

Browsing the Structure of Multimedia Stories

Robert B. Allen and Jane Acheson

College of Library and Information Services
University of Maryland
College Park, MD 20742
rba@glue.umd.edu and achesoja@wam.umd.edu

ABSTRACT

Stories may be analyzed as sequences of causally-related events and reactions to those events by the characters. We employ a notation of plot elements, similar to one developed by Lehnert, and we extend that by forming higher level “story threads”. This notation requires that events and reactions be linked and that the chains of links be terminated back to the beginning of the story. Furthermore, we have built a browser for the plot elements, the story threads, and associated multimedia. We apply the browser to *Corduroy*, a children’s short feature which was analyzed in detail. We provide additional illustrations with analysis of *Kiss of Death*, a Film Noir classic. Effectively, the browser provides a framework for interactive summaries of the narrative.

KEYWORDS: browsing, hypermedia, multimedia, narratives, plot, structure, summaries

1. STORIES AND STRUCTURE

As digital processing of video has become more practical, there has been considerable research into automatic methods for segmenting those videos (e.g., [2, 11]). Typically, this work has focused on detecting shot and scene boundaries. We argue that the best level of analysis for videos is not shot and scene boundaries, but plot segments. A single scene may have one character exiting and another character entering so it may bridge two strands of the plot. We describe a system for tagging plots of multimedia stories; this system becomes the basis for the browser. Plot is the causal sequence of events from a story. Other aspects of the stories such as characterization and mood are minimized.

There are several scholarly traditions for exploring the structure of stories. Here we examine narratology and psychological approaches to narratives. We then de-

scribe related work on tagging systems and multimedia interfaces.

Narratology

Literary and film criticism have long studied the elements that compose a story. A useful distinction, originally described by the Russian Formalist school, divides *fabula*, the story’s themes, characters and main point, from *syuzhet*, the artistic and syntactic structure by which the fabula is conveyed [22]. The separation of these two elements allows for the study of archetypical stories, such as folk tales, independent of the study of techniques in written fiction, storytelling, or filmmaking.

In general, every story raises a problem or addresses an unresolved situation. At the simplest level a story retains a single major topic and ends with some sort of resolution. Propp’s [18] seminal work with Russian wondertales suggests that certain traditional story forms mix and match fabula fragments, drawing from a cultural bank of available scripts such as “hero saves girl” or “hero is bewitched”. While Propp studied a narrow and rigid story tradition, modern stories can still be fit into his framework. Though the available scripts change with the culture, the archetypes generally do not. Hollywood films constantly employ archetypical stories of heroes and those stories are told in a structured way.

Psychological Approaches to Structure and Stories

Structure is integral to human perception and comprehension, from natural language syntax to music appreciation [7]; and structure has been a major theme in psychological theory. For instance, field theory and models of cognitive balance [25] suggest that structure is crucial to cognitive organization.

Of particular relevance is the evidence about the importance of structure for human understanding of stories [5, 6], and Rumelhart has argued that there is a “well-formedness” of stories [19] and he has proposed a formal grammar for describing stories. However, this model’s rigidity has been criticized [24] and it has not been successfully applied to any complex stories. Scripts [20] are

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Digital Libraries, San Antonio, TX.
Copyright 2000 ACM 1-58113-231-X/00/0006. . .\$.50



Figure 1: IBrowse tagger applied to news stories (from [13]).

another approach to the description of narratives. These led to the work of Lehnert [15] which, in turn, leads to the approach we use here.

Tag Systems and Hypermedia Interfaces

A formalism for describing stories, even if it were not entirely validated ethnologically or psychologically, could be helpful for tagging the structure of stories. Narrative tags would produce a network of hypermedia links. In that respect, they would be similar to argumentation systems, such as SEPIA [23], which identify the progression of ideas during a group discussion.

The IBrowse system [3] applied tables-of-contents to digitized videos. In a sense, the IBrowse hierarchies were a set of tags for describing the content of the videos. In addition to the browser, a tagging tool, shown in Figure 1, was developed to support the creation of the tables of contents. The tree editor for the hierarchy is shown in the top right corner of the figure. MAD [4] is a video-scripting which employs hierarchies for video authoring.

