

REAL-TIME CAMERA CONTROL FOR INTERACTIVE STORYTELLING

Fred Charles, Jean-Luc Lugin, Marc Cavazza and Steven J. Mead

School of Computing and Mathematics

University of Teesside

Middlesbrough, TS1 3BA,

United Kingdom

E-mail: {f.charles, j.l.lugin, m.o.cavazza, steven.j.mead}@tees.ac.uk

KEYWORDS

Automatic Camera Control, Interactive Storytelling, AI-based Animation, Computer Games.

ABSTRACT

In this paper, we present a fully implemented prototype of real-time cinematic control for character-based interactive storytelling approaches, where story diversity emerges from dynamic interaction between characters. We describe the specificities of real-time cinematic control within a dynamic virtual environment, where events occur at different locations at the same time. We also present results based on a situation of interaction between characters in the unfolding of a story.

INTRODUCTION

Our interactive storytelling system (Cavazza et al., 2002) is based on the interaction between characters' behaviours during the unfolding of the main characters' plans representing their role-play. The (active) spectator can intervene with situations at any time throughout the story, though not at all time.

Because of spatial and temporal constraints, there is a need for a real-time cinematic system that will present to the spectator the on-going situations that are most relevant to the story at the time.

In traditional film production, the director is responsible for the overall decision-making process. He ensures that the narrative is conveyed effectively using the film techniques at his disposal. In our system, the virtual director has a similar role. The virtual director will query the client narrative application and determine in real-time which idiom will best fit the scene, based on the events specifications. To do this, the director requires some information from the application, namely the type of event, number of participants, and emotional or affective context of the story at the current point in its telling.

This paper refers to the specificities of real-time cinematic sequences, being the dynamic nature of stories instantiated from independent representations of character-based roles, and the difficulties in producing a meaningful montage of on-stage events generated in real-time, opposed to scripted film sequences.

In the next sections, we will introduce the important concepts of character-centred storytelling as well as a brief description of our interactive storytelling system. Then, we describe the

traditional cinematographic elements used in automatic camera control systems. Finally, we give an overview of the implementation of our camera system, illustrated by an example based on a typical situation of a conversation between characters.

CHARACTER-CENTRED NARRATIVE REPRESENTATION

We have chosen to use plans to represent individual roles for the characters, rather than the global narrative structure. Each character is associated a plan corresponding to its role or, more precisely, the set of possible role instantiations according to a given storyline. It can be seen as a resource for story generation. Each plan corresponds to the character's role in a given story instantiation: it represents the plot through a character's behaviour (Figure 1). The plot itself consists in the on-stage integration of the various roles through the situations created by the interactions between characters.

The on-stage performances of characters are the translation of each sub-task element of the character's plan, called primitive actions, and perceived from the spectator's point of view as a sequence of meaningful events. These actions are formally represented as syntactic triples (subject, verb, object). The subject represents the protagonist (the actor), the verb corresponds to the action to carry out, and the object represents either a physical object or another character. Though the action can not exist without the subject (actor), it may be performed without the specification of an object.

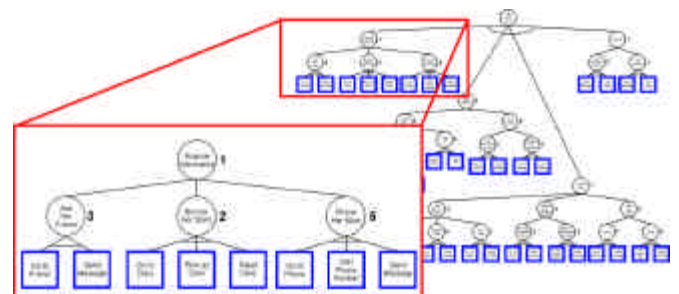


Figure 1: Plan Representation of Character Behaviour

While there are no straightforward rules to convert high-level narrative functions into characters' plans, we have attempted to devise specific rules that could be applicable in the context of the simple genre (sitcoms) with which we are experimenting. The basic hypothesis is that the final story will emerge from the relations that exist between the various characters' plans, these relations being determined from the story genre.

SPACE, TIME AND CAUSALITY

Narrative is a way of comprehending space, time, and causality. Since in film there are at least two important frames of reference for understanding space, time, and causality, narrative in film is the principle by which data is converted from the frame of the screen into a diegesis - a world - that frames a particular story, or sequence of actions, in that world; equally, it is the principle by which data is converted from story onto screen (Branigan 1992).

