Authoring and Navigating Video in Space and Time.

Article · January 1997

Source: DBLP

3 authors, including:

Nitin Sawhney
The New School
58 PUBLICATIONS **1,407** CITATIONS
SEE PROFILE

Ian Emery Smith
Igneous Systems
74 PUBLICATIONS **4,813** CITATIONS
SEE PROFILE

Authoring and Navigating Video in Space and Time

Nitin Sawhney
MIT Media Laboratory

David Balcom
IBM Interactive Media

Ian Smith
Georgia Institute of Technology

A formal methodology is needed to integrate and exchange spatial and temporal properties in hypermedia and hypertext. We propose a generic framework to structure and dynamically present a new form of video- and text-based media called hypervideo. We developed a Hypervideo Engine and produced an experimental hypermedia work, *HyperCafe*, to illustrate the general properties and aesthetic techniques possible in such a medium.

Walter Cronkite once again presents the news, this time to a junior high school student watching a video clip on his computer. Cronkite says something about Vietnam, helicopters, and the ancient city of Hue. Another clip appears just to the right of the first one, and they play side by side for a few seconds. "The Tet Offensive..." Cronkite continues, but for a moment his voice fades as the boy moves his cursor to the new video, a related documentary about Hue. "The ancient walled city of Hue . . ." a new voice begins.

A white frame appears around the documentary clip's frame: make the choice, or it goes away. The clip plays for a few more seconds, then disappears as the subject of Cronkite's news report changes. He continues to the impending NASA Apollo rocket launch. "One of these Apollo rockets will land a man on the moon in just over a year," Cronkite begins, "but there is still much to be done."

A new video detailing the history of the Apollo missions appears in the same place the Hue documentary did, next to Cronkite. Again the white frame surrounds the clip, and again Cronkite's voice lowers in volume when the boy passes his mouse over the new selection, which gets louder. This time he clicks. Cronkite fades to black. The boy follows the new path, the Apollo rockets, and where they'll lead him no one can say. Back to the news?

While the hypervideo application "News of the Past" doesn't exist, it well could. A user could choose a specific era or event and watch video clips of its news coverage—current, historical, foreign, partisan. The news clips would contain opportunities for action—called links in traditional hypertext—that appear at related moments. A separate video clip appearing next to the "main" news clip would offer a related clip: another point of view, a recent news account of the same story, or a commentary on the rhetoric of news reports in the 1960s.

Hypervideo thus functions much like hypertext. It offers its users a path to follow and provides narrative moments that determine what lies ahead and explains what came before. Unlike a Web page, which can contain several static and simultaneous links within the same space, hypervideo opportunities come and go as the video sequences play out in time. Like static hypertext, more than one opportunity can occur at a time—several clips can play at once, or parts of the video frame itself can be active links. Hypertextual commentary flows around the moving image, offering deeper associations. Yet unlike hypertext, these opportunities go away if not selected. The link assumes a new axis along with space: that of time.

This new medium, hypervideo, clearly has implications for educational software, training tools, new forms of creative expression, and filmmaking. Here we consider strategies for authoring and navigating hyperlinked digital video in space and time.

Lurking in a digital cafe

HyperCafe[1] is an experimental hypermedia project we developed to illustrate general hypervideo concepts. *HyperCafe* places the user in a virtual cafe composed mostly of digital video clips of actors involved in fictional conversations. Users can follow different conversations via temporal and textual opportunities that present alternative narratives. Hypertextual elements take the form of explanatory text, contradictory subtitles, and intruding narratives.

We envisioned *HyperCafe* primarily as a cinematic experience of hyperlinked video scenes. The video sequences play continuously, and at no point can the user's actions stop them. The user simply navigates through the flow of video and

1070-986X/97/$10.00 © 1997 IEEE

links presented. This aesthetic constraint simulates the feeling of an actual visit to a cafe. To provide greater immersion in the experience, we employed a minimalist interface with few explicit visual artifacts on the screen. All navigation and interaction occurs through mouse movement and selection. For instance, changes in the cursor's shape depict different link opportunities and the dynamic status of the video.

A user first entering *HyperCafe* sees an establishing shot of the entire scene. Moving text at the bottom of the screen provides subtle interpretations and instructions on the video content. The camera moves to focus on each of the three tables, giving the user 5 to 10 seconds to select any conversation (see Figure 1). Having chosen, the user enters a narrative sequence determined by the conversations at the selected table. Specific conversations trigger video-based previews of related narrative threads that the user could choose to follow. Video scenes and hypertext appear at different locations in the screen space, emphasizing the importance of temporally and spatially active associations in hypervideo.

Based on our work with *HyperCafe*, we defined a framework for hypervideo structures along with the underlying navigation and aesthetic considerations. We later used the framework in developing the Hypervideo Engine. Let's first consider the prior work in this area that served as a basis for our notions in hypervideo.

