Orchestrating Digital Micromovies

Glorianna Davenport, Ryan Evans and Mark Halliday

Thomas: "Hi, what's your name?"
David: "David Kung."
Thomas: "I like that."
David: "What are you doing here?"
Thomas: "I'm a graduate student in the Interactive Cinema Group."
David: "That's true."
Thomas: "What's this light here?"
David: "Well, this light fixture actually comes from the first movie house in Beijing, and my grandfather took it as a souvenir when the theater closed down."
Thomas: "Don't listen to this guy. He's an undergraduate. He doesn't know anything."

The exchange above is taken from an unusual movie called *An Endless Conversation* [1]. The movie is a conversation between two characters. They are talking about the Interactive Cinema Group at MIT's Media Lab, where this movie was made. Although the dialogue may appear to be a normal, if slightly wooden, transcript from a linear movie, *An Endless Conversation* is no ordinary film. It is a simple example of what we call a personalizable movie and a first step toward new kinds of experiences that will be created using digital video technologies. A still frame from the movie can be seen in Fig. 1.

Digital video can now be stored and accessed on a computer just like any other type of data. This has profound implications for the production and playout of cinematic narratives. Instead of one uninterrupted stream of sequences, individual shots are stored as discrete units; these can be played out in numerous distinct progressions, selected via a computer program. A fundamental problem in creating digital movies of this type is how to ensure that the computer will create a coherent and meaningful experience. The Digital Micromovie Orchestrator (DMO) successfully tackles this problem by using a simple keyword-based description strategy and an extensible layered filter structure that suggests an appropriate next shot.

Many interactive experiences are scripted as pre-computed or branched structures that require input (interaction) from the user at specific pre-defined narrative junctures. The imperative of such interactions disrupts the reverie of the experience so important to a traditional cinematic narrative. A viewer should be able to sit back and "enjoy the show." The problem of mandatory interaction is exemplified by the interruptive vote-based decision-making in *Murder, Anyone?* [2] and in the recent theatrical release of *I'm Your Man* [3]. In contrast, the DMO emphasizes fluid interaction as an essential component of digital movies. Viewers can sit back and watch the movie, but they can also subtly change the direction of what appears on-screen at any time. With *An Endless Conversation,* for example, the viewers can personalize the movie by changing the pacing from fast to slow or the rating from "Parental Guidance suggested" (PG) to "Restricted" (R) while the movie is running [4].

Fluid interaction is made possible because the computer uses a set of filters to select the next appropriate shot in real time. Instead of precomputing the entire story in advance, the DMO successively filters clips from the entire database down to a single appropriate shot for the current point in the movie. The DMO's filters are layered to create a narrative structure that can be mutated as the movie progresses.

This real-time orchestration of shots opens the door to new kinds of cinematic experiences. These experiences can be considered multi-threaded stories in that the story plays out differently at each viewing. While the current work exemplifies simple cinematic scenarios, including a conversation and an action sequence, future movies might be considerably more complex. For example, the user might specify, "I have 30 minutes to see this movie. I'm really interested in what Max has to say, but I don't want too much violence," and then halfway into the movie decide to lower the violence level even more through fluid interaction. Thus, a movie can be different each time it is watched, and it can be personalized and changed while it is running.

ABSTRACT

The authors describe how computers can be used to build narrative structures that create simple cinematic sequences from a large database of shots. The Digital Micromovie Orchestrator (DMO) does this by allowing the maker to attach sketchy descriptions to video clips in the database and to build narrative abstractions in the form of layered filter structures. These structures orchestrate shot selection in real time to create a flowing video experience on-screen. The DMO suggests a new direction for creating cinematic narratives using digital technology. This paper analyzes in detail the first movie produced using the DMO and discusses the process of making interactive movies using the DMO or a similar digital movie-construction kit.

A DIGITAL ENVIRONMENT

In a digital environment, video becomes merely another piece of data that can be described, accessed and displayed. The DMO takes advantage of this by making use of micromovies. A micromovie [5] is a short piece of video with descriptive information attached to it that represents a unit of meaning determined by the moviemaker's intent. These micromovies are filtered by the computer according to the machine-readable descriptions that form part of a micromovie. Even though each individual filter might be quite simple (i.e. one for pacing, one for continuity), it is the layering of multiple filters that

Glorianna Davenport (educator, researcher, moviemaker), MIT Media Lab, 20 Ames Street, Cambridge, MA 02139, U.S.A.

