

A Discourse Planning Approach to Cinematic Camera Control for Narratives in Virtual Environments

Arnav Jhala and R Michael Young

Department of Computer Science
North Carolina State University
Raleigh, NC – 27695, USA
919.513.4199, 919.513.3038

ahjhala@unity.ncsu.edu, young@csc.ncsu.edu

Abstract

As the complexity of narrative-based virtual environments grows, the need for effective communication of information to the users of these systems increase. Effective camera control for narrative-oriented virtual worlds involves decision making at three different levels: choosing cinematic geometric composition, choosing the best camera parameters for conveying affective information, and choosing camera shots and transitions to maintain rhetorical coherence. We propose a camera planning system that mirrors the film production pipeline; we describe our formalization of film idioms used to communicate affective information. Our representation of idioms captures their hierarchical nature, represents the causal motivation for selection of shots, and provides a way for the system designer to specify the ranking of candidate shot sequences.

Introduction

Many advances in the design of systems that communicate effectively within 3D virtual environments have focused on the development of communicative elements that operate within the virtual world via conventional means such as natural language dialog, the coordination of gaze, gesture or other aspects of agent embodiment. A virtual camera is one of these elements that has been established as a powerful communicative tool by the cinematographers and directors.

A virtual world's camera is the window through which a viewer perceives the virtual environment. 3D graphical worlds have been developed for applications ranging from data visualization to animated films and interactive computer games. There are various issues that arise in each of these domains with respect to the placement of the camera. The basic objective of the camera is to provide an *occlusion free view* of the salient elements in the virtual world. In addition to this, effective camera placement techniques must also address the issue of choosing the *best position and angle* for the camera out of multiple possible positions and angles, for instance, as shown in Figure 1, a

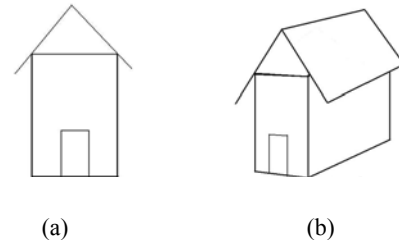


Figure 1. Shots illustrating how change in camera position can convey compositional parameters like depth/perspective.

front view of the house does not convey the sense of perspective. Choosing a different angle conveys these properties more effectively. For applications where the information present in the virtual world changes over time, the view selection needs to maintain *clarity* as well as *spatial and temporal coherence*. For instance, in Figure 2 it seems that two trains are running in opposite directions, while actually it is the same train shot from different locations.

In narrative-oriented virtual worlds, the camera is a communicative tool that conveys not just the occurrence of events, but also affective parameters like the mood of the scene, relationships that entities within the world have with other entities and the pace/tempo of the progression of the underlying narrative. For instance, in the vshots shown in Figure 3, the telling of the narrative is enhanced by selection of camera angles such that the initial low angle shot establishes dominance of the character that later turns submissive with the progression of the narrative as highlighted by the transition to high angle shots¹.

Cinema can be seen as an example narrative-oriented discourse medium where intentions of the director and cinematographer are communicated to the viewer through a sequence of coherent scenes. Filmmakers have developed very effective techniques for visual storytelling (Arijon 1976, Mascelli 1970, Monaco 1981). Although cinematography is an art form, there are certain rules for composition and transition of shots that are commonly followed by filmmakers.

¹ From the movie "A Few Good Men" © Columbia Tristar Entertainment



(a) (b)
Figure 2. Shots illustrating how change in camera position affects the viewer's orientation within the virtual world due to crossing line of action. (a geometric constraint).

In order to use the storytelling expertise developed by filmmakers in game environments we formalize shots and idioms¹ from cinematography as communicative plan operators that change the beliefs of the viewer in order to realize the intentional goals of the director/cinematographer. We demonstrate by an example how the affective elements of a story are communicated through the construction of communicative plans that are generated through a plan space search by a discourse planning algorithm.