Unfortunately, the simple hierarchical outline or table-of-contents model hides much of the structure. For instance, news programs may lead with feature stories and then later in the broadcast they may provide additional stories related to those original features. Effectively, we are calling for a “theory” of a well-structured table of contents. The *relationships* of nodes in the hierarchies are not clearly identified. For instance, the film *Call of the Wild* begins by tracing the life of the main human character and then it turns to a parallel but independent description of the dog’s background. A simple hierarchy has no way to indicate that parallel development. Indeed, in some stories, aspects of the plot are interleaved so that they can be made sense of only by tracing one plot line and then the other.

2. MODELING PLOTS WITH EVENTS AND MENTAL STATES

The basic elements of our notation are events and reactions by characters. We attempt to minimize the subtle details of the content and focus on the structure. We realize that this is an oversimplification in many cases; indeed, we adopt it exactly because it simplifies the complexities of the stories (cf., [10]). We use a system for identifying events and mental states which is derived from Lehnert’s description of plot units [15]. However, her simple plot unit structures such as “Promise Kept” consist of a series of actions that occur in a predictable fashion between pairs of characters. Although we adapted Lehnert’s notation, we did not attempt to isolate the specific plot units she employed because our plots seemed far too complex.

Events

Events (‘E’) are both actions taken by characters and circumstances which happen to those characters. In our approach, an event has to be perceived by a character and based on its effect on that character, events are assigned one of four valences: +, −, ±, ?. One event may cause reactions in several characters, but the event itself is marked only under the character to whom it occurred. In such a case, other characters would receive mental state tags to describe their reactions.

Mental States

Mental states (‘M’) follow an event as a character’s reaction and, in turn, they may motivate a character’s later actions. A mental state is assigned when a character makes a decision, realizes something, or otherwise exhibits an emotion in a classifiable way. In most cases, the analyst must infer the mental states. In part, this is due to the visual nature of film, which does not show abstractions well. Like events, mental states may take one of four valences: +, −, ±, ?. While stories are usually constructed so that the mental states of the major