Related work

A primary influence on hypervideo originated with the hypertextual framework of Storyspace,[2] a hypertext writing environment from Eastgate Systems that employs a spatial metaphor in displaying links and nodes. Users create "writing spaces," or containers for text and images, which they link to other writing spaces. The writing spaces form a hierarchical structure that users can visually manipulate and reorganize. We constructed a hypertextual version of *HyperCafe* in Storyspace. This enabled us to visualize explicit and implicit connections in the work, creating a map of "narrative video spaces" that we used in prototyping *HyperCafe*.

Synthesis,[3] a tool based on Storyspace, lets you index and navigate analog video content associated with text in writing spaces. It could be used for hypervideo production in the design and prototyping stages. Although primarily used for recording meetings, Synthesis provided an early demonstration of text-to-video linking.

***Figure 1. Scenes from* HyperCafe.**

Video-to-video linking was first demonstrated in the hypermedia journal *Elastic Charles*,[4] developed at the Interactive Cinema Group of the MIT Media Laboratory. Micons—miniaturized movie loops—would briefly appear to indicate video links. This prototype relied on analog video and laser disc technology requiring two screens. Digital video today permits newer design and aesthetic solutions, as in the Interactive Kon-Tiki Museum.[5] Here the designers stressed rhythmic and temporal aspects to achieve continuous integration in linking from video to text and video to video, by exchanging basic qualities between the media types. Time dependence was added to text and spatial simultaneity to video.

Our notion of opportunities in hypervideo requires a temporal window for navigating links in video and text, as an intentional aesthetic. Opportunities exist as dynamic video previews and as links within the frame of moving video. Continuity between video-to-video links results from using new camera and interaction techniques called *navigational bridges* (defined in the framework, below).

VideoBook by Ogawa et al. demonstrated time-based, scenario-oriented hypermedia.[6] Here, multimedia content was specified within a nodal structure and timer driven links were automatically activated to present the content, based on the time attributes. Hardman et al.[7] used timing to explicitly state the source and destination contexts when links are followed. Synchronizing media elements is time consuming and difficult to maintain. Buchanan and Zellweger's Firefly system[8] let authors create hypermedia documents by manipulating temporal relationships among media elements at a high level, rather than as timings.

Besides linking strategies and synchronization issues, storytelling is important in many experiences. Agent Stories,[9] developed at the Interactive Cinema Group, fosters creating multi-threaded story structures built on knowledge representa-

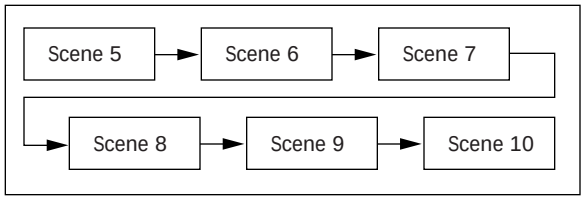


Figure 2. Narrative sequence of scenes.

tions of characters in the story. The Contour system[10] demonstrated an evolving documentary, where viewers could influence the story context. As new media elements were added, they were assigned scores and descriptors. A viewer experience results from changing the relative prominence of these media elements as the viewer browses through the content.

In our approach to hypervideo, the author retains control over the narrative structures presented to the viewer. We focus on the viewer experience generated by navigating and interacting with the video and text-based opportunities that unfold in the course of the narrative. Such opportunities are intentionally structured and presented within a temporal continuity based on aspects of film aesthetic and hypertext fiction (further discussed elsewhere[1]). In developing a structural framework for hypervideo, we considered the design aspects and interaction modalities for presenting such experiences to viewers.

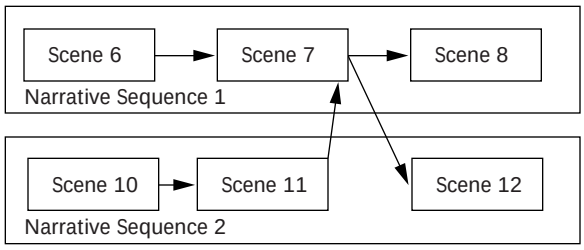


Figure 3. Scenes shared between narrative sequences.

Framework for hypervideo

Hypervideo necessitates a restructuring and rethinking of ideas about authoring and navigating links. In a video-centric medium, the notion of links (as traditionally considered in hypertext) must be redefined to consider the medium's spatial and temporal properties. Authoring hypervideo requires designing video and text-based narratives that appear dynamically in space and time. New interaction paradigms must be developed to navigate a hierarchy of linked, video-driven content. Such a temporal and spatial approach towards hypermedia and hypertext has the potential to change the traditionally confining modes of representation. It may also help us develop richer narrative expressions and innovative applications.