Related Work

Camera Control in Virtual Environments

Computer Graphics researchers (Drucker 1994, Bares et al. 1999) have addressed the issue of automating camera placement from the point of view of geometric composition. The virtual cinematographer system developed by Christianson et al. (1996) models the shots in a film idiom as a finite state machine that selects state transitions based on the run-time state of the world. The idioms are defined using a Declarative Camera Control Language (DCCL). Tomlinson et al. (2003) have used expressive characters for driving the cinematography module for selection of shots and lighting in virtual environment populated with autonomous agents. More recently, the use of neural network and genetic algorithm based approaches to find best geometric compositions (Hornung 2003, Halper 2004) have been investigated. Kennedy (2002) uses rhetorical relations within an RST planner to generate coherent sequences. This approach is limited in the sense that the idiom representation does not consider the relationship between camera shots, and the system only reasons locally about mood at a particular time in the story.

Previous approaches to camera control in virtual environments have been restricted to finding the best framing for and determining geometrically smooth transitions of shots. They have not attempted to exploit the narrative structure and causal relationships between shot or

¹ Stereotypical ways of filming shot sequences



Figure 3 Shots illustrating how camera angles are used to convey dominance of a character in the story

scene segments that affect the selection of camera positions.

Planning Coherent Discourse

Research in generation of coherent discourse in the field of artificial intelligence has focused on the creation of text. Planning approaches (e.g., (Young et al.1994, Maybury et. al.1992, Moore and Paris et al.1989)) have been commonly used to determine the content and organization of multi-sentential text based on models developed by computational linguists (Grosz & Sidner 1986, Mann & Thompson 1987). Communicative acts that change the beliefs of a reader are formalized as plan operators that are chosen by a planner to achieve the intentional goals of the discourse being conveyed.

Camera Planning for Communicating Affective Parameters

In our approach we consider camera shots as intentional planned communicative acts that change the beliefs of the viewer. We draw a parallel between our approach and natural language discourse generation systems as well as the film production process (writing – direction – cinematography – camera control). In this section, we describe the representation of film idioms as plan operators and the camera planning process that performs the functions of a cinematographer in the film-production process. The story to be filmed is input as a sequence of actions executing in the world. This information is stored in the camera planner's knowledge base along with annotations indicating properties of the story and its characters (e.g., mood, tempo). A snapshot into the knowledge base for the planner is shown in Figure 5. Information about the story is used to generate a planning problem for the discourse planner; the goals of this problem are communicative, that is, they involve the user coming to know the underlying story events and details and are achieved by communicative actions (in our case, actions performed by the camera). A library of hierarchical plan operators that represent cinematic schemas is utilized by the discourse planner to generate sequences of camera directives for specific types of action sequences (like conversations and chase sequences). These camera directives are then translated into constraints for a geometric constraint solver implemented in the underlying graphics engine.