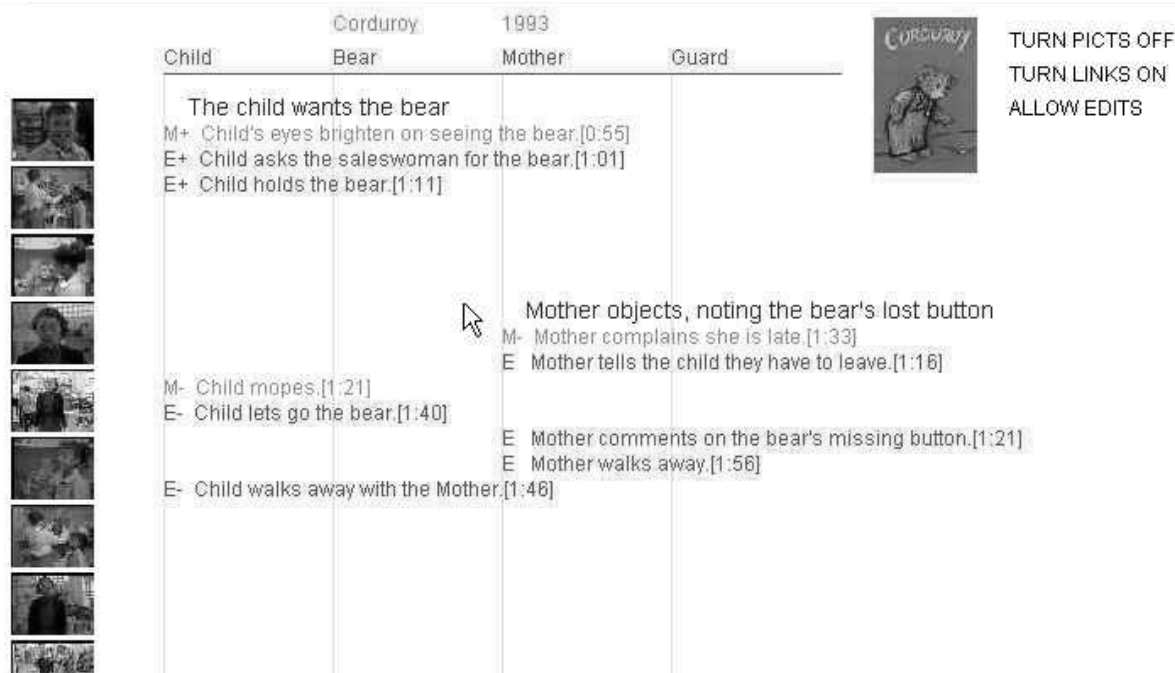


Figure 2: Browser with expanded threads and images. The large font shows Story Thread labels while the smaller M lines show Mental States and the E lines show events.

characters are clear, unknown (?) mental state valences might be found, for instance, in mysteries.

Links

Mental states and events are connected to each other. These links document causation and the importance of one tag to the next. There are four link types: motivation, activation, termination, and equivalence.

Activation (a) generally connects a mental state to a later event. This link provides clear causation from a character's state of mind to the actions he/she takes.

Motivation (m) generally connects forward to a subsequent mental state. An event or previous mental state is marked as causing later mental states.

Termination (t) links back to an earlier event or mental state. This closes a complex event sequence. The act of keeping a promise completes the making of that promise, so the 't' connects them and, in a sense, completes a circuit, balancing 'm' and 'a' links.

Equivalence (e) links are also backward, and they indicate an outcome that is not significantly different from its starting point in moving toward a goal. Equivalence does not, however, necessarily imply termination; it simply marks a value judgement about the state.

Plot elements are generally be part of chain of links. In cases where they are not part of chains, the element may be a candidate for deletion or the viewer is expected to make inferences about possible intervening events that

are not shown.

Similarly, in most cases chains of links need to be concluded with 't' links. Exceptions to this rule might include the denouement at the end of the story. The setting at the beginning of the story may not be included within the terminations. Such cases would be early candidates for deletions in creating summaries.

3. STORY-THREAD STRUCTURES

Once the low-level plot units have been identified we derived higher-level structures (see Figure 2) which we term Story Threads.

Clustering of Plot Elements

Beyond a simple description of the visual elements of the film with which we work, the narrative analysis techniques we use are intended to capture the meaningfulness of events to characters within a film and the clusters of activity or mental states which may denote plot importance.

The mapping of a story into plot notation generally reveals clusters of events; the presence of a cluster may be evidence of the importance of the event – or the series of events – captured in that area. We form the story-thread units from these clusters and label the story thread with a phrase that captures the higher-level concept of the cluster. Some of the heuristics we employ include the centrality and the connectedness of events. If an element is the focus of a termination link, it may be considered a central, perhaps causative element in a long series.



Figure 3: Story-Thread-level display for *Kiss of Death*.

Similarly, an element which is linked to several other elements later in the story may also be deemed significant. At the beginning of the story, smaller clusters are allowed because they are simply presenting the characters and establishing the situation which will be played out.

Deletions from the Plot

Sometimes whole scenes may be dropped from a higher-level description of a film. *Corduroy*, for example, relies on a repetition of the basic bear-hides-from-security-guard unit. Having seen such a unit once, the viewer's understanding is not substantially increased by seeing it repeated as a tool for dramatic tension. A description which relates the mood of the work might benefit from inclusion of the repetition structure, but a plot- and theme-related description would be unnecessarily cluttered. Humor that does not specifically drive the plot is generally dropped; so too would the thrill of a car chase be rendered in its unthrilling plot significance.

Most deletions remove *redundancy* from the description. The ability to infer a state of mind, rather than present evidence of it; the logical cause to a described event; the expository statement of a described event, and foreshadowing later actualized are all examples of redundancy built into the story which are unnecessary in story thread descriptions. Inclusion of these elements might help to create mood, or give a more detailed sense of the plot, but the story can be understood without them.

4. CRITERIA FOR TAGGING AND EXAMPLES

The human tagger must make judgements in assigning elements to characters and describing the plot. Some knowledge of the entire story is required to make many of these judgements.