In a traditional hypertextual framework, nodes, links, and writing spaces provide an essential structure for hypertext documents. In hypervideo, the medium's temporal and spatial nature complicates the framework. We propose several new structural and navigational concepts to provide a unified approach towards hypervideo. The grammar of hypervideo is defined in terms of *scenes*, *narrative sequences*, *navigation*, *link opportunities*, and *navigational bridges*.

Scenes

Scenes represent the smallest unit of hypervideo. A scene consists of a set of digital video frames (possibly including an audio track) presented sequentially. An example from *HyperCafe* illustrates a scene: a video clip plays in the center of the screen. A man approaches a woman in a cafe, sits down next to her, and begins to talk. This scene represents a complete structural unit that encapsulates unique contextual information in such a way that it can be associated with other, related scenes.

Narrative sequences

Narrative sequences represent a potential path or thread through a set of linked video scenes and synchronized hypertext, sometimes dynamically assembled based on user interaction or on the context of the scenes (see Figures 2 and 3). Although a scene may be shared by several narrative sequences, the context may change in each. Hence, the narrative sequence serves as a unifying concept for creating story threads from a collage of video- and text-based elements. The body of narrative sequences can be considered a "map" overlaid onto the graph structure of hypervideo nodes (containing scenes and hypertext). This representation provides authors with a familiar temporal and semantic structure.

Navigation

Traversing the scenes of a hypervideo, within and across narrative sequences, requires time-based links—opportunities for navigation that only exist for a short duration. Traditionally, links imply static associations always available to a reader, whereas opportunities imply a window of time (or space) when an association may be active. Such

opportunities appear dynamically, based on the current scene's context, and provide navigational pathways to related scenes. Several types of opportunities based on text and video-based media exist, with both temporal and spatial properties.

Temporal opportunities. Traditional hypertext presents users with several text- or image-based links simultaneously—the opportunities in any one node are available concurrently. In the narrative sequences of hypervideo, several opportunities can be provided temporally, determined by the context of the events in the current video scene. The user has a brief window of time (say, three to five seconds) to pursue a different narrative path. If the user makes a selection, the narrative shifts to a labyrinth of paths available within the hypervideo's structure. Otherwise, the predetermined video sequence continues to play and the temporal opportunity becomes unavailable.

A temporal opportunity could be generally defined as a time-based reference between video scenes, where a specific time in the source video can trigger (if activated) the playback of the destination video scene. Such opportunities can be presented in two forms:

- *Intra-frame opportunities.* Temporal opportunities within the frame of the video scene are available implicitly (perhaps indicated by changes in the cursor, textual intrusions, sound, or other interaction techniques). If the user selects the current video scene (*source*) at a specific point in time, for example, when a character in the scene makes a particular statement, the narrative moves to a different (but related) destination video scene. Hence as the source video scene continues to play, different opportunities become active at specific periods of time within the current scene.

- *Extra-frame opportunities.* Temporal opportunities may also be made available as video-based previews of the destination scenes. As the source video scene plays out, one or more previews of related video scenes would dynamically appear in different locations on the screen, at different points in time for varying duration. Here the user would have a chance to browse the previews by watching and "hearing" the scenes before making a choice to navigate to the link opportunity's actual destination. Temporal previews may be represented by video scenes with wireframes around the video to indicate links (see Figure 4), and the volume dynamically changes as the user moves the cursor over the previews.

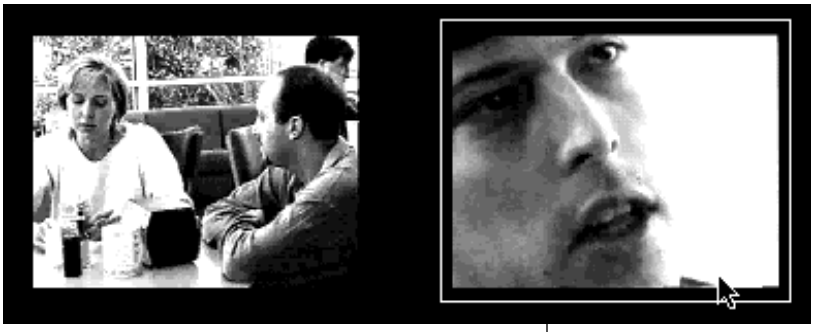


Figure 4. Temporal opportunity as a video-based preview.

Temporal hypertext. Interpretative text in hypervideo can consist of textual narration, explanatory text supporting the scenes, or contradictory/intruding narratives. Such text-based content may be associated (and hence synchronized) with specific scenes or the entire narrative sequence. As a particular scene plays out, a textual narrative may appear for a short time. Such dynamic text may be represented by fading or blurring the text as it appears or lines of text moving across the screen (with varying speed and direction).