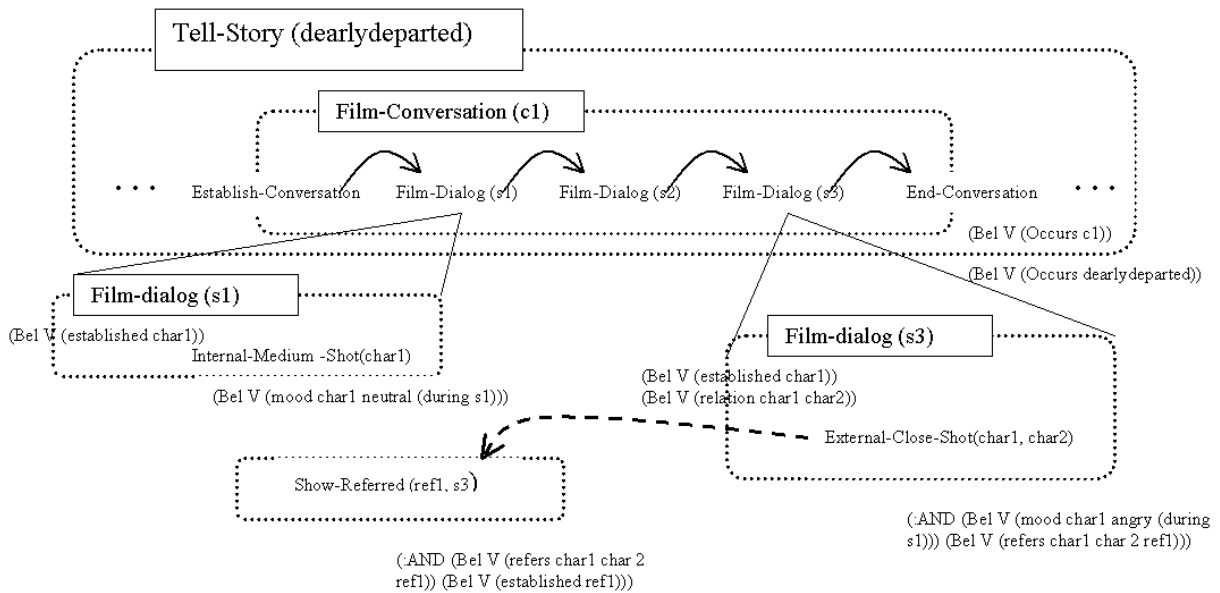


Figure 4 A Discourse plan for high level camera directives. The dotted boxes contain decompositions of abstract camera actions (simplified), the closed boxes contain names of operators and parameters. The predicates at the top left of boxes indicate the preconditions of the actions and predicates at bottom right of the boxes represent effects of actions. Strong lines indicate temporal relationship between actions and dotted lines, causal links.

In the following discussion, we will refer to the discourse shown in Figure 4 to illustrate how the representation of hierarchical idioms is used to generate communicative acts to convey affective information in the story as high level camera directives.

Story and Screenplay: Declarative Representation of the Story World Plan

The story plan that is generated by the story-planning algorithm is input to a plan description generation module. This module translates the plan data structure into temporally indexed plan description predicates and adds this knowledge to the camera planner’s knowledge base. We follow the declarative plan representation proposed by (Ferguson 1992). The story plan is then annotated by hand with additional information about the actions and characters (Figure 5) to indicate the directorial goals for communicating the affective parameters of the story and characters.¹ The information about the story world is also temporally indexed relative to other events/actions happening in the plan. This declarative description of the

¹ Future work will address the automatic generation of these annotations from knowledge of plan structure.

story plan is then used by the camera planner when it generates camera actions for the story. Figure 5 depicts a snapshot of the story-world action description and the affective annotation that are input to the planning algorithm.

Representation of Film Idioms

As noted by Christianson et al. (1996), film idioms can be represented in a hierarchical manner with constraints at the scene level as well as shot level, where shots are constituents of scenes. We represent camera shots and abstract idioms as hierarchical plan operators where the top level actions capture the rhetorical content that is to be conveyed to the viewer and drive the selection of primitive actions. Consider the operators in Figure 6. In this example; we present one of the most commonly used idioms in cinema – the conversation idiom. A conversation between two characters can be filmed through repeated use of three types of shots; the apex shot (2-shot), the internal shot, and the external over-the-shoulder shot. A number of factors like the mood of the characters involved, the relationships between characters, and the topic of the conversation determine how the individual shots are selected to film a particular conversation.

```

(character Royce) (character Marguerite)
(object phone)
(location room_royce) (location kitchen_marge)
(at Royce room_royce)
(at Marguerite kitchen_marge)
(mood Royce neutral (before s1)) (mood Marguerite neutral (before
s1))
(conv-conv c1) (conv-type c1 phone) (conv-start c1 s1) (conv-
end c1 s2)
(conv-steps c1 (s1 s2 s3))
(step s1) (act-type s1 speak) (agent s1 Royce) (secondary s1
Marguerite)
(effect s1 (spoken Royce "What the ... Hello"))
(mood Royce neutral (during s1))
(step s2) (act-type s2 speak) (agent s2 Marguerite) (secondary s2
Royce)
(mood Marguerite neutral (during s2))
(effect s2 (spoken Marguerite "Do I hear lord's name in vain"))
(step s3) (act-type s3 refer) (agent s3 Royce) (secondary s3
Marguerite)
(effect s3 (spoken Royce "Good God Mama what is that
thing?"))
(effect s3 (refers Royce Marguerite object))
(mood Royce angry (during s3))