Ryan Evans (researcher, digital moviemaker), MIT Media Lab, 20 Ames Street, Cambridge, MA 02139, U.S.A.

Mark Halliday (researcher, digital moviemaker), MIT Media Lab, 20 Ames Street, Cambridge, MA 02139, U.S.A.

Received 9 September 1992.

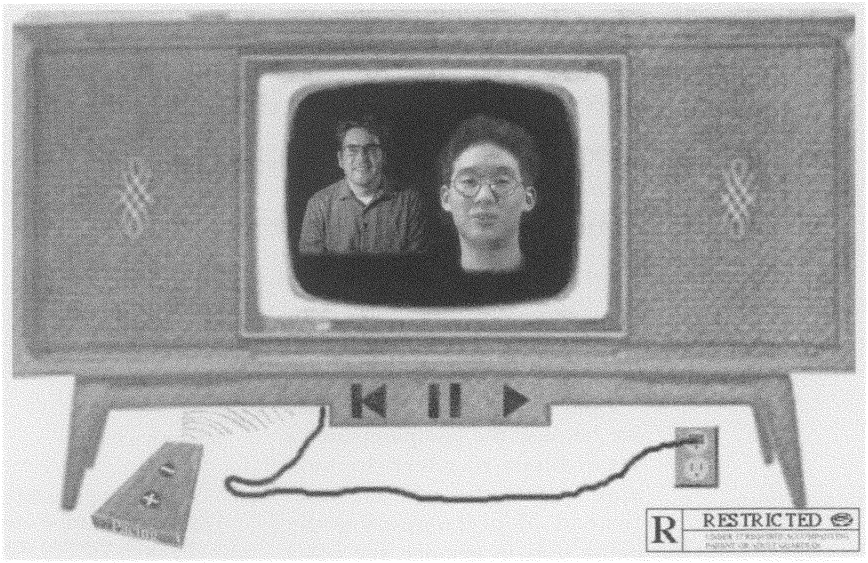


Fig. 1. ***An Endless Conversation*****, still frame, digital video, 1992. This interactive movie has two controls. The pacing of the movie can be changed by clicking the + or - button on the remote control. The rating can be switched between R and PG by clicking on the rating panel. The viewer can initiate these interactions at any time during viewing.**

can be changed in real time (through interaction) that gives the system its power and flexibility.

The DMO can be thought of as a simple machine implementation of an editor. A film or video editor takes a collection of raw footage, logs it and then uses an idea or script as a basis for editing together sequences of shots to form a video narrative. Similarly, the DMO has a template for a story in the form of a succession of layered filters. A group of filters can be tailored to a particular scene: we have developed groups of filters for dialogue (two characters talking), action (climbing stairs) and viewer interaction (controlling pacing and rating). Each time an edit is required, the filters select a shot appropriate for the current point in the story. The most important aspect of the system is that the template can be layered and changed in real time, allowing interaction to take place at any time during the story.

Using this approach, the DMO allows us to explore a new mode of interaction called "fluid interaction." Imagine a narrative that progresses much like a linear film, except that the viewer, in order to change the flow or feel of the experience, is free to manipulate it along certain dimensions [6]. This type of interaction provides an experience that lies somewhere between the constant clicking and decision-making of current interactive narratives (e.g. computer adventure games and interactive training programs) and the lack of interaction inherent in linear films. The viewer can get caught up in the reverie of the experience without the interruption of interaction until he or she desires to change the movie experience. Changing the experience might be as subtle as speeding up the pacing or it might be as drastic as changing the focus of the narrative from Dave's character to Thomas's.