Textual elements may be triggered by specific events in a scene. For example, in *HyperCafe* one scene has an actor discussing death. In another narrative sequence, the same actor, in a different scene, talks about possibilities and changing outcomes. A textual statement appears below the frame that says, "You nearly ran me over with your car today." This textual statement comments on what has gone before and alludes to what may lie ahead. Words in this text, if selected, may trigger related scenes. Several lines of interpretative text overlapping with the video sequences creates aesthetic and narrative "simultanieties"[11] in the overall videotext. The presence of temporal hypertext in the realm of the primarily visual video clips also creates a tension between word and image that further illustrates thematic elements present in *HyperCafe*.

Spatiotemporal opportunities. Moving, video-based scenes have temporal properties (the video changes as it moves in time) and spatial properties (positioning within the overall screen space). These scenes also have implicit spatial semantics attributed to the nature of their content. For example, a scene from *HyperCafe* featuring characters sitting at several tables provides a

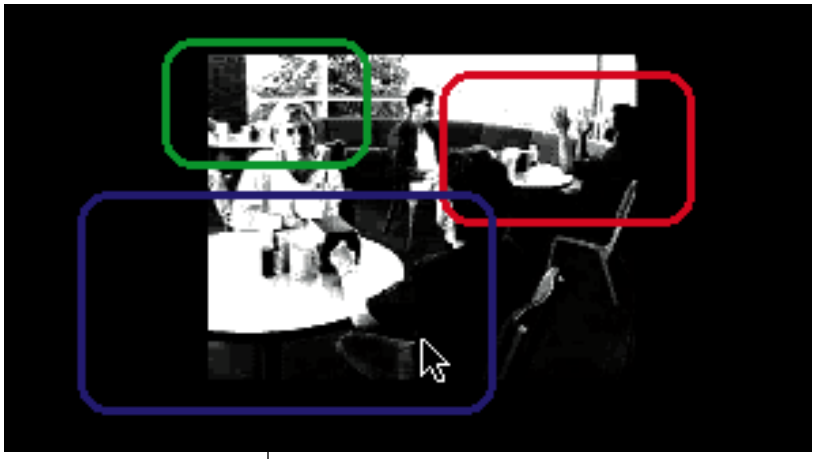


Figure 5. Spatiotemporal opportunities in a scene.

rich source of opportunities to interact with specific tables or characters while the scene plays out, as the camera slowly pans across the room. Such opportunities are found *within* the frame itself, where spatial positioning of the conversants in time recalls or uncovers related interactions when activated. Exploring the filmic depth of the scene reveals these "spatiotemporal" opportunities that potentially trigger other narrative sequences, such as another scene with a close-up of conversations at a table in the background.

Figure 6. Navigational bridges between scenes using **HyperCafe** *pan shots.*

Generally, hypervideo requires a more exploratory form of interaction to effectively employ spatiotemporal opportunities. The user could be made aware of such opportunities by several potential interface modes: flashing wireframes within the video, changes in the cursor, and/or possible playback of an audio-only preview of the destination video when the cursor moves over the link space. Several large and overlapping wireframes (see Figure 5) could detract from the aesthetic of the film-like hypervideo content, yet the use of cursor changes alone requires the user to continuously navigate around the video space to find links.

The cursor-only solution resembles Apple's QuickTime VR interface for still images, yet moving video (and hence dynamic link spaces) complicates navigation in hypervideo. An audio-only preview on the link-space assumes that the destination video has an audio-track that would fade in (played over or through the audio stream for the current video) along with an auditory cue to indicate the preview.

One solution employs a combination of modes: Initially, an audio preview in a specific stereo channel provides a general directional cue indicating spatiotemporal links. Changes in the cursor prompt the user to move around the video space. Actual link spaces are shown by a cursor change, coupled, if necessary, with a brief flash of a wireframe around the link space. Overlapping link spaces are shown only temporally, not simultaneously. We must prototype and evaluate which mode or combination of modes best suits spatiotemporal opportunities.

Other interaction techniques could also work. For example, you could represent opportunities within the scene by creating a visual blur or halo around the characters or by changing the color properties and contrast to emphasize specific regions. Opportunities deeper within the frame of the scene could have different levels of blur or color tone, to provide a visual distinction for simultaneous spatiotemporal opportunities.

Navigational bridges

Hypertext-based narratives can produce a sense of continuity through the textual interpretation and associations created in the process of reading. In a video-centric medium, you may create a continuous aesthetic by letting scenes easily blend together as they progress in a narrative sequence or are activated through opportunities. For example, in a cafe scenario, if a user selects a specific table of conversations in the background of a scene, the navigation to the related scene could be represented by a gradual zoom or pan to the table and its participants (see Figure 6). In addition, specific filler sequences could be shot and played in loops to fill the dead-ends and holes in the narratives. Hence, we need to develop navigational bridges to provide a sense of continuity between video-to-video scene transitions and structural unity for scenes within a narrative sequence.