Selecting Elements

Even in a basic description, several elements are not included. Atomistic dissection of every event would hamper broader understanding of the story, so a human tagger generally chooses to describe a whole event (e.g., 'gets dressed') rather than the individual actions that make up the event ('puts on hat, puts on . . .'). The descriptions employ a standard of *minimum causality* between related events. While an exhaustive tagging might describe the bear seeing a toy and then forming a mental state about that toy as two separate elements, the latter may be assumed to include the former. In addition, minor characters can also be excluded, to the extent that they do not display mental states relevant to the plot. These characters exist in the work only to cause the mental states and events of other characters or to act merely as passers-by.

Annotations for Events and Mental States

It is difficult to abstract a story in a visual medium. A shot loses its meaning when taken out of context, and unlike textual abstracting, film abstracting is forced to rely entirely on the substance of the work in question. Text is simply a more flexible medium: An abstracter can create combinations of words to supply a text description, including condensing and generalizing to a higher level.

Analysis of *Corduroy*

We used the children's story *Corduroy* [16] to test the tagging process and browser interface. It is an 18-minute long live-action film, based on the well-known children's picture book, and it provides a narrative with simple themes and clear goals for the hero character, a stuffed bear in search of a button for his corduroy jeans and ultimately a home for himself. Like folk tales, children's stories tend to be highly normative in theme and narrow in scope. Moreover, in this film adaptation, the

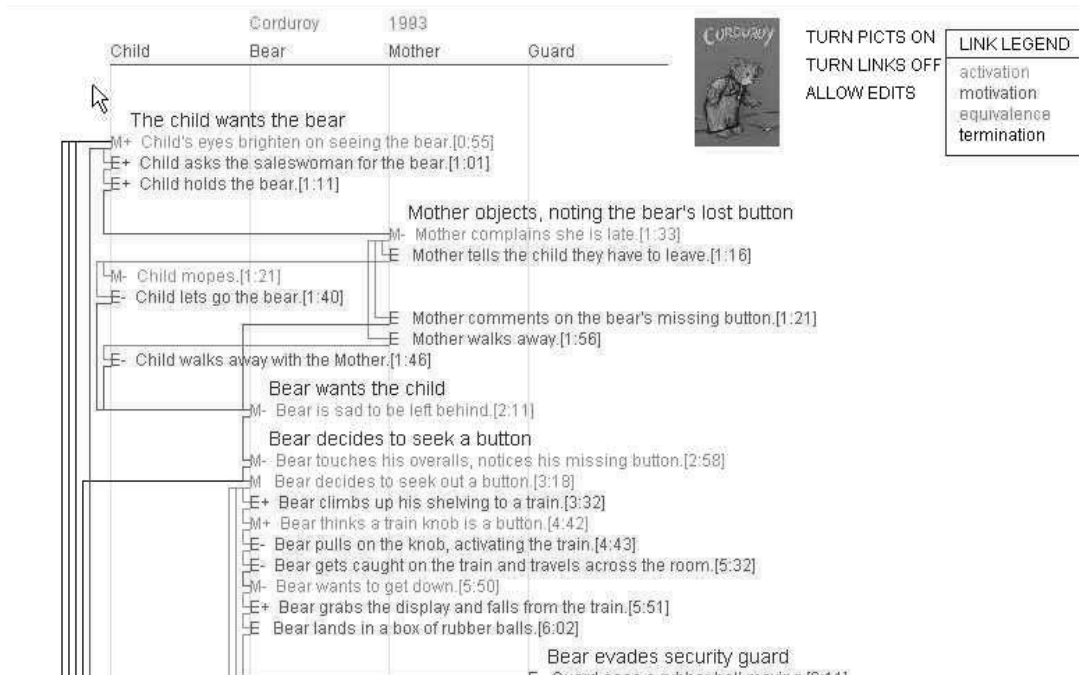


Figure 4: Browser with links-on for *Corduroy*. Note the way the Story Thread labels capture clusters of events that are, generally, centered around one or another of the characters. The dark lines running up the left side are termination links.

bear doesn't talk or have facial expressions; in this film, music is essential for understanding the mental state of the bear. Figure 2 shows examples of the plot elements and Story Threads. Overall, for the plot elements, we assigned 72 events and 51 mental states. The Story Threads analysis suggests that both the girl in the story and the bear have blocked goals that are mutually satisfied when the girl gets the bear.