Design and Implementation

The DMO consists of a video logging module and a filter-based shot selector. The logger allows the author to bind "sketchy" descriptions to fragments of video, creating digital micromovies. A sketchy description contains the minimum amount of information for use in a particular movie. The selector module makes use of the descriptions attached to micromovies to sift out unwanted shots using a set of filters designed by the author and manipulated by the viewer. Several types of filters are chained together to orchestrate a sophisticated personalizable movie. Current filters include template filters, continuity filters, stylistic filters, and interaction filters, but the selector mechanism is designed to easily incorporate new types of filters. At the moment, the intention is that makers will design descriptive elements and filter-based story structures prior to recording the visual material, much as a linear filmmaker writes a script before shooting a movie.

The Logger Module

The DMO logging component (software designed by the authors) guides the moviemaker through a process that is analogous to creating logs for use in editing a traditional linear film. It allows the creator to associate information in the form of keywords with fragments of digital video to create micromovies. The logging component encourages sketchy logs whose depth and sophistication are not sufficient for general use, but which allow the author to quickly record descriptions. The logger and the entire DMO system describe video and construct movies at the clip level. The moviemaker creates a video segment (a clip) by defining beginning and end points. The clip's content is then described with keywords in a manner that is appropriate to the experience she or he is trying to build. Thus the moviemaker decides what constitutes a micromovie, while the system decides (with guidance from narrative abstractions created by the moviemaker) what constitutes an appropriate narrative sequence.

The log is structured around *slots* and *values* that use keywords as descriptive elements. A slot is a general characteristic type (e.g. Characters or Pacing) for describing a collection of video clips. Values are attached to slots (e.g. Characters: Thomas or Pacing: Long) to describe a particular video clip. Each video database has several slots associated with it, and each shot in a database has one or more values within each slot (Fig. 2). Typical slots used in *An Endless Conversation* include character, type of utterance, subject of utterance, rating and pacing. Keyword classes and slot groups provide default values within the logging tool to help maintain consistency across logs [7]. The DMO system has no required slots for descriptions. Instead, the author is able to create and delete slots at will and design filters that depend only on those slots. This free-form description structure provides the creator with the freedom to log values that are important to the movie while ignoring other values.

The Shot-Selection Module

Just as the logger module mirrors the process of creating logs for a traditional video- or film-editing job, the shot-selection module makes use of those logs to mirror the process of actual editing. An editor chooses possible shots based on knowledge of available video, story structure and cinematic style. The selection module chooses shots based on a simple

layered filter model that mimics some of this knowledge. The module takes as its input the entire video database and a set of specifications from the moviemaker. It then returns all of the video clips that meet those specifications. The shot-selection process is accomplished via layered filters that progressively weed out shots that do not qualify. The filters have simple structures. Each contains a set of descriptions that are matched against micromovie descriptions. Since the output of each filter is simply a subset of the original database, the filters are easily chained by applying one filter to the output of another. The filters that select shots for a complete movie can be added, deleted and mutated over the course of the presentation, either programmatically or through viewer interaction.

Filters

To create *An Endless Conversation,* the creator designed four filter layers that were used by the shot selector to create the movie. The first filter layer is based on a template structure that parallels the structure of the conversation between Thomas and David: a question, a response, a reaction, another question, and so on. The second filter layer ensures continuity over the course of the movie. In *An Endless Conversation,* the topic of each response must match the previous question. The third filter layer is stylistic. This filter is created from a rule base that embodies some of the moviemaker's ideas about cinematic conventions. During the course of the dialogue, this layer specifies when to use a close-up, a medium shot or a wide shot of the character, based on the framing of previous shots. The fourth and final layer is the interaction filter. In *An Endless Conversation,* this filter is mutated based on the user's preferences concerning rating and pacing. The filter layers are represented in Fig. 2.

Template Filter. The template filter is perhaps the simplest filter layer in the DMO system. It is also the most powerful authoring tool. The author outlines a template that describes the progression of the movie as a collection of filters arranged along a time line. A driver built into the system applies these filters (layered with other types of filters) to choose each successive shot. The filters are activated and de-activated on a shot-by-shot basis according to the time line. In *An Endless Conversation,* the template was specified such that the first shot must be a question from one character, the second shot must include the other character responding to the question and the third shot shows the first character with a reaction. This template structure is easily translated into a list of filters that are matched against micromovie descriptions.