During a video production, camera techniques can produce navigational bridges between some scenes without breaking the cinematic aesthetic. For example, triggering specifically shot video seg-

ments between the tables can represent movement between them. In *HyperCafe*, generic pans between scenes, such as long shots of the cafe or filming the actor's feet (where continuity is maintained), provided footage for several navigational bridges.

Visual continuity may not be a strict requirement in hypervideo, as well-established cinematic conventions deal with continuous expression, that is, the grammar of visual cuts and pans in the shots and sound-based continuity techniques. We need to develop such a grammar for hypervideo-based scenes, where a user's interaction with the scene also produces an overall sense of continuous expression. For example, two scenes in a cafe may be bridged together by using a continuous soundtrack of the cafe ambiance that gradually fades in or out during the transition. Clearly such contextual continuity can be represented by navigational bridges that use visual, audio, and textual techniques or a combination of such techniques.

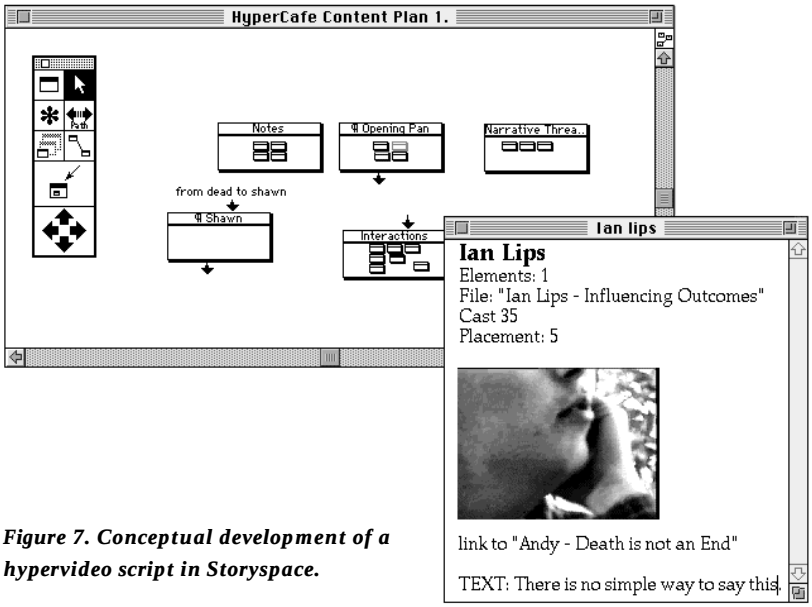


Figure 7. Conceptual development of a hypervideo script in Storyspace.

Hypervideo production

Conceiving and implementing a hypervideo requires several phases of design and production, roughly divided into four parts: *conceptualization*, *video production*, *postproduction*, and *scripting*. The conceptualization (planning) phase requires the author to consider the hypervideo's overall (large-scale) structure. A hypertext authoring tool such as Storyspace can help substantially at this point, in constructing the *video spaces* (analogs of the nodes in a conventional hypertext). Note that since Storyspace has no explicit way to express hypervideo's temporal and spatial nature, the author must use textual notes to serve this need. Storyspace helps the author organize information in hyperlinked structures.

At this stage the author must conceptualize the narrative's link structure prior to actual video production. The result of this phase—a *hyperscript*—should embody sufficient detail about the video scenes, narrative sequences, and opportunities to produce a *shooting script* for the next phase.

The shooting script provides details regarding locations, timelines, order of shooting, actors, and set backgrounds for an actual video shoot. The video production phase requires the author to map the hyperscript onto the process of linear (traditional) production. Now the usual array of specialists are needed to produce the video footage, such as crew for video, sound, and lighting, as well as actors and a director. Some scenes might need two or more cameras to capture the action from multiple perspectives, such as long-shots, close-ups, or reaction shots. These shots can be used in a hypervideo (via the linking mechanism) in new ways unlike traditional video. Additional shots can provide navigational bridges between related scenes, such as top and below-level pans, wandering point-of-view (POV) shots, and rotating shots.

The third phase consists of postproduction or video editing. It involves editing the raw video footage and capturing it in digital form. Multiple takes of the same material can work in interesting ways in a hypervideo, so postproduction lets authors find ways of incorporating alternate takes or camera perspectives of the same scenes.

Once edited, the video must be transcribed and cataloged for later organization into a multithreaded hypervideo. If the author used Storyspace during conceptualization, it might help to associate the video spaces created in Storyspace with the actual digitized video (see Figure 7). This lets the author match conceptualizations with video footage. It also permits better organization of the selected video scenes, which later aids in constructing the hypervideo scripts.