An important concept in interactive storytelling is causality as a story is defined as a sequence of causally related events. The editing must assure that the consistency of such causal chain of events remains (Raskin, 1998). Causality supports the consequences of interaction, whether it be character-character interaction or spectator intervention. Some interactive storytelling systems make causality explicit in their representations (Young, 2000). However, in a task network representation based on actions and sub-goals, causality is not explicitly represented. One form of implicit causality is the enabling of further actions by their predecessors in the task network ordering, but it is not related to interaction and dynamic generation. Other forms of causality are implicit, illustrated by the interplay of choice and causality in the narrative, which has been described by Raskin (1998).

Many on-stage objects have an intrinsic narrative significance as being resources for characters' actions. In modern narratology (Barthes, 1966), they refer to as a "dispatcher": a dispatcher is an object to which choice is associated, triggering narrative consequences.

In the previous section, we described characters' actions as represented by triples (subject, verb, object). The potential influence of action resources on the unfolding story will emphasise the narrative implication of certain object instances. For instance, a knife or a gun hold a stronger narrative value.

In a plot-based approach (Young, 2000) causality can be explicitly represented whereas in a character-based approach, the anticipation of an action's resolution is dominant. The character-plot duality has thus a translation in terms of causal representations.

Editing rules are meant to preserve and reinforce the narrative continuity of a story. The several parts of the story need to refer sufficiently to each other, allowing the spectator to integrate them into a single chronological sequence of events.

FILM IDIOMS

Perhaps the most significant cinematographer invention is a collection of stereotypical formulas to capturing specific scene as sequence of shots. While there are an infinite variety of idioms, film directors have learnt to rely on a small subset of these. Traditional books (Metz 1974) (Arijon, 1976) provide an informal compilation of formula, along with a discussion of the various situations in which the different formula can be applied.

For example, in a dialogue between two people, a filmmaker might begin with an apex view of both actors, and then alternate views and each following the actor's speech direction, at times using internal placement and at times using external placement.

Expert cinematographers have used cinematic idioms for a long time to direct the flow from frame to frame by representing common shots such as the establishing shot and two shot of conversing players (Mascelli, 1965). Virtual 3D cinematography systems adopted idioms to help generate sequences of prototypical shots to film actions such as conversations between a small group of virtual players (Drucker and Zeltzer, 1995) (He et al., 1996). Other systems (Christianson et al., 1996) (Lu and Zhang, 2001) extended the concept of idiom, a sub-unit of cinematographic expertise, as a means of capturing the essence of a scene.

The concept developed a means of encoding techniques for conveying certain scenarios effectively. By creating an assortment of fairly rigid structure to shoot different kinds of scenes, the Virtual Cinematographer paradigm [5] is limiting itself in two ways. Firstly, the system is limited to create effective shots for scenarios that it is familiar with. Secondly, each transition between two idioms will break the continuity of the scene, creating a rupture in the narrative, hence not adapted to real-time generation.

More recently, Amerson and Kime (2001) have proposed a system for real-time camera control in interactive narratives called FILM (Film Idiom Language and Model). This system, inspired from the Virtual Cinematographer, considerably completes and improves it. The FILM model uses the common cinematographic techniques to construct camera placement based on input from the narrative planner. Information about common film idioms is encoded in a scene tree using the FILM Language. Objects within the FILM system use this knowledge in conjunction with the planner input to constrain the location and orientation of the camera for viewing a given action at execution time.

Similar to the FILM system, we propose a "hybrid" system that uses abstractly defined idioms as constraints to choose the best camera placement for any shot at any moment in the unfolding of the on-going story. However, unlike in (Amerson and Kime, 2001), where the narrative planner generates the information that must be conveyed to a spectator during a given scene, the Virtual Director gets its information from the action recognition module which reports all on-going actions in the interactive environment based on each characters' point of view.

In section 4, we describe a fully implemented prototype developed for the Unreal™ engine using its scripting language (UnrealScript).

PARALLEL ACTIONS

Though our interactive storytelling paradigm guarantees meaning to the story unfolding, via cause and effect duality in the interactions of characters' roles, interactions may occur at different locations, but within the same time space.

A parallel action is defined as a device of narrative construction in which the development of two pieces of action is presented simultaneously.

The task of relating two storylines, or two characters, or two different events, or a larger number of storylines, characters and events, is assigned to parallel film editing. These types of parallel editing could be defined as follows:

- The lines of interaction are close together, in the same space.
- The lines of interaction are far apart, in different places, and only a common motivation provides the link.

Therefore, the concepts of parallel film editing are important to consider for our system, when the scaling-up of the characters' roles will bring simultaneous storylines. As each storyline develops separately, the cinematic camera control must account for the information contributing to either separate or concurrent storylines, arising from the characters' roles independence or not.