Continuity Filters. The continuity filter layer changes itself over the duration of the movie, based on the course of the action up to that point. This layer tries to control consistency of image, plot and character. The continuity layer contains simple continuity rules that look at descriptions of previously presented micromovies and filter out clips that are inconsistent with the viewer's experience up to that point. Using the script at the beginning of the paper as an example, a continuity filter makes sure that the topic of the response matches the topic of the question. The moviemaker could design this layer to monitor simple continuity points, such as setting, costume or time of day.

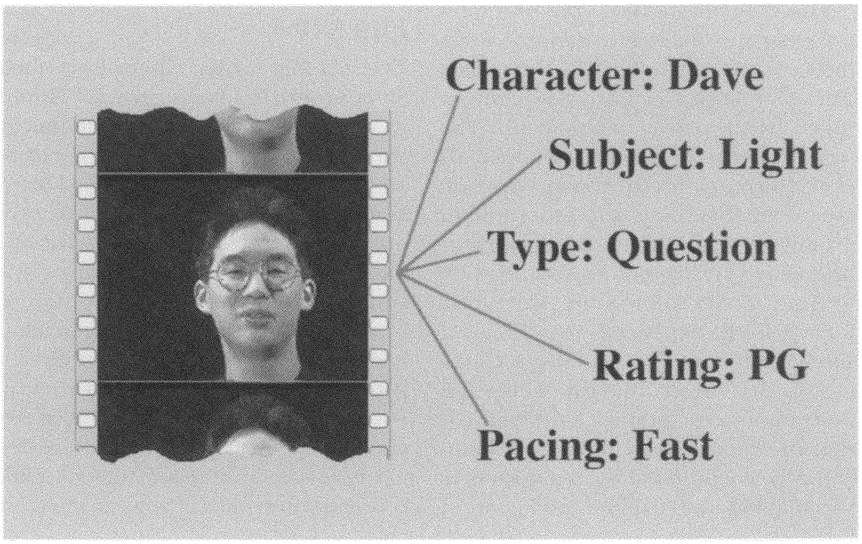
Stylistic Filters. The stylistic filter layer controls cinematic conventions during movie presentation. This layer differs from the continuity filter in that it does not select shots based on their content. Instead it looks at the style of the shot (e.g. framing, camera angle, lighting). The filters in this layer are constructed from a stylistic rule base. This base uses previous shots as input (typically the past one or two shots) and outputs shot selection filters. Two simple stylistic rules might read as follows:

- If this is the first shot of the movie, then the framing should be extremely wide (i.e. an establishing shot for the space).
- If the character was seen in the last shot in extreme close up, then do not show her in extreme close up in the next shot.

The creator can exercise close control over the look and feel of the movie using such stylistic filters.

Interaction Filters. The final filter layer consists of filters that are directly or indirectly manipulated by the viewer. The viewer is presented with various controls before and during the movie presentation. Manipulation of these interface objects changes parameters that modify shot-selection filters in the interaction filter layer. *An Endless Conversation* currently allows the viewer to manipulate the rating (PG, R) and the pacing (fast, slow) as the movie unfolds. This type of interaction is accomplished by translating the user's action on the interface into changes in the filter layer. Each shot is logged as R or PG, and when the rating is set to PG the rating filter simply weeds out the risqué possibilities. The pacing control creates a similar filter that weeds out lengthy shots when the user wants fast pacing and weeds out short shots when the user prefers slow pacing.

Fig. 2. The filter structure created by the moviemaker progressively reduces the entire database of shots down to one appropriate shot for the current point in the story. This process happens every time a new shot is needed. This example has three filters. The power of the system results from the fact that it is possible to layer as many filters as are needed and to change them for different parts of the story.



More sophisticated interaction modes could be implemented using this filter layer, such as a "More" button. This button would allow the user to go more deeply into a particular scene at any point in the movie by modifying a template filter layer to allow more time to be spent with the current idea.