```

Figure 5 Snapshot of the story world description given as an input to the camera planner.¹

In our representation, as shown in Figure 4, each abstract conversation action is expanded into a set of actions including an *Establish-Conversation* and an *End-Conversation*. The planner refines these actions, adding constraints and additional details, until primitive camera actions (e.g., specific shot directives) are produced. These actions direct the geometric constraint solver to establish and maintain the Line-of-Action throughout a single conversation sequence. This provides a way to explicitly represent coherent sequences and scene boundaries.

Affective information within the story is used to constrain the selection of primitive shot-types. For instance, the abstract action *Film-Dialog* is further expanded with a primitive close-up shot instead of a long-shot if the mood of the character in the story is angry through the constraint (*mood Marguerite angry (during s3)*).

Low-angle primitive shots are used to indicate dominance of one character over another in a scene. For a scene where the relationship between the characters is unspecified or where the characters are not closely related, long-shots are preferred over close-ups. In the example shown in Figure 5, close-shots are used, since the conversation's participants are closely related.

By representing idioms as plan operators, the rhetorical structure of the narrative is captured through the causal reasoning carried out by the planning algorithm during the operator selection process. This is elaborated in the next section. As an example, consider the action Show-Referred (ref1, s3) from Figure 4. This introduces a cutaway from the original action of showing the character delivering the dialog to the referred object during the dialog. The planner also automatically adds steps for establishing shots through the same type of causal reasoning.

¹ Modified Act I Scene 2 from the play *Dearly Departed*

```

(define (decomposition tell-story)
:parameters (?story)
:constraints ((story ?story) (story-conv ?story ?scenes))
:links()
:steps (
...
(forall ?scene in ?scenes
(step2 (film-conversation ?scene)))
...
)
:orderings ()
:rewrites (((BEL V (story ?story))
((forall ?scene in ?scenes
(BEL V (Occurs ?scene)))))))

(define (decomposition film-conversation)
:parameters (?c)
:constraints ((conversation ?c) (conv-steps ?c ?slist))
:steps (
(step1 (apex-shot ?c))
(forall ?step in ?slist
(step2 (film-dialog ?step)))
)
:orderings ((step1 step2))
:rewrites (((BEL V (Occurs ?c))
((forall ?step in ?slist
(BEL V (Occurs ?step)
))))))

(define (action film-conversation)
:parameters (?c)
:precondition NIL
:primitive NIL
:constraints NIL
:effect ((BEL V (Occurs ?c))))

```

Figure 6 Representation of film idioms as plan operators (simplified)

Viewer Model

Communicative camera actions affect the viewer's beliefs about the story. We use a simple viewer model representing beliefs about characters, objects, their properties, and the spatial and temporal relationships that hold between them. These are represented in general as: (BEL V ϕ), where V is the viewer and ϕ is a temporally indexed predicate. For instance, one of the beliefs of the viewer, that a certain action a_1 occurred during the execution of another action a_2 then it can be represented by (BEL V (Occurs a_1 (during a_2)))

Camera Planning Algorithm

We use a decompositional partial-order causal link planning system named Longbow (Young et al.1994) for generation of shot sequences. The planning algorithm progressively generates plans by a) adding steps to the existing plan such that the effect of the action directly satisfies a condition in the goal state; b) adding steps to the plan such that the effect of the action satisfies a precondition of an already existing action, and c) refining an abstract action that has not been expanded. Further, the camera planning algorithm adds temporal constraints from camera actions to the story-world actions that they are

responsible for filming. For this purpose, we use the interval temporal relationship predicates defined by Allen (1983).