After editing, the selected video scenes must be digitally captured on a suitable computing platform. Although you could use high-resolution video, a 160×120 resolution at 15 frames per second might suffice for hypervideo, especially for preliminary prototyping. Additional digital editing and effects can be performed with desktop

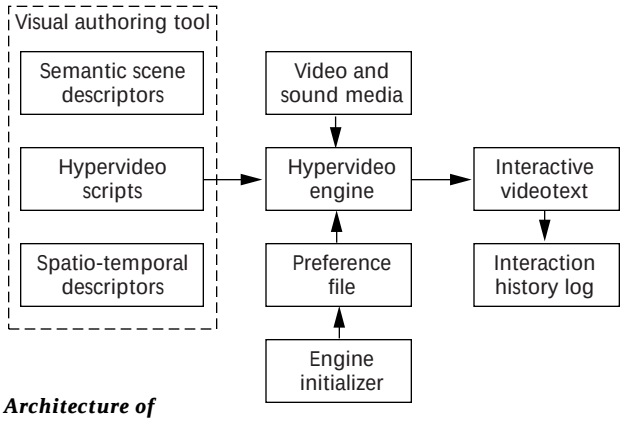


Figure 8. Architecture of the Hypervideo Engine with future extensions.

video-editing software like Adobe Premiere. The author may then choose to trim the length of certain scenes or apply digital filters to change their appearance. In *HyperCafe*, a black and white filter provided a film-like, grainy quality.

Before turning to the final stage of the production process, authoring scripts, we will consider properties of the Hypervideo Engine.

Figure 9. Interaction techniques for a temporal opportunity.

Hypervideo Engine

To demonstrate concepts in the hypervideo framework, we developed a playback tool we call the Hypervideo Engine. We implemented this prototype system in Macromedia Director's Lingo programming language on a Macintosh platform. The development environment met crucial requirements, providing synchronous control of audio, video, and textual media resources and letting us create a high-level scripting interface to the Engine for authoring hypervideo narratives.

The Engine uses scripts that authors can write in any text editor (and perhaps in the future with an authoring tool). The scripts can specify the spatial and temporal placement of hypertext and video clips on the screen. If no spatial position is specified, the Engine will dynamically assign (nonoverlapping) screen positioning for simultaneous video clips or text. At runtime, the Engine first preloads all the necessary media elements to avoid delays while the user navigates the hypervideo. The Engine interprets the scripts and displays the media elements (audio, video, and text) synchronously, based on the space-time parameters specified in the script. While the various media play out, the user might navigate to other parts of the hypervideo, causing the Engine to shift to relevant areas of the script. Authors can also mandate specific runtime settings for the Engine, via a preference file. Figure 8 shows a layout of the current Engine components plus possible future extensions controlled by an interactive authoring tool.

As users interact with the hypervideo and explore the opportunities presented, the Engine continuously generates an *interaction log* file. The log file records the user's navigational strategy by logging the video clips played and the navigational links selected, along with relevant time stamps of each activity. A *bookmarking* feature lets users mark their current state in the narrative script, so the Engine can replay the hypervideo from an earlier state when launched.

Interaction techniques

We used a minimalist user interface for navigation and interaction with the hypervideo narratives generated by the Engine. A cinematic metaphor and design aesthetic results from minimizing buttons or static interface artifacts. This permits a simple language of interaction based on cursor movement and selection with dynamic audio and visual indicators. Several interaction techniques, such as changes in the cursor shape, volume adjustment, and flashing wireframes depict different link opportunities and the dynamic status of the hypervideo. These techniques encourage the user's active exploration.

As the Engine plays a video sequence, the user may watch the narrative or skip to other parts of the video (if skipping is enabled in the preferences). In the video spaces, a change in the moving cursor—from an arrow to a cross-hair—represents a temporal opportunity (see Figure 9). Status messages indicate the destination if the user makes a selection within the currently playing video scene.

At certain times during playback of a video clip, one or more related video scenes may appear

briefly at specific locations on the screen. These represent temporal video opportunities and (from the Engine's point of view) would transfer control to other parts of the script. Temporal videos play for a short duration and disappear if not selected by the user. If the cursor moves over a temporal video, a wireframe highlights the video as a link opportunity, and the video preview's audio volume increases relative to the main video scene. This lets a user preview the video content before deciding whether to follow the link opportunity.

```
on MainVideoScript

     showText "There is no simple way to say this," "bottom," "right"
     playVideo "Andy - Artificial Consciousness," 0, 0, 5, 1, 1
     removeText "bottom"

     showText "Panning," "top," "right"
     tempoVideo 3, "Pan - Shawn to Ian to Kelly," 3, 6, 0, "TempoScript_First"
     removeText "top"

     tempoVideo 5, "Ian Lips - Influencing Outcomes," 3, 6, 0, "TempoScript_Second"
     playVideo "Andy - Death is not an End," 0, 0, random(9), 1, 1

end
```

Figure 10. Example hypervideo script.