SYSTEM OVERVIEW

The cinematography expertise encoded in the system is captured in two main components: The *Virtual Director* and the *Virtual Cinematographer*. Each of those abstract components is composed of two types of modules: low-level module (platform/domain dependant) and high-level module (platform/domain independent) (Figure 2).

In our narrative paradigm, the current state of the world is "wholly" determined, though the director does not have control on its changes over time, unlike (real-world) cinema directors. The director selects scenes based on the subject nature of the shot, e.g. a character going to a café or two characters having a conversation.

For example, we know where a character is going, though the director can not influence or modify its course. The modification of such behaviour would come from the interactive nature of the system. Another character passing by could stop momentarily the characters to have a mundane chat or the spectator could influence his behaviour by telling him that another character knows where what he is looking for is.

Due to the real-time and dynamic natures of the application, the camera control system must constantly reason on the current state of the world. Events (also called tasks) are recorded using the following template (*subject, verb, object*), as described in the *Virtual Cinematographer* (He et al., 1996). The *subject* is always an active character, while the *object* maybe another character, a fixed object (*book, gun, etc...*) or null. The *verb* represents a type of typical action (*move, pickup, talk, idle*). Hence, there are as many events as the number of active characters. Each of the variables for the three template components is associated static heuristic value, called *story weight*. The value of this weight is proportional to the narrative importance of the object it is

associated with. For example, if an object (e.g. a gun) is judged strategic, it will be gratified with a higher story weight than a banal object. The same principle applies for the allocation of weights for actions, as *talking* is more meaningful to the narrative than *walking*. This judgment seems rather artificial and subjective, and is in accordance to the subjective decisions of the film director. Optimally, the system should gather enough information to reason at a higher narrative level. Using this information, the system performs a heuristic classification of the events and extracts the most significant event and so selects the idiom associate to it. This method works upon the assumption that to any typical type of event, there is an appropriate attached idiom. It also selects an adapted pre-set of visual preferences to apply. Once the scene is selected, the system binds any unbound variables in the idiom specification and passes the information to the *Virtual Cinematographer* (He et al., 1996). During this process, the system may query the application for additional information, such as the specific location/orientation and dimensions of the various characters. Moreover, the system will constantly analyse the screen contents to immediately correct the camera settings if occlusions are detected. The scene is then rendered using the animation parameters and descriptions of the current environment sent by the application, and camera specifications (position/orientation) sent by the camera control system.

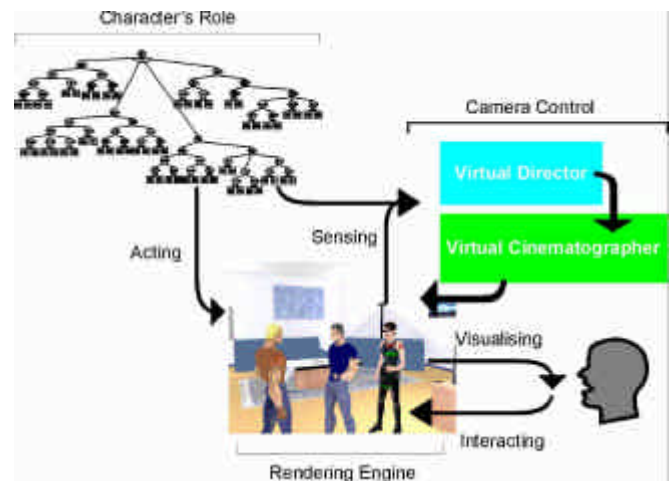


Figure 2: System Architecture

RESULTS

Arijon (1976) states that basic techniques for the coverage of two-or-three person static dialogues are also valid for larger groups. Rarely do four people carry on a dialogue simultaneously. There is always a leader, conscious or unconscious, acting as moderator, and shifting attention from person to person.

The example, presented in Figure 1, illustrates this particular feature. It was introduced within our prototype using a hierarchical finite state machine that handles dialogues between three characters (*idiom_3_talk*). The finite state machine is based on similar characteristics to a lower-level idiom that handles dialogues between two characters (*idiom_2_talk*). The considered idiom includes four states.

The initial state uses an establishing shot (*establish_shot*) of all three characters, while the second state relies (*a_b_talk*) on the lower-level idiom *idiom_2_talk*. The other two states (*c_talks*) and (*c_reacts*) capture the reactions from the remaining character.

Following further tests, the camera control prototype confirmed its capabilities in dealing with dialogues of four or more characters, though hardly ever encountered in plot instantiations from our interactive storytelling system.

CONCLUSION

We have presented the specificities of cinematic control in interactive storytelling where a story is generated dynamically in real-time, and described a fully implemented prototype of cinematic camera control. Future works will include extending the abilities of the system to manage separate storylines unfolding at the same time and providing the virtual director with the ability to choose between styles of montage according to different movie genres.