The DPOCL planning algorithm searches through the space of all possible plans representing combinations of shots for communicating the goals of the problem. Initially, the planning algorithm picks an open (unsatisfied) goal e and selects a camera action operator C whose effect unifies with the goal. The operator -- with all necessary variable bindings -- is added to the empty initial plan as a step (S_i). The bound preconditions pre_{S_i} of the newly added step are added to the plan's list of open conditions. This process is repeated until all open conditions are satisfied through effects of actions. If there is no action from the operator library that satisfies a condition, the algorithm terminates indicating a failure. The plan algorithm executes the following steps during the addition of steps.

Causal Reasoning. If the effect eff_{S_i} of a step S_i in the plan satisfy the precondition pre_{S_j} for the execution of another step S_j in the plan then the planner adds a *causal link* from step S_j to step S_i . If a step S_k occurring between S_i and S_j in the plan negates the effect of S_i then S_k is considered to be a *threat* to the plan. Threats are handled by changing the order of execution of the threatening action S_k to occur before S_i or after S_j or by adding binding constraints to the steps involved to avoid conflict. In Figure 4, the dashed arrow indicates a causal link.

Temporal Reasoning. In our implementation, each camera action C_i is associated with two temporal markers: (starts-at C_i ? t_p) and (ends-at C_i ? t_p). Here t_p is a temporal reference to a story world action (e.g. (*during* ? s)). These temporal references are added to relate the camera actions with the corresponding story actions that they film during execution. For instance, in Figure 4, each camera shot is temporally marked for execution with the respective speech act. Further, the camera action *show-referred* occurs during the execution of the camera action that is filming the referring speech act.

Decomposition. Steps added in a plan are either abstract steps or primitive steps. Abstract steps added into the plan structure are decomposed into their constituent actions. Each step within a decomposition is added to the plan and the step's preconditions are added as open conditions. Figure 6 shows examples of abstract actions in the planner's operator library. In the example shown in Figure 4, the abstract action *Film-Conversation* is refined with primitive shots for filming the speech acts occurring in the conversation. Each constituent action satisfies a sub-goal that combines with other sub-goals to satisfy the goals for the abstract action. In a conversation, the orientation of the participants is established prior to the start of conversation by a wider shot. This is captured in the schema by the hierarchical operator shown in figure 6 with the (*apex-shot* ? c) action followed by (*film-dialog* ? $step$) steps.

After the plan space is created by the planner, a ranking function is then used to rank all the generated plans and the best plan is chosen according to the preferences of the user

encoded in the ranking function. Complete details of the planning algorithm are described in (Young *et al.* 1994).

Ranking Cinematic Sequences

A number of parameters (like tempo, importance of characters) are responsible for the selection of particular sequences of shots within idioms. To guide the search process during planning, the DPOCL algorithm uses a best-first approach in which a heuristic search function ranks nodes at the fringe of the search space, ordering the unexpanded nodes most promising to least promising. In our approach, we take advantage of the heuristic search function to rank plans not only based on estimations of how close the plans are to being complete, but also based on the match between the structure of the plan and desirable features of the plan's narrative structure.

For instance, the cinematic rule that is governed by the tempo of the story is that in parts of the narrative where tempo is high; choose shorter actions and multiple cuts. This can be incorporated in the ranking function by ranking camera plans with a greater number of primitive steps over other candidate plans. The ranking function can also be used by designers to specify their preference for cinematic style. For example, if a user prefers extreme-close-up shots, then plans containing decompositions that use extreme-close-ups are ranked higher than other plans with comparable structure. In the given conversation sequence, the selection of camera angles is governed by the mood of the character that is the object of the speech act being filmed. Even though there are other factors that affect camera angles, this simplification works for our restricted domain. Our representation does provide a means for specification of more complex idioms and cinematic rules, since it is based on a generic discourse planning algorithm.

Implementation

Our work is implemented using Mimesis (Young et al. 2004), a service-oriented architecture for intelligent control of narratives in virtual environments. The Longbow discourse planner is implemented in LISP. The camera control code executes within the Unreal Tournament 2003 game engine.

