Leveraging Machinima to Characterize Comprehension of Character Motivation

Kara Cassell[1] and R. Michael Young[2]

[1] 3C Institute, Durham, NC, 27713, USA
cassell@3cisd.com
[2] University of Utah, Salt Lake City, UT, 84112
young@eae.utah.edu

Abstract. Deliberation-driven reflective sequences, or *DDRSs*, are cinematic idioms used by film makers to convey the motivations for characters adopting a particular course of action in a story. We report on an experiment where the cinematic generation system Ember was used to create a cinematic sequence with variants making different choices for DDRS use around a single decision point for a single character.

Keywords: Experimental evaluation · cinematic generation · machinima.

1 Introduction

One important aspect of visual narrative is character development. Understanding the deliberations that a character undergoes when deciding between competing goals or courses of action is one key aspect of the insight viewers gain on a character's personality and growth. Conveying the mental processes involved in those deliberations is key to portraying character personality to a narrative consumer. In written narratives this can be done through the use of an internal monologue or by a narrator explicitly describing a character's internal thought processes. In film, storytellers typically use *internal shots*, close ups of characters, to signal to the viewer that characters are thinking. Cinematographers use specific shots, usually *cut-away shots* – shots that briefly cut to some other content and cut back to the previous context – to indicate what it is that the character is thinking about. Cut-away shots help to break up otherwise long static shots, make the sequence more interesting, and (most relevant to this paper) display relevant information to the viewer or audience [1].

The work described in this paper seeks to provide experimental insight into the ways that visual narrative making use of internal shot/cut-away filmic idioms may impact the comprehension of viewers. Previous work by Cassell [4] has developed a computational method for generating these types of idioms in the context of comic panels and machinima, but questions remain about when a generation system should make use of these idioms and what the likely cognitive consequences on viewers would be from their use.

The system *Ember* [3], a visual discourse generator developed by Cassell, is a narrative discourse generation system that reasons about character decision making to construct a discourse that can communicate character decision-making to the viewer. Ember's generative capability was specifically developed to be able to create cinematic sequences that effectively convey these decision making processes using what they call a *reflective sequence*, the pattern of an internal shot, some number of cut-away shots, followed by another internal shot. When the internal shots occur during a deliberation by the character being filmed, and the content of each cut-away shows some aspects of the story world relevant to the deliberation, Cassell calls these shot sequences **deliberation-driven reflective sequences, or DDRSs**. While Ember has demonstrated the ability to generate shot sequences that meet the structural definition of a DDRS, no clear empirical evidence exists that shows the impact on a viewer's comprehension of the story that results from the inclusion of DDRSs in Ember cinematics.

In order to better understand how DDRSs play a role in audience inferences around action choice, we designed an experiment that would compare viewer reports of their inferences across a range of similar machinima sequences, where each cinematic varied only by the presence/absence/content of a DDRS. As we describe below, that experiment showed little differentiation in effect between the various DDRSs that were generated. While the experiment does not shed light on the impact of DDRSs, it does a) provide a methodology for leveraging machinima to explore narrative film comprehension and points to a number of specific revisions to needed follow-on experiments that will provide better experimental control.

2   Background and Related Work

Current cinematic generation research typically falls into one of two main categories: 1) work that focuses on the low-level problems of camera placement, direction, focus, and angle and 2) work that views cinematic generation as a narrative discourse construction problem. Systems that focus on the low level problems typically use constraint solvers or intelligent agents [6, 2, 13, 8]. They make use of knowledge around film-making developed and articulated by cineamtographers (e.g., [11]), such as the rule of thirds or not crossing the line, to guide camera shot selection.

One such system is the one developed by Christianson et. al. [6] that encodes cinematic idioms, or standard sequences of shots, into a declarative camera control language (DCCL). Christie and Normand [7] designed an interactive system which allows the viewer to see similar shots that are created based on cinematic principles. Lino [10] created a system that automatically generated cinematics by breaking a virtual space into what was called *director volumes*. The goal was to encode knowledge of cinematic principles into spaces in the world so the system would be able to place the camera accurately for low level shot manipulation. Work by Wu et. al. [14] has created the *Film Editing Patterns* (FEP) language.

This was designed to formalize standard cinematographic techniques and styles. The language specifies FEP constructs that constrain shot sequence features.

Systems that view cinematic generation as narrative discourse construction use methods similar to those in the natural language generation community. Darshak [9], for example, views cinematics as a hierarchical structure similar to how the work of Moore and Paris [12] views discourse. This hierarchical structure uses abstract and base shots to encode cinematic principles. Darshak then uses a decompositional planner to create a shot sequence using these abstract shots and their decompositions into base shots[5].

The Ember system [3, 4] expands on ideas developed in Darshak, consisting of a partial-order, causal link planner [15] extended with the capability to reason about character decision making during cinematic generation. This reasoning leads to discourse structure intended to help maintain the salience of references to past events, characters or objects that can contribute to the explanation of characters' action-centered decision-making processes. In the experiment described below, we leverage Ember's ability to generate cinematics containing shot sequences that convey DDRSs. The shot specifications created by Ember make assumptions about the inferences drawn by human viewers, and the evaluation seeks to characterize those inferences as a means to further support Ember's expressivity.

3 Brief Summary of the Experiment

The experiment described here sought to gauge the extent to which inclusion of DDRSs in cinematic sequences can affect a viewer's comprehension of characters' decision-making processes. In this experiment, we showed viewers a cinematic sequence constructed by hand but whose story and discourse structure was created automatically by planning systems (In the case of the discourse, the structure was created by Ember). Subjects were divided into five treatment groups, and each treatment group viewed a cinematic story that differed from other groups only in the inclusion of a DDRS that focused on a decision made by the story's protagonist around his future plan for action. After viewing the cinematic, we asked subjects to characterize their judgment about why the protagonist decided on his course of actions. We then compared their judgments with the presence/absence/content of the DDRS they saw to gauge the DDRS impact.