A New Approach to Moviemaking

Making a movie within the constraints of the DMO system is very different from making traditional linear movies. The whole process has to become open-ended. For example, micromovies have to be scripted and shot in such a manner that the filtering system has a variety of shots to choose from for each part of the movie. If there are several appropriate shots, the exact response to a question by one character may not be known in advance. It may be long and involved. It may be short and to the point. The response will be decided by the computer and will depend on the interactions made by the viewer. To filmmakers accustomed to controlling every aspect of a production, this apparent loss of control may be slightly disconcerting. We need a new type of moviemaker—one who is willing to relinquish some of the control to the computer and the viewer. Rich, multi-threaded narratives will require that some of the authoring be done by the computer and that the author create algorithms to control the computer. The filter-based system we propose allows the moviemaker to shape the structure of what appears on-screen through the creation of filters. The important point to remember is that both the filters and the filming are ultimately controlled by the creator.

The DMO has evolved out of a long progression of alternative narrative structures created within the MIT Media Lab. These structures explore new content suited to interactive environments. The real challenge for interactive moviemakers is to come up with content that will create a compelling and entertaining experience within the framework of an interactive environment. An early pioneering application was the Aspen project, which explored the idea of surrogate travel using videodisc technology. In that project, the experience was built upon a geographic environment that the viewer could navigate around [8]. *New Orleans Interactive* is a documentary that began to explore the import of attaching computer-readable content descriptions to video clips. While watching a linear movie about New Orleans as the city prepared for, held and concluded the 1984 Louisiana World Exposition, viewers, using keywords attached to people, places and themes, can get more information relative to what they see in the linear movie [9]. Ben Rubin explored the use of layered filters for building a personalized movie. His constraint-based editing system selects shots based on limited input from the viewer. It builds an edit list to create a linear piece for subsequent viewing. However, the filters in the system are not extensible, and they cannot be mutated during the presentation [10].

The DMO extends this line of research. It allows researchers to look at how the computer can use descriptions attached to video clips within a framework where the filtering is done in real time. There are two significant advances in the DMO approach. First, the story can be re-directed through interaction at any point since the filtering is done in real time. Second, the DMO takes us away from the need to create a large, fixed structure where every possible strand of a narrative has been pre-scripted. The layered filters create a simple, yet flexible story structure into which micromovie clips can be "dropped." The rate of growth of a fixed-structure, multi-threaded, interactive narrative can sometimes be exponential [11]. Using the DMO, it is possible to start thinking about scripting, shooting and editing a movie in a much more flexible and open way. A look at the methodology used in the creation of *An Endless Conversation* will demonstrate the approach required for making a personalizable movie.

Filtering

The first step for the author is to create an appropriate filter structure. The filters have to be created at the outset so that the appropriate video clips can be scripted to fit them. In the case of *An Endless Conversation,* we decided that the template filter would be comprised of a question, followed by a response, and finally a reaction. This template is repeated ad infinitum, resulting in a limited but coherent exchange between the two characters that goes on forever (hence, its title). The other filters will also affect the scripting of a movie. For example, the interaction filters for the conversation require that pacing and rating be taken into account. Clips of varying length and levels of offensive language will be needed.

Scripting

Once the filters have been defined, the next step is to create a shooting script defining shots that are appropriate for each possible combination of filter layers through the course of the presentation. In *An Endless Conversation,* this involves many different types of statements from each character. For the question filter in the template layer, we include utterances such as "What's your name?," "What are you doing here?" and "What is this light?" The response template filter requires clear responses to each of these questions from both characters. A question will have more than one possible response. By scripting several possibilities for each type of shot, the movie can be personalized along the exact lines required by the author. For example, a question will have long and short responses, allowing the pacing to be controlled. Only short responses will be chosen when the pacing knob is set on "Fast." The reaction template filter is a much more generic slot. It includes possible rebuttals of varying length and tone by each character to what he has just heard: "I agree," "I disagree," "Don't listen to this guy" and so on.