Limitations and Future Work

We have begun formalizing the language of film for cinematic presentation of narratives in virtual environments. This approach requires knowing the details of the story for pre-planning camera moves and thus cannot be guaranteed to work in real-time for large story spaces. Geometric constraint solvers can be optimized by utilizing the context-aware high-level directives for reducing search space of possible geometric solutions. These concepts and techniques have potential in development of automated and semi-automated tools for

creative designers that aid in rapid development of interactive narrative based virtual worlds. We are working towards strategies for empirical evaluation of cinematic camera planning systems. Comparative evaluation with previous approaches is complicated since no other approaches have attempted to take into account the situational parameters of the underlying story.

Acknowledgements

This research was supported by NSF CAREER Award #0092586.

References

- Allen, J.F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832-- 843, 1983.
- Arijon, Daniel (1976). *Grammar of Film Language*, Los Angeles, Silman-James Press, 1976
- Bares W. and Lester, J. (1997). Cinematographic User Models for Automated Realtime Camera Control in Dynamic 3D Environments, *Proceedings of Sixth International Conference, UM-97*
- Christianson David, Anderson Sean, He Li-wei, Salesin David, Weld Daniel, Cohen Michael (1996). Declarative Camera Control for Automatic Cinematography, *Proceedings of AAAI, 1996*
- Drucker Steven, Zelter David (1997) Intelligent Camera Control in a Virtual Environment, *Graphics Interfaces 97*
- Ferguson, George (1992). Explicit Representation of Events, Actions, and Plans for Assumption-Based Plan Reasoning, *Technical Report 428, Department of Computer Science, University of Rochester, NY, June 92.*
- Grosz, B. and Sidner, C. (1986). Attention, Intention and Structure of Discourse, *Proceedings of ACL 1986*
- Halper, N., Helbing, R. and Strothotte, T. (2001). A Camera Engine for Computer Games: Managing the Trade-Off Between Constraint Satisfaction and Frame Coherence. *Proc. Eurographics 2001*
- Hornung, A., Lakemeyer, G., and Trogemann, G. (2003). An Autonomous Real-Time Camera Agent for Interactive Narratives and Games. *IVA 2003*
- Hovy, E (1993). Automated Discourse Generation Using Discourse Structure Relations. *Artificial Intelligence* 63, pp. 341-385, 1993
- Kevin Kennedy, Robert Mercer (2002). Planning animation cinematography and shot structure to communicate theme and mood, *Proceedings of Smart Graphics*, 2002.
- Mascelli Joseph (1970) *The Five C's of Cinematography*, Cine/Grafic Publications, 1970
- Maybury M (1992). Communicative acts for explanation generation, *IJMMS* (1992) 37, 135-172
- W. C. Mann, S. A. Thompson (1987). Rhetorical Structure Theory: A Theory of Text Organization. *TR- ISI/RS-87-190*, USC ISI, Marina Del Rey, CA., June 1987
- Monaco James (1981). *How To Read A Film*, New York, Oxford University Press, 1981
- Moore, J.D., Paris, C.L. (1989). Planning text for advisory dialogues. *In Proceedings of the 27th Annual Meeting of the ACL*, pg. 203--211, Vancouver, B.C., Canada, 1989.
- Tomlinson Bill, Blumeberg Bruce, Nain Delphine (2000). Expressive Autonomous Cinematography for Interactive Virtual Environments *Fourth International Conference on Autonomous Agents*, Barcelona, Spain 2000.
- Young R M., Moore, J. (1994). DPOCL: A Principled Approach To Discourse Planning, *Proceedings of the INLG workshop*, Kennebunkport, ME, 1994
- Young, R. M., Riedl, M., Branly, M., Jhala, A., Martin, R.J. and Saretto, C.J. (2004). An architecture for integrating plan-based behavior generation with interactive game environments, *The Journal of Game Development 1*, March 2004.