Questions presented to subjects were written to specifically ask why actions had happened in the story. The possible answers were presented as still frames from the cinematic that conveyed specific actions, and subjects were asked to rank the set of still images in order of best answer to worst answer. Ten questions were asked. The first question was concerning the character decision. The next nine were a combination of three causal ancestor questions, three temporal questions, and three causal descendant questions.

87 paid participants were recruited on Amazon Turk and randomly assigned to one of five treatment groups. 25 recruited subjects were excluded due to incomplete surveys for a total of 62 participants.

Each participant viewed the cinematic assigned to their treatment group. After viewing the cinematic, participants were asked to complete a survey. The survey questions asked subjects to select actions that pertained to causal and/or motivational relationships between actions they had seen occur in the cinematics. These actions were presented as still frames showing a representative moment of each relevant action as portrayed previously in the cinematic they viewed.

In this study, we sought to evaluate whether the participants would rank answers that relate to the actions present in the conveyed decision sequences higher than the alternate actions that were not. To do this, we ran a Wilcoxon rank sum test between groups for the ranks of each relevant answer to the first question. We also wanted to evaluate whether the participants would rank answers that relate to the actions present in the conveyed decision sequences higher than their rankings for the same actions if no DDRS was conveyed. To do this, we ran a Wilcoxon rank sum test between groups for the ranks of the same answers.

Interestingly, for all but one of the comparison groups, the comparisons between groups resulted in no significant differences.

4   Discussion and Future Work

We found the lack of significant difference between groups surprising, as we felt in particular that the two of our treatment groups conformed to Hollywood conventions that clearly indicate distinct interpretations about the cause of characters' choices for action.

Two reasons why none of the different DDRS situations, including both control groups, resulted in any significant differences may be:

1. Uncontrolled structural or semantic aspects of the stories designed for the experiments may influence subjects' attribution of a character's motivation.
2. DDRS do not play the role in prompting inferences that we assumed they do.

It may be the case that there are underlying semantics around the sources for the character's motivation that are uncontrolled for in our design. Also, it may be that DDRSs are not able to shift a viewer belief of a character's decision making.

References

1. Arijon, D.: Grammar of the Film Language. Silman-James Press (1976)
2. Bares, W., McDermott, S., Boudreaux, C., Thainimit, S.: Virtual 3D Camera Composition from Frame Constraints. Proceedings of the Eighth ACM International Conference on Multimedia pp. 177–186 (2000)

3. Cassell, B.A., Young, R.M.: Ember, Toward Salience-Based Cinematic Generation. In: Workshop on Intelligent Narrative Technologies at the Ninth Annual AAAI Conference of Artificial Intelligence and Interactive Digital Entertainment (2013)
4. Cassell, K.: Dynamic Generation of Narrative Discourse that Communicates Character Decision Making. North Carolina State University (2019)
5. Cheong, Y.G., Jhala, A., Bae, B.C., Young, R.M.: Automatically generating summary visualizations from game logs. In: AIIDE. pp. 167–172 (2008)
6. Christianson, D.B., Anderson, S.E., He, L.w., Salesin, D.H., Weld, D.S., Cohen, M.F.: Declarative Camera Control for Automatic Cinematography. In: Proceedings of the Thirteenth National Conference on Artificial Intelligence and the Eighth Innovative Applications of Artificial Intelligence Conference. pp. 148–155 (1996)
7. Christie, M., Normand, J.M.: A Semantic Space Partitioning Approach to Virtual Camera Composition. Computer Graphics Forum **24**(3), 247–256 (2005)
8. Elson, D.K., Riedl, M.O.: A Lightweight Intelligent Virtual Cinematography System for Machinima Production. In: Artificial Intelligence and Interactive Digital Entertainment (2007)
9. Jhala, A., Young, R.M.: Cinematic Visual Discourse: Representation, Generation, and Evaluation. IEEE Transactions on Computational Intelligence and AI in Games **2**(2), 69–81 (Jun 2010)
10. Lino, C.: Virtual Camera Control using Dynamic Spatial Partitions. Ph.D. thesis, University Rennes 1 (2013)
11. Monaco, J.: How to read a film: The art, technology, language, history, and theory of film and media. New York: Oxford University Press (1981)
12. Moore, J.D., Paris, C.L.: Planning Text for Advisory Dialogues: Capturing Intentional and Rhetorical Information. Computational Linguistics **19**(4), 651–694 (1993)
13. Tomlinson, B., Blumberg, B., Nain, D.: Expressive Autonomous Cinematography for Interactive Virtual Environments. In: Proceedings of the Fourth International Conference on Autonomous Agents. pp. 317–324. ACM (2000)
14. Wu, H.Y., Palù, F., Ranon, R., Christie, M.: Thinking like a director: Film editing patterns for virtual cinematographic storytelling. ACM Trans. Multimedia Comput. Commun. Appl. **14**(4), 81:1–81:22 (Oct 2018). https://doi.org/10.1145/3241057, http://doi.acm.org.prox.lib.ncsu.edu/10.1145/3241057
15. Young, R.M., Pollack, M.E., Moore, J.D.: Decomposition and causality in partial order planning. In: Proceedings of the Second International Conference on AI and Planning Systems. pp. 188–193 (1994)