Shooting

After the script has been laid out, shooting can begin. In a traditional film, the director captures a variety of performances from the actors to allow the editor a choice of the best "take" for a particular point in the story. Likewise, the more takes of a particular shot the DMO has to work with, the more possibilities the layered filters have at any point to personalize the presentation based on user manipulation. For *An Endless Conversation,* the shooting process was similar to shooting an interview. The main difference is that all possible character utterances—questions, responses, reactions—were filmed. The most important aspect of shooting for a personalizable movie is to get a large variety of different takes (long, short, animated, subdued, wild, controlled, and so on). The director must capture a variety of utterances that are also generic enough to be used in different situations. For example, many of the reactions in *An Endless Conversation* can be used after any character's response.

LOGGING

Once the shooting is over, the footage has to be logged by the moviemaker, just as in a traditional film. Each shot has to have descriptions attached to it so that the program can use the layered filters to choose the correct shot. As the raw video is broken into coherent shots by the author, each one is logged into the database, creating a collection of micromovies that can be accessed by the orchestration program. The descriptions used by the DMO system for *An Endless Conversation* are character name, shot length, type of utterance, subject of utterance, rating, and pacing. Based on only these descriptions, the computer can put together a sequence of shots that recreates a conversation between two characters.

DIRECTIONS

In addition to *An Endless Conversation,* a second movie, *Stairwell to Nowhere,* has been completed with the DMO system. The movie presents a short action sequence in which a man runs up a stairwell and exits through a doorway. It explores the automatic generation of an action sequence template from a physical model of a space. Before the movie begins, the viewer can specify which stairwell landing the character should begin on and which landing the character should end on. A simple algorithm uses a physical model of the stairwell to generate a template filter specific to the viewer's preferences. The algorithm makes sure that both the space and the action played out within it are clearly communicated to the viewer through cinematic conventions. In such a system, the role of the DMO as a viewing environment can be extended to previsualization. The moviemaker can plan and shoot simple action sequences while the DMO monitors sketchy descriptions to make sure that spatial and action continuity will be understandable to the viewer.

Another presentation recently created with the DMO system is *CyberCritics.* This movie presents two movie critics who review movie clips on-screen and have exchanges about their reactions to the movie. The viewer is able to choose which movies she wants reviewed and if she wants to hear more from a particular critic. The *CyberCritics* project expands the creative options open to the maker in the digital realm. Two of the more exciting ideas are positional editing and automatic compositing. Positional editing refers to placing the characters within the screen space so that they can refer to each other and other objects within the space by looking, gesturing or pointing. This involves not only logging each micromovie with directional information, but also creating a rule base that allocates screen space to micromovies as they are presented. Automatic compositing explores how two or more digital video streams might be combined in real time. Traditional compositing is the process whereby a character is filmed against a blue background and seamlessly inserted into a background that represents a location. The *CyberCritics* movie creates a location where the characters can be automatically positioned and talk to each other as in a television show. This system relies on shot-selection filters that take into account not only the layout of the screen space but also what action is required of a character for her or him to refer to other on-screen objects.

Currently we are continuing work on a fourth interactive DMO movie called *Train of Thought.* In this movie, we are beginning to explore two aspects of more sophisticated cinematic structures. First, we are incorporating parallel action as an interactive mode. Second, we are working with pre-edited sequences that correspond to specific stages within a story rather than with shots that build up to a sequence. The movie is centered around a romance between an amateur filmmaker, Jack, and a woman, Nicole, who is moving overseas in a few days to take a good job. The viewer can mediate between the "dream" thread and the "reality" thread at any point in the story. The dream thread explores Jack's subconscious view of the situation, while the reality thread presents a more objective view. Depending on the viewer's path through the experience, different parts of the narrative are played out. Inevitably, the viewer must experience the movie more than once to gain a full understanding of the implications of the events. The video database for *Train of Thought* includes cinematic sequences often ranging up to 3 min in length and made up of several separate shots. By working with pre-edited sequences, we have been able to focus on higher-level story constructs (e.g. character development, chronological storytelling) rather than the shot-level sequencing of the earlier experiments. A still frame from *Train of Thought* appears in Color Plate B No. 1.