REFERENCES

Amerson, D. and Kime, S., 2001. "Real Time Cinematic Camera Control for Interactive Narratives." In Proceedings of AAAI SSS 2001.
 Arijon, D., 1976. "Grammar of the Film Language". Communication Arts Books, Hastings House, Publishers, New York.

Barthes, R. 1966. "Introduction a l'Analyse Structurale des Récits" (in French), Communications, 8, pp. 1-27.
 Branigan, E. 1992. "Narrative Comprehension and the Fiction Film". London: Routledge p.32
 Cavazza, M., Charles, F. and Mead, S.J., 2002. "Character-based Interactive Storytelling". IEEE Intelligent Systems, special issue on AI in Interactive Entertainment.
 Christianson, D. B., Anderson, S. E., He, L., Cohen, M. F., Salesin, D. H., Weld, D. S., 1996. "Declarative Camera Control For Automatic Cinematography". In Proceedings of AAAI '96, 148-155.
 Drucker, S. and Zeltzer, D. 1995. "CamDroid: A system for implementing intelligent camera control". In Proceedings of the 1995 Symposium on Interactive 3D Graphics, 139-144.
 He, L.; Cohen, M.; and Salesin, D. 1996. "The virtual cinematographer: A paradigm for automatic real-time camera control and directing". In Proceedings of the ACM SIGGRAPH '96, 217-224.
 Lu, R. and Zhang, S., 2001. "Automatic Generation Of Computer Animation".
 Mascelli, J. V. 1965. "The Five C's of Cinematography". Cine/Grafic Publications, Hollywood.
 Metz, C., 1974. "Film Language: A semiotics of the Cinema". New York: Oxford University Press.
 Raskin, R., 1998. "Five Parameters for Story Design in the Short Fiction Film". P.O.V., n. 5.
 Young, R.M., 2000. "Creating Interactive Narrative Structures: The Potential for AI Approaches". AAAI Spring Symposium in Artificial Intelligence and Interactive Entertainment, AAAI Press.

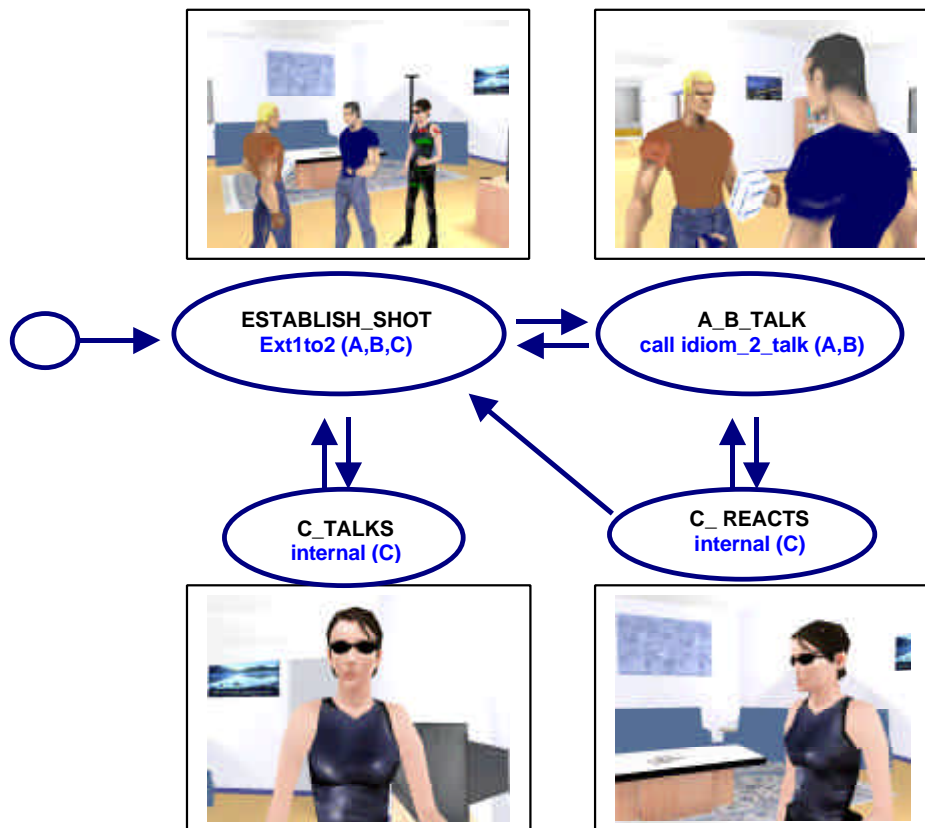


Figure 3: Idiom of a Dialogue Between Three Characters.