Analysis of *Kiss of Death*

We analyzed parts of *Kiss of Death* [14], a full-length Film Noir classic from 1947. It was chosen on the basis of its adherence to classical Hollywood style, a formalized set of visual narrative expectations. It contains a single hero undergoing an internal struggle; the hero resolves his trouble in the last reel and even gets the "girl". The narrative of *Kiss of Death* is considerably more complicated than that of *Corduroy*. The action is driven forward intermittently by the steps of a deal being offered, accepted, and completed. A map of the deal does not, however, capture every action and mental state in the film, nor does it go forward in linear fashion. The deal brings the characters together. Scenes describing the hero's goals and manifestations of threats interrupt the deal.

Figure 3 shows the Story Threads for *Kiss of Death*. Three threads may be identified from the left, center, and right columns of the figure: (1) the threats to the hero's happiness on the left, (2) the deal by which the hero triumphs in the center, and (3) the goals to which

the hero aspires on the right. These threads overlap in time through the story and may coexist in a single scene. The description for *Kiss of Death* is more complex than that of *Corduroy*. It captures only the theme whereas *Corduroy* captured both theme and plot. To a certain extent, a longer film simply requires more extreme abstraction to be manageable. By stressing theme, the multiple-character focus is also minimized: instead of representing the villain, Figure 3 shows the threat to the hero's success which the villain represents.

5. STORY-THREADS BROWSER

The Story Threads browser is implemented in Java and runs as an applet. The functionality described in this paper is operational, but some extensions such as playing video clips with the Java Media Framework are only partially implemented.

Plot Elements and Story Threads

Figures 2 and 3 are screen dumps from the browser. Much like the table-of-contents browsers such as IBrowse, the story threads labels are active. Mouse clicks on them open and close the associated plot elements.

In Figure 2, the Story Thread labels have been expanded to show the plot elements. The horizontal dimension is also significant. For *Corduroy* we have chosen character as the horizontal dimension while for *Kiss of Death* we used goals. We could have used other properties such as setting.

Linking

Linking to show causality is integral to the plot element framework; the links give the users a way to trace the events and reactions that form the story. However, the links are complex and may distract from other uses of the interface so they were set to be toggled by a button. Figure 4 shows the story with the links displayed. The links clearly show the complexity of causation for events even in this relatively simple story. For instance, we can see the many implications of the bear's deciding to seek a button at 3:18 into the story.

Still Images

Still pictures were extracted from the video. To the extent possible, these were selected as representative frames for each of the plot events. If the user elects to view them (as shown in Figure 2) they are tiled down the left edge of the interface and the story-thread labels are aligned with them.

6. LIMITATIONS OF THE APPROACH

Features Ignored Due to Simplification

We have intentionally simplified the structure of stories in the course of implementing our system. By focusing on plot and to a lesser extent, theme, we have excluded other aspects in a story such as settings and character development. Some of these limitations could be remedied with extensions to the notation. This notation does not describe the relationships among people. In addition, the subtexts of such stories are not addressed.

The viewer must assume the knowledge and mental state of the characters he/she describes from the evidence given in the film. However, we may not know how a given character reacts to an event, or even whether he/she is aware of that event. Similarly, a character may labor under a mistaken assumption that the viewer does not share.

Intentional manipulations in the story's presentation, such as flashbacks and misleading events, have not been addressed. Further study will be needed for stories in which the presentation of the story differs in order, truth value or other means from the story's own core events.

20th Century literary criticism downplays the role of the author as the primary decipherer of a text. Any culturally-competent reader or viewer may make as legitimate a reading of a work as that of the author. If each reader offers his or her own reading, then a single text may sustain several different patterns of narrative coherence at the same time, especially in stories with ambiguous elements. We have not yet tested cases of stories which have more than one popular reading; we suspect they would yield mutually-exclusive visual maps and comparison of these could have educational value.