As more interactive experiences are created with the DMO system, it becomes apparent that research into filter-based interactive movies must move beyond the realm of cinematic sequences (e.g. a dialogue between two characters or an action sequence) to the realm of cinematic structure (e.g. parallel action, conflict resolution, flashback, character development). This becomes possible as authors of interactive movies become more familiar with narrative abstraction and building sketchy video descriptions. Creating narrative abstractions involves picking out and formalizing the essential constructs of a particular story and its intended modes of interaction. This becomes more difficult as cinematic structures become more complex. One possibility for simplifying this process is the use of *metafilters.* Metafilters orchestrate high-level story constructs by selecting filter layers from a descriptive database. Essentially metafilters are filters for filters. This story-structure paradigm allows the creator to build low-level abstractions for cinematic sequences separately from abstractions related to higher-level story structures.

CONCLUSION

Only recently has it become possible to use computation to control the structure of movies. The Digital Movie Orchestrator is a system that gives the computer some control over the content appearing on-screen. A moviemaker breaks the video down into coherent shots, attaches descriptions to these shots to form micromovies, and creates a filtering system that controls the sequencing of shots that make up a personalizable movie experience. The significance of the system is that the filtering is done on-the-fly, thereby allowing real-time control over the direction of the narrative. With real-time shot selection, personalizable movies and fluid interaction modes can be created.

Unlike existing interactive systems, which put the onus for forward navigation on the viewer, the DMO allows the viewer to interact with the story in a fluid manner. It is possible to sit back and enjoy the show or to change the direction of what appears on-screen at any time. Fluid interaction helps promote the reverie of the storytelling experience, an element often lacking in interactive programs.

An Endless Conversation, Train of Thought and the other examples outlined in this paper point to new ways of thinking about building movie experi-

ences. The DMO's descriptive approach to shot selection offers moviemakers an extensible methodology for building multi-threaded narratives. As we establish more extensive video databases and gain more experience with the structure of layered filters, we will build even more sophisticated interactive movies.

Acknowledgments

The work described herein was realized at the Interactive Cinema Group of the MIT Media Laboratory. It was supported in part by directed research contracts from Kansa Corporation, Bellcore and British Telecom.

References and Notes

1. *An Endless Conversation* is included on the CD-ROM *QuickTime™: The CD* (San Francisco: Sumeria, 1992). *An Endless Conversation* and the other movies created with the DMO can be run on any Apple Macintosh™ computer. The DMO system is implemented in Macintosh Common Lisp and Apple QuickTime™.

2. *Murder, Anyone?* (Cincinnati, OH: Vidmax, 1982). Released on videodisc for home entertainment, viewers select alternative chapters and soundtracks to further the story.

3. *I'm Your Man* was designed in 1992 by Bob Bejan of Controlled Entropy Entertainment of New York. It was released in specially equipped Loews theaters where viewers could vote at branch points by pressing buttons at their seats.

4. The Motion Picture Association of America rates movies to restrict their viewing to certain age groups. The PG rating indicates that all ages may attend but that parental guidance is advised. The R rating indicates that viewing is restricted to adults and children accompanied by adults.

5. The term *micromovie* is thought to have been used informally in the Architecture Machine Group in the late 1970s and early 1980s. It refers to short movies whose length and content are appropriate for interactive video applications. We are extending the definition to include attached descriptions.

6. T. Galyean, "Continuous Variables for Interactive Film," Computer Graphics and Animation Group working paper (Cambridge, MA: MIT Media Lab, 1992).

7. T. Aguierre Smith, "If You Could See What I Mean . . . Descriptions of Video in an Anthropologist's Video Notebook," Master's thesis (Cambridge, MA: MIT, 1992).

8. A. Lippman, "Movie-Maps: An Application of the Optical Videodisc to Computer Graphics," *Proceedings of SIGGRAPH* (Seattle, WA: 1980).

9. G. Davenport, "New Orleans in Transition, 1983–1986: The Interactive Delivery of a Cinematic Case Study," transcript of remarks (Boston: The International Congress for Design Planning and Theory, 1987).

10. B. Rubin, "Constraint-Based Cinematic Editing," Master's thesis (Cambridge, MA: MIT, 1989).

11. A. Bruckman, "The Combinatorics of Storytelling: Mystery Train Interactive," Interactive Cinema Group working paper (Cambridge, MA: MIT Media Lab, 1990).