As video scenes play, text can be displayed at a specified time, duration, and position on the screen. Specific words can serve as hypertext links to other parts of the narrative. The user explores the lines of text by moving the cursor over them to reveal link opportunities, indicated by a changing cursor and status messages showing the destination. Thus a line of text may contain several active words with different hypertext links. Another aesthetic would favor a visual indication of hypertext opportunities by letting them fade in and out, *aging* over time.

Now that we have introduced the Hypervideo Engine, let us return to the production process. The final stage in producing a hypervideo involves authoring the necessary scripts interpreted by the Engine. We developed a hypervideo scripting convention with commands to play video, text, and sound, and define their temporal and spatial positioning. Specific commands define link opportunities between video, text, and related scripts. The author must define a "main video script" where the narrative may be initiated, plus specific scripts representing narrative sequences containing video scenes. Temporal links in the video, temporal videos, and hypertextual links defined in the scripts provide mechanisms for users to navigate through the hypervideo. Figure 10 shows an example of a hypervideo script.

Storyspace-based representations of the hypervideo provide a good basis for writing the actual scripts. An authoring environment integrated into the Hypervideo Engine would permit visual representation and automatic generation of hypervideo scripts.

Future work

We constructed the Hypervideo Engine to demonstrate the feasibility of our hypervideo model's navigational and structural concepts. While we feel it demonstrates the framework, the current implementation of the prototype can be extended in some areas.

Scene descriptors

In the current implementation, authors must specify an explicit ordering of scenes and hypertext to create meaningful and continuous narrative sequences. This arises from the nature of the video-based content, where the scenes' visual properties cannot automatically generate narrative sequences (at least without using recent image-processing techniques). The scenes can be augmented by text-based scene descriptors, making the system aware of each scene's narrative context. Scene properties, such as location, characters, transcripts of conversations, and metatextual interpretations, would let the Engine generate narrative sequences based on a specified scene browsing or presentation strategy. Hence user queries on specific characters and conversational elements may reveal undiscovered and emergent patterns in the narrative.

Spatiotemporal descriptors

Although we have prototyped techniques for spatiotemporal links, the current release of the Hypervideo Engine includes no high-level mechanism to let authors define such descriptors in video scenes. Such opportunities can be implemented as dynamically available (perhaps transparent) hot spots associated with specific aspects in the video frame. These hot spots can be defined in a scene by a set of position (x, y) and time coordinates. This approach, although feasible, requires tedious coding. Moreover, if a hot spot must be

registered with an object in the scene and the object moves within the frame (or the camera moves with respect to that object), this approach becomes unworkable. However, with the recent progress in video segmentation and content analysis techniques, conceivably objects in the moving video frame could be detected and tracked automatically. Such techniques would minimize (if not eliminate) explicit coding of spatiotemporal links by the user. For a good discussion of approaches to parsing and abstraction of visual features in video content, see recent work by Chang et al.[12] Several innovative solutions to automatically define hotspots in digital video have been developed recently by companies like Ephyx Technologies (the V-Active system) and Arts Video (the MOvideo system). It's important to consider the effective use of such techniques within the aesthetic possibilities of video-based narratives.

Interactive authoring tools

The hypervideo scripting techniques we provide allow high-level control of the Engine, yet a visual authoring tool can further simplify this approach. Note that this scenario echoes the use of HTML scripting for producing WWW content; only recently have visual authoring tools become available for the Web. A hypervideo authoring tool should support the different phases of hypervideo production, as well as the design of narrative sequences and link opportunities. Such an authoring environment would let authors visually define their linked video structure, like hypertext authoring in Storyspace. Narrative possibilities could then be textually simulated before shooting any video. After the video production, the scenes could be captured and transcribed using the authoring tool. The scenes would be relinked visually to form threads of narrative sequences, and their spatial and temporal properties would be manipulated directly. Textual, temporal, and spatiotemporal opportunities could be added to the scenes. The tool could then create views of the hypervideo space at varying levels of granularity (scene level, narrative sequence, and overall) and organization by keywords, links, or specified paths. The presentation delivered by the Hypervideo Engine would simply serve as a view generated by such an authoring tool.

Spatial representation

Hypervideo, as we have defined it, provides navigation in a 2D space; it seems clear that it can extend into another dimension. Motion-picture time can be presented as a 3D block of images flowing away in distance and time. For hypervideo, a 3D space would permit representation of multiple (and simultaneous) video narratives, where the added dimension could signify temporal or associative properties. Video sequences already visited could be represented by layers of (faded) video receding into the background. A series of forthcoming links and prior (unrealized) links could also appear in the foreground or background of the 3D space.