Difficult Cases

Some more complex cases are harder to fit into the relatively linear model for presenting plot elements adopted here. Stream-of-consciousness stories such as *Finnegan's Wake* and post-modern narratives such as Brecht's *Mother Courage and her Children* are unlikely to be easily analyzable. They actively and intentionally violate the traditions and 'rules' of story. An unreliable narrator – or worse, an unreliable camera – offers miscues, misleading or even outright lying to the audience. In such a case, the 'truth' of the story becomes clear only at the denouement, if ever. The narrative map, then, might not be faithful to the events as seen while being faithful to the 'real story'.

7. SUMMARIES

The development of tagging system and the browser are intertwined with summarization. The Story Thread labels can be used directly as a summary. The fact that we deal with multimedia content (e.g., [9, 17]) and the interactivity of the interface add other dimensions to summarization. For instance, an interface such as ours might be useful for browsing return lists in searches to video libraries (e.g., [12]).

The quality of the abstract depends in part on the perceived needs of the user. In the case of *Corduroy*, the children's book is likely to be at least vaguely familiar to a user; a cursory examination of cluster descriptions may remind the user of the story's main elements

8. EXTENSIONS

While the examples here, have focused on describing the core of the plot structure, it would be desirable to extend the notation and the browser to show better how the plot structure relates to the finished production.

The selection and linking of plot elements is a labor-intensive process. A large number of revisions is often required to find the best set of events and reactions. Just as the IBrowse tagger (Figure 1) was developed to aid in the creation of the hierarchies, it would be helpful to develop a tagging tool for adding plot elements and determining the story threads.

While this work has focused on plot structures, expository structure is important for content such as educational videos [21]. Although expository text has been widely studied (e.g., [8]), expository multimedia has not. Moreover, plot structure and expository structures often are interwoven so the two types of descriptions need to be coordinated.

This work could provide a framework for developing dynamic stories and personalized stories. Moreover, beyond simply tagging and interacting with existing stories, the tools could be useful for the construction of new stories. For instance, they could be used to detect

and notify the author about unresolved or “dangling” elements.

9. IMPLICATIONS AND CONCLUSION

There are many possible applications of a system for describing and presenting story structures. One would be to provide students with an aid to understanding stories and their development in different media. Thus, this would support “media literacy” initiatives. The work also has implications for the the emerging MPEG-7 standard [1]. For instance, the most significant parts of story could be given priority for transmission on a congested network.

Narratives underlay a wide range of content from news to history to educational videos. We have presented a new model for conceptual browsing of narratives. Our notation and the browser seem to have worked well for the traditional story forms explored here.

ACKNOWLEDGMENTS We thank Robert P. Kolker from the English Department for many useful discussions. We also thank Laura Slaughter for discussions about story grammars and for help in digitizing the films. We thank Dagobert Soergel and Joshua Raditz for comments on the manuscript.

