 12: Shot duration distributions

Despite the two distributions being similar and both rel-

atively close to the computed log-normal distribution (over the whole sequence), the cumulative cost of the pacing is four times larger with Zemeckis' version than our automatically generated sequence. The explanation appears in Figure 13, which shows the computed pacing cost for each shot of the sequence. Two categories of "bad" pacing can be identified in the graph: very short takes and very long takes.

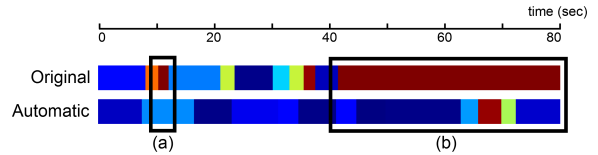


Figure 13: Pacing costs computed through the whole sequence for the original movie and the automatically generated sequence. Important costs are computed in the original version for very short (a) and very long (b) shots.

The high cost highlighted in Figure 13(a) is due to a very small shot duration (see Figure 14(b)). This shot breaks the rhythm of the sequence to show the short reaction of the character. In the automatically generated version, the same sequence is handled using two longer shots that cover several actions.



Figure 14: Shot sequence from the original movie. A very short shot (b) is inserted to show *Marty's* reaction.



Figure 15: Shot sequence from the generated version. Only shots with all characters involved in the occurring actions are used.

In Zemeckis' movie, the last two shots of the sequence are long takes lasting 27 seconds and 11 seconds with elaborate panning camera motion (see Figure 16). Due to their deviation from the ASL, the cost of these two shots is very high, as shown in Figure 13(b). In the computer-generated version, those same 38 seconds are handled with ten different shots from six different viewpoints. This gives a different dynamic to the scene, but does not make it a better solution. By preventing large deviation from the ASL, the computer-generated version sometimes fails to find better solutions. Future work is needed to better understand how to handle such cases, where the quality of extended shots

with elaborate camera movements should probably be given more weight.



Figure 16: Very long take from the original movie handled with an elaborate panning camera motion.

4. Conclusions

In this paper, we gave a thorough analysis of the results of an automatic film-editing technique. The comparison with the original sequence gave many leads for future work on film-editing style. This comparative study showed that the sequences generated by the system minimize violations of editing and cinematographic rules at the expense of stylistic decisions. It confirms the conclusion of the user study: the system generates valid solutions avoiding common editing mistakes and constitutes a strong basis for automatic editing. Nevertheless, future work is yet to be conducted to handle more complex situations and generate more sophisticated and stylistic sequences. This analysis emphasizes the need to use more complex models to better understand and use the narrative discourse. These elaborated models are indeed necessary to improve the implementation of the Hitchcock principle. The need to adapt the pacing for specific situations is also highlighted. It should account for camera motions to consider variation of the rhythm, as illustrated in the original sequence of *Back To The Future*.

Finally, we have found that the high-quality 3D reconstruction of a movie scene has provided useful insights into the art of film-editing. We would like to invite other researchers in the field to create similar benchmark scenes. Indeed, further analysis should be conducted over different sequences from different movies styles to compare computer-generated editing solutions with professionally produced movies using other styles.

References

- [CAH*96] CHRISTIANSON D. B., ANDERSON S. E., HE L.-W., WELD D. S., COHEN M. F., SALESIN D. H.: Declarative camera control for automatic cinematography. In *AAAI* (1996), pp. 148–155. 2
- [CDN10] CUTTING J. E., DELONG J. E., NOTHELPER C. E.: Attention and the Evolution of Hollywood Film. *Psychological Science* 21, 3 (Mar. 2010), 432–439. URL: <http://dx.doi.org/10.1177/0956797610361679>, doi:10.1177/0956797610361679. 5
- [DeL09] DELOURA M.: *Real Time Cameras, A Guide for Game Designers and Developers*. Morgan Kaufman, 2009. 2

- [ER07] ELSON D. K., RIEDL M. O.: A lightweight intelligent virtual cinematography system for machinima generation. In *Artificial Intelligence and Interactive Digital Entertainment (AIIDE '07)* (2007). 1, 2
- [GRLC15] GALVANE Q., RONFARD R., LINO C., CHRISTIE M.: Continuity Editing for 3D Animation. In *AAAI Conference on Artificial Intelligence* (Austin, Texas, United States, Jan. 2015), AAAI Press. URL: <https://hal.inria.fr/hal-01088561>. 1, 2, 3, 5
- [Haw05] HAWKINS B.: *Real-Time Cinematography for Games*. Charles River Media, 2005. 2
- [HCS96] HE L.-W., COHEN M. F., SALESIN D. H.: The virtual cinematographer: a paradigm for automatic real-time camera control and directing. In *SIGGRAPH* (1996), ACM, pp. 217–224. URL: <http://doi.acm.org/10.1145/237170.237259>, doi:10.1145/237170.237259. 1, 2
- [JY11] JHALA A., YOUNG R. M.: Intelligent machinima generation for visual storytelling. In *Artificial Intelligence for Computer Games*. Springer New York, 2011, pp. 151–170. 2
- [KM02] KENNEDY K., MERCER R. E.: Planning animation cinematography and shot structure to communicate theme and mood. In *Proceedings of the 2Nd International Symposium on Smart Graphics* (New York, NY, USA, 2002), SMARTGRAPH '02, ACM, pp. 1–8. 2
- [LRGG14] LINO C., RONFARD R., GALVANE Q., GLEICHER M.: How Do We Evaluate the Quality of Computational Editing Systems? In *AAAI Workshop on Intelligent Cinematography And Editing* (Québec, Canada, July 2014). URL: <https://hal.inria.fr/hal-00994106>. 1
- [Sal09] SALT B.: *Film Style and Technology: History and Analysis* (3 ed.). Starword, 2009. 5
- [Smi05] SMITH T. J.: *An Attentional Theory of Continuity Editing*. PhD thesis, University of Edinburgh, 2005. 2
- [TBN00] TOMLINSON B., BLUMBERG B., NAIN D.: Expressive autonomous cinematography for interactive virtual environments. In *Proceedings of the Fourth International Conference on Autonomous Agents* (New York, NY, USA, 2000), AGENTS '00, ACM, pp. 317–324. 2
- [TS67] TRUFFAUT F., SCOTT H. G.: *Truffaut/Hitchcock*. Simon & Schuster, 1967. 2
- [ZKSC85] ZEMECKIS R., KERAMIDAS H., SCHMIDT A. F. E., CUNDRY D. C.: *Back to the future*. Universal Pictures, 1985. 2





Figure 17: Comparison of Human and computer generated edits of the same sequence from 25 cameras. Each camera is assigned a color based on its id (in this figure, colors are not related to the cost of the shots). The scenario describes the actions involving Marty (M.), George (G.), Goldie (Go.), Lou (L.) and the Cashier (C.)