Collaborative work

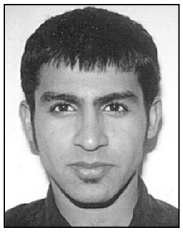
The nodal nature of hypervideo makes it conducive to collaborative work over a network. However, until bandwidth increases and video streaming technologies improve, users must retain video-based media locally (perhaps distributed on CD-ROM) and use the Internet to distribute, share, and modify the videotext structure (currently stored as textual scripts). Users could add their own links or textual annotations and make their videotexts available via WWW or MUD, a multiuser doman. Some form of version control could let users access previous narrative sequences, link opportunities, and annotations. Users could even navigate temporally through the historic state of the videotext, revisiting previous sequences and spaces while investigating what lies ahead—another path among many.

The techniques and methodologies we have described for navigating in "video space" are experimental and untested, yet provide an appealing minimalist approach. Applications of hypervideo in film conceptualization, training, simulation, and electronic news delivery demonstrate its potential appeal beyond the experimental expressions of our *HyperCafe* scenario. We hope that developers and authors can employ the basic hypervideo framework and approach in constructing their own applications, authoring tools, and creative expressions. MM

References

1. N. Sawhney, D. Balcom, and I. Smith, "*HyperCafe*: Narrative and Aesthetic Properties of Hypervideo," *Proc. Hypertext 96*, ACM, 1996, pp. 1-10. Also see the *HyperCafe* Web site: http://www.lcc.gatech.edu/gallery/hypercafe.
2. J.D. Bolter, *Writing Space: The Computer, Hypertext, and the History of Writing*, Lawrence Erlbaum and Associates, Hillsdale, N.J., 1991.

3. C. Potts, J.D. Bolter, and A. Badre, "Collaborative Pre-Writing With a Video-Based Group Working Memory," Tech. Report 93-35, Graphics, Visualization, and Usability Center, Georgia Institute of Technology, Atlanta, Ga., 1993.
4. H.P. Brøndmo and G. Davenport, "Creating and Viewing the Elastic Charles—A Hypermedia Journal," in *Hypertext: State of the Art,* R. Aleese and C. Green, eds., Intellect, Oxford, U.K., 1991, pp. 43-51.
5. G. Liestøl, "Aesthetic and Rhetorical Aspects of Linking Video in Hypermedia," *Proc. Hypertext 94*, ACM Press, New York, 1994, pp. 217-223.
6. R. Ogawa et al., "Design Strategies for Scenario-Based Hypermedia: Description of Its Structure, Dynamics, and Style," *Proc. Hypertext 92*, ACM Press, New York, 1992, pp. 71-80.
7. L. Hardman, D.C.A. Bulterman, and G.V. Rossum, "The Amsterdam Hypermedia Model: Adding Time and Context to the Dexter Model," *Comm. of the ACM,* Vol. 37, No. 2, Feb. 1995, pp. 50-62.
8. M.C. Buchanan and P.T. Zellweger, "Specifying Temporal Behavior in Hypermedia Documents," *Proc. Hypertext 92*, ACM Press, New York, 1992, pp. 262-271.
9. K. Brooks, "Do Story Agents Use Rocking Chairs? The Theory and Implementation of One Model for Computational Narrative," *Proc. Multimedia 96*, ACM Press, New York, 1996, pp. 317-328.
10. G. Davenport and M. Murtaugh. "ConText: Towards the Evolving Documentary," *Proc. Multimedia 95*, ACM Press, New York, 1995, pp. 381-389.
11. J. Rosenberg, "Navigating Nowhere/Hypertext Infrawhere," *ECHT 94*, *ACM SIGLINK Newsletter*, Dec. 1994, pp. 16-19.
12. H.J. Chang et al., "Video Parsing, Retrieval, and Browsing: An Integrated and Content-Based Solution," *Proc. Multimedia 95*, ACM Press, New York, 1995, pp. 15-24.



Nitin Sawhney is a graduate student in the Speech Interface Group at the MIT Media Laboratory. He received an MS in Information Design and Technology from the Georgia Institute of Technology, Atlanta, in 1996. He worked on hypermedia and mobile computing projects at the Graphics, Visualization, and Usability Center at Georgia Tech and the Fuji-Xerox Research Laboratory in Palo Alto. His current research interests include wearable audio computing, adaptive user interfaces, and audio-based communication environments.

Readers can contact Sawhney at The Media Laboratory, MIT E15-352, 20 Ames Street, Cambridge, MA 02139-4307, e-mail nitin@media.mit.edu.

David Balcom is a producer with IBM Interactive Media in Atlanta, Georgia. He holds an MS degree in Information Design and Technology from the Georgia Institute of Technology. He, along with Sawhney and Smith, won the Egelbart Best Paper Award at the 1996 ACM Conference on Hypertext for "HyperCafe: Narrative and Aesthetic Properties of Hypervideo."

Ian Smith is a PhD candidate at the Graphics, Visualization, and Usability Center in the College of Computing at Georgia Institute of Technology. His primary research interests are user interface software, constraint satisfaction systems, and multimedia. He is currently completing his PhD dissertation, *Support for Multi-Viewed Interfaces,* and will graduate in March 1998.