REFERENCES

1. MPEG-7 Context and objectives. <http://www.darmstadt.gmd.de/mobile/MPEG7/index.html>.
2. AIGRAIN, P., AND JOLY, P. Medium knowledge-based macro-segmentation of video into sequences. In *Intelligent Multimedia Interfaces*, M. Maybury, Ed. AAAI/MIT Press, 1996.
3. ALLEN, R. B. Interface issues for interactive multimedia documents. In *Digital Libraries: Current Issues*, N. Adam, B. Bhargava, and Y. Yesha, Eds. Springer-Verlag, 1995.
4. BAECKER, R., ROSENTHAL, A. J., FRIEDLANDER, N., SMITH, E., AND COHEN, A. A multimedia system for authoring motion pictures. In *Proceedings ACM Multimedia Conference* (1996: Boston), pp. 31–42.
5. BAGGETT, P. M. Structurally equivalent stories in movie and text and the effect of the medium of recall. *Journal of Verbal Learning and Verbal Behavior* 18 (1979), 333–356.
6. BARTLETT, F. C. *Remembering: A Study in Experimental Social Psychology*. Cambridge University Press, New York, 1932. [1972 paperback edition].
7. BERNSTEIN, L. *The Unanswered Question: Six Talks at Harvard*. Harvard University Press, 1981.
8. BRITTON, B. K., AND BLACK, J., Eds. *Understanding Expository Text: A Theoretical and Practical Handbook for Analyzing Explanatory Text*. Erlbaum, Hillsdale, NJ, 1985.
9. CHRISTEL, M. G., SMITH, M. A., TAYLOR, C. R., AND WINKLER, D. B. Evolving video skims into useful multimedia abstractions. In *Proceedings ACM SIGCHI: Human Factors in Computing Systems* (1998: Los Angeles), pp. 171–178.
10. CRAMPES, M., VEUILLEZ, J. P., AND RANWEZ, S. Adaptive narrative abstraction. In *Proceedings ACM Hypertext* (1998: Pittsburgh), pp. 97–105.
11. DAILIANAS, A., ALLEN, R. B., AND ENGLAND, P. Comparisons of automatic video segmentation algorithms. In *Proceedings SPIE Photonics East'95: Integration Issues in Large Commercial Media Delivery Systems* (1995: Philadelphia).
12. DING, W., MARCHIONINI, G., AND SOERGEL, D. Multimodal surrogates for video browsing. In *Proceedings ACM Digital Libraries* (1999: Berkeley), pp. 85–93.
13. ENGLAND, P., ALLEN, R. B., DAILIANAS, A., SULLIVAN, M., BIANCHI, M., AND HEYBEY, A. IBrowse - The Bellcore video library toolkit. In *Proceedings SPIE Photonics West'96: Electronic Imaging Science and Technology '96: Storage and Retrieval for Image and Video Databases IV* (1996: San Jose).
14. KOLMAR, F. (PRODUCER) AND HATHAWAY H. (DIRECTOR). *Kiss of Death*, 1947. 20th Century Fox.
15. LEHNERT, W. G. Plot units: A narrative summarization strategy. In *Strategies for Natural Language Processing*, W. G. Lehnert and M. H. Ringle, Eds. Erlbaum, 1981, pp. 223–244.
16. MANTEGNA J. AND TEMPLETON, G. (PRODUCERS), TEMPLETON, G. (DIRECTOR). *Corduroy* (video). In *Corduroy and Other Bear Stories*. Children's Circle, Weston, CT, 1993. [Based on “Corduroy” by D. Freeman, Viking Press, 1968].
17. O'CONNOR, B. C. Access to moving image documents: Background concepts and proposals for surrogates for film and video works. *Journal of Documentation* 41 (1985), 209–220.
18. PROPP, V. *Morphology of the Folktale*. University of Texas Press, Austin, 1968. [Originally published in Russian in 1928].

19. RUMELHART, D. E. Notes on a schema for stories. In *Representation and Understanding*, D. Bobrow and A. Collins, Eds. Academic, New York, 1975, pp. 211–236.
20. SCHANK, R. C., AND ABELSON, R. P. *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Erlbaum, Hillsdale, 1977.
21. SEMPLE, P. P., ALLEN, R. B., AND ROSE, A. Developing a multimedia video library: Content preparation, indexing, and usage. In *Proceedings of EdMedia* (2000: Montreal, accepted).
22. STAM, R., BURGOYNE, R., AND FLITTERMAN-LEWIS, S. *New Vocabularies in Film Semiotics: Structuralism, Post-structuralism and Beyond*. Routledge, New York, 1992.
23. STREITZ, N., HAAKE, J., HANNEMANN, J., LEMKE, A., SCHULER, W., SCHUTT, H., AND THURING, M. SEPIA: A cooperative hypermedia authoring environment. In *Proceedings ACM Hypertext* (1992), pp. 11–22.
24. WILENSKY, R. Story grammars versus story points. *Behaviour and Brain Science* 6 (1983), 579–623.
25. ZAJONC, R. B. Cognitive theories in social psychology. In *Handbook of Social Psychology (vol 1)*, G. Lindzey and E. Aronson, Eds., 2nd ed. Addison Wesley, 1968, pp. 320–411.